Department of Informatics
University of Fribourg, Switzerland

Human Factors as Constraints and Solutions in Document Processing

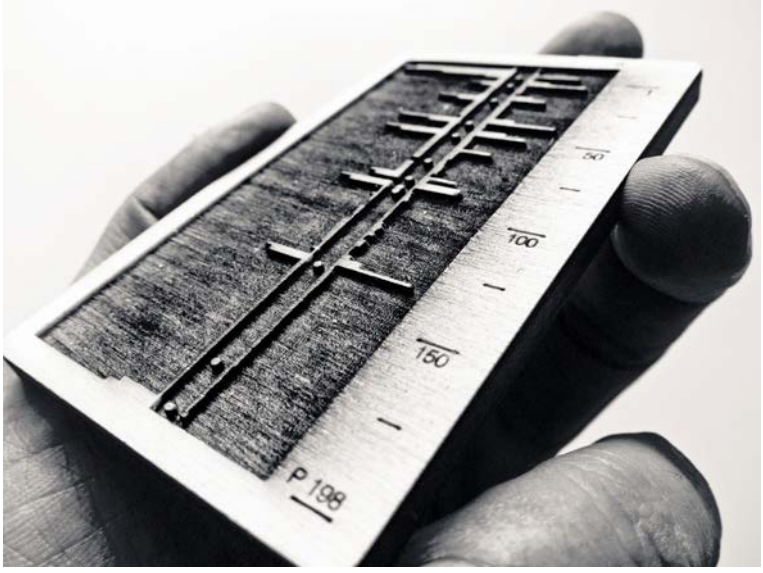
---

THESIS

presented to the Faculty of Science and Medicine
of the University of Fribourg, Switzerland,
in consideration for the award of the academic grade of
Doctor of Philosophy in Computer Science

by Vlad Atanasiu
from Paris, France & Timișoara, Romania

Thesis No. 8159
Fribourg, 2023



Accepted by the Faculty of Science and Medicine
of the University of Fribourg, Switzerland,
upon the recommendation of:

Prof. Dr. Edy Portmann
University of Fribourg
President of the Committee

Prof. Dr. Rolf Ingold
University of Fribourg
Thesis supervisor

Prof. Dr. Andreas Fischer
Haute École d'Ingénierie et d'Architecture, Fribourg
Co-Examiner

Prof. Dr. Prof. Peter Stokes
École Pratique des Hautes Études, Université PSL, Paris
Co-Examiner

Fribourg, 21 June 2023

Thesis supervisor
Prof. Dr. Rolf Ingold

Dean
Prof. Dr. Ulrich Ultes-Nitsche

Abstract

The works presented here challenge a host of common notions in the area of document processing by demonstrating [1] understandable handwriting expertise, [2] personalized legibility enhancement, [3] serendipitous document visualization, and [4] layout information potentials. The underlying conceptual theme raises awareness of human factors as both constraints and solutions in script and layout analysis, and consequently of the necessity for systemic thinking. The methods employed rely on image processing, statistics, spectral analysis, and information visualization, and integrate them with visual psychology, color science, user experiments, and graphic arts domain knowledge, as a means of innovating through interdisciplinarity. Proposing practical solutions and theoretical perspectives rooted in surprising empirical findings, the works further distinguish themselves by having generic value and applications beyond documents.

[1] The script expertise research consisted in creating perceptual-mathematical correlates to improve writer identification and the understandability of computational methods, in estimating handwriting variability and the impact of dataset scrambling on misclassification, and in surveying computational expertise in forensics documents.

[2] In the legibility area, we developed enhancement methods using color processing and visual illusions, produced a theory of image enhancement for critical applications, empirically demonstrated the benefits of personalized, contextualized, and complementary image enhancement, reviewed the utility of color for shape analysis, and studied the limits of legibility optimization.

[3] Regarding document visualization, we created a technique for the visual exploration of document structures using the architectural metaphor of documents-as-buildings, and surveyed its rich applications; through user evaluations, we highlighted the role of serendipity in information-seeking, along with the impact of socioprofessional context on the perceived usability of information systems.

[4] Concerning layout information, we introduced a measure of informativeness, namely structural information potential (SIP), which combines aspects of the Shannon entropy, Fourier transform, fractal dimension, and golden ratio, and defines a perceptually coherent pattern phase space ranging from uniform to clustered to random to regular. SIP helps to determine objects with minimal structural redundancy that complement regular graphs and the Platonic solids, and is specifically appropriate for document triage, as well as a general information-theoretical principle.

Keywords
---
document processing,
document triage,
misclassification,
digitization quality control, document visualization, layout analysis, handwriting expertise, handwriting variability, human factors, perceptual image processing,
personalized image enhancement,
color processing,
serendipity,
spectral analysis,
structural entropy,
fractals

Sommaire

Les présents travaux questionnent certaines notions courantes dans le domaine du traitement automatique des documents, contribuant à (1) la compréhensibilité de l'expertise de l'écriture manuscrite, (2) l'amélioration de la lisibilité par personnalisation, (3) l'utilité de la sérendipité pour la visualisation des documents, et (4) l'analyse du potentiel d'information des mises en pages. Le thème conceptuel sous-jacent vise a valoriser les facteurs humains, en tant que contraintes et solutions, pour l'analyse de l'écriture et de la mise en page, et a défendre la nécessité d'une pensée systémique. Les méthodes employées reposent sur le traitement des images, la statistique, l'analyse spectrale, et la visualisation, ainsi que leur intégration avec la psychologie visuelle, la science des couleurs, l'expérimentation humaine, et les connaissances pratiques dans les arts graphiques, comme moyen d'innovation par l'interdisciplinarité. Les travaux proposent des solutions pratiques et des perspectives théoriques ancrées dans des découvertes empiriques surprenantes, tout en se distinguant également par leur valeur générique et des applications au-delà des documents.

(1) La recherche sur l'expertise de l'écriture manuscrite consiste dans la création des corrélations perceptuelles-mathématiques pour l'amélioration de l'identification de scripteurs et la compréhension des méthodes computationnelles, dans l'estimation de l'impact de la variabilité manuscrite et de la désorganisation des jeux de données sur les erreurs de classification, et dans un panorama de l'expertise computationnelle des documents forensiques.

(2) Dans le domaine de la lisibilité, nous avons développé des méthodes d'amélioration utilisant le traitement des couleurs et des illusions optiques, produit une théorie de l'amélioration des images pour les applications critiques, démontré empiriquement les bénéfices de la personnalisation, contextualisation, et complémentarité pour l'amélioration des images, passé en revue l'utilité de la couleur pour l'analyse de formes, et étudié les limites de l'optimisation de la lisibilité.

(3) Pour la visualisation des documents, nous avons crée une technique pour l'exploration visuelle des structures de documents utilisant la métaphore architecturale des documents-comme-bâtiments, et répertorié ses riches applications; à travers des évaluations utilisateurs, nous avons mis en valeur le rôle de la sérendipité en traitement des documents, et des contextes socioprofessionnels pour l'usabilité des systèmes d'information.

(4) Concernant la mise en page, nous avons introduit une mesure de l'informativité, le potentiel d'information structurel (SIP). Elle combine des aspects de l'entropie de Shannon, la transformée de Fourier, la dimension fractale, et le nombre d'or, et définit un espace perceptuel cohérent des motifs allant de l'uniforme au fragmenté, à l'aléatoire et au régulier. SIP aide a déterminer les objets avec redondance structurelle minimale, en complément aux graphes réguliers et solides de Platon. SIP est particulièrement approprié pour le triage de documents, de même qu'en tant que principe général dans le cadre de la théorie de l'information.

Mots clés
---
traitement des documents, triage de documents, erreur de classification, contrôle de qualité de la numérisation, visualisation des documents, analyse de la mise en page, expertise de l'écriture manuscrite, variabilité de l'écriture manuscrite, facteurs humains, traitement perceptuel des images, personnalisation de l'amélioration des images, traitement des couleurs, sérendipité, analyse spectrale, entropie structurelle, fractals

Contents

Acknowledgments – 6
Foreword – 7
List of Publications – 8

SYNTHESIS OF WORKS

Conceptual Perspectives – 12
Intellectual Background – 16

Script Expertise – 18
Script Legibility – 21

Layout Visualization – 23
Layout Information – 26

Future Outlook – 27
Overall Contributions – 28

Publications – 34
Curriculum Vitae – 346

Acknowledgments

Rolf Ingold
– students' champion

Andreas Fischer
Jean-Luc Bloeche
Maurizio Rigamonti
– happy fellows

Laurence Likforman-Sulem
Nicole Vincent
Isabelle Marthot-Santaniello
– ever surprised

Irvin Schick
Bernice Rogowitz
Ruth Rosenholtz
Aries Arditi
Jean-Marie Nicolas
– breadcrumb scatterers

Evaluation studies participants
– cat herd

Fund in Support of Innovation of the Canton Fribourg
Hasler Foundation
– life bonuses

Jean-Marc Ménadier
Ezio Ornato
Enrico Natale
– open ears

Lucia & Gheorghe Atanasiu
– always there

& cetera

Foreword

This is a special kind of doctoral thesis. Apart from being the author's second, it is a cumulative dissertation: that is, constituted by a thematically loosely connected set of publications, unlike the narrow focus of a monograph, but not precluding a common approach to problem-solving. To give the reader an idea of the work's ecclectism, this volume contains over five hundred pages from twelve of the author's publications (including two books, five journal articles, and five conference papers) spanning the years 2011 to 2022 (while at the University of Fribourg, Switzerland, and ParisTech Télécom, Paris, France), in the field of computer science (with excursions into visual psychology, and the sociology of science and technology), and the domain of document analysis and visualization (at the level of script and layout), mainly using image processing and information visualization methods (as well as statistics, color processing, and visual illusions), and relying on usability studies and domain knowledge (graphic arts). The technical articles are accompanied by less technical derivative publications, which, taken together, better showcase how the wider potential of the research was exploited.

The quantity and variety of the work imposes certain constraints on the matter and form of any introduction to it. Rather than simply presenting a research summary (for which one may directly consult the abstracts), we instead provide a synthesis highlighting two aspects not found (or found only implicitly) in the publications: the overall concepts linking individual publications, and the author's overall contributions. The overview also helps to explain the work in terms of its past and future, and situate it among scientific disciplines and methods.

The author hopes this document brings the reader inspiration.

Vlad Atanasiu
Bern, February 2023

List of Publications

Vlad Atanasiu, Laurence Likforman-Sulem, Nicole Vincent

EXPERTISE

P1. Writer Retrieval: Exploration of a Novel Biometric Scenario Using Perceptual Features Derived from Script Orientation

---

11th International Conference on Document Analysis and Recognition, 18-21 September 2011, Beijing, China, pp. 628-632.
https://doi.org/10.1109/ICDAR.2011.132
–
DEMO: https://glyph.telecom-paristech.fr/
SOFTWARE: https://ch.mathworks.com/matlabcentral/fileexchange/54471-local-shape-outline-orientation-descriptor
https://www.mathworks.com/matlabcentral/fileexchange/54396-alphonse-rex-handwriting-and-print-expertise

Vlad Atanasiu

P2. Forensic vs. Computing writing features as seen by Rex, the intuitive document retriever

---

First International Workshop on Automated Forensic Handwriting Analysis, 17-18 September 2011, Beijing, China, pp. 16-20.
https://ceur-ws.org/Vol-768/Paper4.pdf

Vlad Atanasiu

P3. Project presentation: Rex Online

Gazette du livre médiéval, 2011, 56-57: 131-135.
https://doi.org/10.3406/galim.2011.1988
Includes one conference poster.

Vlad Atanasiu

P4. Cuckoos among Your Data: A Quality Control Method to Retrieve Mislabeled Writer Identities from Handwriting Datasets

---

IS&T International Symposium on Electronic Imaging 2016, Document Recognition and Retrieval XXIII, 15-18 February 2016, San Francisco, CA, USA, paper 056, 16 p.
https://doi.org/10.2352/ISSN.2470-1173.2016.17.DRR-056

Vlad Atanasiu

P5. Expert Bytes: Computer Expertise in Forensic Documents – Players, Needs, Resources, and Pitfalls

---

Boca Raton, FL: CRC Press, 2014, 198 p.
https://doi.org/10.1201/b15467

Vlad Atanasiu, Isabelle Marthot-Santaniello

LEGIBILITY

P6. Personalizing image enhancement for critical visual tasks: improved legibility of papyri using color processing and visual illusions

---

International Journal on Document Analysis and Recognition, 2022 [25]: 129-160.
https://doi.org/10.1007/s10032-021-00386-0
–
SOFTWARE: https://hierax.ch

Vlad Atanasiu, Peter Fornaro

P7. On the utility of Colour in shape analysis: An introduction to Colour science via palaeographical case studies

---

Heliyon, 9 (10): e20698, 25 p.,
https://doi.org/10.1016/j.heliyon.2023.e20698.

Vlad Atanasiu

P8. Ugraphia: The Pursuit of Perfect Legibility

"Graphemics in the 21st century: From graphemes to knowledge",
14-15 June 2018, Brest, France, Extended Abstracts, pp. 113-140.
–
MONOGRAPHY: Brest: Fluxus Éditions, 2024, 186 p.
https://doi.org/10.36824/2024-atanasiu

Vlad Atanasiu, Rolf Ingold

VISUALIZATION

P9. Document Towers: A MATLAB software implementing a three-dimensional architectural paradigm for the visual exploration of digital documents and libraries

---

SoftwareX, 2021, 14: 100684, 7 p.
https://doi.org/10.1016/j.softx.2021.100684
–
SOFTWARE: https://github.com/ElsevierSoftwareX/SOFTX-D-21-00009

Vlad Atanasiu

P10. Applications of the Document Towers method of representing document structures
–
Annex: Dreams, Architecture, and Information Systems: A brief review of the utility of document exploration, visualization, and metaphors

---

Information Design Journal, 2022, 27[1]: 1-47.
https://doi.org/10.1075/idj.20026.ata

Vlad Atanasiu

P11. Improving the Quality Control and Exploration of Digital Documents Using Visualization and Serendipity

---

Journal of Information Architecture, 2022, 7[2]: 7-81.
https://doi.org/10.55135/1015060901/222.011/2.046

Vlad Atanasiu

INFORMATION

P12. The Structural Information Potential and Its Application to Document Triage

---

IEEE Access, 2022, 10: 13103-13138, & IEEE Access, 2023, 11: 85711-85715.
https://doi.org/10.1109/ACCESS.2021.3133654
https://doi.org/10.1109/ACCESS.2023.3299594
–
SOFTWARE: https://ch.mathworks.com/matlabcentral/fileexchange/100099-topological-entropy

Synthesis of Works

Perspectives

The conceptual coherence of the work presented herein may be considered from the triple perspective of application domains, scientific disciplines, and methodological paradigms.

TOPICS

The work deals with issues relating to the computational processing of digital documents, at the level of the script and layout of offline handwritten and printed contemporary and historical documents, in both image and electronic formats, using both automated and interactive methods. The specific topics are writer identification and writer label misclassification detection for handwriting expertise, legibility enhancement of critical documents, layout-based estimation of the information potential of document images, and the visual analysis of electronic document structures. Thus, the research covers a broad spectrum of document aspects, specific issues, and applications. Furthermore, some of the work is applicable beyond the class of paginated documents, and can be extended to pictures and videos.

DISCIPLINES

The above topics belong, from an applicative point of view, to various parts of the computational document processing ecosystem, in particular graphonomics [the study of script, e.g., for forensic expertise], document image analysis, and document engineering. In terms of theories and methods, the work is grounded mainly in image processing and visualization. Classical, non-machine-learning methods are used, such as processing in the frequency domain [e.g., to quantify the scale space distribution of layout patterns] and kernel filtering [e.g., to extract edge orientation from script], which closely ties them to signal processing [e.g., for defining stochastic patterns of document pages]; other leveraged fields include information theory [e.g., to link layout pattern types to informativeness], fractals [e.g., as a model of layout design and photographic framing techniques], color science [e.g., for chromatic contrast enhancement], and statistics [e.g., for quantifying document structures and evaluating experimental results]. Information visualization [itself at the crossroads of mathematics, programming, graphic design, and visual psychology] is used both as an analytic method within the developed applications, and for the purpose of scientific communication of the research results.

Other contributions to this work come from outside computer science, such as from psychology, and science and technology studies [i.e., the study of sociocultural factors in scientific research and technological development]; despite this, they are all familiar to engineers, who know them as “human factors” [7; 17; 21].

Visual psychology has inspired a number of computational document-processing methods developed by the author [e.g., leveraging visual illusions to enhance legibility], guided experimental setups [e.g., selecting evaluation criteria in usability studies of document structure visualizations], and provided explanations for some of the results [e.g., gender differences in the choice of enhancement algorithms].

Domain knowledge in paleography [i.e., the study of old scripts], sci-

ence and technology studies, and other fields of humanities, has helped the author to make informed choices in selecting study topics (e.g., the absence of adequate legibility enhancement methods in papyrology), define requirements and evaluation criteria (e.g., the best reading results are obtained with personalized enhancement methods, rather than the best method on average), and identify the sociocultural basis of some methods (e.g., representing documents using an architectural metaphor is effective because the metaphor is widely culturally understood).

The research has also directly benefited from the author's practical experience with typography, calligraphy, and photography, which motivated the pursuit of specific research topic (e.g., the document visualization was developed in light of the perceived difficulty of overviewing document structures in document layout software and e-readers) and providing underlying concepts (e.g., homogenizing the scale space distribution of patterns, and implicitly their information potential, is a technique commonly used to increase aesthetic attractiveness in many arts, such as typeface and layout design, painting, photography, music, or flower and garden arrangement) [25: 316–317; 20: 208–211; 23; 27; 26].

PARADIGMS

While the topics, aims, and methods of the presented works may differ, they are linked by certain commonalities regarding the ways in which problem solving is approached.

The goal of studying forms in this work (i.e., character shapes and layout patterns) is that of determining not their semantics (i.e., text meaning), but their functions (e.g., modulating information bandwidth through layout design), the multifarious information they may reveal (e.g., writer identity), and other types of potential utility (e.g., increasing document accessibility for persons with low vision). In other words, a core tenet of this research is the consubstantiality of form and function (see the Document Towers further below), and by extension, of the information potential of structures (see the structural information potential section), along with the definition of meaning in terms of functionality (see the future outlook section). These are also fundamental epistemological concepts in pattern analysis, information theory, graphonomics, embodiment (in psychology), design, and the arts, all of which strive to take advantage of the expressive power of forms. From this point of view, an appropriate thesis title would be "Structure, Information, and Action Potentials".

There exists, however, another perspective to the work, more revealing of its peculiarity, which may be articulated as "Perceptual Document Processing" [13; 2; 14]. By this, we mean that it is possible to create computational methods that are effective not solely because of the mathematics and algorithms involved, but primarily because they leverage psychological characteristics of the users; these are the so-called human factors (e.g., our use of visual illusions to improve legibility), a subfield of which is perceptual image processing. Data analysis may be further shared between machines and humans through their visual representation (which is the domain of visual analytics), where machines interactively process, quantify, and represent data, and humans interpret them (e.g., our approach to exploring document structures in the Document Towers paradigm). The perceptual component of the work is also situated at the level of the domain from which concepts were sourced: e.g., in visual communication and graphic design, asymmetrical layout is a common strategy for expressing semantic hi-

erarchies (such as large-sized titles, medium-sized text, and small-sized footnotes) [24], and the technical challenge consisted in elaborating the appropriate mathematical formula to measure the information potential of documents solely from their layout. The choice of features on which to base document analysis may also be driven by perceptual considerations; in the case of our handwriting expertise system, Rex, we developed a family of perceptual features that are intuitively familiar to users (e.g., slant, roundness, and variability). In conclusion, the term "perceptual document processing" covers perceptual aspects at various technological levels, from inspiration and selection to algorithm and representation; it is also broader than visual perception, and includes such elements as social and cultural factors, gender, use context, and so forth. The relevance of the perceptual approach to document processing is that it is not mainstream, despite the potentialities of opening several scientific domains to each other.

Another overall methodological paradigm of this research, which may have become clear by now, is an approach to technological development as holistic, systemic, user-centered, and contextualized. These conceptual frames imply great variability in observers, instruments, and observations, explaining why the term "potential" (both as a hypothetical outcome and its corollary of multiple alternatives) watermarks each of our works, and encourages us to pursue subjective, even serendipitous, solutions, in addition to universal ones. To give two examples, the usability of the Document Towers visualization was rated radically differently by librarians and academics, and almost every user found unpredictably different and useful insights; similarly, when evaluating legibility enhancement methods, user and document variability was taken into account, and the use of several methods was recommended rather than the one with the best average performance (as a statistical mindset might suggest), which is optimal in few if any cases (of average utility, indeed). Thus, a further possible title would be "A Systems Perspective on Document Processing".

In this work, holism is understood as a mental attitude that has been applied for purposes additional to thinking of technologies as situated in a complex sociotechnological web. The research is cross-disciplinary, in that it draws on various knowledge fields and subfields (as mentioned above); an effort has been made to develop methods that generalize beyond primary goals (see in particular the concept of structural information potential); equal importance is given to theoretical and applicative contributions (e.g., structural information potential has various practical applications, as well as being innovative from the points of view of image processing, information theory, and graph theory); and the extent of conceivable applications is deliberately and systematically explored (e.g., a dozen applications have been inventoried for the Document Towers visualization).

Is it possible to integrate the above kaleidoscopic perspectives into a single concept? Certainly! Human visual stereopsis, and even more so the compound eyes of insects, are proofs of not only the possibility, but also the benefits of multiperspectival approaches to problem-solving. The final title chosen for this work, "Human Factors in Document Processing", stands as a good compromise. While retaining the reference to the myriad human-related factors that affect technologies, we use the term "human factors" in a much broader sense than is conveyed by its standard application to the optimization of human-machine interaction alone (which is the goal of ergonomics), in order to refer to the large spectrum of modalities in which human factors may be leveraged to advance science and technology. In the specifics of this work, these comprise research concept identification (structural informa-

tion potential is a case of knowledge transfer between graphic arts and computer science), dataset and feature selection (choosing to work with quantifications of slant, roundness, and regularity as handwriting expertise features, rather than with inscrutable neural networks), algorithm development (computational image enhancement based on visual illusions), system design (proposing multiple legibility enhancement methods to achieve personalization, rather than using the method that is most effective for the average person), and analytic methods (visualization combined with quantification and automation in document exploration). In brief, the term "human factors" is used in the double sense of the opportunities that human factors represent for computational document processing, and the necessity of taking human factors into account in the development of document processing technologies. "Document processing" is used in a broad sense, similar to text and image processing.

In fact, the human dimension of our technical work played other important roles. It drove the research by rooting it in the author's practical domain-specific experience, which became the source of pertinent (if sometimes unexpected) questions and approaches. It also opened up new research fields: each user study did more than validate our methods, with the participants' feedback resulting in discoveries that went beyond our initial goals and challenged our implicit assumptions. These were among the most exhilarating parts of our research, given the unforeseeableness of the outcomes. The quest for legibility enhancement methods, for example, led via the user evaluation to a theory of image enhancement, spawned additional publications, and unveiled the possibility of a broader research program in perceptual image processing.

Intellectual

Background

Just as we defend the view that computational document processing should not be abstracted away from the impact of human factors (and that form and function are interdependent), we consider that a judicious appreciation of the author's current works cannot be dissociated from his intellectual biography. In this spirit, we propose here a review of his past research related to the topics discussed in this thesis. For references, please consult the adjoined curriculum vitae.

The latent mindsets that permeate my work can be traced back to a fascination with the power of visual arts and fiction writing to elicit whole worlds using the simplest means – speech and pencil –, and that of the more sophisticated technology of computer programming – incidentally also belonging to the linguistic sphere – to put things in motion.

Although programming in my teens in Assembler on a Commodore 64 (inter alia, to participate in a mathematical Olympiad with a visualization of fractals), and being destined by fatherly dictate to become an engineer (by passing the entry examinations to the Faculty of Telecommunications of the University of Timișoara, Romania), I took advantage of the opportunities arising from some well-known and rather unexpected historical developments around the year 1989 to become what in past times was called an "orientalist" (i.e., to study the languages, cultures, histories, arts, etc. of realms to the East of Europe, in my case the Middle East), and alternate between study years in France (at the University of Provence, Aix-en-Provence, and the École pratique des Hautes Études, Paris), and research "in the field" (Egypt, Syria, Turkey, and Iran). During this period, I learned Arabic-script calligraphy from masters in Istanbul and Damascus; my production of scientific publications led to practicing typography and graphic design, at times professionally.

I graduated with a thesis in linguistics, consisting in a statistical analysis of lexicographical density that identified correlations between sounds and semantic fields in Arabic (the onomatopoeic thesis of language evolution is ancient, but the novelty was in the quantitative, systematic, and empirical approach). Joining lexicology with calligraphy, my transition from the spoken to the written form of language took place via research (published as a book) on the influence of letter frequencies on the evolution of Arabic script. The PhD thesis was an erudite exercise showing the mutual shaping of calligraphy and sociocultural aspects during the late Middle Ages in Mamluk Egypt and Syria, Mongol Persia, and Ottoman Turkey (here, the contribution was to integrate two visions of paleography: one focused on historical aspects, and another dealing with the morphological description, classification, and expertise of handwriting). This work was partially based on various computational handwriting analyses, including the study of the contextual use of allographs for script expertise, restoration and typographic styling, and drew on explanations of handwriting variability from neurocognitive research. I also freelanced as a programmer, e.g., developing the first image processing software for the expertise of paper structures, along with the first Java Script-based database of dated Middle Eastern manuscripts, and producing scientific transliteration fonts of hieroglyphs and other ancient scripts.

I pursued my interest in psychology by taking visual perception and

somatosensory classes during a postdoc at the Massachusetts Institute of Technology (MIT), in parallel with a class on visualization in the field of science and technology studies (in addition to classes on Japanese language and anime), which, taken together, created a logical link between visual design, technology, and human factors. I was introduced to the research of low vision during an internship in New York at the Lighthouse Guild organization for the rehabilitation of people with vision impairments.

Subsequently, at the Austrian Academy of Sciences, Vienna, I co-coordinated a European project on image-based paper expertise and history, during which I also developed online and offline applications in the domain of geographical information systems (GIS) for the visual analysis of historical data. The GIS work was continued with a commission by the National Scientific Research Center (CNRS) of France, Paris, to develop the front- and backend of the online Atlas of Iran. At the University of Graz, Austria, I taught information visualization for the humanities.

I returned to script processing, first by collaborating with forensic document experts on a project proposal for an integrated European system for the visualization of the geographical distribution of handwriting variability, and then at the École Nationale Supérieure de Télécommunication (ENST), Paris, with research on writer retrieval, before starting a PhD in computer science at the University of Fribourg, Switzerland, on layout analysis and visualization. Here, I also taught information visualization for scientists, and produced a book on the technological and sociocultural state of the art of computational expertise in forensic documents. My work on the legibility enhancement of papyri is a collaboration with the University of Basel, Switzerland.

Expertise

The body of work presented in this section deals with some underdeveloped topics in handwriting expertise, namely understandable methods, datasets overview, and misclassification, in addition to a survey of the profession.

UNDERSTANDABLE METHODS

P1.
Writer Retrieval: Exploration of a Novel Biometric Scenario Using Perceptual Features Derived from Script Orientation

Based on experience in paleography and familiarity with forensic document expertise, we made the simple observation that the features used for the expertise of handwriting by experts and computer scientists are differing substantially, and that computational handwriting expertise systems are often not sufficiently understood by the end users. With the recent increased adoption of methods based on neural networks, the issue of explainable methods has also become a concern for the computer scientists themselves. The lack of understandable methods results in many otherwise promising research efforts not making it into usable applications, as well as in applications that are difficult to trust; this is problematic for scholarly research, and even more so for defending handwriting expertise in courts of law.

Inspired by the concept of "image retrieval" from a database, we devised the "writer retrieval" handwriting expertise scenario as a means of experimenting with the issue of understandable methods. Writer retrieval is the task of retrieving from a set of handwritten documents all those written by a specific writer. We then took a classic morphological feature to quantitatively characterize handwriting, namely the probability density function (PDF) of the writing contour orientation. The novel step consisted in showing that various statistical properties of the orientation's PDF have perceptual correlates, which are thus not only intuitive to experts, but also established descriptors in their methodological "toolboxes". For example, the PDF mode is indicative of the overall script slant, the entropy of the PDF vector may be used to characterize script roundness, and the PDF variance reflects writing regularity. We experimentally demonstrated the effectiveness of these perceptual-analytic features in the context of the writer retrieval scenario, and using a well-curated collection of contemporary Swiss handwriting samples, the open-source IAM dataset [15]. We have shown that it is possible to select expertise features that are both mathematically sophisticated and intuitively comprehensible. Cross-disciplinary compatibility has, in this case, been achieved by a concerted cross-disciplinary effort to develop a conceptual translation interface between disciplines.

DATASET OVERVIEW

P2.
Forensic vs. Computing writing features as seen by Rex, the intuitive document retriever

P3.
Project presentation: Rex Online

One characteristic of handwriting experts, such as paleographers and forensic document examiners, is their thorough and lengthy examination of datasets; this results in a data-centric knowledge that is lacking from the typical education of computer scientists, who are more likely to develop a methods-centered domain knowledge. To address and experiment with this issue, we built a handwriting dataset visualizer, Rex, using the previously developed perceptual-analytic features. The demonstrator allows users to browse datasets and visualize individual handwriting samples, along with graphic representations of their analytical features. It is also a tool to think with about possible future

developments. For example, to account for differences within datasets, tasks, and users, it may be useful to allow for multiple overviews and multiple data search modalities: visual (using reference images), linguistic (describing script features), haptic (by drawing), and semantic (from metadata descriptions).

MISCLASSIFICATION

P4.
Cuckoos among Your Data: A Quality Control Method to Retrieve Mislabeled Writer Identities from Handwriting Datasets

Our claims regarding the utility for expertise of dataset knowledge and overview were serendipitously substantiated by the application of our visualization tool to the evaluation dataset IAM, in which we discovered through interactive exploration that a few samples were labeled with incorrect writer identifiers. Misclassification affects the quality of datasets, and could have real-world repercussions for biometric and forensic expertise in the absence of proper quality controls. In fact, misclassification research is purportedly the most important topic in pattern recognition [10]. Despite its relevance, however, the topic remains practically unexplored in the area of handwritten documents, barring a few publications.

Using the same orientation feature and dataset of writer identifiers as in the previous works, we set out to investigate misclassification, from both a generic perspective and in the specific context of handwriting, in order to detect samples attributed to the wrong writers. To this end, we developed a dozen methods to compare the orientation PDF vectors of individual handwriting samples (the frequency domain and ensemble methods performed best) and rank them by similarity, with the goal of determining the best way to cluster misclassifications (called "cuckoos") in the top ranks. Based on these experiments, our first practical contribution is an algorithmic method for misclassification detection in handwriting samples, based on a particular handwriting descriptor.

Our second contribution is a theoretical investigation into both the domain-specific and generic causes of data misclassification, and the development of methods for misclassification quality control. We empirically create a typology of handwriting variability factors and a generic typology of misclassification detectability. This enables us to distinguish between misclassification factors at the item level (the handwriting sample) and those arising from the assemblage of items into a structure (the dataset). We then extend the findings to show that quality control performance is the result of interactions between the sensitivity of measurement instruments, the data's proneness to misclassification, and use case constraints. Finally, we propose generic quantitative methods to estimate dataset misclassification proneness, along with methodological approaches for the overall performance evaluation of quality control systems.

SURVEY

P5.
Expert Bytes: Computer Expertise in Forensic Documents – Players, Needs, Resources, and Pitfalls

For researchers who are not familiar with the field of forensic documents examination (FDE) [12], but wish to start a project involving computational expertise, the author wrote a book that surveys the profession. Its point of distinctiveness is its focus on the ecosystem of variegated factors affecting the development of technologies, rather than listing existing computational methods. The reader is first introduced to the fundamentals of FDE, such as the nature of forensic documents, the goals of expertise, its subfields, and the training of experts, as well as less-discussed aspects such as stress and the gender gap. Current and emerging research topics are discussed in depth, from resource networking and practical solutions to geographical informa-

tion systems and multi-cultural expertise. One chapter is devoted to a performance assessment of the human and computational expertise, data representativeness, admissibility in court, scientific value, and the public perception of FDE. The field is also situated within the wider context of global scientific forensic research, the market of forensic services providers, and governmental support for forensics. Reviews of handwriting datasets, basic descriptors, and expertise software are found in the extensive annexes, which also include data on standards, organizations, and other information resources.

Legibility

The possibility of script expertise, and even more so reading, depends on the ability to adequately distinguish text from background. Notably, this is not possible for all documents, especially historical documents. Our research addresses different aspects of legibility enhancement: the need for personalization, the use of color and perceptual image processing, the interaction between color and shape, and the limits of legibility optimization.

PERSONALIZED LEGIBILITY ENHANCEMENT

P6.
Personalizing image enhancement for critical visual tasks: improved legibility of papyri using color processing and visual illusions

Given that one main source on the history of Antiquity is constituted of papyri, and considering the quasi-absence of appropriate computational tools for their digital image editing, we developed methods for enhancing the legibility of papyri imaged in the visible light spectrum (the most common of image types) [18]. However, our work also makes broader generic contributions to the field of image enhancement, and is some of our most multidisciplinary.

Based on an empirical user evaluation of our methods and existing ones, along with a detailed analysis of factors contributing to the results, we found that there is no single optimal enhancement method, that different methods provide complementary information, and that the choice of methods varies with the nature of the imaged scene, the characteristics of the imaging and display devices, and the physical environments, tasks, and users. As a measure of the staggering combinatorial variability of these systemic configurations, we note that with regard to user factors, one variability dimension that we explicitly encountered was gender. Our conclusion, implemented in the enhancement software, was that users should be offered a range of image enhancement methods, from which they can freely choose, and be able to combine insights from one or several processed images. In other words, we promote personalized, contextualized, and complementary image enhancement, similarly to how reading glasses are fitted to individual needs. This outlook contrasts with the typical attitude in image enhancement research of selecting the method best suited to the average person.

In terms of enhancement methods, we relied on color and perceptual image processing, for example, by increasing chromatic contrast (i.e., saturation). The utility of this procedure is that it is applied in addition to increasing achromatic contrast (i.e., stretching the dynamic range of the lightness image channel), and that lightness and saturation are mathematically orthogonal dimensions in the CIELAB color space. Furthermore, several perceptual and sociological aspects are leveraged to make this computational enhancement method successful: on one hand, males are more sensitive to achromatic spatial variation of stimuli, while females are more sensitive to chromatic differences [1; 8]; on the other hand, females are today prevalent in papyrology, our application domain. Despite their patent desirability, color and perceptual image processing are not mainstream methods in computer vision curricula, demanding specialization in the broad fields of color science and visual psychology. An implicit aim of our work is to showcase the potentialities of cross-disciplinary research in computer science.

At the theoretical level, our contribution is a paradigm of image en-

hancement for “critical vision”, defined in opposition to “casual vision”, for which the interpretation of images has serious implications; one needs to remain skeptical of these interpretations, not take them as final, and consider multiple ones concomitantly. The two vision tasks command different approaches to image enhancement, e.g., whether to consider, or not, the average method as the best, noise as potential information, uncertainty as needing to be made explicit, and approach enhancement as a systems problem. The theory we elaborated goes beyond explaining the choices in our specific application case, and should act as a general guide to the development and use of image enhancement applications for critical vision. In this respect, we point out the need for explicit theories of image enhancement, in addition to the mathematical and algorithmic apparatus, while developing such applications. The concern with a clear epistemological foundation also permeates several of our other works.

Our open-source software has already been successfully used for enhancing the papyri of the online catalog of the University of Basel, Switzerland, and we have received numerous commendations for our work from papyrologists worldwide [11].

SHAPE AND COLOR INTERACTION

P7.
The Utility of Color for Shape Analysis

This work makes two contributions in pursuit of its two goals: it provides a compact, textbook-like introduction to color science (which is still not well known to many paleographers and script-processing computer scientists), by way of demonstrating the utility of color for shape analysis (which is not necessarily intuitive). The exposition covers the entire image formation process, from physics, imaging, color management and processing, to psychology, culture, and the environment. Each stage is presented via an illustrated case study: e.g., color physics by the advantage of cross-spectral coloring on image enhancement, computational color processing via the use of color-based shadow removal for improved shape segmentation, and the role of culture through differences in color naming, along with its impact on shape classification and machine learning.

LIMITS OF LEGIBILITY ENHANCEMENT

P8.
Ugraphia: The Pursuit of Perfect Legibility

This work is an encyclopedic essay on the socio-cultural history of legibility and the efforts to optimize it. It investigates the limits of legibility optimization in view of the prolonged toil of many scribes, type designers, psychologists, and engineers, that seem sometimes too optimistically oblivious of the many varied and severe constraints to which legibility is subjected. More than two dozen such constraints are investigated, among them epistemology (it is difficult to recognize the perfectly legible script when it is unclear how legibility should be measured), power (legibility is the effectiveness of social control and command exerted by way of normative pattern production, while illegibility is a defense mechanism), entropy (shape distinctiveness increases legibility), kinematics (scripts are optimized for production facility, not solely visual appearance), and personalization (legibility is optimal when tuned to a specific reader). The study concludes that little progress has been made throughout the long-term evolution of the legibility of handwriting, but that there has been ample optimization of artificially produced writing, as well as potential improvements in readers’ performance due to script proliferation and diversification.

Visualization

With the works reported in this section we transition at many levels from our previous research: from script to layout, from images to electronic documents, from computational to visual analysis, and from image processing to design. Because of the dual computational and human essence of graphical user interfaces, the disciplinary requirements also shift from technical to a more verbose emphasis on theories and user evaluations.

THE DOCUMENT-AS-BUILDING METAPHOR

P9.
Document Towers:
A MATLAB software implementing a three-dimensional architectural paradigm for the visual exploration of digital documents and libraries

Today's digital documents cannot be browsed with the same ease as flipping through a physical book, and this problem is only compounded for digital libraries, whose contents remain basically invisible to the user beyond the top-ranking titles of a textual list generated by a search engine. Aside from ergonomics, the concept of exploration itself is eluded by contemporary information systems in favor of targeted searches (one will find only what one is looking for) and language-based interactions (the ability to express oneself in specific words is required). Such paradigms restrain the acquisition of novel information and the diversity of cognitive modalities of information seeking [9].

At the same time, digital document files are rich in spatial information on the position of visual elements. Putting together the above observations, we realized that it is possible to represent the spatial structure of documents in a manner that recalls architectural models, and solve the dual problems of exploration and visualization. We called the resulting paradigm the Document Towers.

The implementation of the Document Towers is generic with respect to representable information and file formats, as long as coordinates are supplied. In practice, our software supports some of the most common document formats for general purpose (PDF), graphic design (InDesign IDML), and digitization (Alto); therefore, the visualized objects are page boundaries, text paragraphs, bitmap images, and vector graphics. Measurements may be performed on the collection of resulting bounding boxes (e.g., quantity, object-to-page-area ratio) for further automated or human appraisal. The quantitative description of patterns of spatial distribution, along with their link to the perceived informativeness of document pages, are a research topic pursued in our work on structural information.

A CORNUCOPIA OF APPLICATIONS

P10.
Applications of the Document Towers method of representing document structures

Some typical benefits of the Document Towers include improved document overview, navigation, and comparison, detection of misclassification and hidden objects, quality control, and information fusion. As we discovered, the range of insights is, however, both staggering and unforeseeable. The rendering media may be on-screen, and allow interactions with the data (e.g., selection, zooming, rotating, transparency, linking visualization and document), or be printed, and thus extend the persistence and available real estate of the visualization; information may be accessed through multiple sensory modalities, such as by engraving the representation on a hard substrate for the benefit of persons with low vision; the representation may include information

ranging from paragraph location to the distribution within the document of the emotions described in paragraphs; the potential tasks cover the entire document lifecycle, from conception and design to digitization, reading, and forensics; while the prospective users might potentially include all the stakeholders in the preceding tasks. The Document Towers are indeed a generic visualization technique for paginated documents, just as the histogram is a representation of scalars independent of their meaning. It is not that some of the other representations similar to the Document Towers lack these properties, but rather that we made a deliberate and systematic effort to identify the broadest band of potential applications for the interested technology developers and end users.

EVALUATIONS AND SURPRISES

P11.
Improving the Quality Control and Exploration of Digital Documents Using Visualization and Serendipity

The performance evaluation of the Document Towers as a tool for document exploration led to a number of unintended findings, which cast it as an effective albeit serendipitous conduit for scientific discovery. Our qualitative case study, which tested the utility of the Document Towers for quality control of digitization workflows and was carried out on a dataset of historical Swiss newspapers, resulted in the discovery of the first military map in a Swiss newspaper, missing tax stamps in the layout analysis, and a lack of labeling in the digital documents of censored paragraphs. These findings have value for historians who use digital libraries, mark weaknesses in digitization requirements, and point to future research areas; more generally, they reframe quality control as a dialog between different stakeholders, rather than in narrow technical terms; furthermore, they underscore the role of serendipity in knowledge-seeking, and illustrate the personal variability of insights obtained from information systems.

We sought to replicate, assess, and generalize these findings using other evaluation methods. A thought experiment suggested that under certain conditions, user experiments (contrary to common engineering expectations) are not necessary, or may even be impossible. In support of such cases, we developed a diagrammatic method to perform an effort/reward analysis of serendipitous discoveries. We finally carried out a user satisfaction and psychometric evaluation of the Document Towers, which demonstrated sizable usability variation across socio-professional contexts, namely, better usability was reported by academics than by librarians. These results dovetail with our work on legibility enhancement, as they demonstrate from different perspectives the need for personalized, contextualized, and complementary technologies to address the great variability in human factors.

THE ONEIRIC DIMENSION OF INFORMATION SYSTEMS

P10.
Annex: Dreams, Architecture, and Information Systems: A brief review of the utility of document exploration, visualization, and metaphors

The concepts at work in the Document Towers have a rich history within and outside and before computer science. Visual analysis, as distinct from numerical analysis, has been a driver of invention from Euclidean geometry to the fractals, and metaphors of information-as-{place, document, architecture} have been cognitive tools from times immemorial up to the era of cyberspace and information architecture systems. The question of exploration vs. search, coupled with that of serendipity, has been debated by librarians since libraries first existed, and are linked with societal issues of freedom of knowledge [16; 28]. We have reviewed these disparate strands, and explained how their intersection is relevant to computational information systems.

But what are the epistemic foundations of technologies that implement metaphor-aided visual exploration information strategies, and why do

they work (or fail)? Referring to the Document Towers, we note some of their characteristics: agnosticism, parsimony, spatialization, holism, and even playfulness (Document Towers look like dollhouses), and mystification (fuzzy images call for close investigation). Overall, such information systems have a hedonic component, eliciting a mental state analog to web surfing: that is, procuring mental flow, a creative state primed to seize upon the unexpected and create a rich net of links – in short, dreamlike [6].

Since many of these technologies remained experimental, we may ask if they are not only fascinating, but also impractical. We consider that if they have failed as everyday methods of information, they continue to represent visions of possible futures that have a motivational role as important as their actual realization. The ultimate value of Vannevar Bush's memex document-processing machine, to cite an example from before the development of digital computers, surpasses its technical achievements, and consists in contributing to the ancient collective dream of intellectual communion (Bush's World War Two war-related objectives aside) [3].

CONTRIBUTIONS

Each research stage of the Document Towers contributes in specific ways to document processing, and may be expressed as follows: detection of a need (visual exploration of digital documents) and seizing of an opportunity (present mass-digitization, and availability of spatial coordinates in digital documents) to develop a solution (the Document Towers), whose design is refined for optimal affordance (using visual simplicity and the metaphor of the document-as-architectural-model), which is evaluated under various points of view (qualitative case study, thought experiment, diagrammatics, and user satisfaction and psychometrics), in which emphasis is placed on its epistemological tenets (agnosticism, essentialism, spatialization, mystification, holism, and enaction) and the great range of potential applications (from document design, quality control, and classification, to improving navigation and access, and creating art), and comparison with the state of the art (on exploration, serendipity, visualization, and metaphors). The integrated value of these elements in turn enhances the distinct and original value of this research.

Information

The work presented in this section continues our research on document layout analysis with a different approach, based on the quantitative analysis of raster images, and a specific layout feature, its "structural information potential". In the author's view, it also represents some of the most far-reaching aspects of our research.

P12.
The Structural Information Potential and Its Application to Document Triage

THE STRUCTURAL INFORMATION POTENTIAL

Structural information potential (SIP) is a measure of information based on the potential of structures to be informative about their content. An example of this concept is the clustered appearance that typically characterizes the first page of scientific articles, which summarizes the article's contents and provides additional data (such as author names, keywords, publication references, and so forth), yielding potentially the largest and most diverse amount of information from a single page in the shortest time with the least effort. This characteristic makes SIP particularly well-adapted to triage tasks (i.e., rapid decision-making under conditions of uncertainty and limited resources), an application illustrated by means of a real-world case study on classifying document images, and that is validated theoretically and experimentally. The SIP method consists in combining aspects of the Shannon entropy, the Fourier transform, the fractal dimension, and the golden ratio into a single equation and several algorithmic components. While the application domain is document images, the concept has a generic character. The method results in a mathematically and perceptually coherent pattern space, characterized by continuous transition between uniform, clustered, random, and regular configurations, and corresponding to a structural information potential with a well-defined maximum. The maximum SIP leads to the identification of shapes and patterns with minimal structural redundancy, termed "fluorescent objects" within the concept of chromatic graphs, as a complement to regular graphs and the Platonic solids [4; 5].

The main contribution of this work is the SIP measurement method, which offers a practical solution to a long-standing problem encountered in a great variety of fields, including information theory, spectral analysis, graph theory, and combinatorics, to name only a few of those surveyed. The newly defined pattern phase space has desirable qualities in terms of informativeness and perceptual affordability, making it relevant to domains of application that are commensurately broad, extending from the analysis of paginated documents to sounds, videos, and three-dimensional objects, and expected to encompass telecommunication, physics, and psychology – in short, everywhere there is structure. In terms of the overarching theme of our work, namely human factors in document processing, SIP is a good example of how the development of a formalism (the measurement method) may be motivated and driven by empirical experience (in graphic design), and preoccupation with psychological factors (the pattern phase space). SIP also has a mathematically aesthetic dimension that sets it apart, in that it fuses distinct major concepts into a single equation from which emanates a spectrum of applications.

Outlook

Let us consider some of our future research directions imparted by the work presented here.

We are currently engaged at the University of Bern, Switzerland, in the project "Meaning is Action", which translates us from the study of forms and information to that of meaning. Specifically, we define meaning as the action to which an entity may lead – in other words, its perceived utility, or "action potential" (to use an analog to "information potential"). We intend to develop this concept into a machine learning method within the framework of reinforcement learning. From this perspective, a document is a system made up of various functions that produce some action (e.g., the title summarizes, the index provides links, and the sentences impart information and emotions); we may say that a document is a "meaning-making machine". Our goals are to provide a formalism to the fundamental problem of meaning in linguistics and computer science, as well as to find alternatives to the insatiable need for data in currently dominant machine learning paradigms. At the crossroads of document and natural language processing, our case study consists in labeling the 150-years-rich digitized archive of Swiss phone books according to entity names (person, profession, address, telephone number) in view of an online historical directory [19].

In terms of the human factors theme, we aim to systematically explore the leveraging of visual psychology for image processing, particularly with regard to the impact of individual and cultural variability on technology design. A related research strand is that of "culturally inspired technology", which consists in deliberately developing procedures and artifacts by observing cultural phenomena, analogous to bio-inspired robotics [22]. We may use as experimental material the rich and well-studied worldwide body of children's stories, which serve as ready-made behavioral algorithms, and ask if something like the shrewdness of Till Eulenspiegel may not be converted into image processing techniques.

A number of spin-off projects were started during the thesis and should be completed. One is an example of "spurious information potential", and one of the last microtypography issues in document engineering: the detection and measurement of vertical alignments of spaces that produce meaningless streaks in the page texture and detract from the reading experience. Another is a pattern catalog for scaling visualizations, a perennial issue in the domain of information representation due to limited display space, and related to the difficulties of overviewing large image datasets.

Finally, we wish to see the effort and time invested in our research fulfill its potential by being implemented in products and used by people. This relates in particular to structural information potential and the Document Towers, which were conceived from the outset with concrete applications and users in mind. Ideally, we would like to see these methods implemented in e-readers, and believe that they would serve persons with low vision well.

Contributions

The concrete outputs of this work are computational methods and software packages that address specific practical issues in the field of document processing, concerning expertise, legibility, informativeness, and visualization. Some of the tools have already been used with success (e.g., for editing difficult-to-read papyri).

The work contributes methodologically to computer science, by promoting the general concepts exposed above, in particular for integrating human and other contextual factors in technology design, and pursuing innovation through inter-disciplinarity. These are well-established notions, but there is still much to be gained from them in the fields of our works.

During research, it was repeatedly observed that the relevant literature lacks a theoretical framework (other than computational) for the developed solutions. For example, legibility enhancement methods are developed without a clear understanding of what enhancement means, for whom, what for, and in which contexts, and what the integrated impact of these technological, psychological, social, and other factors is on methods and evaluation datasets. The result is that the field still struggles to comprehend why there is no optimal image enhancement method, which is in itself a complex notion. To counter such problems, we grounded our methods in the wider contexts of their applications, which led to the development of elaborate sets of principles that guided our research. As a case in point, this approach enabled us to make the operationally important distinction between casual and critical computer vision, as well as to advocate personalized image processing.

While each publication deals with a specific problem, the research consistently resulted in spin-off projects that are valuable in themselves. Examples include the statistical evaluation methods that we had to develop ourselves to meet our specific needs: e.g., a measure of order disturbance (in the structural information potential article), a method to assess dataset scrambling severity (in the publication on the misclassification of handwriting samples), and the use of a frequency domain method borrowed from signal processing and applied to descriptive statistics (in the legibility enhancement article).

*

We may assess the presented work with respect to how well it addresses specific issues, but we could also consider its merits on the basis of the questions it raises and the doors it opens, namely its generative potential.

There is much in the concept of structural information potential to explore in terms of its generalization to media formats and applications, and regarding objects with minimal redundancy of arbitrary dimensions and structure. The Document Towers beseech to be implemented in a variety of technologies. The questions of personalization and contextualization, along with the leveraging of psychological effects in the conception of computational methods, have only been “scratched at” in the work on legibility. The role of perceptual–mathematical correlates in scientific communication highlighted in the work on script expertise is also worth pursuing further.

In summary, a general characteristic of this work is that while the initial research foci are well delimited, they open up to broad application fields. Beyond the metaphors of open questions and open doors, the generative potential of our work also lies in cross-disciplinary bridge-building, specifically by delving deeper into the link between technologies and human factors.

¶

Selected

References

1. Abramov Israel, James Gordon, Olga Feldman, Alla Chavarga [2012] – "Sex and vision II: color appearance of monochromatic lights", Biology of Sex Differences, 3: article 21, 15 p., https://doi.org/10.1186/2042-6410-3-21.

2. Beghdadi A., M.-C. Larabi, A. Bouzerdoum, K. M. Iftekharuddin [2013] – "A Survey of Perceptual Image Processing Methods", Signal Processing: Image Communication, 28: 811–831.

3. Burke Colin B. [2018] – America's Information Wars: The Untold Story of Information Systems in America's Conflicts and Politics from World War II to the Internet Age, Lanham, MD: Rowman & Littlefield.

4. Chartrand Garry, Paul Erdős, Ortrud R. Oellermann [1988] – "How to Define an Irregular Graph", The College Mathematics Journal, [19] 1: 36–42.

5. Chartrand Garry, Ping Zhang [2009] – Chromatic Graph Theory, Boca Raton, FL: CRC Press.

6. Csikszentmihalyi Mihaly [2014] – Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi, Dordrecht: Springer.

7. Dillon Andrew [2004] – Designing Usable Electronic Text: Ergonomic Aspects of Human Information Usage, Boca Raton, FL: CRC Press, 2nd ed.

8. Elliot Andrew J., Mark D. Fairchild, Anna Franklin [eds.] [2015] – Handbook of Color Psychology, Cambridge: Cambridge University Press.

9. Fisher Karen E., Sanda Erdelez, Lynne E. F. McKechnie [eds.] [2006] – Theories of Information Behavior, Medford, NJ: Information Today.

10. Hand D. J. [1986] – "Recent advances in error rate estimation", Pattern Recognition Letters, 4 [5]: 335–346.

11. HIERAX: Software for Enhancing the Legibility of Papyri, https://hierax.ch.

12. Huber Roy A., A. M. Headrick [1999] – Handwriting Identification: Facts & Fundamentals, Boca Raton, FL: CRC Press.

13. Journal of Perceptual Imaging, http://www.imaging.org/site/IST/IST/Publications/JPI.aspx.

14. Le Moan Steven, Farup Ivan, Blahová Jana [2018] – "Towards exploiting change blindness for image processing". Journal of Visual Communication and Image Representation, 54: 31–38.

15. Marti Urs-Viktor, Horst Bunke [2002] – "The IAM-database: an English sentence database for off-line handwriting recognition", International Journal on Document Analysis and Recognition, 5 [1]: 39–46. Online dataset, https://fki.tic.heia-fr.ch/databases/iam-handwriting-database.

16. McCay-Peet Lori, Elaine G. Toms [2018] – Researching serendipity in

Digital Information Environments, San Rafael, CA: Morgan & Claypool, https://doi.org/10.1007/978-3-031-02312-5.

17. Pettersson Rune (2016) – Information Design Theories. Tullinge: Institute for infology.

18. Piquette Kathryn E. (2018) – "Revealing the material world of ancient writing: digital techniques and theoretical considerations", in Francisca A. J. Hoogendijk, Steffie M. T. van Gompel (eds.), The Materiality of Texts from Ancient Egypt: New Approaches to the Study of Textual Material from the Early Pharaonic to the Late Antique Period, Leiden: Brill, pp. 94–118.

19. PTT-Archiv, https://www.mfk.ch/forschen/ptt-archiv/recherche.

20. Richaudeau François (1969) – La lisibilité: Langage, typographie, signes... lecture, Paris: Gauthier.

21. Salvendy Gavriel, Waldemar Karwowski (2021) – Handbook of Human Factors and Ergonomics, Hoboken, NJ: Wiley, 5th ed.

22. Stoll Wilfried (ed.) (2012) – Bionik: Impulsgeber der Technik, Mainz: Verlag der Universitätsdruckerei Hermann Schmidt.

23. Taylor Richard P., Branka Spehar, Paul Van Donkelaar, Caroline M. Hagerhall (2011) – "Perceptual and physiological responses to Jackson Pollock's fractals", Frontiers in Human Neuroscience, 5, art. 60, 13 p, https://doi.org/10.3389/fnhum.2011.00060.

24. Tschichold Jan (1995) – The New Typography: A Handbook for Modern Designers, Berkeley: University of California Press.

25. Unger Gerard (1981) – "Experimental No. 223, a newspaper typeface, designed by W. A. Dwiggins", Quaerendo, 11 (4): 302–324.

26. Van Tonder Gert J., Michael J. Lyons (2002) – "Visual structure of a Japanese Zen garden", Nature, 419: 359–360.

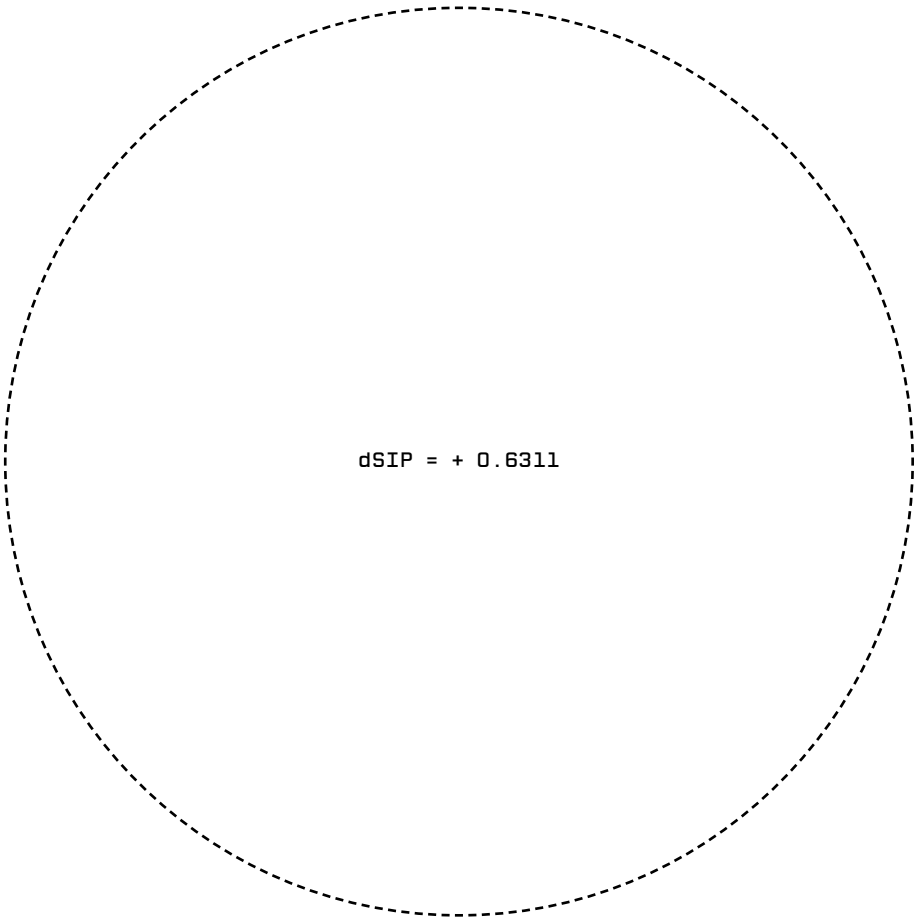
27. Voss Richard F., John Clarke (1978) – "'1/f noise' in music: Music from 1/f noise", Journal of the Acoustical Society of America, 63 (1): 258–263.

28. Wertheim Margaret (2000) – The Pearly Gates of Cyberspace: A History of Space from Dante to the Internet, New York, NY: W. W. Norton.

Publication the First

Copyright © 2011 IEEE. Reprinted, with permission, from Vlad Atanasiu, Laurence Likforman-Sulem, Nicole Vincent [2011] - Writer Retrieval: Exploration of a Novel Biometric Scenario Using Perceptual Features Derived from Script Orientation, 11th International Conference on Document Analysis and Recognition, 18-21 September 2011, Beijing, China, pp. 628-632, https://doi.org/10.1109/ICDAR.2011.132.

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of University of Fribourg's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.



Proc. 11th Intl. Conf. on Document Analysis and Recognition, Beijing, China, September 18–21, 2011 © IEEE

Writer Retrieval—Exploration of a Novel Biometric Scenario Using Perceptual Features Derived from Script Orientation

Vlad Atanasiu
Signal and Image Processing Dept.
Institut Télécom; Télécom ParisTech
Paris, France
atanasiu@alum.mit.edu

Laurence Likforman-Sulem
Institut Télécom; Télécom ParisTech;
CNRS LTCI
Paris, France
likforman@telecom-paristech.fr

Nicole Vincent
Laboratoire LIPADE
Université Paris Descartes
Paris, France
nicole.vincent@mi.parisdescartes.fr

Abstract—We propose a novel scenario called "writer retrieval" consisting in the retrieval from a set of documents all those produced by the same writer. The retrieval is based on a set of ten features correlated to perceived characteristics of the writing (orientation, regularity…). These features are extracted from the probability density function of the orientations of the writing contour. We evaluate the overall efficiency of each feature and the relative importance of features according to queries. Evaluation is conducted on the publicly available IAM database.

Keywords—writer retrieval; local shape orientation; perception of handwriting; handwritten documents; IAM database

I. Background, Definition and Study Goals

(Hand)writing analysis finds many applications in various areas such as forensic sciences, paleography or recognition. Authenticating or identifying a writer from script samples is the core of handwriting forensics, presumably since the invention of writing itself [15]. Establishing the style of modern or old documents is a basic paleographic endeavor and typographic obsession the world over [8, 16, 29]. Clustering handwritings can optimize channeling of documents towards dedicated style-based recognizers [9]. Retrieving documents by appearance is helping navigating the human written memory [6, 28]. We present here a novel scenario related to handwriting analysis, namely the "writer retrieval" scenario.

"Writer retrieval" is the task of retrieving from a set of handwritten documents all those written by a specific writer. It is neither "writer identification," because the task output is documents, not names of writers; nor is it "document retrieval," because the classification criterion is the writers' names—writer retrieval is a hybrid.

Fig. 1 shows the progressive relationship between common handwriting processing tasks and reveals them as instances of more general classification tasks. To our best knowledge "writer retrieval" is used in the sense defined here for the first time in the specialty literature.

Possibly one of its most exciting applications is for archival research: archives have vast amounts of documents, large portions thereof are summarily classified or not at all, and often researchers are looking for documents produced by a specific writer. Historians could use writer retrieval to hunt for rare documents of famous individuals. Forensic experts need writer retrieval if their document repositories lack comprehensive writer classification metadata, or if several handwriting datasets are to be fused or networked. Needless to recall that the interest for the writer retrieval scenario scales up with the ever increasing amount of handwritten documents becoming available in digital format.

The goals of this preliminary study are structuring the paper as follows. After having explored in Section I the nature of writing retrieval, we choose in Section II a script property and a measurement instrument on the basis of which we can test writing retrieval; in Section III we investigate the appropriateness for this classification task of statistics of a probability density function (pdf) rather than the pdf values; in Section IV we identify in the computed contour orientations measurements some correlates natural to human perception and descriptive practices of writing; in Section V we develop a writer retrieval method and carry out a performance test.

II. Script Property and Measurement Instrument

Property. The property selected to perform writer retrieval is the local orientation of the contour of the handwriting trace and the rationale is threefold. First, orientation has been shown to be an effective handwriting discriminator [6, 7, 26]. Second, it has a rich palette of perceptual correlates, allowing visual interpretation of the measurements, an important factor for such applications as forensics or paleography where human expertise is the norm. Third, an important corollary of orientation, "slant," enjoys eighty years of research in the psychology community [13, 19], while orientation has become popular among handwriting computer scientists [4, 5], thus warranting heightened interest.

Instrument. A fair body of research exists outside the handwriting applications fields in respect to the evaluation of the local orientation of shape contours. The main computational instruments (methods) for its measurement are: "*polygonal,*" where orientation is derived from an arc of fixed length defined over the chain-coded representation of the contour [6, 7, 22, 18]; "*scale-space,*" obtained from a progressive smoothing of the contours' coordinates [14, 21, 10]; "*gradient*" of the grayscale image at the locations of the contours, extracted with "straight"-shaped filters [11, 24] or curvelets [23, 17]; "*linear filters bank*" (Gaussian, Gabor...) with varying orientations, where the local shape orientation is given by the orientation of the filter with highest response when convolved with the shape image [9, 26].

A dilemma common to these instruments is parameterization: the polygonal arc length must be adapted to the script

P1
-
35

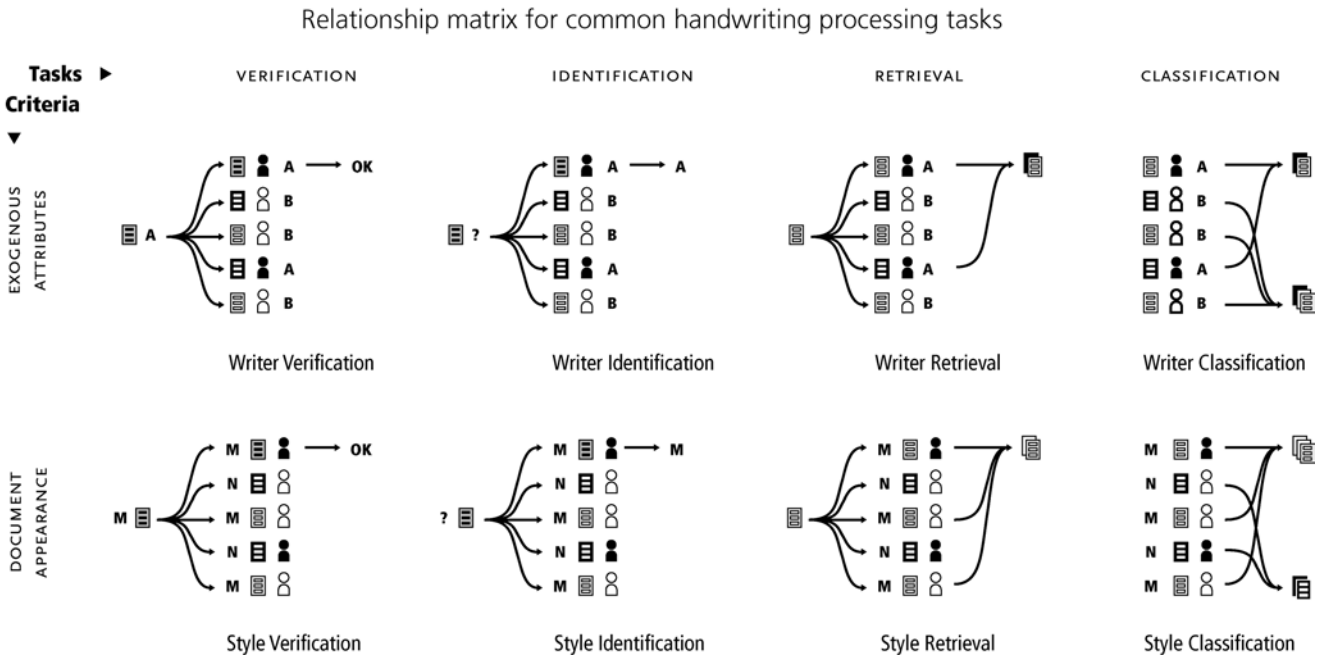


Figure 1. Common handwriting processing tasks fit in a progressive and generalizable relationship matrix. The fundamental difference is that of the number of items in the task output (one statement for verification or name for identification, a documents subset for retrieval, all documents for classification), as well as the nature of the output (logical statement for verification, identifier for identification and document for retrieval and classification). For each task type a distinction arises from whether the classification operates on the "document appearance" (as in "style identification") or on an attribute of the document exogenous to its appearance (the "writer" in the case of "writer identification"). The tasks can be generalized by choosing other classification criteria than the usual "writer" and "style," e.g. "the writer's geographic curriculum," "sex," "age," "health," "handwriting slant," or "roundness."

P1 – 36

size, when using scale-space one has to decide which scale-level to select, for the gradient approach the question is "how far does the 'local' extend?," and the linear filter has to be shaped. We too had to address this and other issues during the design phases of our orientations measurement instrument.

1. *Instrument design*. We found that a linear filter allows the use of a single filter shape for all resolutions of digital handwriting images beyond a threshold of about 25 pixels for a Latin script x-height. If necessary supersampling brings low resolution images over this threshold. In the present implementation we used a two-dimensional Gaussian filter of standard deviations 2 and 0.5, defined over a kernel of 30×30 pixels, convolved in the frequency domain with the binary image of the handwriting contour.

2. *Measurement resolution*. While it is possible to take measurements at only a few orientations, higher resolutions make possible a finer characterization of the script, as we shall see in the next section, which is particularly beneficial for writer identification and forensic expertise. Our orientations resolution is 1 degree in the [0, 179] interval.

3. *Measurement format*. Histograms are easy to generate but have fundamental and well-know instability issues arising from the choice of bin width and their centering location [25]. When carefully crafted, pdf-s are more robust—their wider use in the handwriting processing field can bring real benefits. After experimenting with several estimators for the pdf kernel, we found that the "optimal" bandwidth described in [27] is indeed optimal for our handwriting task—sufficiently smooth, while retaining detail.

4. *Visualization*. Because orientation data is circular it feels rational to present them in polar coordinates. However such diagrams distort to a human eye the represented values by underemphasizing amplitudes in the pdf shape [12]. While conventional, the Cartesian coordinates avoid this optical illusion. In line with common practice in forensics and optometry [15: 107–109], orientations are defined over the (−90, +90] degrees range, with 0 degree at the vertical. For transparency and because they offer different vantage points on the data, we included both the orientation frequency count and the pdf in the measurement visualization.

III. Computed Features and Perceptual Correlates

The orientations instrument generates a measurement vector (the pdf) that represents the observed handwriting globally. By looking at specific statistical properties of the values we gain access to individual script characteristics, whose mixture generates the global pattern. Furthermore, these derived features are more prone to have perceptual—or intuitive—correlates when compared to the source feature. For example the values of a orientations pdf taken as an undivided vector are a script feature that can't be related as such by a human subject to any characteristic of the script. Instead the pdf becomes meaningful if we consider, say, the mode and we gather that this measure is related to what is called "slant." There is also a question about the scientific expertise required from the end users, namely the capacity to understand what a given computational measure represents. "Slant" is a notion both perceptually and cognitively accessible to most observers—not so a "pdf."

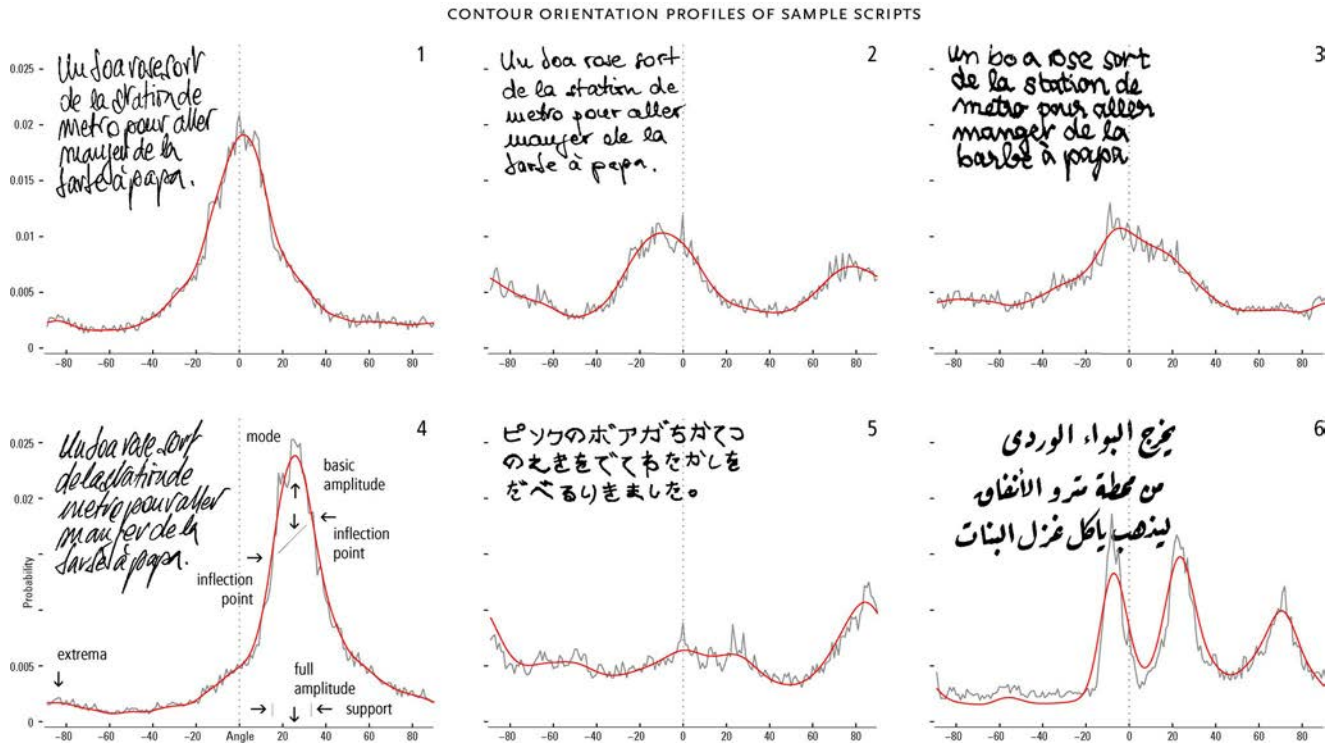


Figure 2. The diagrams show the orientation profiles of sample scripts selected to visualize the wealth of perceptual correlates of computational features. Note how a shift in mode means a difference in slant (1 vs. 4); a widening or narrowing of the support or change of entropy a variation from round to linear script (2 vs. 4); an asymmetry a variability of the slant angle (3 vs. 4); a tri-modality due to the use of a broad-tip Arabic calligraphic reed (6); a relatively flat central plot line contrasting with the extremities reinforced by the mix of curly hiragana and linear katakana Japanese scripts (5). All samples are by author V.A.; 1 & 4, respectively 2 & 3 are written at the same time; all but 3 are in usual writing styles, 3 being written by this dextral writer with his left hand. To facilitate visual comparison, pdf-s (smooth lines) and frequency counts (jagged lines) are superposed by equalizing their means and variances. Probability values (y-axis) are at the same scale in all diagrams.

Therefore we work with computed features derived from the pdf, something we call a "Swiss knife approach" since the same instrument can be used for multiple tasks. Additionally we identify perceptual correlates. Both features and correlates are listed hereafter and graphically summarized in Fig. 2. Note that most relate to landmarks on the pdf shape and depend in particular on the peak of the principal mode. Also, they mimic the moments of a distribution, but needed to be defined ad hoc because of the inherent multimodal nature of the handwriting orientations (even when assuming normal distribution good results can sometimes be obtained, as shown by [4] for handwriting recognition).

1. "*Modality*" represents the number of pdf modes and relates perceptually to the presence of multiple strong writing directions.

2. "*Mode*" refers to the principal mode and is indicative of the preferential slant direction of the script.

3. "*Support*" measures the extent of the principal mode between the inflection points of the pdf delimiting the peak (zero-crossings of the second derivative). The measure is indicative of the degree of slant variability and the density of vertical script strokes.

4. "*Full amplitude*" is the value of the pdf at the location of the mode.

5. "*Basic amplitude*" measures the height of the peak between the inflection points and the mode value.

6. The "*basic*" and 7. "*full aspect*" are the ratio between the "*basic,*" respectively the "*full amplitude*" and the support. Amplitudes and ratios change with the roundness of handwriting on the continuum "round–linear" (compare Fig. 2.1 and 2.2).

8. "*Symmetry*" is for a positive (right) skewed peak the ratio of right and left side of the support (between the location of the mode and that of the right, respectively left side inflection point), and for a negative (left) skew the minus of the ratio between left and right support segments. Symmetry shows a bias in handwriting production away from the principal slant direction.

9. "*Mass*" is the integral of the pdf between the inflection points and relates to shear, slant and stroke density.

10. "*Entropy*" is the Shannon entropy of the pdf and varies mostly with roundness and stroke density.

The reader might have understood at this point an unexpected implication of the use of perceptual correlates: such a terminology makes possible to retrieve documents by describing their appearance in natural language, removing the need for a reference image, not always available, but the usual interaction method with written documents retrieval systems [6: 107–115, 28]. To explore the behavior of a tool implementing the ideas presented in this section, Rex, an online perceptual features browser was developed [2, 3].

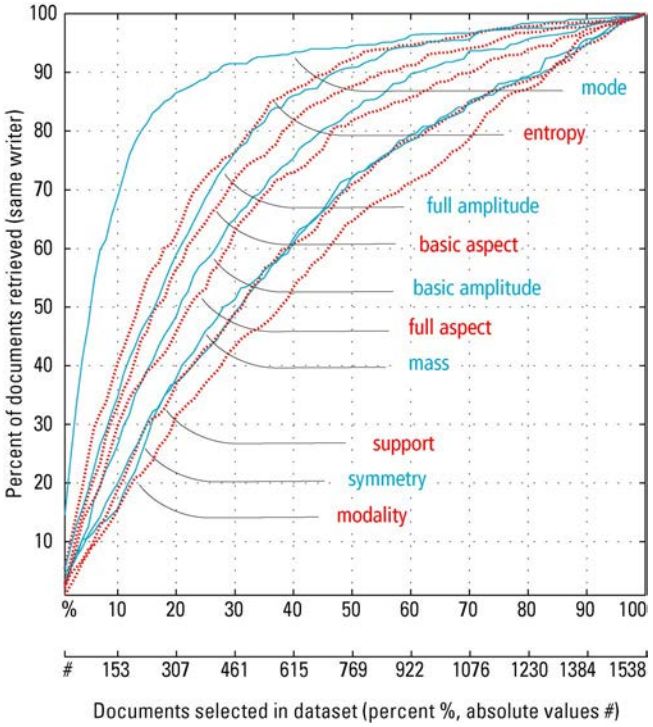


Figure 3. Writer retrieval performance according to features, and percentage of retrieved documents according to the proportion of selected documents. Selected documents are closest to the query in respect to the value of a single feature.

P1
–
38

IV. Experiments

Dataset. The IAM Handwriting Database 3.0, developed by the Research Group on Computer Vision and Artificial Intelligence at the University of Bern, Switzerland, contains 1539 document images written in English [20]. The documents contain 4–5 printed text lines, which the writers were asked to copy. There is a total of 657 distinct writers and the number of documents written by each varies from a single document to 59 documents. This dataset created initially for the purpose of handwriting recognition is also useful for our writer retrieval task. Images are clean so that extracted features are not corrupted by noise. We have used all text lines within a document, except when they overlap the forms' printed text. Such lines have been removed.

Overall efficiency of features. Figure 3 shows the overall performance of each extracted feature (see section III). In this experiment, each curve corresponds to a document feature taken in isolation for the writer retrieval task. To each writer corresponds a query document where the feature value is extracted. An increasing number of documents is selected (a percentage of the whole IAM). The selected documents are those nearest to the query, according to the feature value. We then evaluate the efficiency of the selection by counting the proportion of the same-writer documents which have been selected. The proportion is then averaged over the queries.

Curves are built using 301 queries, which corresponds to one document for each of the 301 writers which are represented by more than one document in the database. However, for each writer the search is performed on 1539 minus 1 documents. We assume that the remaining document(s) from each writer are similar to those of the document query. Thus the percentage of retrieved documents from a same writer provides us with a measure of the efficiency of each feature.

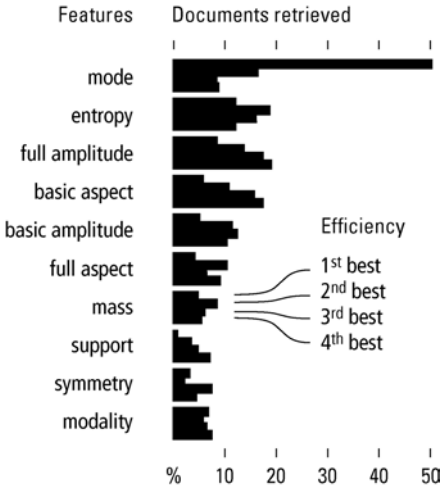


Fig. 4. Features sorted according to their efficiency on query documents.

Features can be divided into several groups according to efficiency (Fig. 4). The principal mode of the orientations pdf is the best overall feature for writer retrieval. Entropy, which takes into account all pdf values, is the second best overall feature. We also notice that the full amplitude of the pdf performs better than the basic amplitude, but the basic aspect performs better than the full aspect.

Feature efficiency according to query. The principal mode of the orientation pdf is a good overall feature for writer retrieval. We now search for a given query whether other features may be useful. In this experiment, we sort the features of a document query according to the following procedure. For each query (associated as previously with a writer) we retrieve all remaining documents of the same writer by expanding as necessarily the size of the set of closest documents. The size of the closest-document set depends on each type of feature. The feature corresponding to the smallest set size is the most efficient feature since the variance of the feature is the lowest for the writer.

Not surprisingly, the principal mode of the pdf is found to be the best feature for 50% of the query documents. However, the entropy feature and features related to the shape of the peak (basic and full amplitudes, basic aspect) are good alternate or complementary features.

The curve obtained when the best feature can be predicted for each query document is shown in Fig. 5. This curve is thus an upper-limit of the writer retrieval system using our set of ten features. Fig. 5 also shows that when using a single feature, not all documents of a writer are retrieved (unless all documents are selected). While when using several features, and choosing the best one, one could retrieve all documents of a writer by selecting about 70% of the database documents.

VII. Conclusions and Perspectives

We have proposed a novel scenario called "writer retrieval" for the selection of all documents authored by a writer.

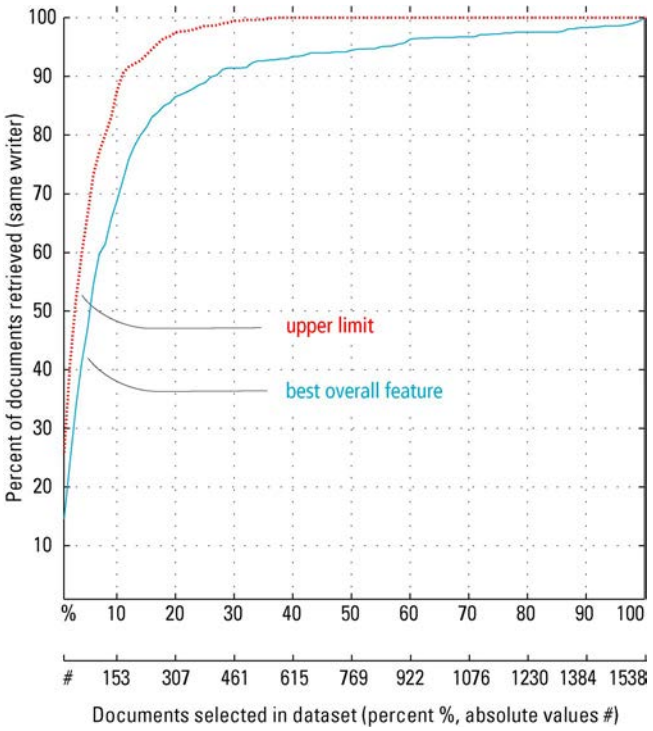


Fig. 5. Upper limit of writer retrieval performance obtained when the best feature for a query document is used. Comparison with the performance of the best overall feature (principal mode of the orientation pdf).

For this purpose we have defined ten features extracted from the orientations pdf of the script contour. We were able to correlate them to perceptual properties of the script.

We have studied the overall efficiency of each feature and their importance according to queries. We have shown the upper limit of retrieval performance obtained when using the best feature for a query. Future work will consist of finding an automatic process to select the best feature for a given query. This may be performed using pdf features for the whole writers population. Combining several features may also improve performance [7, 1].

In these experiments, documents of a single writer are retrieved along with documents from other writers. These documents often share the same script characteristics as the writer's documents, according to the chosen feature. To judge the validity of the image retrieval, future work will focus on psychophysical experiments.

References

[1] V. Atanasiu, "Allographic biometrics and behavior synthesis," *TUGboat*, vol. 24 (3), 2003, pp. 998–1002.

[2] V. Atanasiu, L. Likforman-Sulem, N. Vincent, "Rex, a description-based retriever for written documents," April 2011 [Online]. Accessible: http://glyph.telecom-paristech.fr [Accessed: June 30, 2011].

[3] V. Atanasiu, L. Likforman-Sulem, N. Vincent, "Talking Script. Retrieval of written documents by description of script features," *Gazette du Livre Medieval*, to be published.

[4] C. Bahlmann, "Directional features in online handwriting recognition," *Pattern Recognition*, vol. 39, 2006, pp. 115–125.

[5] A.A. Brink, R.M.J. Niels, R.A. van Batenburg, C.E. van den Heuvel, and L.R.B. Schomaker, "Towards robust writer verification by correcting unnatural slant," *Pattern Recog. Lett.*, vol. 32, 2011, pp. 449–457.

[6] M.L. Bulacu, "Statistical pattern recognition for automatic writer identification and verification," PhD thesis. Groningen: Artificial Intelligence Institute, University of Groningen, 2007.

[7] M.L. Bulacu and L.R.B. Schomaker, "Text-independent writer identification and verification using textural and allographic features," *IEEE Trans. on Pattern Analysis. and Machine Intelligence*, vol. 29 (4), 2007, pp. 701–717.

[8] A.-M. Christin (ed.), *History of Writing*. Paris: Flammarion, 2002.

[9] J.P. Crettez, "A set of handwriting families: style recognition," *Proc. 3rd Intl. Conf. on Doc. Analysis and Recog.*, vol. 1, 1995, pp. 489–494.

[10] I.L. Dryden and K.V. Mardia, *Statistical Shape Analysis*. Hoboken, NJ: John Wiley & Sons, 1998.

[11] H. Farid and E.P. Simoncelli, "Differentiation of discrete multidimensional signals," *IEEE Trans. Image Proc.*, vol. 13 (4), 2004, pp. 496–508.

[12] N.I. Fisher, *Statistical Analysis of Circular Data*. Cambridge, UK: Cambridge University Pres, 1996.

[13] H.R. Flock, "Three theoretical views of slant perception," *Psychological Bulletin*, vol. 62 (2), 1964, pp. 110–121.

[14] C.L. da Fontoura and R.M. Cesar Jr., *Shape Analysis and Classification: Theory and Practice*. Boca Raton, FL: CRC, 2000.

[15] R.A. Huber, A.M. Headrick, *Handwriting Identification: Facts and Fundamentals*. Boca Raton, FL: CRC, 1999.

[16] N. Journet, J.-Y. Ramel, R. Mullot and V. Eglin, "Document image characterization using a multiresolution analysis of the texture: application to old documents," *Intl. J. on Doc. Analysis and Recog.*, vol. 11 (1), 2008, pp. 9–18.

[17] G. Joutel, V. Eglin, S. Bres, and H. Emptoz, "Curvelets based queries for CBIR application in handwriting collections," *Proc. 9th Intl. Conf. on Doc. Analysis and Recog.*, vol. 2, 2007, pp. 649–653.

[18] L. J. Latecki and R. Lakämper, "Application of planar shape comparison to object retrieval in image databases," *Pattern Recog.*, vol. 35 (1), 2002, pp. 15–29.

[19] F. Maarse and A. Thomassen, "Produced and percieved writing slant: Difference between up and down strokes," *Acta Psychologica*, vol. 54, 1983, pp. 131–147.

[20] U. Marti and H. Bunke. "The IAM-database: an English sentence database for off-line handwriting recognition," *Intl. J. on Doc. Analysis and Recog.*, vol. 5, 2002, pp. 39–46. Accessible: http://www.iam.unibe.ch/fki/databases/iam-handwriting-database

[21] F. Mokhtarian and M. Bober, *Curvature Scale Space Representation: Theory, Applications and MPEG-7 Standardization*. Dordrecht, Kluwer Academic Publisher, 2003.

[22] R. Plamondon and C.M. Privitera, "The segmentation of cursive handwriting: An approach based on off-line recovery of the motor-temporal information," *IEEE Trans. Image Proc.*, vol. 8 (1), 1999, pp. 80–91.

[23] D.D–Y. Po and M.N. Do, "Directional multiscale modeling of images using the contourlet transform," *IEEE Trans. on Image Processing*, vol. 15 (6), 2006, pp. 1610–1620.

[24] R.A. Ravishankar, *A Taxonomy for Texture Description and Identification*. Heidelberg: Springer Verlag, 1990.

[25] D.W. Scott, *Multivariate Density Estimation. Theory, Practice, and Visualization*. New York, NY: John Wiley & Sons, 1992, pp. 109–110.

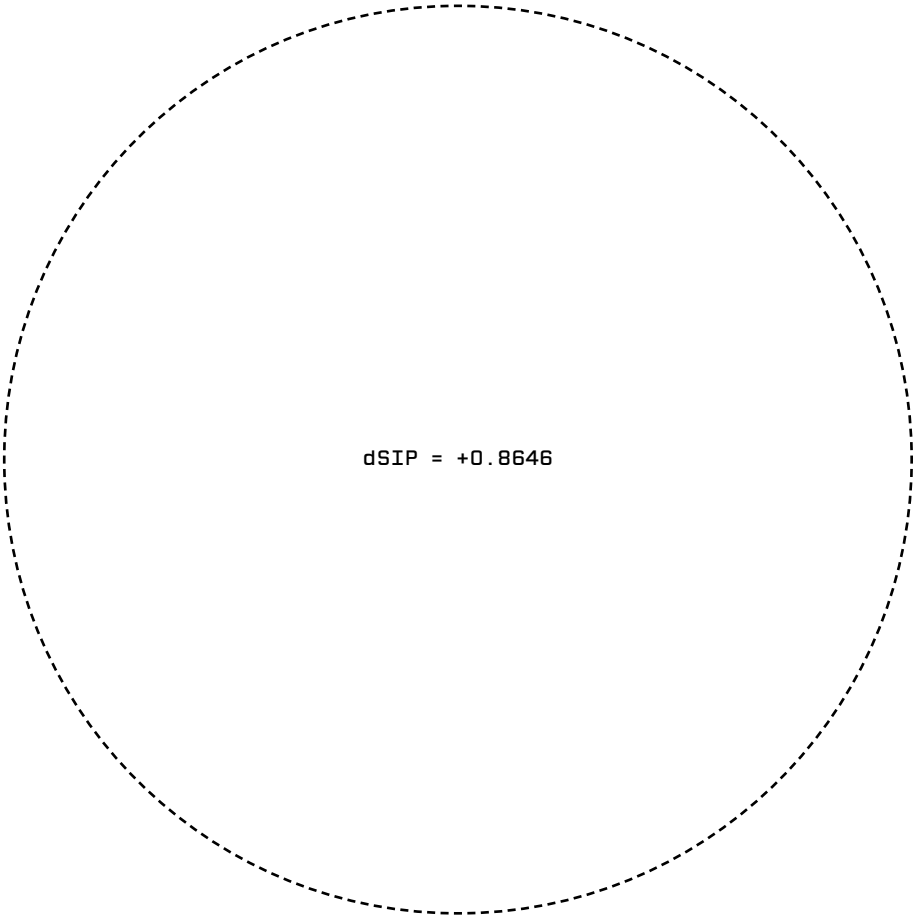
[26] I. Siddiqi, N. Vincent, "Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features," *Pattern Recog.*, vol. 43, 2010, pp. 3853–3865.

[27] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*. London / Boca Raton, FL: Chapman & Hall / CRC, 1998, p. 48.

[28] S.N. Srihari, C. Huang, and H. Srinivasan, "A Search Engine for Handwritten Documents," *Proc. of 12th Conf. on Doc. Recog. and Retrieval (DRR XII)*, 2005, pp. 66–75.

[29] L. Wolf, R. Littman, N. Mayer, T. German, N. Dershowitz, R. Shweka and Y. Choueka, "Automatically Identifying Join Candidates in the Cairo Genizah," *Intl. J. of Computer Vision*, October 2010, pp. 1–18.

Publication the Second


dSIP = +0.8646

Forensic vs. Computing writing features as seen by Rex, the intuitive document retriever

Vlad Atanasiu

Abstract—The paper reveals the superficial matching between script features as understood by forensic experts and computer scientists and advocates the development of computational instruments tailored to fit the features traditionally used by the forensic community. In particular, and including other areas of graphonomics and the general public, there exists a demand for software for the analysis of intuitive features, think "slant" or "roundness," as opposed to analytical features, like "Fourier transform" or "entropy." Rex, a software with such a capability, is introduced and used to explore the potentialities of this approach for script forensics. An investigation of properties of the script contour orientation, the feature used by Rex, is also presented.

***Index Terms*—script features, contour orientation, computational graphonomics, handwriting forensics,**

I. Introduction

In this paper I wish to discuss the distinction between the typical forensic and computer science writing features (section II), introduce a software that takes into account their specifics (section III) and investigate the behavior of the feature used by the said software (section IV). The overall goal of the paper, beside the immediate benefits derived from the individual topics, is to provide thinking material about the challenges building software adapted to forensic applications.

II. Forensic vs. Computing writing features

Semiotics — That much forensic handwriting expertise is subjective and would profit from mathematics and computing in its quest for objectivity and replicability is publicly admitted [1], but the less advertised side of reality is that of software insisting to treat the users on feasts of mathematics and technology without actually meeting their needs [2]. At the root of this dialogue of the deaf lies, among other interesting factors of the sociology of science, the very words "writing feature." For forensic experts the "feature" is usually intuitively comprehensible, such as "slant" [3], while for computer scientists the most powerful "features" are mathematical concepts, like "Fourier components" or "fractal dimension," which need specialized knowledge for their properties to be understood. Developing measurement software for intuitive features not only gives forensic professionals tools which they know how to handle, but also allows them to communicate about their work—an essential aspect in respect to testimony in court. Intuitive features additionally benefit the design of computer systems, improving the ergonomy of user interfaces as exemplified in section III.

Table I
Divergence in use of writing features across graphonomics areas

intuitive ◄	**Writing features**	► analytical
pen pressure, stroke width, character slant, allographs, line justification, baseline shape, regularity, aesthetic quality...		Fourier components, wavelet coefficients, entropy, fractal dimension, hidden Markov model...
▼ primary use	**Graphonomics areas**	▼ primary use
✘	forensics, paleography, typography, art, education, medicine	.
.	writing recognition, digital libraries, biometrics, neurocognitive sciences	✘

Cognition — An interesting viewpoint on the debate over intuitive and analytic features is to consider mathematics as an evolutionary outcrop of the neural computing capacities of the brain. Intuition is evolutionary unconscious learning by interaction with the environment to which conscious analysis supplements when novelties arise. Thus the two can be envisioned as a continuum, mathematics progressively becoming intuitive.

Sociology — To think that the divergence of the two feature types is a function of mathematical educational level is overlooking a fundamental distinction. Writer identification and verification are main mobiles of computational handwriting forensics, and because here only results count, it can use any method without even the need of thorough understanding insofar as it is better. This evolutionary mindset of a goal-focused black box approach is faced by the knowledge-oriented crystal ball attitude seen in the traditional graphonomical research, which adds to the control tasks mentioned above a considerable interest in the handwriting ecosystem, i.e. the structures and dynamics of handwriting features across populations and the underlying factors: material, cognitive, biomechanical, sociocultural.

Linguistics — The issues with the term "feature" extend to a further worldview cloaking inconspicuously its users. The proposition "This font is Roman" is considered in philosophy either as an expression on a property owned by the font (objectivism) or attributed to the font by an observer (subjectivism) [4], [5]. The difference is one of lifestyle: the world is there for

Manuscript received August 15, 2011.

V. Atanasiu was with Télécom ParisTech, Paris, France. He is now with Kabikadj, Paris, France (phone: +33 143 254 811, email: atanasiu@alum.mit.edu, website: http://alum.mit.edu/www/atanasiu/).

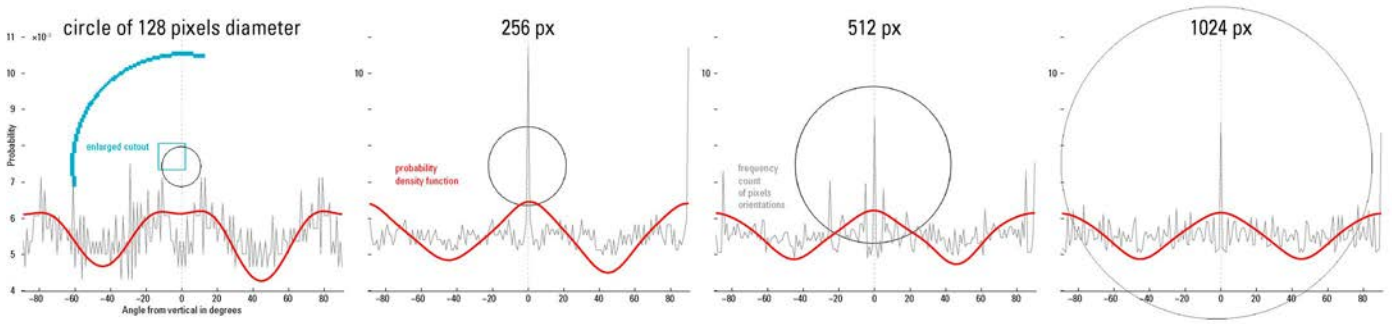


Fig. 1. Contour orientation profiles — Look carefully at the enlargement of the 128 pixels diameter circle and you'll see four horizontal and vertical pixels in a row: a bitmap shape representation has more pixels in these directions than warranted by the ideatic shape. The distortion decreases with object size. The ordinate values reveal that bias is small: its amplitude is ~0.002, while for a typical written document the mean is ~0.025 and the maximum ~0.04 (see Fig. 3 and [16]–[18]).

truth to be discovered or for models to be invented. Translated at lexical level this is what defines the terms "feature" and "descriptor," among their numerous handwriting related synonyms [3]. To this author "descriptor" seems more appropriate since it doesn't presuppose anything about the object (it just is) and it's easier and more fun to be critical about a model than a truth. Incidentally, while "feature" prevails in graphonomics, "descriptor" has a foothold in the wider pattern recognition community, as witnessed in a wording like "shape descriptor."

Implications — Computer scientists have to consider in common intelligence with forensic experts three issues worth mentioning because they bear an influence on how the software presented later in the paper is to be used. The issues are the desired precision of the analysis, the definition of the features and the affordability to analyze them in the current state of the art. I will illustrate this through two visual examples.

P2 – 44

Precision — Fig. 1 presents three bitmap circles of various sizes for which the orientation along their contour is measured (details in section III). Being circles, we would expect that all orientations be equally well represented, but due to the discrete nature of the underlying raster in which the shapes live the distribution is biased towards the orthogonal direction — the distribution will peak at 0 and 90 degrees ([6], [7], for hexagonal grids see [8]). Making a model of the distortion and applying it to arbitrary orientation profiles should solve the issue, but it turns out that the distortion is shape specific. For example, a vertical line has no distortion at all, so there is no need for correction. A somewhat better choice is to increase the image resolution at capture time or after, with the drawback of generating voluminous files and knowing that often only low resolution images are available. This digital geometry problem is compounded upstream by the design of discrete Gaussian filters for orientation measurement [9], and downstream by digitization, the same physical document producing at pixel level different shapes depending on its alignment with the digital grid of the imaging system, hence affecting the replicability of results [10], [11]. A number of techniques address these issues [12]–[15] but the implications for handwriting analysis have yet to be fully explored, starting with the question of how much precision is needed for which application. High accuracy graphonomics is therefore an area open to investigation.

Definition — I discuss now the slant of three Roman script characters as perceived by a human and raise the question of how this simple feature should be defined. In the case of **l** the slant is vertical and corresponds to the shape's axis of equilibrium through its center of gravity — here the slant is a physical property of the object. For an **o** there is no way to tell how the character is oriented would the baseline be unknown — slant is here a property of the object relative to the surrounding. The slant of **y** can be considered as upright only if we are able to identify the shape as character "y" and be aware of the convention that this lower case letter has to be considered vertical despite its physical right-leaning — this is a case of semantic slant. A deeper examination might reveal even more criteria. In conclusion, a slant analysis algorithm implementing human expert behavior appears to be more challenging than suspected, given first the very difficulty to define the feature, and secondly due to the mix of perceptual and cultural considerations to model.

Afordability — The last sentence leads to the issue of affordability: do we have the technological means to perform comprehensive slant analysis since we need to recognize unconstrained handwritten characters? This task not being presently solved, a positive answer can be given only if we are happy with a certain degree of imprecision, its exact amount having to be determined. Some of the fine computational forensic expertise that we would wish to attain is thus yet out of reach.

III. Rex, the intuitive document retriever

Rationale — Written documents in databases can be retrieved by appearance by one of the following methods: visual (using a reference document), semantic (describing script features), haptic (by drawing) and exogenous (from document ecosystem metadata). Semantic retrieval is convenient because it is intuitive (it takes place via a graphical and natural-language interface), free of any preexisting model (not always available) and can describe aspects of a script (contrary to the holistic approach of visual retrieval). The software that grew out of these considerations, called Rex, suits the demand for tools supporting forensic specific features as described above (Fig. 2) [16]–[18].

Technicalities — The software measures the local orientation along the writing contour, a popular computational graphonomics feature [19]. This is done by applying on the binary image

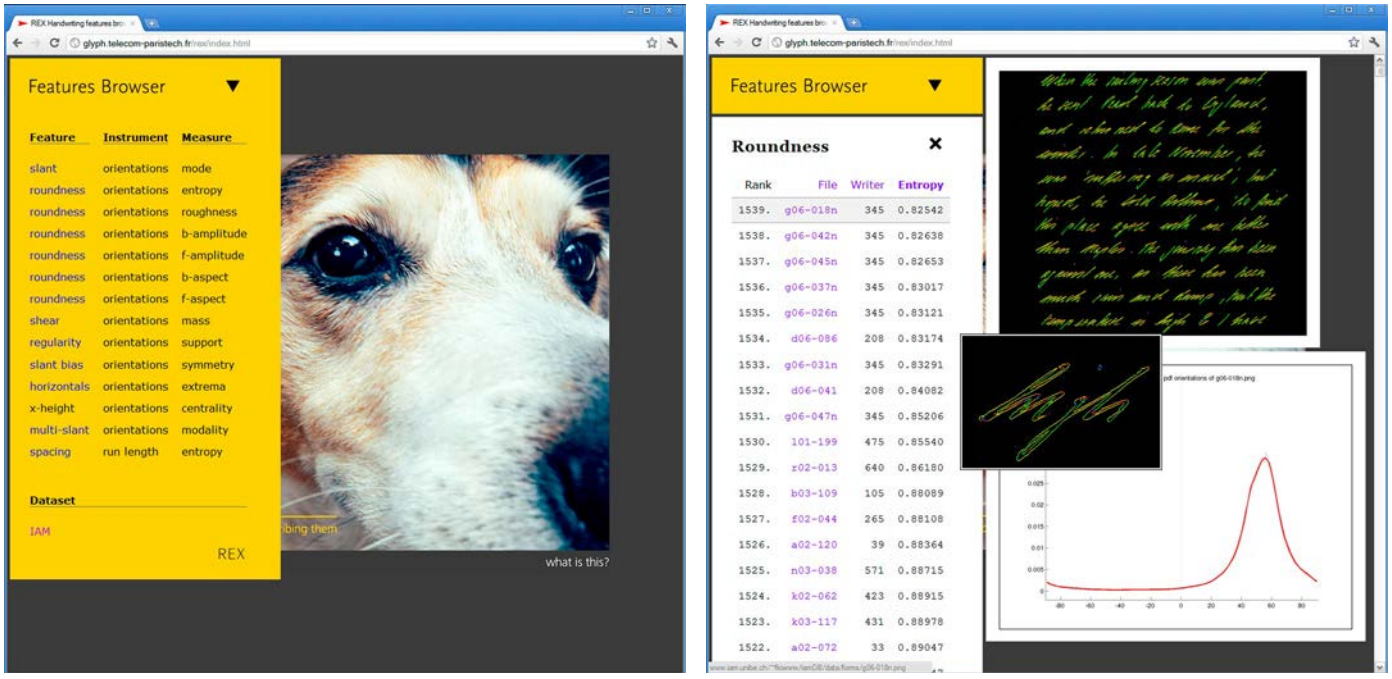


Fig. 2. Rex screenshot — After selecting an intuitive script feature (left picture, showing also the underlying mathematical measurements and instruments), users obtain a list of documents ranked according to the quantitative value of the feature, in this case "roundness" (right picture, giving the file and writer id too). The document and a mouse-over zoom with pixels colorcoded by orientation is presented, as well as the orientation profile and a hyperlink to the original document.

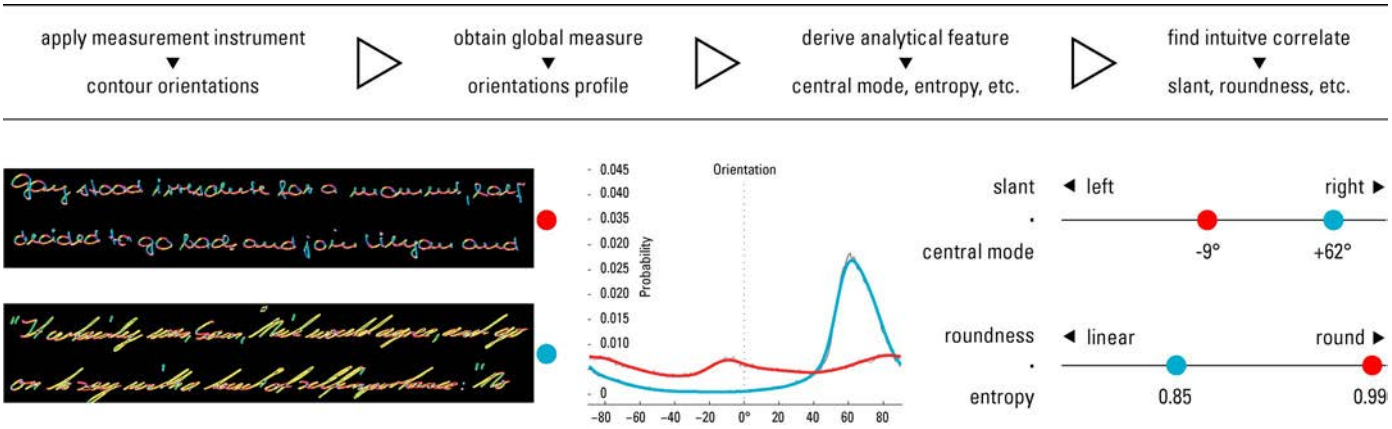
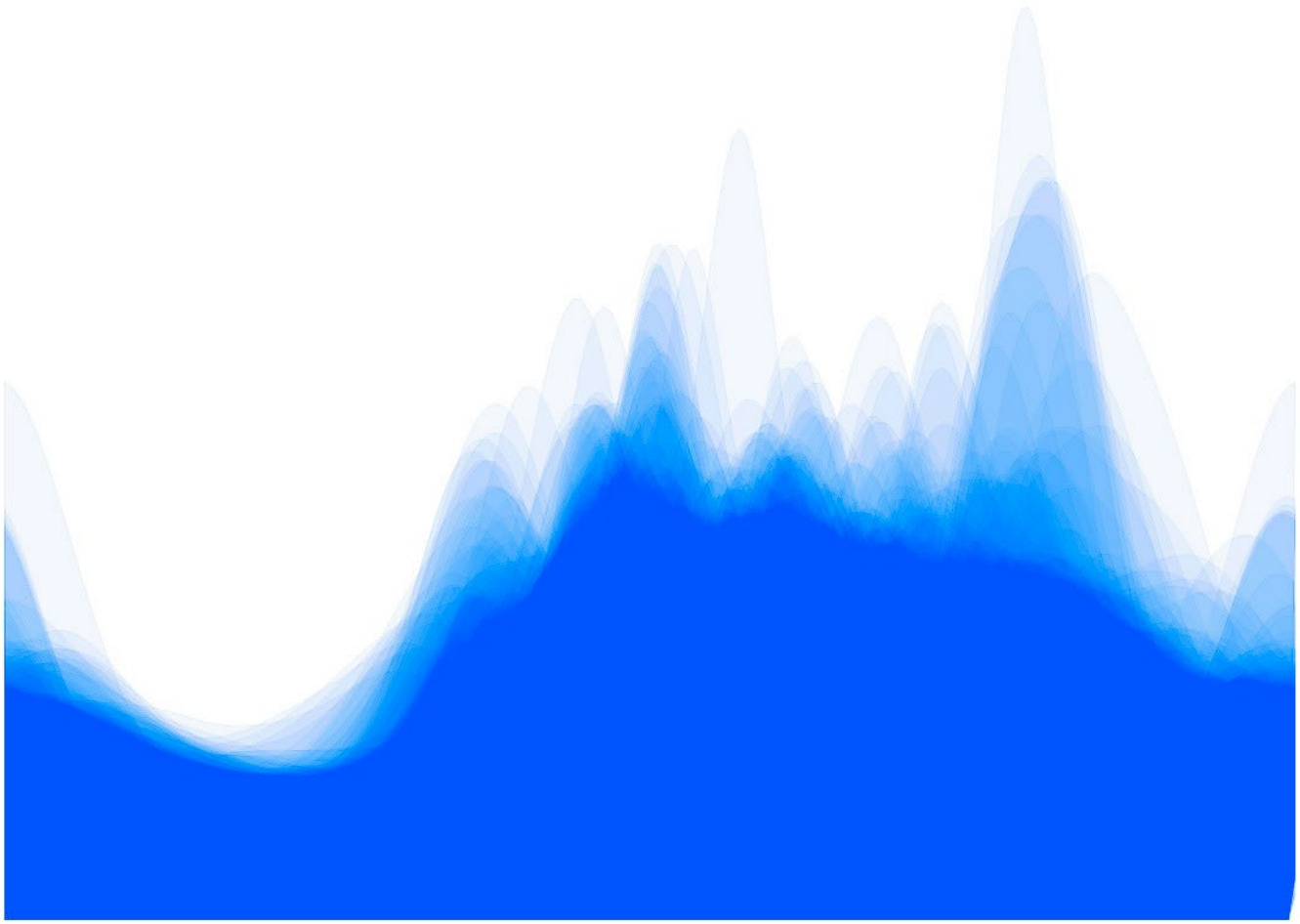


Fig. 3. Pixels to vectors to scalars to concepts — Prospecting for intuitive writing descriptors by extracting various statistical parameters of a global measurement. The colormap of the script samples (P02-081 and L01-199 of [8]) encodes the contour orientation at each pixel location — red for example being horizontal.

of the contour an anisotropic Gaussian filter bank with one degree of radial displacement. At this stage of this well-known approach two innovations are introduced, in addition to the fine grained resolution. First, after deriving the probability density function from the orientations' frequency count, statistical properties of the distribution are obtained. Second, it was discovered that these statistics correlate with various script features of the intuitive type, perceived as distinct one from another, such as "slant," "roundness" or "density" (Fig. 3). To sum up, Rex behaves like a handy, multipurpose Swiss army knife.

Applications — The Swiss reference is not fortuitous, since the handwriting documents presently used by Rex originate in that country (IAM Handwriting Database 3.0 [20]). This shows again the surprising versatility of the tool in that it is not only a document browser, but also a teaching tool about handwriting. In addition to learning about individual documents, Rex provides an insight in the make-up of a population of writers — that of the canton of Bern from where most of the dataset writers hail (Fig. 4). The question that immediately springs to mind — "Do writers from other parts of the world have similar characteristics?" — is typical of the richness of research and pedagogical possibilities opened by such an instrument (indeed, the few Greek, Chinese and other foreigners among the contributors show scriptural characteristics apart form the Swiss majority). If the present usage of Rex is rather limited to a browser of a specific dataset and much development can be imagined, it is nevertheless also an intriguing tool to experiment with as a testbed for other computational forensic applications.



P2
–
46

Fig. 4. "1539 Swiss Mountains" — Superpositon of all orientation profiles of the Swiss IAM Offline Database 3.0. We might be aware that few people have a left-slanted handwriting, but this visualization makes the phenomenon visible and measurable: there are few and only small peaks on the left-hand side of the image. It is also apparent that peaks on the right are narrower, hence slanting a handwriting reduces its roundness, making it more linear: a shear transform takes place.

IV. Properties of the orientation feature

While contour orientation is a concept easy enough to grasp, it has a number of less apparent properties with implications for the expertise work. They reveal why studies find orientation not the best performing biometric instrument [19].

Rotation — The feature is evidently not rotation invariant, meaning that the same document will have different measurement profiles depending on, for example, the skew of the paper in a scanner (Fig. 5.1–2). However the difference is only a translation of the profile, thus the bias can be corrected.

Organization — Contour orientation exhibits some unusual cases of shape invariance, all deriving from its low sensitivity to the spatial organization of pixels, due to the fact that, by definition, the measure is done locally. It is thus possible to have perceptually different shapes with the same orientation profile. Fig. 5.5 demonstrates scrambling invariance.

Localization — The various informations that can be read in the global orientation profile can't be traced to specific locations in the written document. If there is, say slant variation in a particular line, we see it in the profile, but can't localize the given line and even not know if the variation is concentrated in one line or spread over the entire document.

Convexity — For 180° shape rotations the profiles are identical, leading to shape confusion (Fig. 5.3–4).

Neighborhood — Fig. 5.6 shows that lines and circles in certain configurations can look the same to the orientation instrument: it is unaware about the neighborhood.

Additivity — Shapes contribute linearly to profiles, facilitating combinatorial pattern simulations from primitives.

V. Conclusions

I conclude by reminding that forensic and computational script features are usually not identical, that they need to be thoroughly explored to be safely used, and that public software, like Rex, introduced here, are excellent learning opportunities.

References

[1] B. Found and D. Rogers, "The Probative Character of Forensic Document Examiners' Identification and Elimination Opinions on Questioned Signatures," in *Proc. 13th Conf. of the Intl. Graphonomics Society, Melbourne, Australia*, 2007, pp. 171–174.

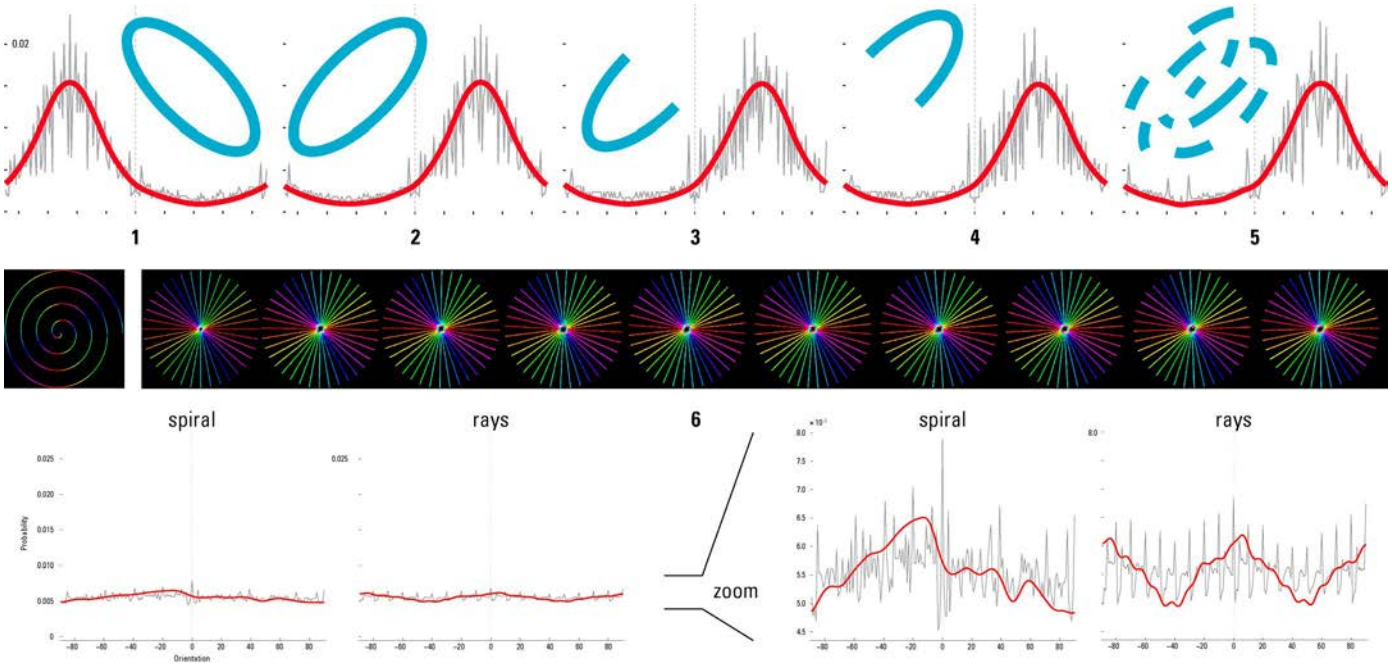


Fig. 5. Cases of confusion — (1, 2) A rotation of shapes (in blue) is equivalent of a translation of the orientation profile (in red). (3, 4) Rotation by 180° or (5) breaking up a shape doesn't affect the orientation profile beyond quantization errors. (6) The bitmaps row shows a spiral and 10 ray bundles, each bundle being rotated by 1° in respect to its neighbor, covering the entire angular sensitivity spectrum of the measurement instrument. Despite the perceptual pattern difference — one linear, the other curly — the orientation profiles are similar, especially when seen at the scale of the writing of Fig. 3 (the differences become visible when zooming in).

[2] R.J. Verduijn, C.E. van den Heuvel, R.D. Stoel, "Forensic Requirements for Automated Handwriting Analysis Systems," in *Proc. 15th Conf. of the Intl. Graphonomics Society, Cancun, Mexico*, 2011, pp. 132–135.

[3] R.A. Huber, A.M. Headrick, *Handwriting Identification: Facts and Fundamentals*. Boca Raton, FL: CRC, 1999, pp. 89–91.

[4] D.H. Mulder, "Objectivity," *The Internet Encyclopedia of Philosophy*, September 6, 2004 [Online]. http://www.iep.utm.edu/objectiv/ [Acessed: August 3, 2011].

[5] Ch. Swoyer and F. Orilia, "Properties," *The Stanford Encyclopedia of Philosophy*, July 2011 [Online]. Acessible: http://plato.stanford.edu/entries/properties/ [Accessed: August 3, 2011].

[6] G. Gonzato, F. Mulargia and M. Ciccotti, "Measuring the fractal dimension of ideal and actual objects: implications for application in geology and geophysics," *Geophysical J. Intl.*, vol. 142, 2000, pp. 108–116.

[7] R. Klette and A. Rosenfeld, *Digital Geometry*, San Francisco, CA: Morgan Kaufmann, 2004.

[8] R.C. Staunton and N. Storey, "A Comparison Between Square and Hexagonal Sampling Methods for Pipeline Image Processing," in *SPIE Conf. Optics, Illumination, and Image Sensing for Machine Vision, Philadelphia, PA, USA*, 1989, pp. 142–151.

[9] E. R. Davies, "Design of optimal Gaussian operators in small neighbourhoods," *Image and Vision Computing*, vol. 5(3), 1987, pp. 199–205.

[10] B. Nagy, "An algorithm to find the number of the digitizations of discs with afixed radius," *Electronic Notes in Discrete Mathematics*, vol. 20, 2005, pp. 607–622.

[11] M.N. Huxley and J. Žunić, "Different Digitisations of Displaced Discs," *Foundations of Computational Mathematics*, 2006, pp. 255–268.

[12] F. de Vieilleville and J.-O. Lachaud, "Comparison and improvement of tangent estimators on digital curves," *Pattern Recognition*, vol. 42(8), 2009, pp. 1693–1707.

[13] D. Coeurjolly and R. Klette, "A Comparative Evaluation of Length Estimators of Digital Curves," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26(2), 2004, pp. 252–258.

[14] L.J. van Vliet and P.W. Verbeek, "Curvature and bending energy in digitized 2D and 3D images," in *Proc. 8th Scandinavian Conf. on Image Analysis, Tromsø, Norway*, 1993, pp. 1403–1410.

[15] S.-Ch. Pei and J.-W. Horng, "Fitting digital curve using circular arcs," *Pattern Recognition*, vol. 28(1), 1995, pp. 107–116.

[16] V. Atanasiu, L. Likforman-Sulem, N. Vincent, "Rex, a description-based retriever for written documents," April 2011 [Online]. Accessible: http://glyph.telecom-paristech.fr [Accessed: June 30, 2011].

[17] V. Atanasiu, L. Likforman-Sulem, N. Vincent, "Talking Script. Retrieval of written documents by description of script features," *Gazette du Livre Medieval*, to be published.

[18] V. Atanasiu, L. Likforman-Sulem, and N. Vincent, "Writer Retrieval–Exploration of a Novel Biometric Scenario Using Perceptual Features Derived from Script Orientation," in *Proc. 11th Intl. Conf. on Document Analysis and Recognition, Beijing, China*, 2011, to be published.

[19] M.L. Bulacu, "Statistical pattern recognition for automatic writer identification and verification," Ph.D. dissertation, Artif. Intell. Inst., Univ. of Groningen, The Netherlands, 2007.

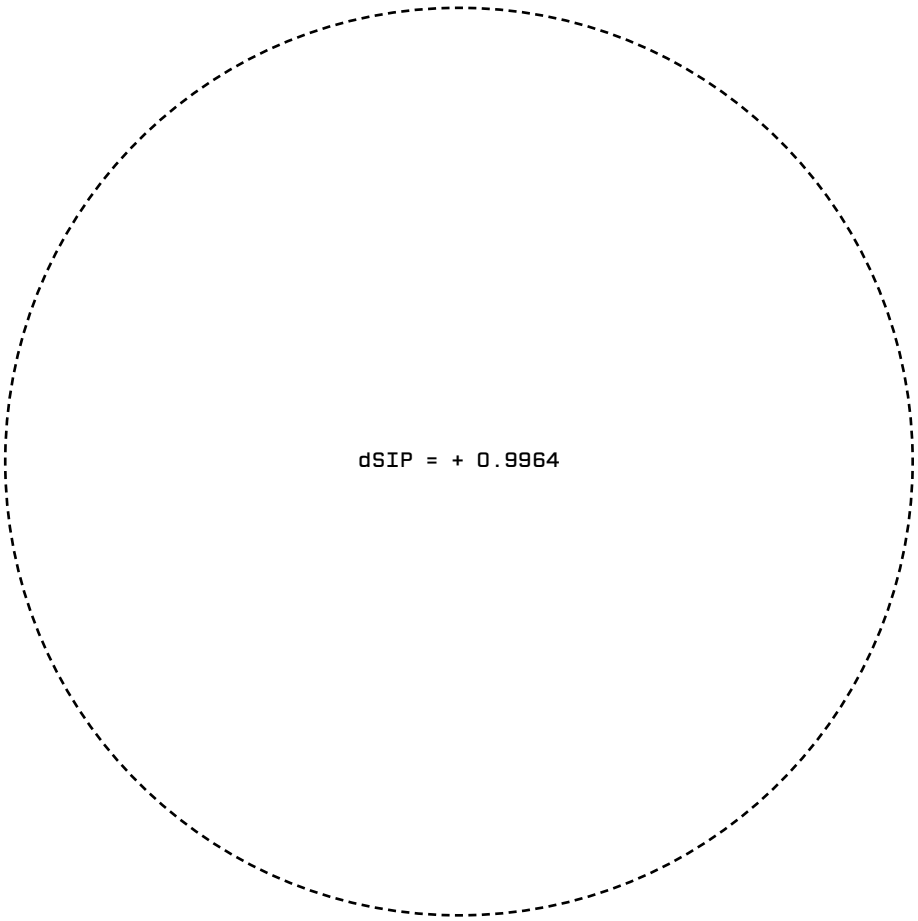
[20] U. Marti and H. Bunke. "The IAM-database: an English sentence database for off-line handwriting recognition," *Intl. J. on Document Analysis and Recognition*, vol. 5, 2002, pp. 39–46. Available: http://www.iam.unibe.ch/fki/databases/iam-handwriting-database

Vlad Atanasiu was born in Timişoara, Romania, 1970. He passed the entry examination to the Telecommunication Faculty, Polytechnic University, Timişoara, in 1989 and went on to switch careers thanks to the fall of the Iron Wall, earning a B.A. in Middle Eastern Studies, an M.A. in Linguistics, from University of Provence, Aix-en-Provence, France in 1994, respectively 1996, and a Ph.D. in Paleography, Art History and History from École pratique des Hautes Études, Paris, France, in 2003. He took classes in Cognitive Science at the Massachusetts Institute of Technology, Cambridge, USA during post-doctoral studies, 2003–2005.

He was Intern in the Vision Science psychophysical department of Lighthouse International, New York, USA, 2004–2005; Senior Scientist for Image Processing and Geographical Information Systems as well as Coordinator of the European project "Bernstein. The Memory of Papers. Image-Based Paper Analysis and History" at the Austrian Academy of Science, Vienna, Austria, 2006–2009; and Research and Development Engineer for handwriting analysis and document retrieval at Télécom ParisTech, Paris, France, 2010–2011. He is now CEO of Kabikadj, Paris, France. He published the book *Letter frequencies and their influence on Arabic calligraphy*, Paris: L'Harmattan, 1999.

Publication the Third

Reprinted with permission of the Gazette du livre médiéval.



REX
glyph.telecom-paristech.fr

Talking Script

RETRIEVAL OF WRITTEN DOCUMENTS BY DESCRIPTION OF SCRIPT FEATURES

Documents in databases can be retrieved by appearance by one of the following methods: visual (using a reference document), semantic (describing script features), haptic (by drawing) and exogenous (from metadata about the writers). Semantic retrieval is convenient because it is intuitive (it takes place via a graphical and natural-language interface), free of any preexisting model (not always available) and can describe parts of the script (contrary to the holistic approach of visual retrieval). We present here a prototype of a semantic script retriever.

Vlad Atanasiu
Laurence Likforman-Sulem

Signal and Image Processing Dpt.
Télécom ParisTech
Paris, France
atanasiu@alum.mit.edu
likforman@telecom-paristech.fr

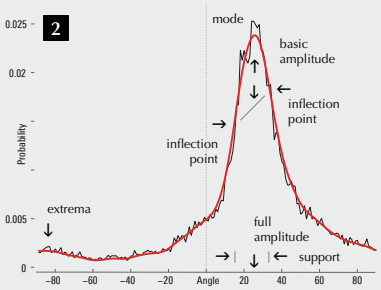
Nicole Vincent

Laboratoire LIPADE
Université René Descartes
Paris, France
nicole.vincent@mi.parisdescartes.fr

1

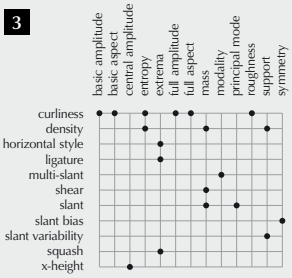
Contour measurement

2



Features of the orientations distribution

3

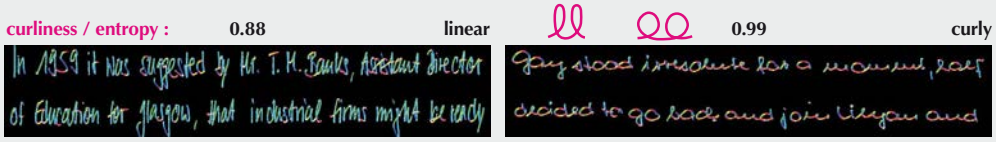


Perceptual correlations

[1] An anisotropic Gaussian filter bank with one degree radial resolution is applied to the binary contours of the script image to obtain the local orientation at each contour pixel. [2] Statistical properties of the resulting distribution are computed. [3] Experimentally it is found that these computational measurements relate to perceptually meaningful individual script features.

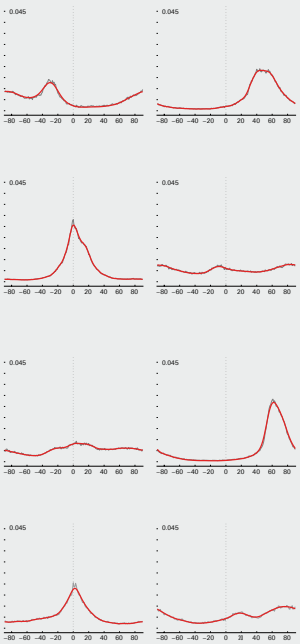
slant / mode : –30 backwards slant — 45 forward slant

curliness / entropy : 0.88 linear — 0.99 curly



shear / mass : 0.11 less shear — 0.58 more shear

density / support : 8 more dense — 48 less dense



Conference poster, Analyse d'image et paléographie systématique, Institut de Recherche et d'Histoire des Textes, 14–15 April 2011, Paris

Handwriting samples from IAM OffLine Handwriting Database 3.0 http://www.iam.unibe.ch/fki/databases/iam-handwriting-database/

Poster manufacturing courtesy of Graphem http://liris.cnrs.fr/graphem/

Kumagoro Design http://alum.mit.edu/www/atanasiu/

Table ronde

APPLICATIONS ACTUELLES DE L'INFORMATIQUE À LA PALÉOGRAPHIE :

QUELLES MÉTHODES POUR QUELLES FINALITÉS ?

Modération : Marilena Maniaci

Intervenants

Vlad Atanasiu (*), Signal and Image Processing Dpt., Télécom ParisTech, Paris
Irene Ceccherini, Facoltà di lettere e Filosofia, Università di Firenze
Thomas Falmagne, Bibl. nationale de Luxembourg / Handschriftenzentrum de la DFG, Frankfurt am Main
Alain Guerreau, Centre de recherches historiques, Paris
Maria Gurrado, Institut de recherche et d'histoire des textes, Paris
Marilena Maniaci, Facoltà di Lettere e Filosofia, Università di Cassino.
Alessandra Scotto di Freca (*), Facoltà di Ingegneria, Università di Cassino.
Peter Stokes (*), Department of Digital Humanities, King's College, London
Dominique Stutzmann (*), Institut de recherche et d'histoire des textes, Paris
Georg Vogeler (*), Zentrum für Informationsmodellierung in den Geisteswissenschaften – Karl-Franzens-Universität, Graz
Teresa Webber, Trinity College, Cambridge

(Les intervenants marqués d'un astérisque ont présenté un projet de recherche à l'occasion de cette journée. Voir ci-après, p. 131.)

M. Maniaci :

Agli organizzatori di questo incontro parigino va il merito di aver voluto abbinare, in un unico evento, la giornata conclusiva del programma Graphem e il quinto appuntamento annuale del *Seminario permanente sulla corsività*. La giustapposizione fra l'approccio "tradizionale" e quello "sperimentale", da molti ancora percepita come una contrapposizione epocale assume in questa cornice una funzione di stimolo a riflettere sul passato, sul presente e sul futuro della paleografia, da una prospettiva – quella della progressiva, e direi inevitabile, diffusione delle tecnologie digitali – portatrice di stimoli e opportunità nuove, ma anche di interrogativi ancora largamente irrisolti, a volte neppure chiaramente focalizzati e formulati. Vale la pena di riassumere i principali, per introdurre il dibattito.

Gazette du livre médiéval, nº 56-57 — 2011.

des informaticiens. Pourtant, la volonté initiale des équipes d'informaticiens, cherchant d'abord à appliquer leurs méthodes, n'a été tournée au profit d'une méthode plus strictement paléographique que dans la dernière partie du projet. À ce point, deux remarques doivent être faites. Primo, même dans le projet Graphem, l'illusion d'une possible objectivation du fait historique à travers l'analyse d'image par ordinateur n'était pas absente, et il y avait bien l'espoir d'arriver à une classification « objective » des écritures, qui a fait refuser pendant longtemps de fournir une table des types d'écriture aux équipes STIC. C'est sans doute l'un des apports importants du projet Graphem à la communauté paléographique : l'écriture à travers les âges est un phénomène trop complexe et trop variant pour qu'une analyse assistée par ordinateur puisse, de soi-même, sans analyse supervisée et sur un ensemble d'échantillons, fournir une classification pertinente historiquement. Secundo, l'obtention de subvention doit être regardée plus largement : ce n'est pas seulement une différence générationnelle, c'est le produit même de la science paléographique telle qu'elle s'est construite, comme savoir-faire et expertise. Les projets en paléographie dite « traditionnelle » nécessitent une expertise développée sur le long terme, alors que les financements concernent des projets de court terme. Aussi y a-t-il inadéquation entre les projets de paléographie et les modes de financement. Bien plus, même si les organismes nationaux ou européens accordaient des montants équivalents aux projets de paléographie sans le volet numérique, l'on peut affirmer qu'il y aurait trop peu de paléographes formés pour pouvoir utiliser utilement ces financements.

V. Atanasiu :

I wish to highlight some critical aspects of the collaboration between palaeographers and computer scientists for those who engage in the common enterprise of digital palaeography.

Don't blame the computer. — We often hear that "the computer can't do this" as an excuse to not using informatics for palaeography, but in a sense the computer doesn't exist, the computer is just an algorithm conceived by a human and its inability to perform certain tasks related to the individual's ingenuity.

The professional gap. — Computer scientists deride the lack of scientific method in palaeographic expertise and palaeographers complain that they don't understand what computer scientists are talking about. Clearly one side has to shed its subjectivity and adopt an analytical and quantitative vision, while the other has to become pedagogically more approachable, be willing to learn more about palaeography and see if its software products really address the needs of palaeography.

Avoiding the asymmetric handshake. — For a successful collaboration to happen a lot of talk between palaeographers and computer scientists has to take place, in the process of which each will be transformed a little bit into the other. The basic requirement for this are regular meetings, best – and in this case maybe only – through physical meetings. In my experience, barring such interaction, nobody will be happy at the end of

the collaboration. Palaeographers have also to understand that to solve their problems, there is a lot of research to be done, which implies the availability of adequate financial resources. Conversely computer scientists have to make clear the limitations of their products, understand that they will work with ill-defined problems, and that features not present in the initial specifications will pop up all along the project and will need to be implemented, making the software design challenging.

Conflicting expectations. — Most computer scientists see the collaboration with palaeographers as an opportunity to advance the state of the art in handwriting processing, and their main output, to which their professional branch is pushing them, are publications rather than practical software. They are doing research. Palaeographers however expect computer scientists to produce functioning tools, which is development activity. In this respect digital palaeography projects would be better served by partnering palaeographers and companies instead of research labs.

Different interests. — Related to the previous paragraph is where the primary interest of the partners lies: palaeographers are more interested in the data and computer scientists in methods for processing it. Evidently a solid cross-domain understanding (and interest) is expected for successful products.

Technical aspects. — Digital palaeography can exist only if the necessary software is available, which is not the case today. Except for a handful of software, most of the digital palaeography work is performed with research level software, that is with minimal interfaces (ergonomy is artistic make-up anyway, isn't it?) and documentation (but we know what the software does!), usually not compiled (let the palaeographer compile the source code), rickety interoperability (on line mash-up's anybody?), and short life-span (who remembers yesterday's software?). Besides software issues there is also the ground-truthed datasets problem, datasets which are difficult to obtain given that there are no two palaeographers that will agree on the exact interpretation of a document (a possible definition of humans).

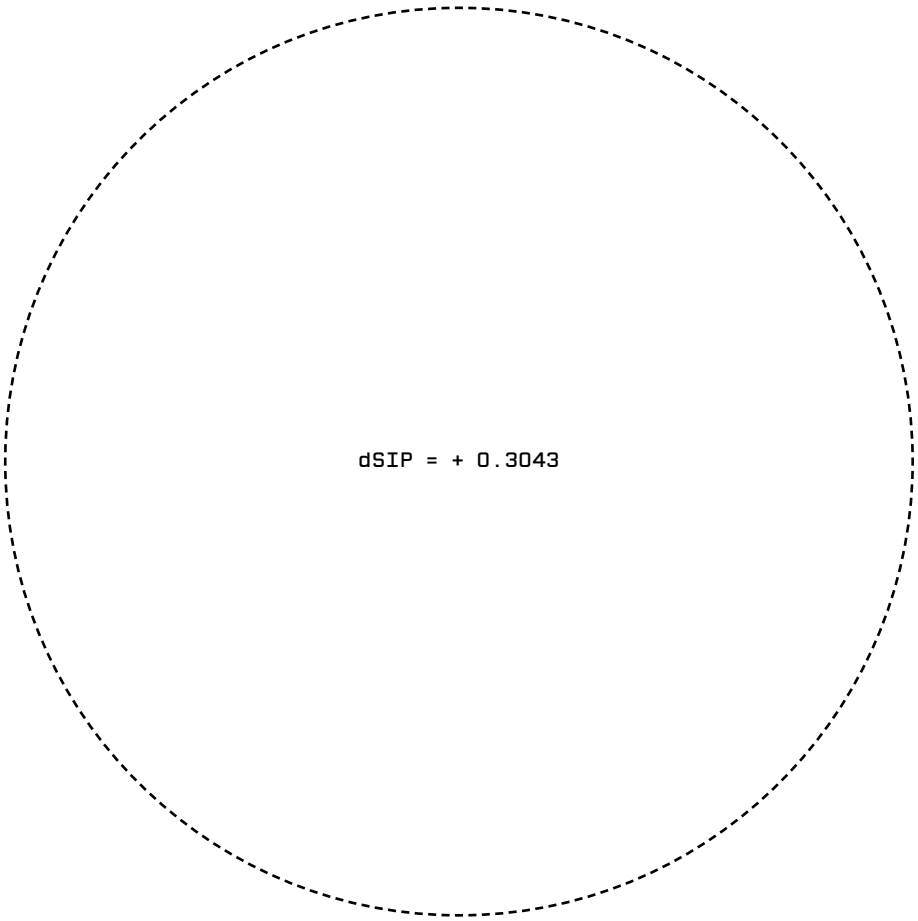
Financial aspects. — Opportunities to finance digital palaeography projects are rare to say the least, so one way forward is to get the field out of its ivory tower and design products that appeal to a wider public. Where's the Photoshop of handwriting?

G. Vogeler :

I'm sometimes a bit puzzled that the computer engineers seem to have an advantage in the communication between palaeographers and computer engineers. Is it really the case that they can decide what can be done and what can't be done? Isn't it the other way around, that only the palaeographer can decide which method is fruitful and which should be altered or abandoned because it didn't produce any useful results? But certainly the main issue is that it is wrong to oppose traditional palaeography to digital palaeography. The question, as I stated earlier is: Where can we apply digital technologies to palaeographic research interests and where should we not?

Publication
the Fourth

Reprinted with permission of IS&T: The Society for Imaging Science and Technology, sole copyright owners of Electronic Imaging: Document Recognition and Retrieval XXIII.



Cuckoos among Your Data: A Quality Control Method to Retrieve Mislabeled Writer Identities from Handwriting Datasets

Vlad Atanasiu, Department of Informatics, University of Fribourg, Bd de Pérolles 98, 1700 Fribourg, Switzerland

Abstract

Motivation: *Handwriting datasets may contain specimens assigned to the wrong writer. A little discussed problem, such misclassifications, "cuckoos", can bias recognition, retrieval, identification, and other expertise systems, with serious consequences in biometric and forensic applications. Indeed, misclassification research has been purported as the most important topic in pattern recognition.* Objective: *We describe the design of a generic semi-automatic method for detecting possible misclassifications and illustrate it by way of an exemplary classification criteria (writer identity), measurement feature (contour orientation), and document distance metrics combination.* Method: *The core of the method consists in automated ranking of writer classes by stylistic variability, using the open source software Alphonse, followed by visual inspection of a limited number of top ranking classes, using an interactive handwriting datasets visualization tool, Rex. The method is independent from dataset producers and does not necessitate training. It is the result of empirical and theoretical research, and its performance demonstrated on the Swiss IAM offline handwriting dataset.* Findings: *We show that to evaluate the performance of a quality control it is necessary to consider the interdependency between system sensitivity and task difficulty. We propose a dataset-independent measure of the scrambling severity of a dataset and its proneness to misclassification. We find that in a broad writer population the variability of the contour orientation approaches a log-normal distribution, increasing the amount of genuine outliers.*

1. Introduction

Proposition 0: Rationale. A document recognition and retrieval system is only as good as its training and live data.

Corollary: Implication. The user does not care if the wrong document was retrieved because of the low quality of the system or that of the data, it only sees a wrong answer to his or her request, thus possibly spelling the demise of an otherwise good product.

Proposition 1: Solution. A writer verification system turns into a dataset quality control method by verifying the identity of every sample in a dataset.

Proposition 2: Generalization. Any system verifying the membership of an item to a class becomes a method for controlling the classification quality, once applied to every data item.

Proposition 3: Specificity. Writer identity is just one among many instances of handwriting classification criteria to which Proposition 2 applies, such as classification by style, handness, sex, age, or country.

Proposition 4: Constraints. The method given here can be applied indistinctly to any of the above classification criteria, beyond writer identity, insofar as the classes are stylistically homogeneous and the measuring instrument is sensitive to the classification criteria.

Definition: Terminology. We call a mislabeled data item a "cuckoo", and a class containing cuckoos "cuckold". The task of retrieving misclassified data is "cuckoo retrieval".

*

Paper objective — The objective of this paper is to describe the implementation of the generic principles of quality control expounded in the above Propositions, by way of an exemplary application to the specific case of writer identity. To this end, a semi-automatic software system, Alphonse [6], was created to experiment with an equally exemplary, yet effective, combination of image processing techniques using contour orientation as measuring instrument of the handwriting style, a cocktail of statistical metrics for distances between documents, and an interactive visualization tool for handwriting datasets, Rex [8].

Quality control in light of document retrieval — Quality control can be viewed as a special task in document retrieval, insofar as it means retrieving all mislabeled documents in a dataset. To the difference of document verification, often performed on a single document, quality control is not mere verification repeated as many times as there are documents in a dataset. The challenge of quality control is to group all misclassification candidates at the top of a ranking list, have an estimate of how many there are, and have the tools for visual confirmation by the human users of the system. In a previous publication we minted the concept of "writer retrieval" – the retrieval of all documents in a dataset produced by the same writer –, which is itself an extension of writer identification [7]; here we present the retrieval of misclassifications as an extension of the verification task.

Topic relevance — Misclassification research has been purported as the most important topic in pattern recognition [82, 38]. Specifically, data quality control contributes to document recognition and retrieval by improving the performance and security of systems [74, 14, 36]. In parallel, it serves other document processing tasks, such as identification and verification, by ensuring a healthy data basis. As for the relevance of writer identity, it is essential to the broadest spectrum of fields in which handwriting plays a role: biometrics, forensics, paleography, art dealership, medicine, education, digital libraries, and so forth. Indeed, it is the availability of writer identities what makes handwriting-based biometrics and forensics possible. When assignments of handwriting samples to the wrong writer occur, its impact varies in nature and intensity. As an example, the performance of writer identification systems evaluated on the dataset tested in this paper, IAM OffLine Handwriting database 3.0 [53], popular with academics, might be better than reported, or worse, due to a small amount of misclassifications. Quality control is also time consuming and tedious: it took about six months for one person to manually label the 382,200 characters of NIST SD3, another widely used handwriting dataset [86: 209; 35: 1; 87]. Thus, it seems desirable to have at disposal methods for the quality control of handwriting datasets.

Paper contribution and originality — The paper points toward an important, but little discussed topic. Its *theoretical* contribution is the insight that dataset quality control can be cast as an extension of item verification, which leads to a generic framework for the quality control of handwriting datasets, valid for various application domains, tasks, and classification types. The *operative* novelty of the paper's method is to create a quality control layer independent of

the dataset production workflow and software, where usually quality control takes place. At the level of *implementation*, we propose an original system utilizing the synergy between machine and human capabilities on one hand, and image processing, statistics, and interactive visualization on the other hand. Our choices are based on empirical and theoretical research, in the favor of which arguments their good performance. Interviews with the dataset producers were conducted to understand its design process and history. We also refine the *evaluation concept* to account for the interdependence between system performance, data complication, and use case requirements. From a *practical* point of view, the system doesn't demand training on preexisting data models, hence its simplicity and generality. The selected handwriting feature is versatile, robust, and its characteristics understood. Method implementation and reusability are facilitated by the publishing of the handwriting analysis software and visualization tool. An *unexpected* contribution of the paper is the discovery of a log-normal type distribution of variability in a writer population, a topic that begs for further research.

Paper structure — Having clarified the object of the paper, stated the principles guiding us, and stressed the need for quality control of handwriting datasets, we will dwell into the state of the art of quality control (Section 2) and explore the typology, factors, and implications of misclassification (Section 3). We examine the proposed detection method (Section 4) and evaluate experimentally its performance (Section 5) and task difficulty (Section 6). We conclude with perspectives on future work (Section 7) and a retrospective of the accomplishments (Section 8).

2. Research on quality control

Fields — Quality control is addressed in document processing by *normed workflow protocols* for data acquisition [86, 75], *data cleaning* of digital repositories [51], *ground truthing* of training data [86, 68, 12], *error-rate estimation* of classification systems [1], and results *evaluation* [86, 42]. More generally, the subject relates to a number of broader scientific fields: *statistical process control*, devoted to quality control in industrial and management settings, a field emerged in the United States before World War Two and later contributed to the success of the Japanese industry [64, 48, 65]; *data quality*, crystallized during the "dotcom boom" of the late 1990s and specifically concerned with digital data [74, 40, 57]; *outliers theory*, elaborated in statistics [10, 62]; and *misclassification research*, fundamental to pattern recognition [82, 38]. Ad hoc solutions to misclassification have been developed in a wide range of areas, beyond the dedicated fields just mentioned, providing interesting inspirational material: medicine (diagnosis reliability) [50], industrial inspection (surfaces defects) [59], geographical information sciences (imprecise coordinates) [76], biometry [52], surveillance (abnormal behavior) [43], satellite imagery (feature detection) [17, 33], and many more.

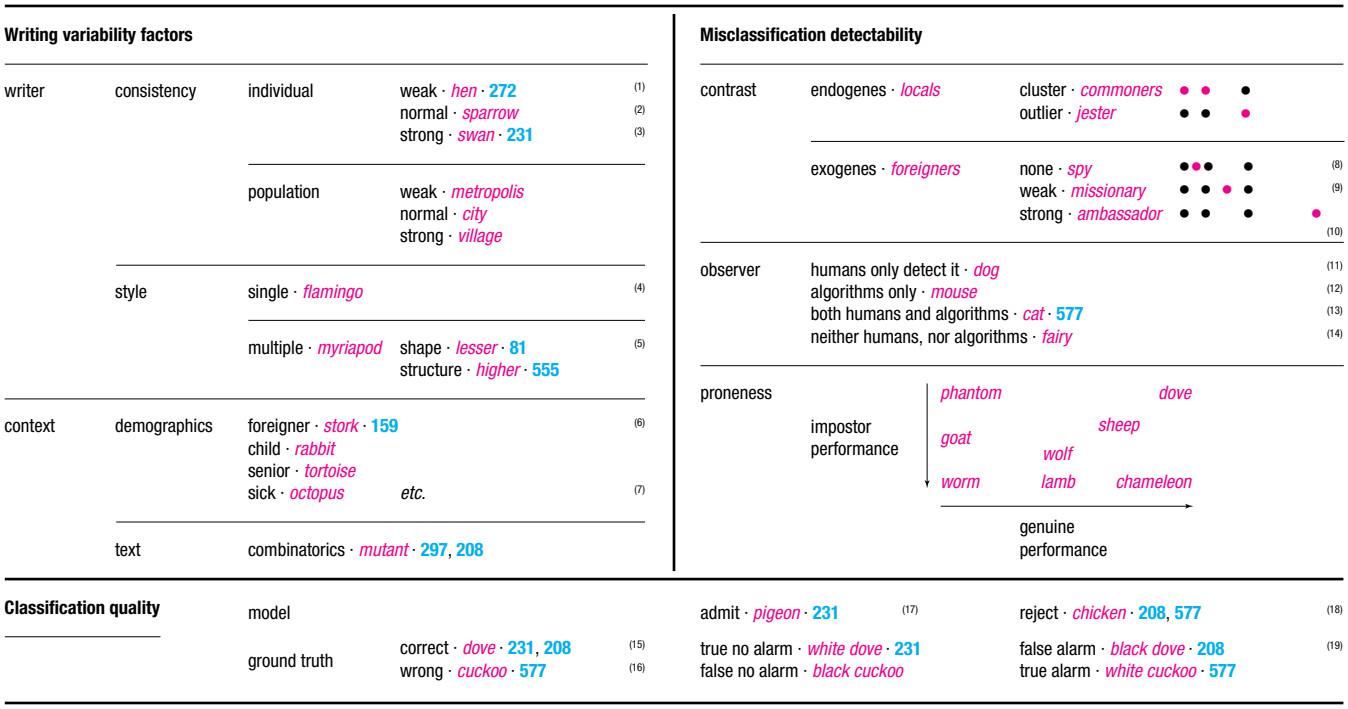
Challenges — Regarding writer identity, most quality controls are consubstantial with the data production process or classification algorithm [86, 75], creating a need for independent control capabilities once datasets and classification outputs reach end users. Also, control methods usually employ machine learning, which depends on the availability of training data and implementation efforts [37, 35, 14]. Finally, some datasets are more prone to errors than others [87], an information that could improve control systems and evaluation methods (like in scanning software, where the use of image degradation models are common [9]). These three issues – *independence from dataset producer*, *learning-free quality control*, and *prediction of misclassification proneness* – are addressed in these pages.

Prior work — Despite our best efforts we discovered only two sets of publications on the identification of misclassifications in handwritten documents, both from the 1990's [56, 36; 37]. (As a testimony of the topic's pedigree, a historical note: The first publications originate in the AT&T Bell Laboratories research group who invented the Support Vector Machines and, later, Deep Learning [21, 78]. The second publication was part of a research that sparked a fruitful era in computational handwriting processing at the Institute of Applied Mathematics, Bern, Switzerland, of which the IAM handwriting dataset used in this paper was one outcrop.) The difference between our method and the two others starts with a difference in *goals*: writer identification vs handwriting recognition. It results in a *structural difference of the input data*: on one hand classes with good *homogeneity*, due to the relative stability of an individual's handwriting, and on the other hand a high intra-class *heterogeneity*, generated by mixing allographic characters written by many writers. A pure *statistical* approach involving no learning is sufficient in the first case, while *learning* is necessary in the second. An ensuing requirement is that of preexisting *training data*. Both methods contend, however, partially with the *same problems*, such as whether to include a *manual check* of the machine's output (the answer from all researchers was "yes") and how to determine when the check should *stop* (here, the solutions are framed by the general approach of the respective quality control methods: using measures of statistical variance, of information theoretical surprise, and of Bayesian probabilities of error rates, respectively). In conclusion, *when intra-class variability is low, frequentist statistics are sufficient to detect misclassifications, otherwise machine learning is necessary*.

Alternative — Anticipating the method description, we mention that the underlying principle of the proposed handwriting dataset quality control consists in ranking writer classes by their internal stylistic variability, ideally resulting in an accumulation of the misclassifications in the top ranks. There exists a straightforward alternative to our *class variability ranking* method: reclassify anew the dataset, using a concurrent number of classifiers and compare their results with the original classification. *Concurrent classification* is a well-known technique, e.g. Web search engines use it for labeling images (sometimes by human volunteers) and it is even embodied in hardware (multi-viewer microscopes used by biologists, medical doctors, and others, to classify specimens by consensus [88]). The idea sounds attractive for automating handwriting quality control, were it not for the lack of information on the amount of cuckoos, which makes it a task of classification with unknown number of classes, a much harder problem than ranking existing classes by variability [29, 80, 26]. At the same time one has to be sure of the quality of individual components and the fine tuning of their interaction in this system of systems. Last, there is the practical difficulty in acquiring multiple systems: do they exist? how much do they cost? what resources do they demand? The interest of the class variability ranking method appears, in these circumstances, as a compelling solution.

3. Fundamentals of misclassification

Genesis — Cuckoos are typically created when labeling data with metadata (*labeling errors*) and when clustering the resulting entities in classes (*categorization errors*). Responsibility lies principally with the dataset producers, but software, the document writers themselves, and even environmental factors can play a role in misclassification. Cuckoos are mostly *natural*, emerging by accident, due to *attentional* and *procedural issues* (e.g. typos, merging of incompatible data, natural language or script ambiguities). Another cuckoo specie is *artificial*,



(1) "Hen writing" stands for bad, illegible handwriting. Expression attributed to the Latin playwright Plautus, 2nd c. BC, still in use in contemporary Italy [84, 85]. — (2) An epitome of commonness. — (3) Geese were mascots of Taoist calligraphers in Ancient China [47: 27]; swans emblems of grace in Western cultures. — (4) Stands on one foot / script style. — (5) Caterpillars master multiple "hands" / styles. — (6) Celebrated migratory bird. — (7) It drank ink. — (8) Hides well. — (9) Adopts local customs / script styles, but ends modifying them / the data distribution. — (10) Stands out. — (11) Man's best friend. — (12) Of the electronic pointing device sort. — (13) Even algorithms watch lolcatz on the Internet. — (14) We known domestic fairies are there, but can't see them. — (15) Symbol of peace. — (16) Lays its eggs in other birds' nests. — (17) Admit: unremarkable sight. — (18) Reject: is a bird, but doesn't fly. — (19) Alarming, although harmless, like a black sheep.

Table 1. *The handwriting and quality control bestiary — In Roman characters the formal terms, in Italics the informal, and in bold writer ids of illustrative samples from the IAM dataset.*

Writing variability factors — *Handwriting variability results from complex interactions of a profusion of writer specific and contextual factors [5: 55–57, 41]. "Bad" writing is just the most familiar cause, the number of distinct "hands" mastered by an individual is another: polystylistic writers can vary both formal features, such as slant, and structural features, like allographs. Of particular interest for misclassification are within and inbetween writer variability, the consistency of the demographic make-up of writer population and dataset determining the amount and detectability of misclassifications. A case in point is the presence of foreigners, remarked by both this author and author Ha [26] in their respective datasets. Authorship verification poses peculiar problems when dealing with data obtained over long periods of time, such as in biometric, medical, or historical applications, because the transformation of script with biological and cognitive development. Conversely, short-term events, such as substance intake, stress, and illness, create outlier samples prone to misclassification. Little, if anything, in the metadata of handwriting datasets records, however, such events, complicating quality control. · Given that, by design, our method identifies writers with high script variability, it can double as a tool for retrieving specific demographics. Though, the capability raises ethical and legal questions. For instance, our system has thrust out of the mass of documents items ostensibly written by two foreigners, non-native Latin script writers. Would this have been a real life system, its potential for indiscriminate use would be blatant.*

Misclassification detectability — *Besides within writer variability, the second most important factor for the detectability of misclassifications is the disparity, or "contrast", between the endogenous and exogenous samples in a writer class. Cuckoos are more difficult to detect if genuine within writer variability is low and endogenous-to-exogenous contrast is weak. Similar difficulties arise for populations or datasets with low between writer variability. Different observers have also different detection capabilities. Documents, individuals, and populations can be further categorized according to their propensity at being correctly identified in terms of authorship. The schema, called the "zoo plot" and based on the "biometric menagerie" typology [27: 166–167], is useful to predict the performance of identification systems or take particular workflow and managerial measures, such as avoiding some demographics or submitting them to human expertise.*

Classification quality — *By introducing the term "cuckoo" for a misclassified item, one might feel curious, or in need for, the nomenclature for what is not cuckoo and the different sorts of cuckoos. This typology is provided here in the form of a confusion matrix for misclassification types.*

produced intentionally, for legitimate reasons (scientific experiments, such as in this paper) or out of malice (e.g. forgeries, vandalism). For other origins of cuckoos see also [57: 5–22, 48: 157, 10: 26–28].

Typology — A typology of writer misclassifications enriches our understanding on their genesis, prevention, and rectification. Table 1 provides such a classification schema and Fig. 3 supplies examples. To explore the dataset used in developing the typology, the reader is encouraged to use the online handwriting features browser Rex, created by the author for navigation within handwriting datasets. In regard to the adopted nomenclature, we were guided by the principle of keeping terms short, clear, and memorable. Insofar as many concepts have no established names, we shadowed the formal, cumbersome terms, with names of (mostly) animals, following in this a custom from the early days of speech processing [25], recently adopted in biometrics [27: 161–180] and with echoes in criminalistics [71: 150].

Detectability — The detectability of cuckoos depends on the coupling effect of the sensitivity of the control *instrument* to misclassifications, the *scrambling* pattern of which and how many documents are misclassified to which classes, the *dataset* structure, the *producer* workmanship [11: 35], the *sample* representativeness, the variability of the underlying demographics of the writer *population*, and the *text combinatorics*, i.e. the visual pattern resulting from the interaction of language, script, orthography, and content [2]. Section 6 will further discuss how these layers affect performance evaluation.

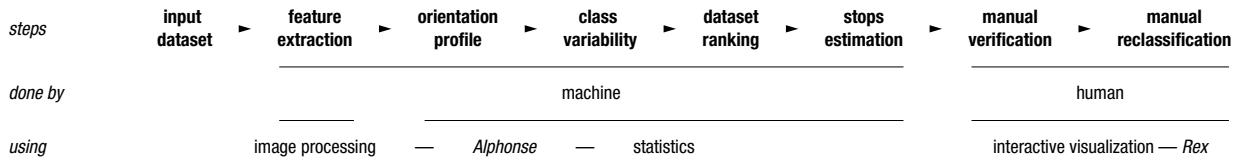


Figure 1. *Process diagram, elements, actuators, and capabilities of the Alphonse Rex quality control system.*

Severity — The primary aspects to consider when assessing the implications of identity misclassification are the *stringency* to detect them, the *tolerance* range of accepted failure, the *affordance* to meet these requirements with the available resources, the *proneness* of populations and producers for misclassifications, and the *prevalence* of the phenomenon [83]. The importance of the issue is area and task dependent, ranging from critical for biometric, forensic, and medical applications, to benign for art historical studies.

Workflow — Data quality control is managed at the *organizational* level of data producers or consumers, *architectural* level of the adopted control strategy, and *computational* level of software implementation [74: 8–10]. By combining the various references in the quality control literature, a generic procedure for the control of misclassifications emerges, comprising five steps: (*1*) *assessment* of the errors' impact for the given use case, (*2*) *procurement* of necessary control methods and tools, (*3*) *preventive* action by integrating quality control in the dataset production process, (*4*) *post-production* error detection and correction, and (*5*) *reporting* the quality control procedure and performance. Steps 4 and 5 are relevant to this paper and are expanded upon hereafter.

P4 – 60

4. Method description

Method summary — The backbone of our quality control method consists in using image processing and statistical analysis to automatically rank writer classes according to within writer variability, then manually inspect a number of the top-most classes for misclassified items with an interactive visualization tool. Fig. 1 gives a schematic overview of the quality control process: extraction of a writing feature, measurement of class variability, estimation of stops, check for misclassifications within stop limits of those classes with high variability, and reassignment of misclassified samples to new writer classes.

Feature measurement — The feature is the local orientation along the writing contour, an angular value, given for each pixel of the contour, in respect to the handwriting's horizontal baseline, assumed parallel to the image abscissa (Fig. 2). It is obtained by convolving the binarized contour image with a bank of gaussian filters of size 30×30 pixels, sigma 2 and 0.5, and 1 degree orientation step [7]. The probability density function of the dominant orientation angles is a vector, the *contour orientation profile* of a script. It tells us how long a writer moved the writing instrument in a given direction. Fig. 2 provides an example of two documents attributed in the IAM dataset to the same writer, a classification to which we were alerted by the difference of their orientation profiles, leading to the discovery of distinct signatures on the original documents, and thus to a cuckold writer class.

The rationale for having selected orientation as feature is as follows. The scope of this paper is to prove a concept – quality control by exhaustive writer verification –, not find the best possible feature for this task. So from this point of view any well performing feature would do. As evident from Table 2, there are anyways to many features

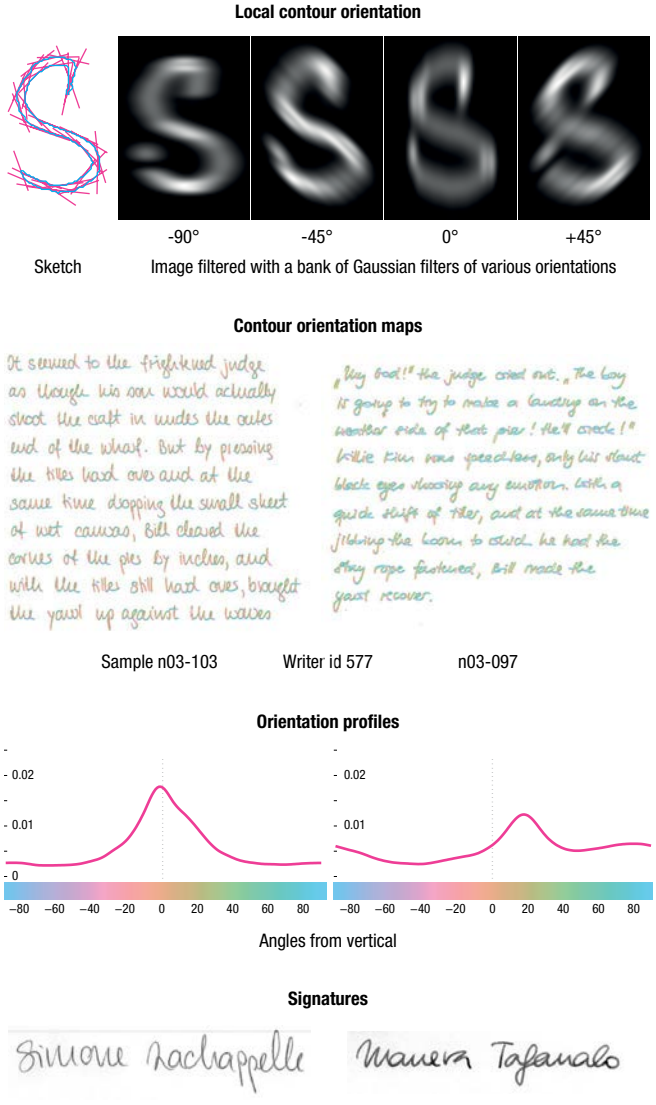


Figure 2. *The orientation feature, its profile, and its analysis — A sketch illustrates the concept of local contour orientation for a single character. The anisotropic filtered images offer a peek into the process of extracting the orientation angles. Once they are obtained for each pixel along the contour, the result can be visualized by color coding the document. The orientations distribution gives the probability density function shown by a graph, where the zero angle is relative to the vertical axis of the image. — These specific handwriting samples are an example of cuckoos. The quality control system has detected a substantial difference in the measured signal, the orientation profile; visual comparison of the two samples reveals, indeed, a stylistic distinctiveness. A verification of the signatures apposed on the original forms proves the machine and human assumption: despite being labeled with the same writer id in the dataset metadata, the samples belong to different authors.*

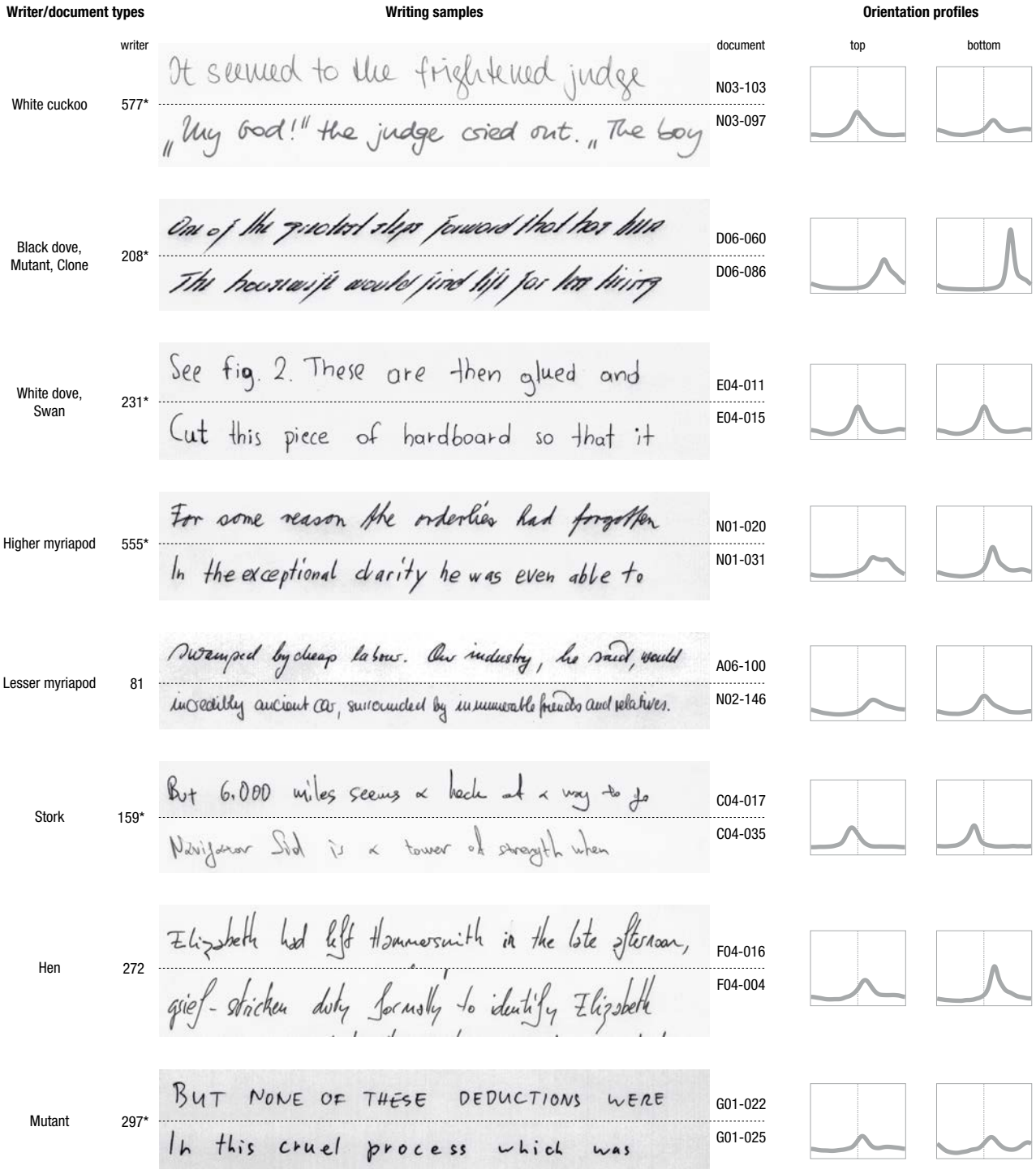


Figure 3. *Typology samples — The dataset is IAM 3.0; samples are cut-out of larger paragraphs; writer class id left of the samples, document id to the right; on the left hand is the nomenclature defined in Table 1; on the right hand the orientation profiles of the paragraphs to which the top and bottom samples belong. An asterisk marks documents signed by their respective writers. E.g., the documents of class 577 are signed with two different names, proving that this is a cuckold class; while class 555 exhibits the same name in its items, comforting the graphonomical expertise hypothesis of a polystylistic writer class. When writer 555 makes structural style changes, such as using different allographs, writer 81, also a myriapod, morphs superficial parameters, in this case the slant. Class 208 appears to the algorithm to be written by two writers, possibly due to a difference in text combinatorics. Class 272 has one of the largest natural within writer variability in the dataset (rank seven in Fig. 6, with method M2), while class 231 has one of the smallest variability. Class 159 is signed with a Greek name – note the Latin "a" written as a Greek "α" alpha. The disparity registered by the instrument in class 297 is due to a change in writing case, not writer.*

to implement, not a few beyond present capabilities in computational graphonomics. Nevertheless, contour orientation was not chosen arbitrary. First, it has good descriptive performance, being one of the top writer identification and verification instruments of the state of the art [16, 77]; it is a widely used handwriting expertise feature and its properties studied [4]. Second, it is versatile: various statistical features of the orientation profile have perceptual correlates. The mode, for example, is indicative of the script slant, the profile's entropy relates to the script roundness, and the spread around the mode reflects the writer's consistency [7]. Under statistical scrutiny, contour orientation is more than a numerical vector, it is a prism that splits an abstract numeric datum into a variety of meaningful perceptual concepts. In this sense, there are more than one handwriting feature which are employed in this paper. The advantage of analytical–perceptual correlates is to make the quality control system usable beyond the computing community, by users who don't necessarily have a mathematical background.

Within writer variability — The orientation profiles allow us to estimate the stylistic variability (or consistency) of a writer, by using some measure of distance between the handwriting samples. An example of profiles for a single writer is illustrated in Fig. 4a. In the following we explain the process by which we chose our measures and outline the algorithm for their use.

Choice of methods — Our goal being to find exogenous items in a collection of distribution vectors, we need a measure that emphasizes the variability within classes. This can be variously approached as a problem in outlier detection [10], distance measurement [23, 18], or shape description [30, 58, 28]. From the aforementioned literature we selected three basic descriptive statistics of dispersion (*summation*, *standard deviation*, *l2-norm*), four well known pair metrics (*chi-square distance*, *cosine distance*, *earth mover's distance*, and *dynamic time warping*), and a data preparation technique (*Fourier transform*).

The choices are motivated by their adequacy to the nature of our data and the goal mentioned above, and their profusion is a sign of complementarity. Observing that orientation profiles are circular, suggests the use of the frequency domain, where regularities are easily analyzed. Given that the data represents handwriting and handwriting is deformed in an "elastic" fashion, makes dynamic time warping and earth mover's distance natural choices to measure variability, the very reason for which the two methods were created. The sensitivity to outliers of summation, standard deviation, and l2-norm is why they are convenient in quality control. The chi-square and cosine distances are chosen for the more prosaically reason of being standard methods of vector distance measurement.

To design an integrated method we proceed along two ways. On one hand we find the best methods by experimentation. On the other hand we use ensemble methods to fuse the results of the base methods (with the *Borda count* for sum of ranks and normalized magnitudes [13, 49, 69, 70, 81, 46]). This allows to both circumvent the dilemma of method selection and improve performance.

Algorithm (Fig. 4 bottom) — (*1*) *Domain selection*: Use as starting point the probability density function orientation vectors or the magnitude of their Fourier transform. (*2*) *Matrix to vector reduction*: *Point-wise variant*: Make a matrix whose rows are the orientation vectors; then apply to each column the standard deviation or l2-norm statistics to obtain a vector. *Pair-wise variant*: Compute the pair distance between orientation vectors using chi-square, cosine, earth mover's, or dynamic time warping. (*3*) *Vector to scalar reduction*: Apply summation, standard deviation, or l2-norm to the result of step 2, to get the writer class variability.

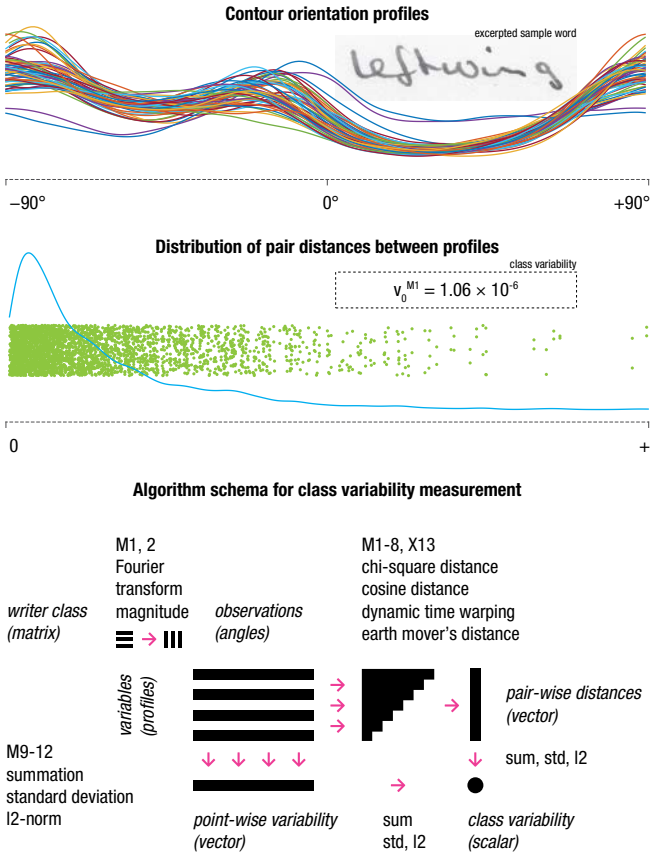


Figure 4. *Within writer variability* — Top: *Contour orientation profiles of the 59 handwriting samples of writer id 0 of the IAM dataset, an unusual case of left-slanting.* — Middle: *Distribution of pair distances between the profiles obtained with method M1 of Fig. 6; values as green dots, density as blue line, numerical value of class variability.* — Bottom: *Graphical representation of the algorithm for class variability measurement, consisting in a matrix to vector to scalar reduction.*

Inbetween writer variability — Once we got the variability for each writer using the above methods, we simply rank them and obtain the writer variability profile of the entire dataset (Fig. 5). We expect the cuckoos to lie in the top classes, on the presumption that writers have consistent writing styles, i.e. low variability.

Ranking stops — Before we start to manually check for misclassifications, there is a last automated step: determine when to stop the manual check of top ranking classes. We devised a generic rule, highlighted the role of interactivity, and sketched a model-based method.

Generic stops — Looking at the dataset writer variability distributions (Fig. 5) we observe that, except for the low performing method M12, few writer exhibit very high handwriting variability. By taking as stops analytically meaningful values, e.g. multiples of the standard deviation, we hope to isolate high variability writers from the rest of the population. Fig. 6 confirms that cuckoos in the IAM dataset fall within the range of the first two statistical stops, again, with the exception of M8 and M12.

Depending on the available amount of time and desired quality control precision, the user can chose at which level to stop the manual check – knowing that one should ideally check as many classes as possible. The stops suggested here are an approximation of the 3 sigma rule used in process control applications [20: 69–70].

Subjective stops — It is critical that the user be not blindly guided by inflexible thresholds, but takes advantage of the information on the whole dataset variability distribution revealed by its visualization. Note the distinction between "when" and "were" to stop. The more a class has members, the more time it takes time to verify them and less long will be the affordable total number of checked ranks.

Model based stops — Two types of information could help us refine setting the stops: knowledge of the natural writer variability in the sampled population and knowledge of the misclassification propensity of the human or machine classifier having produced the dataset.

The former demands sufficient data for inferring a model. Sampling theory could be used to estimate how much exactly (one study estimated at 1,200 the minimal sampling number representative of the US writer population for forensic writer identification [79]). The real problem, however, is not one of quantity, as of sampling method, given the rudimentary knowledge of handwriting demographics [5: 185].

This issue leaves us unsure about if and when a variability model based on one dataset realistically models any other. For instance, the qualifier "Swiss" for the IAM dataset has to be understood in the light of what it really samples. The unsuspecting user might think it as representative of the Swiss population, while someone aware of the diversity characteristic of this tiny country, will approach it with more watchfulness. Indeed, the styles of scripts taught to children in Switzerland varies with canton, school, and even individual teacher [22]. To clarify the matter, the author conducted interviews with the dataset producers [54]. It appears that specimens were collected mostly from residents of canton Bern, friends and colleagues, their extended families (id 472–475), some foreigners, including one "Gorbatchev" (id 51) and "Zorro" himself (*sic*, id 607). Furthermore, sizable batches were written outside Switzerland, in Greece, France, and Germany, as well as a piece "written on the train" (id d05-021). The language of the text to copy, English, was with few exceptions not the first language of the participants, and, for some, even Latin was not their everyday script. Production was guided by a ruled page positioned under the writing form, adding to the graphonomical constraints. In conclusion, the dataset is heterogeneous in terms of participants origins, motivations, and acquisition settings. The question is then what value has its comparison with other existing handwriting collections, Dutch, Greek, American, Arabic, or Chinese?

Manual check — The stop ranking allows, in signal theory terms, to *detect* the presence of misclassified documents in writer classes; the visual inspection carried out in this step, *confirms* and *identifies* specific documents as such. Sometimes, clues demanding no handwriting expertise are enough to take a decision: signatures on the original documents of the IAM dataset weighted strongly in favor of handwriting samples being considered cuckoos, or not. Formal handwriting expertise makes, however, better informed results. In this case, the author is a trained paleographer and calligrapher. Table 2 lists handwriting features examined during the manual check, with their use frequency. Inspection was greatly facilitated by the tool at our disposal, the handwriting dataset visualizer Rex.

Reclassification — Reassigning discovered cuckoos to their genuine membership class could be considered an activity beyond quality control and, insofar as it is a identification task, a new start in the classification–verification loop. For automated methods, see the appropriate writer identification literature, for example [15].

Here, reclassification was performed semi-automatically, using Rex. We first ranked the IAM dataset by slant, which corresponds to the statistical mode of the orientation profile and is the most discriminant among the profile's features [7: 631]. Then, we searched

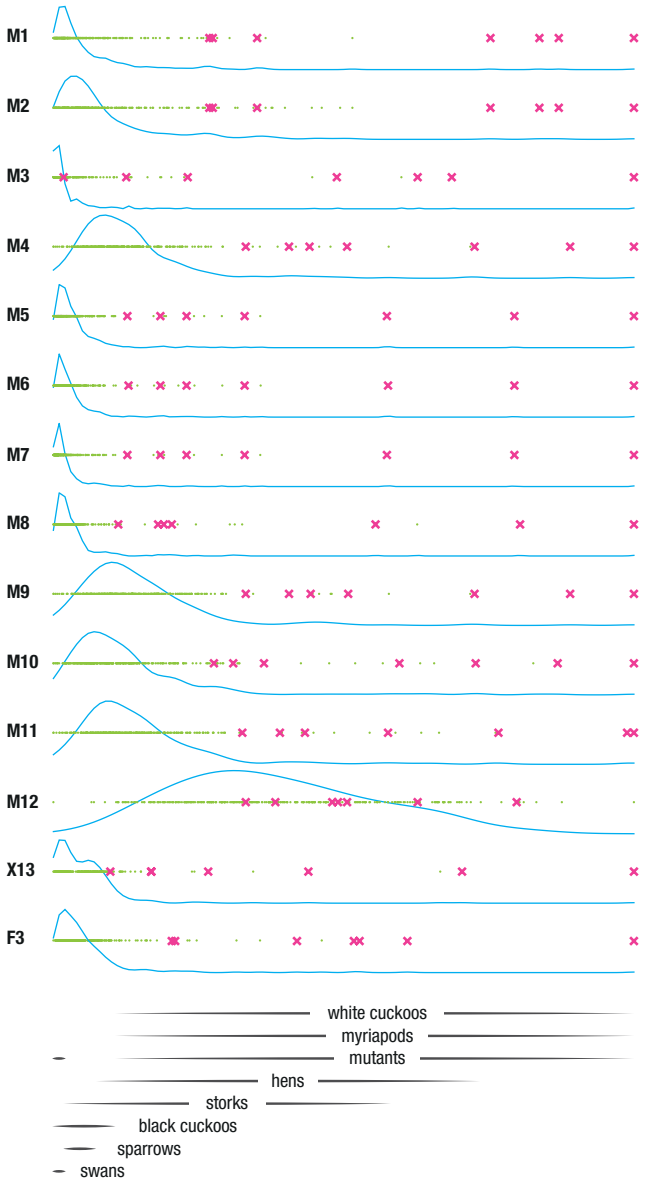


Figure 5. *The handwriting variability of the IAM dataset writers — The graphs represent the distribution of within writer variability according to the experimental methods listed in Fig. 6. The x-axis gives the variability magnitude for the 301 classes with at least two handwriting samples (green dots); the blue line is the probability density function of the distributions, normalized to fit the same area for illustrative purposes. Cuckoos are marked by red crosses. — At the bottom the approximate location of various demographic strata, according to the nomenclature of Table 1, which concur to give the distributions their particular shapes.*

P4 – 63

in the neighborhood of each detected cuckoo for possible identical writers. None were found and we conclude that there are, in fact, 664 writers in the IAM dataset, instead of 657 as suggested by its metadata.

Software — Image processing and statistical analysis implemented in the Alphonse system are custom functions written in the Matlab programming language. The interactive visualizer Rex is a web application written server-side in PHP and utilizing a MySQL database, and client-side in HTML, CSS, and JavaScript, with hyperlinks to the online IAM dataset.

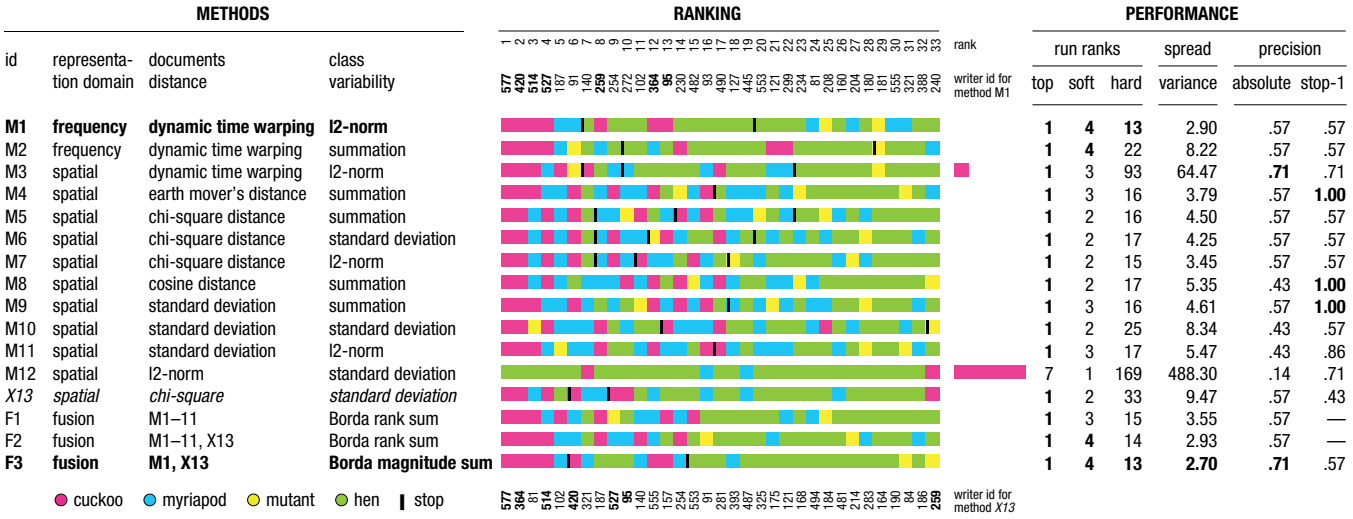


Figure 6. *System performance — The figure shows the methods tested in this paper, with the data representation domain and metric names for measuring the distance between documents and the class variability; the best performing methods are in bold. Method X13 represents a third party system. The writer ids are given for method M1; note that they change across methods; cuckoos beyond rank 33 are aggregated left of the main bars. The writer typology is indicated by color codes; refer to Table 1 for definitions. Proposed check stops are marked by bars between ranks. In the performance table are given the extremities of the soft and hard runs in term of ranks. The spread is the ratio of observed and optimal rank variances of cuckold classes. The absolute precision indicates the number of cuckold classes observable within the top-7 classes, where the value "7" is given by the amount of genuine cuckold classes. The precision for the 1st stop is an estimation of how well the stops were defined, i.e. the ratio of cuckold classes found if stopping the manual verification at the proposed stop.*

5. Performance evaluation and insights

P4 – 64

Performance criteria and indicators — A quality controller has to know in which order to check writer classes for misclassification and how many of them to check. The better a system responds to these questions, the better is its performance, hence the choice of the following evaluation criteria. (*1*) *Rank* of genuine cuckold writer classes in the system output, expressed as rank of the first genuine cuckold class (*top*), rank of the last class in the first consecutive run (*soft stop*), and rank of the last ever genuine cuckold class (*hard stop*). (*2*) *Rank spread* of cuckold classes $RS = \sum_{i=1}^{n} r_i^2 / \sum_{i=1}^{n} i^2$, were n is the number of genuine cuckold classes and r their rank. The closer RS is to 1, the better the system performs. The denominator normalizes the actual system performance by the maximal performance, attained when all cuckoos are contingent at the top of the ranking. The measure of rank spread, a ratio of variances with mean zero, palliates to the independence of the precision measures from the distribution of classes within the considered range. In a single value it makes a synthesis of the other performance measures given here. Note its sensitivity to outliers, which favors control methods yielding shorter hard stops and faster completion of the quality control task. (*3*) *Absolute precision*, indicating the number of cuckold classes observable within the top-7 classes, where the value "7" is given by the amount of genuine cuckold classes. (*4*) *Distinctiveness* of cuckold classes, conveyed by the distance between ranks (Fig. 5) and measurable by the *precision for the first stop*, an estimation of how well the stops were defined, i.e. the ratio of cuckold classes found if stopping the manual verification process at the proposed stop. (*5*) *Separability* between explainable and non explainable class variability (i.e. cuckoos, myriapods, and mutants vs hen writers). It is preferable to have explainable handwriting variability among classes in the top ranks of the output, because such classes take less resources to discard during verification. (*6*) *Difficulty* of the quality control task, the special topic to which Section 6 is devoted.

Ground truthing — Seven cuckoos were manually identified in the IAM dataset (either of the two documents of writer classes id 95, 259, 364, 420, 514, 527, and 577). The hypothesis that classes reported here are cuckolds was confirmed by two police forensic experts and two graphonomics computer scientists collaborating with the author (but see [31] for the reliability of human experts).

Inquiries with the dataset producers revealed how the errors came about. Initially, the IAM dataset was meant to serve the training of a handwriting recognition system, so recording the writer identity was secondary. After it became of interest, it was realized that writers do not systematically sign their samples, making necessary a manual post-hoc classification for part of the dataset. At that point, we surmise, appeared the cuckoos. Forms were distributed to participants according to the id sequence and the distance between cuckoos of same class in that system is 0, 0, 2, 3, 6, 16, and 268. This logical and, possibly physical proximity in the document stack, might have been a confusing factor during classification.

The success of our system at detecting misclassifications demonstrates the utility of an additional quality control layer, independent of the dataset creators. Likewise, it portrays the advantages of combining human and machine verification, supported by interactive visualization.

Results — Fig. 6 supplies numerical performance indicators and permits comparative visual analysis of the experimental methods. The best method is F3, which gives the longest soft stops, shortest hard stops, and lowest spread, closely followed by M1. As it can be seen, the system ranks classes containing misclassifications in reasonable stop ranges, given the amount of writers with high variability in this dataset. The quality of the performance can be even better apprehended when looking not at the rankings (Fig. 6), but at the magnitude distribution (Fig. 5), where the cuckoos clearly appear as outliers.

Insights — (a) *In a handwriting expertise system the script feature is not to always the most important factor.* Some features have, both for human and machine experts, stronger discriminative power

than others, e.g. allographs vs run lengths [15, 16, 61, 3]. So we compared our system, which uses a good but not the best feature, contour orientation, with an award winning system (X13), based on the multiscale distribution of interest points in a handwriting sample, such as corners, junctions, and extremities ([32], best student paper at the International Graphonomics Society conference 2015). The fact that we fare better in performance, prompted the ensuing two remarks.

(b) *All elements of a handwriting expertise system are important.* Our experiments underscore the importance of the document distance metric to the expense of handwriting features. Additionally, the experiments M11 and M12 show that just by choosing different permutations of the same metric, performance can be significantly affected. Besides ranking, the order of classes is also modified.

(c) *Understanding the properties of the elements of a handwriting expertise system helps fine tuning its performance.* This truism is explained when we consider that even such a lowly feature as run lengths has its place in the expert's arsenal: it can model word spacing, what the mighty allographs can not (for examples, select "run lengths" in Rex). In respect to our experiments, the good performance of method M1 was achieved after realizing (α) that the circular nature of the orientation pdf allows for vector comparison in the frequency domain, which we knew to be a competitive shape distance measurement technique (the MPEG-7 multimedia standard incorporates frequency domain based shape description [58]), and (β) that translation invariance, a property of the frequency domain, has no adverse impact. A translation in the orientation profile corresponds to a change in handwriting slant [4: Fig. 5.1–5.2]. Due to the biomechanics of handwriting, slant is correlated to shear, which keeps profiles distinct (e.g., $\mathbf{O} \rightarrow \boldsymbol{O}$). Only mirrored styles will look the same to the frequency domain distance measures ($\mathbf{O} = \boldsymbol{O}$), but most Latin script is right slanted [4: Fig. 4].

(d) *Importance of manual check.* Long tailed distributions, as exhibited by handwriting variation, have natural outliers [60, 10: 37–38], due to polystylistic writers and random text combinatorics. A more imaginative handwriting expertise than usual is required to establish their genuineness and separate them from cuckoos (Fig. 6), strengthening thereby the rationale for manual check.

(e) *System overfit.* Studying the similarities and disparities of the performance indicators, the question arises whether the system is not overfitting the data, and what is the part of randomness in the apparent fluctuation. One way to approach the question is *brute force* method evaluation with many systems, on many datasets. The comparison with a third party system, X13, alleviates overfitting apprehensions; while multiplying the tested datasets has the drawback of using handwriting samples about which it is unknown what exactly they represent within the wide variety of writer populations and writing conditions. Another way, presented in the next section, is by *simulation* of misclassifications in a controlled setting in a single dataset and studying the system's response.

(f) *Limits.* A "masking effect" between a cuckoo and members of its host class occurs when the stylistic *contrast* between handwritings is too low for the misclassification to be detected [10: 40] (Table 1). Such cases are difficult to sort out, but the issue is mitigated by the fact that it is more stringent to detect outlier misclassifications, since they reduce more strongly the accuracy of expertise systems, than exogenous items with low within class contrast.

(g) *Handwriting variability for both a single writer and a population appears to follow roughly a log-normal distribution.* It is well known that many types of rapid human motions, such as produced during handwriting and signature, can be modeled with log-normal

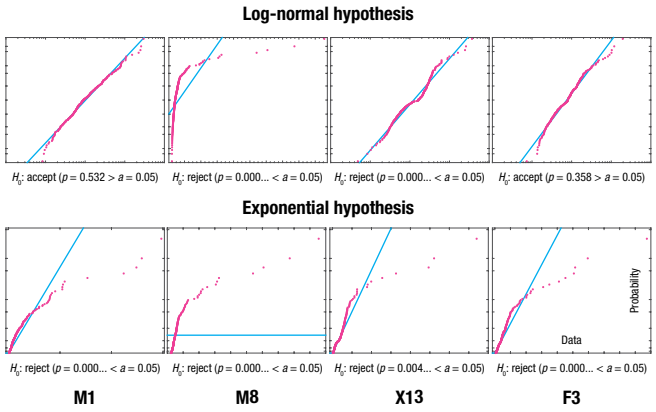


Figure 7. *The nature of the inbetween writer distribution — Probability plots of data from Fig. 5 tested against fitted log-normal and exponential distributions. The more the data dots fit the line, the better the adequacy between distributions.*

distributions [66, 67, 24, 73]. Looking at Fig. 4b and 5, it is interesting to observe that, for entirely other reasons, this distribution type manifestly applies equally to within and inbetween writer variability of contour orientation. To test the hypothesis, we fitted the data to a log-normal and an exponential distribution [55], respectively, and performed an Anderson-Darling goodness-of-fit test between the empirical and fitted data [63]. The null hypothesis is supported for the log-normal, with the exception of methods M8 and X13, and rejected for the exponential. Visual confirmation is obtained from probability plots (Fig. 7). How can this phenomenon be explained and what are its implications?

Writing by hand is similar to physical sports in that the writer strives to achieve and maintain biomechanical consistency. Assuming a statistical normal performance distribution, its left side is positively skewed away from zero mean variability because of the difficulty of near-perfect results, reserved to the – nowadays vanishing – elite of penmanship and calligraphers. Because it is easy to loose concentration on the quality of script while writing, especially long texts, a process of decay stretches the right side of the distribution. The two writer specific forces are supplemented by external ones. The left-hand skew is amplified because identical results can't be expected when the written text, the writing implements, and the contextual distractions perpetually change. Similarly, rare events such as deletions, introduce outliers on the right of the distribution. When considering now the handwriting variability in a population, its distribution shape is influenced by the mixture of demographic niches. While the mass of writers maintain a certain regularity in writing, there are as well individuals with extreme consistency, masters of multiple styles, adult learners of foreign scripts, or simply sloppy writers, sometimes involuntary subjects of random events, like writing produced on a train.

A practical implication of the log-normal handwriting variability distribution among a broad writer population is that in expertise applications the human or machine expert should be aware of the real possibility of genuine outliers [60]. At a theoretical level, the above insights allow the creation of models of handwriting variability. In respect to quality control, the log-normal distribution explains some aspects of our system performance, namely the ease of cuckoo detection (note in Fig. 5 that with little exception cuckoos are outliers) and their coexistence with genuine outliers in the top ranks (see the mix of color coded writer types in Fig. 6).

P4 – 65

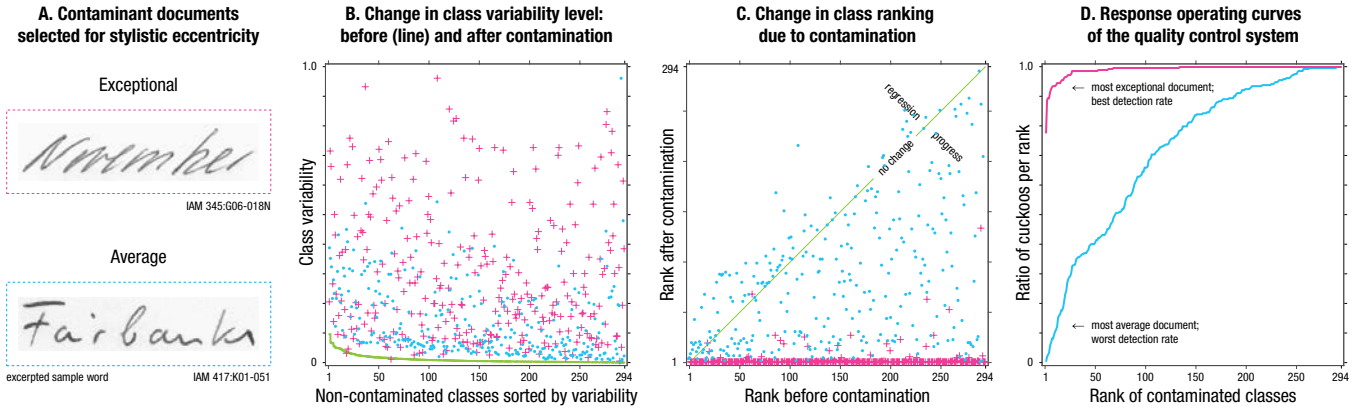


Figure 8. *Impact of dataset contamination with high and low contrast documents on quality control performance — Unusual documents (a, top) induce a high class variability (b) and push up class ranks (c), improving quality control performance (d) more than average documents (a, bottom). — The continuous line in the (b) diagram represents writer classes with more than two samples in the IAM dataset corrected for misclassifications and ranked by level of variability. The markers give the amount of within class variability in the same ranking order, after contamination with the documents shown in (a). Rearranging the classes according to the new variability, lets us compare the two distributions (c). When a class has "progressed" in rank, it means that the cuckoo it contains can be found more quickly. The degree of improvement in quality control performance can be judged by the vertical distance of markers form the diagonal. Clearly, contamination with highly individual samples produces far more detectable cuckoos (the markers are close to the ordinate). The analysis is confirmed by the response operating curves of the system, which compute the cumulative amount of cuckoos per rank using method M1 (d).*

6. Task difficulty and prediction

In the previous section we evaluated the quality control system Alphonse relative to an a priori unknown number of "wild cuckoos" occurring naturally in the IAM dataset. Would the performance be the same if the misclassified documents were other? In other words, *how difficult was the task*? To answer the question, controlled experiments of "artificial insemination" with known "laboratory cuckoos" will be carried out in the following. They will allow to investigate the possibility of a more generic performance measure, that proves the soundness of our concept beyond its realization in the specifics of a system and a dataset.

Performance in a coupled environment — We postulate that the performance of a quality control system can be fully understood only in reference to the task difficulty. To exemplify this, let us look at the dependence of quality control on dataset structure, by performing an *endogenous contamination* experiment with selected "Eastern Eggs" (Fig. 8). First, we retrieve the two stylistically most exceptional and average, respectively, handwriting samples, by finding, with method M1, the documents with the maximal and minimal sum of distances to all other documents in the dataset. They represent the best and the worst performance we can expect from our system on this dataset. Subsequently, these two Eastern Eggs are hidden in turn in each class except their own and the change in class variability and rank measured. We build a response operating curve (ROC) that for each rank k gives the precision p_k of the quality control as the number c of Eastern Eggs appearing in the top-k ranks weighted by the total number n of writer classes minus one: $p_k = (\sum_{i=1}^{k} c_i) / (n-1)$. As expected, the performance curve of the average document is inferior to that of the exceptional document, confirming that contamination with high contrast documents is easier to detect than with low contrast items (Fig. 8d).

Please consider that the "exceptional document" is so from the vantage point of the measurement instrument. A human observer might have selected, for example, writer id 122 as the most idiosyncratic: its stretched intercharacter ligatures (include) are typical of Arabic script, not Latin.

Layered performance evaluation — Given the interdependence demonstrated above, the question arises about what exactly is evaluated: the system sensitivity, the data complication, or something else? If we consider the quality control system to be the articulation between data and user, then it is surrounded by a sequence of layers modulating its performance (Fig. 9). Data-side, the performance depends on the *dataset "complication"* (in the horology sense of the term), which is a result of a number of factors contributing to the emergence of misclassifications: dataset *producers excellence*, dataset *scrambling severity*, *writer variability*, and *text combinatorics*. User-side, performance is framed by the *stringency* of the use case, its *tolerance* range, and the *affordance* of the system and user to meet these requirements. (These factors determine the selection of check stops, i.e. how many classes are to be manually checked for cuckoos.) Finally, at the center of data–user interaction lies the *sensitivity* of the quality control system to detect misclassifications. The observed quality control performance fluctuates with all these parameters. In conclusion, *it seems necessary to distinguish between system sensitivity, data complication, and use case constraints, when addressing quality control performance*.

Relative grading — To grade a quality control system we would need a set of reference criteria. However, the class variability can vary considerably given the number of layers affecting it – especially due to the randomness of scrambling – and there is no ready list of criteria telling why a handwriting is more easy to expertise than another. A solution to the *first problem* proceeds by "peeling" away the contribution of each layer. After performing quality control we obtain a version of the dataset hopefully unbiased by scrambling. The amount of scrambling is indicative of the producers excellence, an information also derivable from the quality of other datasets of the same producer. The level of masking in the unbiased dataset characterizes the difficulty of any system to deal with this specific set of handwriting samples, and, in extenso, the difficulty created by a writer population, assuming that the sample is statistically representative. A solution to the *second problem* – grading performance without a reference scale – consists in comparing the performance

P4 – 66

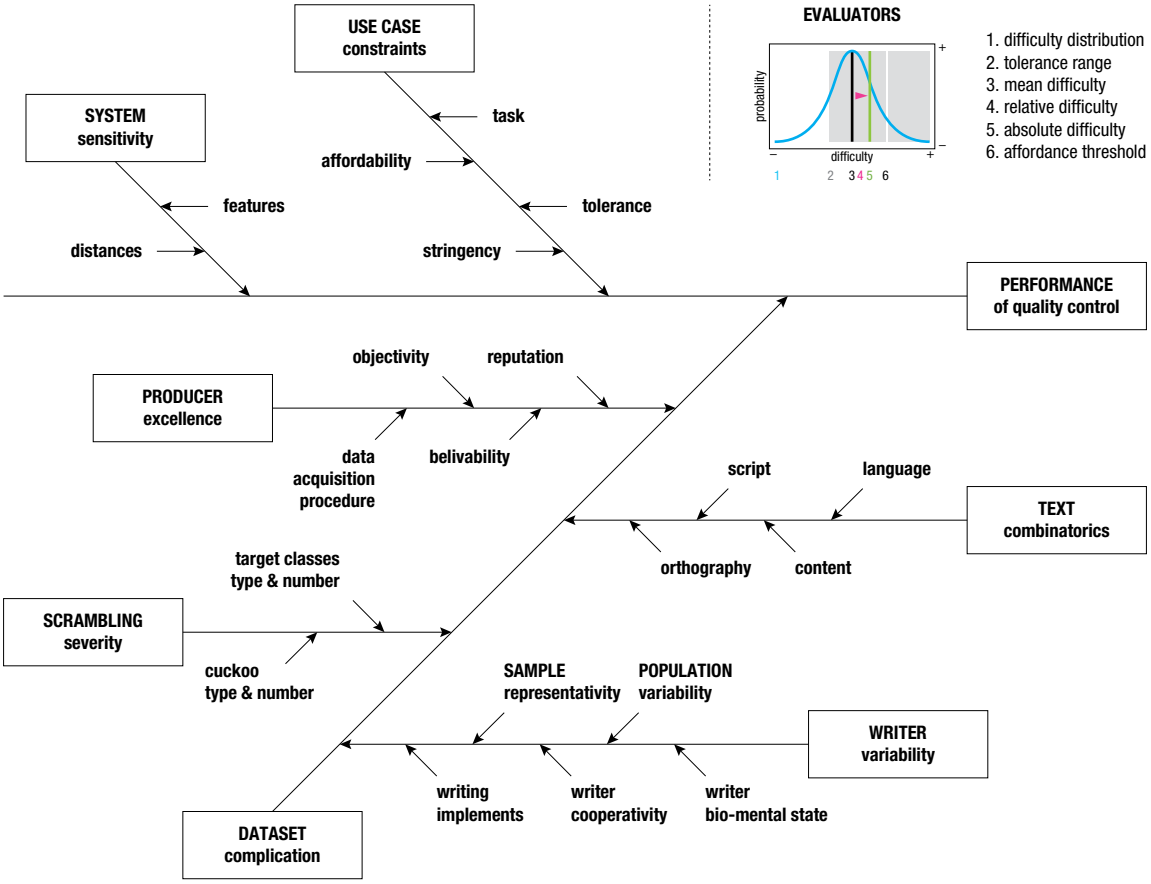


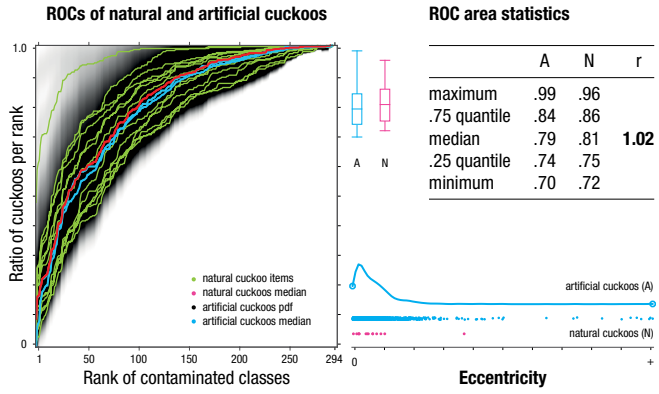
Figure 9. *Performance evaluation factors — This cause-and-effect diagram shows how quality control performance is affected by a multitude of factors. The concept of the inset diagram can be applied to each factor to evaluate its degree of "difficulty" (the data is fictitious). To evaluate, for example, the relative performance of a quality control system, we estimate the distribution of the performance of other systems (blue curve) and compute the distance (red arrow), in standard deviation units, between its mean (black line) and the absolute performance value of the evaluated system (green line). Its performance can also be evaluated in regard to whether it is within the use case tolerance range (gray area) and below the affordance threshold (white line).*

of the evaluated system with that of many others: a well-performing system is one that is at least as good as the majority of other systems, within the frame of the same set of data and use case parameters. Fig. 9 shows, for fictitious data, how the relative grading could be visually represented. Next we turn to obtaining numerical indicators for the scrambling severity and sample difficulty.

Prediction — What does it mean that a population or dataset is "prone to misclassifications"? When there is high in-class variability and low between-class variability, then cuckoos are easily masked by genuine outliers, humans and machines can easier make classification mistakes, and it is more difficult to establish good quality control stops. These characteristics create conditions favorable to the emergence of cuckoos. The question to consider is how to predict the misclassification proneness. We also have to keep in mind that this is a distinct question from the measurement and prediction of the dataset scrambling severity.

In order to evaluate the proneness for misclassification of a dataset unbiased by cuckoos, it would have to be scrambled in all possible ways [10: 37–38]. As an approximation, a sample could be drawn from the totality of permutations. For instance, we could extend the Eastern Eggs experiment by systematically contaminating *each* writer class with *each* item of the dataset, not just the most exceptional and most common (Fig. 10). The precision of quality control after this operation is inverse-proportional to the *masking potential of the sampled population* and represents the *absolute quality control performance* to expect from this system–dataset pairing. Because the method scrambles only one cuckoo egg at any given time, it produces an *optimistic* estimation: its value would decrease with an increase of contaminant documents and contaminated classes. If the performance of other systems is known, these values let us compute the *relative challenge to quality control*. Since outcomes depend on the selected script features and variability measurements, the same instruments have to be used for benchmarking the performance of different systems and difficulties of different dataset.

To compare two datasets X and Y we rely on the ratio r of the medians m_x and m_y, respectively, of the set of areas A_x and A_y under the ROCs: $r = m_x / m_y$. Depending on the nature of the compared datasets, a result greater than unit indicates (*1*) *a low scrambling severity* or (*2*) *a lower difficulty for quality control of dataset* X *against* Y. The IAM dataset appears to have both a medium scrambling severity (r = 1.02) and a medium structural difficulty (large spread of ROC curves) (Fig. 10). A look at the location of cuckoos on the centrality distribution in the same figure suggests that, with one exception, they are not outliers, that is, they could have occurred with relative ease.



	A	N	r
maximum	.99	.96	
.75 quantile	.84	.86	
median	.79	.81	**1.02**
.25 quantile	.74	.75	
minimum	.70	.72	

Figure 10. *Scrambling severity and dataset difficulty* — Left: *The grayscale bitmap represents the distribution of 1539 × 294 ROCs for the exhaustive scrambling of the IAM dataset (contamination by each document of all writer classes having more than two samples). It is obtained from pdfs of ROC values for each rank and informs us about the dataset difficulty. Superposed is the median ROC (blue) and the 14 ROCs of natural cuckoos (green) and their own median (red). By comparing the ROC distribution of artificial (A) and natural (N) cuckoos we can estimate the scrambling severity.* – Right top: *Statistics of the area under the ROCs.* — Right bottom: *Eccentricity of individual documents in respect to all others, as the sum of pairwise distances measured with method M1 (blue). The two documents in Fig. 9 are the most central and eccentric and coincide with the graph extremities. Cuckoos appear in the densest part of the distribution (red).*

SCRIPT	**general**	feature variability — feature use context — likelihood: **converging indices are mutually reinforcing**
	shape	**allographs of characters** — **allographic sets** (e.g. upper- & lower case) — size — vertical alignement lines (e.g. baseline, **x-height**): variability, proportions between lines — affine transform: **slant**, stretch — stroke: shape, weight, contrast, chroma — inline brokenness — outline irregularity — extremities shape — roundness — trace fluidity — **ligaturing** — spacing: characters, **words**; kerning, tracking — surface: texture, color — 3D imprint of writing implement — paragraph: margin alignment, indentation, interline spacing, columns amount & spacing — rivers — crossing-outs — hyphenation (rules, number of consecutive hyphens) — **regularity** — quality
	layout	shape outline — location — dynamism
	text	an abundance of personal names increases the number of upper case letters, modifying the script texture
CONTEXT	**physical**	paper quality explains script quality
	semantic	word meaning explains its shape
LOGIC	**heuristics**	demographics explain script style
	likelihood	converging indices are mutually reinforcing

Table 2. *Elements of human handwriting expertise — Lists of graphonomic expertise elements [41, 5: 128]; in red those effectively used by us for groundtruthing.*

7. Future research

Multiple features systems — What would it take to fully automate quality control? For handwriting recognition, the creators of the NIST dataset have convincingly argued in favor of semiautomatic solutions as being faster and more reliable. Their work brought them to the startling conclusion that "a quoted accuracy rate for a set of segmented characters is meaningless without reference to human performance on the same set of characters" [86: 1, 6–10]. To address the question for writer verification, we looked at the elements that went into the human expertise deployed for cleaning datasets infested with cuckoos.

We found that all decisions were based on a limited set of handwriting features and our decisions were strengthened when several indices concurred (Table 2). Only a single document could not have been possible to attribute to the correct writer given only sample dissimilarities, were it not for an identical signature on two forms (Fig. 3, id 555; a fact check revealed two homonymous students by that name at IAM, but from different points in time). Logic and contextual knowledge (e.g. on the dataset demographics) didn't play any role in our decisions (but it might do for other datasets and use cases). There were instances a writer using exclusively uppercase letters in one sample and mixed cases in another (Fig. 3, id 297), or a prevalence of some letters due to the content that made two samples by the same writer look dissimilar to our measuring instrument (Fig. 3, id 208). By reasoning we understood from where the disparity came, but the factors where not something not detectable by state of the art algorithms. *We conclude that it seems possible to substantially improve the quality of authorship control using a multidimensional handwriting features space.* Indeed, multiple classifier systems gave good results in other handwriting expertise domains, such as writer identification [44, 39], and had in this very paper positive impact on performance.

Probabilistic approach — The layered performance factors discussed in the last section and graphically represented in Fig. 9, strike as apt to be modeled by a probabilistic network. The probabilistic approach to quality control – following the Bayesian [45] or likelihood paradigms [72] – would indeed allow the evaluator to take into account, in a flexible way, the richness of factors that shape performance. Presently, however, the paucity of data and theory precludes such promises. For example, our knowledge of handwriting demographics is too limited to define priors and likelihoods, as is our ability to quantify the difficulty of various expertise tasks, such as quality control, verification, identification, classification, or retrieval.

In truth, this work does have a Bayesian twist. Our belief in the quality of the system's performance was updated to include information ("para-data") not present in the data measured by the contour orientation instrument or in the dataset metadata. First, the signatures revealed non-native Latin writers, whose high handwriting variability masked true cuckoo classes. This contributed to the necessity of developing a writer typology and prompted a reconsideration of how performance could be measured. Second, interviewing the dataset producers illuminated the unexplained complication of the IAM dataset and what to expect from it in terms of sample representativity. Last, the development of the quality control method itself, predicated by the potentialities of the various technical choices tributary to the subjective approach of the author and feeding a lengthy stream of preprint versions of this paper, was an iterative Bayesian optimization process, reflected by the performance of the proposed method. The above remarks are meant to recall that the psychology and sociology of scientific research are also part of the solution to an engineering problem.

8. Conclusion

We presented a generic method to detect mislabeled script samples in handwriting datasets and an exemplary software system for quality control of writer identities. The method is appropriate when no machine learning is desired or possible. The procedure's essence consists in an automated ranking of writer classes by within-class handwriting variability, given some statistical measure of spread, followed by interactive manual inspection of a number of classes deemed potential harborers of misclassifications. Experiments were

conducted and misclassifications successfully detected; the performance of our system compares well with that of other. Performance is best evaluated considering system sensitivity, dataset complication, and use case requirements. Methods are discussed to measure and predict the difficulty of performing quality control on arbitrary datasets.

Acknowledgments

The dedication of the IAM dataset producers and their generosity in making it publicly available is what made our study possible. The few classification errors that were found are only a tribute to the quality of their work. — We wish to thank Marcel Würsch for sharing the document distance data and Andreas Fischer for fruitful discussions.

References

[1] J. Arlandis, J.C. Perez-Cortes, and J. Cano, "Rejection Strategies and Confidence Measures for a k-NN Classifier in an OCR Task," in *Proc. 16th Intl. Conf. on Pattern Recognition, Quebec, Canada, Aug. 11–15, 2002*, vol. 1, pp. 576–579.

[2] V. Atanasiu, *On letter frequencies and their influence on Arabic calligraphy* [*De la fréquence des lettres et de son influence en calligraphie arabe*], Paris: L'Harmattan, 1999.

[3] V. Atanasiu, "Allographic biometrics and behavior synthesis", *TUGboat* [*Proc. 14th European TEX Conference, Brest, France, June 24–7, 2003*], vol. 24 (3), pp. 998–1002, 2003.

[4] V. Atanasiu, "Forensic vs. Computing writing features as seen by Rex, the intuitive document retriever," in *Proc. First Intl. Workshop on Automated Forensic Handwriting Analysis, Beijing, China, September 17–18, 2011*, pp. 16–20.

[5] V. Atanasiu, *Expert Bytes. Computer Expertise in Forensic Documents: Players, Needs, Resources, and Image*, Boca Raton, FL: CRC Press, 2014.

[6] V. Atanasiu, "Alphonse: Handwriting Expertise Tools", *Matlab Central File Exchange*, software, 2015.11.25, http://www.mathworks.com/matlabcentral/fileexchange/?term=vlad+atanasiu.

[7] V. Atanasiu, L. Likforman-Sulem, and N. Vincent, "Writer Retrieval–Exploration of a Novel Biometric Scenario Using Perceptual Features Derived from Script Orientation," in *Proc. 11th Intl. Conf. on Document Analysis and Recognition, Beijing, China, Sep. 18–21, 2011*, pp. 628–632.

[8] V. Atanasiu, L. Likforman-Sulem, and N. Vincent, *Rex, a description-based retriever for written documents*, web application, April 2011, http://glyph.telecom-paristech.fr.

[9] H.S. Baird, "The State of the Art of Document Image Degradation Modeling," in B.B. Chaudhuri (ed.), *Digital Document Processing: Major Directions and Recent Advances*, New York: Springer, pp. 261–279, 2007.

[10] V. Barnett and T. Lewis, *Outliers in Statistical Data*, Chichester: John Wiley & Sons, 1978.

[11] C. Batini and M. Scannapieca, *Data Quality: Concepts, Methodologies and Techniques*, Berlin: Springer, 2006.

[12] J. van Beusekom, F. Shafait, and Th.M. Breuel, "Automated OCR Ground Truth Generation," in *Proc. 8th IAPR Workshop on Document Analysis Systems, Nara, Japan, September 16–19, 2008*, pp. 119–125.

[13] J.-Ch. Borda, "Mémoire sur les élections au scrutin," *Mémoires de l'Académie Royale des Sciences*, pp. 657–658, 1781.

[14] C.E. Brodley and M.A. Friedl, "Identifying Mislabeled Training Data", *J. of Artificial Intelligence Research*, vol. 11, pp. 133–169, 1999.

[15] M.L. Bulacu, "Statistical pattern recognition for automatic writer identification and verification," Ph.D. dissertation, Artificial Intelligence Institute, Univ. of Groningen, The Netherlands, 2007.

[16] M.L. Bulacu and L.R.B. Schomaker, "Text-independent writer identification and verification using textural and allographic features," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29 (4), pp. 701–717, 2007.

[17] M.C. Burl, U.M. Fayyad, P. Perona, and P. Smyth, "Automated Analysis of Radar Imagery of Venus: Handling Lack of Ground Truth," in *Proc. IEEE Intl. Conf. on Image Proc., Austin, TX, Nov. 13–16, 1994*, pp. 236–240.

[18] S.-H. Cha, "Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions," *Intl. J. of Mathematical Models and Methods in Applied Sciences*, vol. 1 (4), pp. 300–307, 2007.

[19] S.-H. Cha and S.N. Srihari, "Multiple Feature Integration for Writer Verification," in *Proc. 7th Intl. Workshop on Frontiers in Handwriting Recognition, Amsterdam, The Netherlands, September 11–13, 2000*, pp. 333–342.

[20] Committee E-11 on Quality and Statistics, *Manual on Presentation of Data and Control Charts Analysis*, West Conshohocken, PA: ASTM. 2002.

[21] C. Cortes and V. Vapnik, "Support-Vector Networks", *Machine Learning*, vol. 20 (3), pp. 273–297, 1995.

[22] Deutschschweizer Erziehungsdirektoren-Konferenz, Arbeitsgruppe Schrift, *Entscheidungsgrundlagen zur Zukunft der Schweizer Schulschrift*, Luzern, 2013, http://www.basisschrift.ch/kantonale-informationen.

[23] M.M. Deza and E. Deza, *Encyclopedia of Distances*, New York: Springer, 2009.

[24] M. Djioua and R. Plamondon, "Studying the variability of handwriting patterns using the Kinematic Theory", *Human Movement Science*, vol. 28, pp. 588–601, 2009.

[25] G. Doddington, W. Ligget, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, Goats, Lambs and Wolves. A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation," in *Proc. Intl. Conf. Spoken Language Processing, Sydney, Australia, Nov. 30 – Dec. 4, 1998*, paper 0608.

[26] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, Hoboken (NJ): John Wiley & Sons, 2000.

[27] T. Dunstone and N. Yager, *Biometric System and data Analysis. Design, Evaluation, and Data Mining*, New York: Springer, 2009.

[28] I.L. Dryden and K.V. Mardia, *Statistical Shape Analysis*, Hoboken (NJ): John Wiley & Sons, 1998.

[29] G.A. Fink and Th. Plötz, "Unsupervised Estimation of Writing Style Models for Improved Unconstrained Off-line Handwriting Recognition," in *Proc. 10th Intl. Workshop on Frontiers in Handwriting Recognition, La Baule, France, Oct 23–26, 2006*, pp. 429–434.

[30] C.L. da Fontoura and R.M. Cesar Jr., *Shape Analysis and Classification: Theory and Practice*, Boca Raton, FL: CRC Press, 2000.

[31] B. Found and D. Rogers, "The Probative Character of Forensic Document Examiners' Identification and Elimination Opinions on Questioned Signatures," in *Proc. of the 13th Conf. of the Intl. Graphonomics Society, Melbourne, Australia, November 11–14, 2007*, pp. 171–174.

[32] A. Garz, M. Würsch, and R. Ingold. "Training- and Segmentation-Free Intuitive Writer Identification with Task-Adapted Interest Points," in *Proc. 17th Conf. of the International Graphonomics Society, Pointe-à-Pitre, Guadeloupe, June 21–24, 2015*, pp. 109–112.

[33] N. Ghoggali and F. Melgani, "Automatic Ground-Truth Validation With Genetic Algorithms for Multispectral Image Classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47 (7), pp. 2172–2181, 2009.

[34] M. Gomez-Barrero, J. Galbally, J. Fierrez, J. Ortega-Garcia, and R. Plamondon, "Variations of handwritten signatures with time: A sigma-lognormal analysis", in *Proc. 6th Intl Conf. on Biometrics, Madrid, Spain, June 4–7, 2013*, pp. 1–6.

[35] P.J. Grother, *NIST Special Database 19. NIST Handprinted Forms and Characters Database*, NIST, technical report, Gaithersburg, MD, 1995.

[36] I. Guyon, N. Matić, and V. Vapnik, "Discovering Informative Patterns and Data Cleaning", in U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (eds.), *Advances in Knowledge Discovery and Data Mining*, Cambridge, MA: The MIT Press, pp. 181–203, 1996.

[37] Th.M. Ha, "Efficient Detection of Abnormalities in Large OCR Databases," in *Proc. 4th Intl. Conf. on Document Analysis and Recognition, Ulm, Germany, Aug. 18–20, 1997*, pp. 1006–1010.

[38] D.J. Hand, "Recent advances in error rate estimation," *Pattern Recognition Letters*, vol. 4, pp. 335–346, 1986.

[39] C. Hertel and H. Bunke, "A set of novel features for writer identification," in *Proc. 4th Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, UK, Guildford, June 9–11, 2003*, pp. 679–687.

[40] K. Hildebrand, M. Gebauer, H. Hinrichs, and M. Mielke, *Daten- und Informationsqualität*, Wiesbaden: Vieweg+Teubner, 2011.

[41] R.A. Huber and A.M. Headrick, *Handwriting Identification: Facts and Fundamentals*, Boca Raton, FL: CRC Press.

[42] J.J. Hull, "Performance Evaluation for Document Analysis," *Intl. J. of Imaging Systems and Technology*, vol. 7, pp. 357–362, 1996.
[43] I. Ivanov, F. Dufaux, Th.M. Ha, and T. Ebrahimi, "Towards Generic Detection of Unusual Events in Video Surveillance," in *Proc. 6th IEEE Intl. Conf. on Advanced Video and Signal Based Surveillance, Genoa, Italy, Sep. 2–4, 2009*, pp. 69–74.
[44] R. Jain and D. Doermann, "Combining Local Features for Offline Writer Identification," in *Proc. 14th Intl. Conf. on Frontiers in Handwriting Recognition," Heraklion, Greece, Sept. 1–4, 2014*, pp. 583–588.
[45] F.V. Jensen and Th.D. Nielsen, *Bayesian Networks and Decision Graphs*, Berlin: Springer, 2007.
[46] J. Kittler *et al.* (eds.), *Multiple Classifier Systems Intl. Workshop*, Berlin: Springer, 2001–2015.
[47] R.C. Kraus, *Brushes with Power: Modern Politics and the Chinese Art of Calligraphy*, Los Angeles: University of California Press, 1991.
[48] H. Kume, *Statistical Methods for Quality Improvement*, Tokyo: The Association for Overseas Technical Scholarship, 1985.
[49] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Hoboken, N.J.: John Wiley & Sons, 2014.
[50] N. Liyanage, "Misclassification bias in epidemiologic studies," MS dissertation, Dept. of Mathematics and Statistics, McGill Univ., Montreal, QC, 1995.
[51] D. Lopresti, "Models and Algorithms for Duplicate Document Detection," in *Proc. 5th Intl. Conf. on Document Analysis and Recognition, Bangalore, India, Sep. 20–22, 1999*, pp. 1006–1010.
[52] E. Marasco, A. Ross, and C. Sansone, "Predicting Identification Errors in a Multibiometric System Based on Ranks and Scores," in *Proc. of 4th IEEE Intl. Conf. on Biometrics: Theory, Applications and Systems, Washington DC, Sep. 27–29, 2010*, pp. 1–6.
[53] U.-V. Marti and H. Bunke. "The IAM-database: an English sentence database for off-line handwriting recognition," *Intl. J. on Document Analysis and Recognition*, vol. 5, pp. 39–46, 2002, http://www.iam.unibe.ch/fki/databases/iam-handwriting-database.
[54] U.-V. Marti, Bern, Switzerland; H. Bunke and V. Frinken, Institute of Computer Science and Applied Mathematics, Bern; M. Lewicki, German Research Center for Artificial Intelligence, Kaiserslautern, Germany; personal communications, March–April 2011, August, October 2015.
[55] MathWorks, "fitdist", *Matlab Documentation*, 2015.11.24, http://ch.mathworks.com/help/stats/fitdist.html.
[56] N. Matić, I. Guyon, L. Bottou, J. Denker, and V. Vapnik, "Computer aided cleaning of large databases for character recognition", in *Proc. 11th Intl. Conf. on Pattern Recognition, The Hague, The Netherlands, Aug. 30 – Sep. 3, 1992*, vol. 2, pp. 330–333.
[57] A. Maydanchik, *Data Quality Assessment*, Bradley Beach, NJ: Technics Publications, 2007.
[58] F. Mokhtarian and M. Bober, *Curvature Scale Space Representation: Theory, Applications and MPEG-7 Standardization*, Dordrecht: Kluwer Academic Publisher, 2003.
[59] A. Monadjemi, "Towards Efficient Texture Classification and Abnormality Detection," Ph.D. dissertation, Univ. of Bristol, Bristol, UK, 2004.
[60] J. Neyman and E.L. Scott, "Outlier Proneness of Phenomena and of Related Distributions", in J.S. Rustagi (ed.), *Optimizing Methods in Statistics*, New York: Academic Press, pp. 413–430, 1971.
[61] R. Niels, "Allograph based writer identification, handwriting analysis and character recognition," PhD dissertation, Donders Inst. for Brain, Cognition and Behaviour, Radboud Univ. Nijmegen, The Netherlands, 2010.
[62] V. Niennattrakul, Ch.A. Ratanamahatana, and E. Keogh, "Data Editing Techniques to Allow the Application of Distance-Based Outlier Detection to Streams," in *Proc. Intl. Conf. on Data Mining, Sydney, Australia, Dec. 13–17, 2010*, pp. 947–952.
[63] NIST/SEMATECH, "Anderson-Darling Test", *e-Handbook of Statistical Methods*, 2015.11.24, http://www.itl.nist.gov/div898/handbook/eda/section3/eda35e.htm.
[64] J. Oakland, *Statistical Process Control*, Amsterdam: Butterworth-Heineman, 2008.
[65] W.S. Peters, *Counting for Something. Statistical Principles and Personalities*, New York: Springer, pp. 159–170, 1987.
[66] R. Plamondon, "A kinematic theory of rapid human movements: Part I. Movement representation and generation", *Biological Cybernetics*, vol. 72, pp. 295–307, 1995.
[67] R. Plamondon, Ch. O'Reilly, C. Rémi, and T. Duval, "The lognormal handwriter: learning, performing, and declining", *Frontiers in Psychology*, vol. 4, art. 945, 2013.
[68] S. Pletschacher and A. Antonacopoulos, "The PAGE (Page Analysis and Ground-truth Elements) Format Framework," in *Proc. 20th Intl Conf. on Pattern Recognition, Istanbul, Turkey, Aug. 23–26, 2010*, pp. 257–260.
[69] L. Rokach, *Pattern classification using ensemble methods*, Singapore: World Publishing, 2010.
[70] A.R. Ross, K. Nandakumar, and A.K. Jain, *Handbook of Multibiometrics*, Berlin: Springer, 2006.
[71] D.K. Rossmo, *Geographic Profiling*, Boca Raton, FL: CRC Press, 2000.
[72] R. Royall, *Statistical Evidence: A likelihood paradigm*, Boca Raton, FL: Chapman & Hall / CRC Press, 1997.
[73] V.I. Rupasov, M.A. Lebedev, J.S. Erlichman, M. Linderman, "Neuronal Variability during Handwriting: Lognormal Distribution", *PLoS One*, vol. 7 (4), art. e34759, 2012.
[74] Sh. Sadiq (ed.), *Handbook of Data Quality Research and Practice*, Berlin: Springer, 2013.
[75] L. Schomaker and L. Vuurpijl, *iUF Firemaker: A benchmark dataset for writer identification*, technical report, Nijmegen: Nijmegen Institute for Cognition and Information, Univ. of Nijmegen, 2000.
[76] W. Shi, *Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses*, Boca Raton, FL: CRC Press, 2010.
[77] I. Siddiqi and N. Vincent, "Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features," *Pattern Recognition*, vol. 43, pp. 3853–3865, 2010.
[78] T. Simonite, "Teaching Machines to Understand Us", *MIT Technology Review*, vol. 120 (5), pp. 72–79, 2015.
[79] S.N. Srihari *et al.*, "Individuality of Handwriting," *J. of Forensic Sciences*, vol. 47, pp. 856–872, 2002.
[80] J. Subrahmonia, "Similarity Measures For Writer Clustering," in *Proc. 7th Intl. Workshop on Frontiers in Handwriting Recognition, Amsterdam, The Netherlands, September 11–13, 2000*, pp. 541–546.
[81] V. Torra and Y. Narukawa, *Modeling Decisions: Information Fusion and Aggregation Operators*, Berlin: Springer, 2007.
[82] G.T. Toussaint, "Bibliography on Estimation of Misclassification," *IEEE Trans. on Information Theory*, vol. 20 (4), pp. 472–479, 1974.
[83] R.Y. Wang and D.M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *J. of Management Information Systems*, vol. 12 (4), pp. 5–34, 1996.
[84] Wikipedia contributors, "Roman cursive", *Wikipedia*, 2015.11.24, https://en.wikipedia.org/wiki/Roman_cursive.
[85] Wiktionary contributors, "zampa di gallina", *Wiktionary*, 2015.11.28, https://en.wiktionary.org/wiki/zampa_di_gallina.
[86] R.A. Wilkinson, M.D. Garris, and J. Geist, "Machine-Assisted Human Classification of Segmented Characters for OCR Testing and Training," in *Proc. SPIE Conf. Character Recognition Technologies, San Jose, CA, 1 Feb. 1993*, vol. 1906, pp. 208–217.
[87] R.A. Wilkinson *et al.*, *The First Census Optical Character Recognition Systems Conference*, NIST, technical report, Gaithersburg, MD, 1992.
[88] Zeiss, *Did You Know... that up to 21 People Can Look Through a Microscope at the Same Time?*, webpage, http://www.zeiss.com/corporate/en_de/about-zeiss/news/did-you-know/58nd-edition-of--did-you-know----.html.

P4 – 70

Author Biography

Vlad Atanasiu is doctoral candidate in computer science at the University of Fribourg, Switzerland. He holds a PhD in paleography (ÉPHÉ, Paris, 2003) and an MA in Arabic linguistics (University of Provence, 1995). He was postdoc in the Architecture and Cognitive Science Departments at MIT, worked on applications to Digital Humanities of image processing, visualization, and geographical information systems at the Austrian Academy of Sciences, and on handwriting analysis at Telecom ParisTech. atanasiu@alum.mit.edu, http://alum.mit.edu/www/atanasiu/

Supplementum

On *data complexity* in pattern recognition, including handwriting analysis, the reader might want to consult the rich collection of articles in M. Basu and T.K. Ho (eds.), *Data Complexity in Pattern Recognition*, London: Springer, 2006. [Many thanks to George Nagy for pointing out this reference.]

A further *contribution to misclassification research*, in the context of data streams, is M.-R. Bouguelia, Y. Belaïd, and A. Belaïd, "Stream-based Active Learning in the Presence of Label Noise", in *Proc. 4th Intl Conf. on Pattern Recognition Applications and Methods, Lisbon, Portugal, Jan. 10–12, 2015*, pp. 25–34. Inspiration on the study of misclassification can be gathered also from the topic of *rare categories*: Jingrui He, *Analysis of Rare Categories*, Berlin: Springer, 2012.

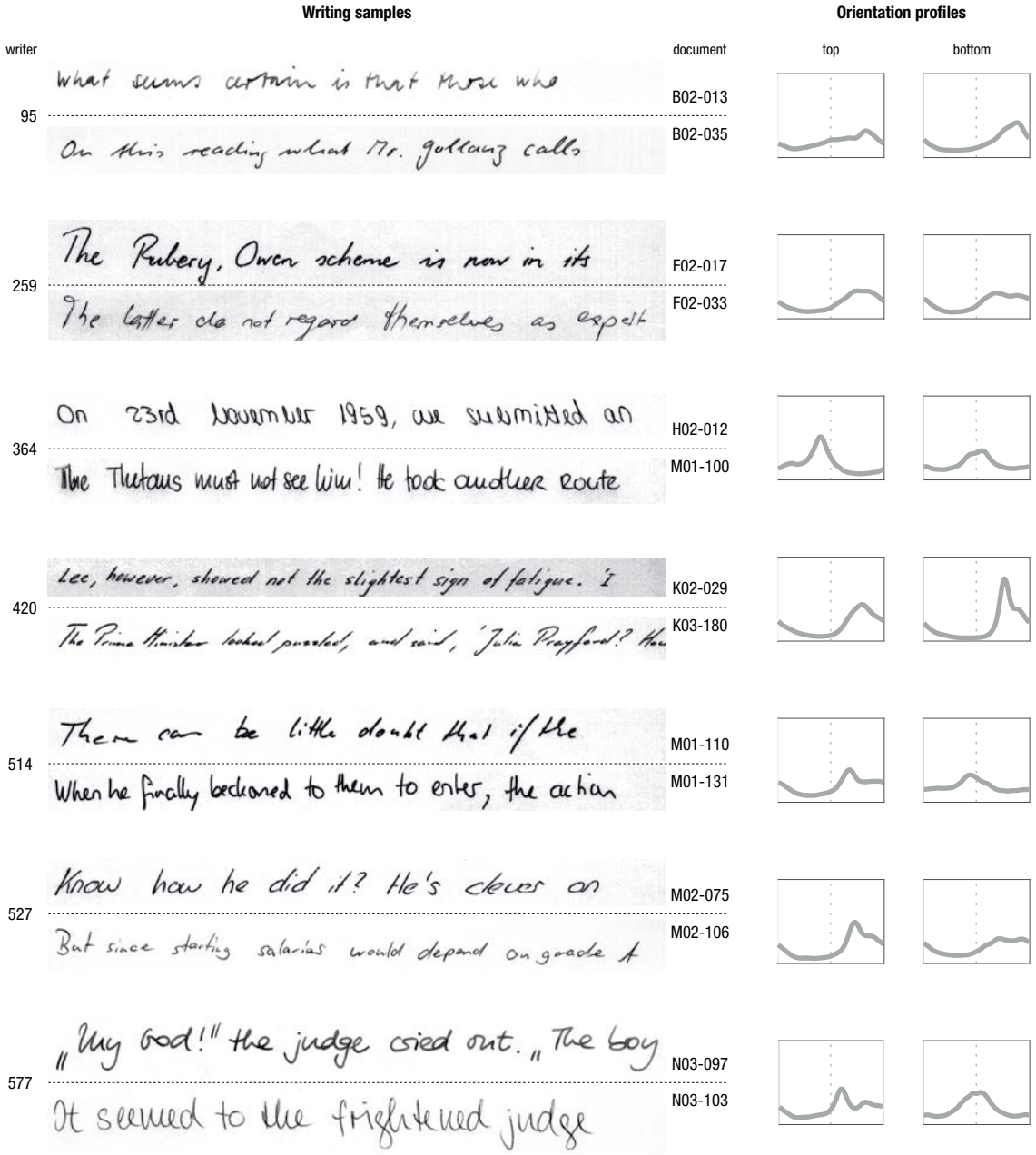


Figure S1. *The Seven IAM Cuckoos — Complete list of the cuckoos discovered in the IAM dataset, with sample handwritings of each.*

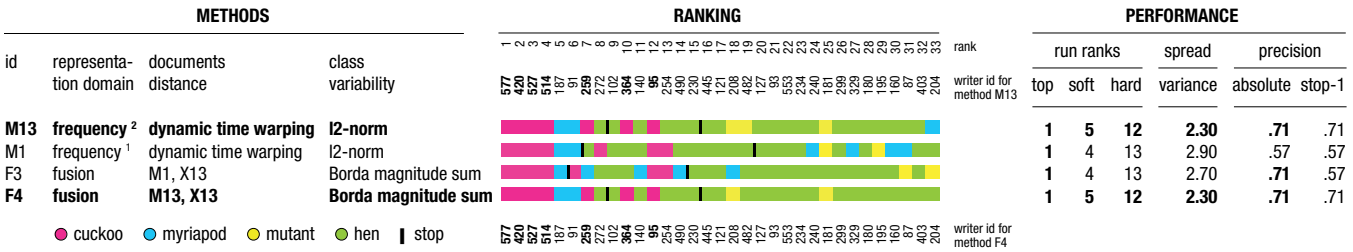
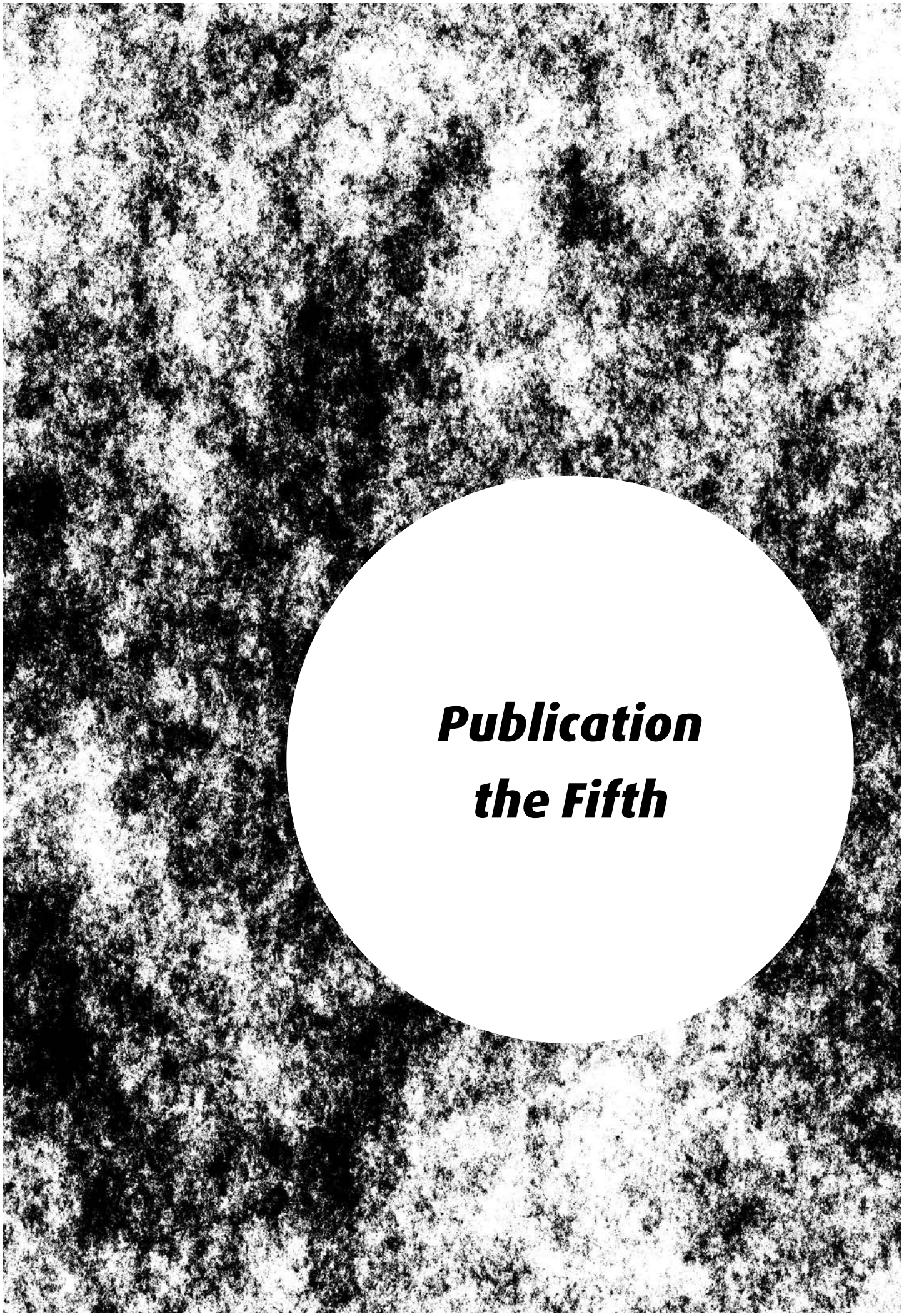


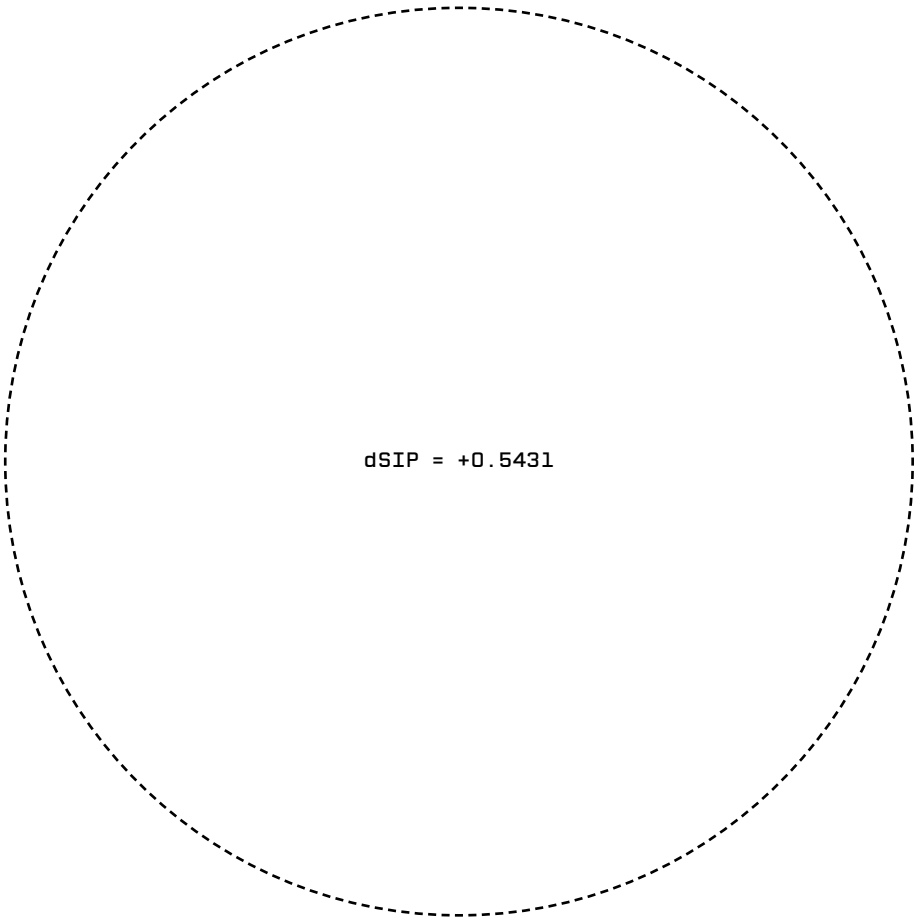
Figure S2. *Improved system performance — (1) Magnitude of Fourier components. (2) Magnitudes weighted inverse proportionally to their frequency.*

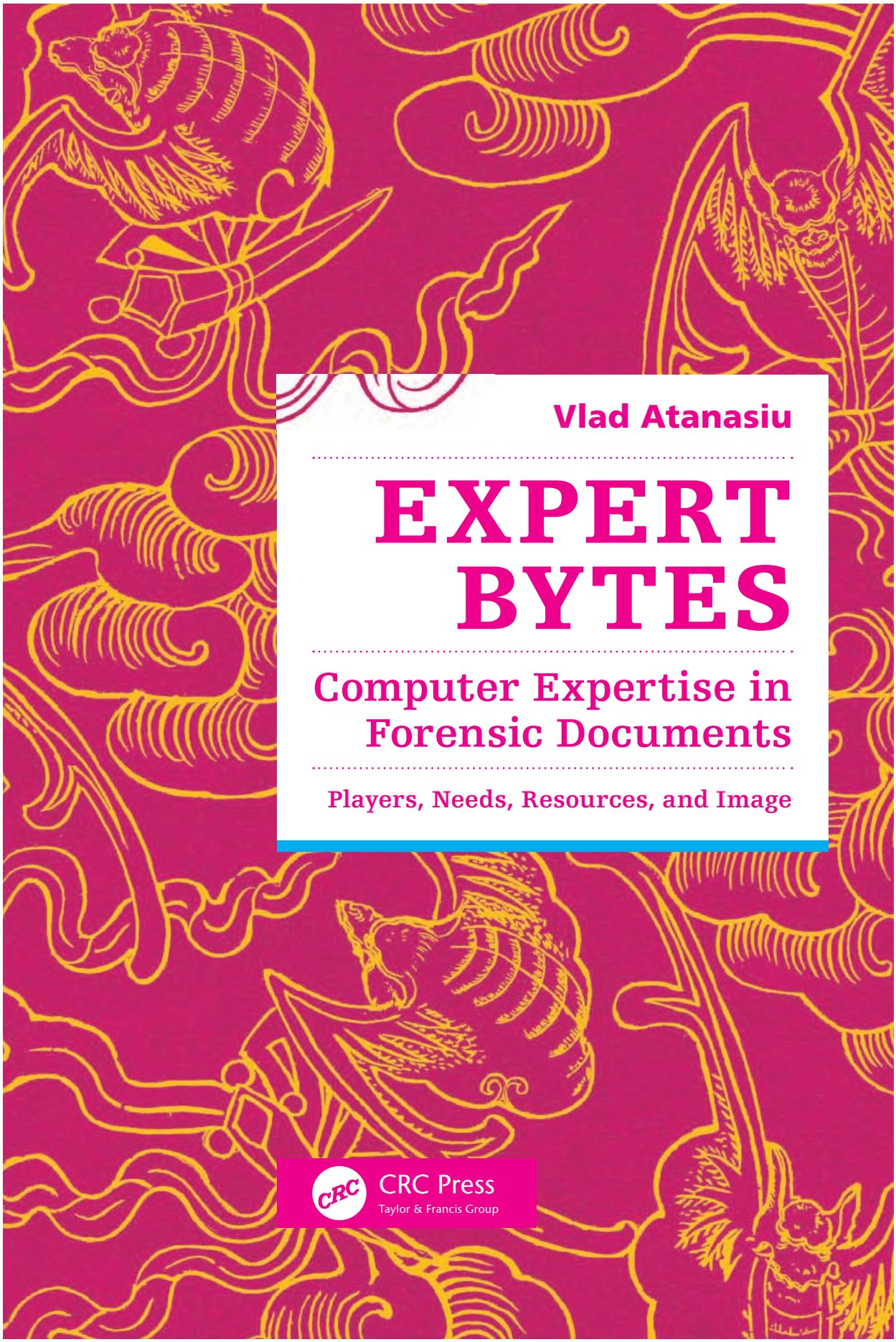
The best method to measure the distance between two vectors was based on the magnitude of these vectors in the frequency domain (Fig. 6). Fig. S2 shows an improved method (M13), implementing a low-pass filter in which the magnitude is weighted by $1/f$, were f represents the frequency associated to the indices of the magnitude vector. Let us note that the phase distortion between vectors produced much worse results (calculated as sum of the arclengths of the circular distance between phase vectors).

A recent addition to metaphoric nomenclatures comes from political sciences and concerns the voters bestiary (Jason Brennan, *Against Democracy*, Princeton: Princeton University Press, 2016, pp. 4–5); older ones concern typologies of museum visitors, teacher attitudes towards high school dropouts, and hikers' attitudes towards outdoors travel (Michael Q. Patton, *Qualitative Evaluation Methods*, Beverly Hills, CA: Sage, 1980, p. 460, 469, 458). A host of individual metaphorical terms related to statistics exist, some having become popular idioms: "butterfly effect", "black swan", "gray rhinoceros", "dragon kings", "the ostrich effect", even "the elephant in the room", "the skeleton in the closet".


Publication
the Fifth

For copyright reasons this publication is not included in the present document. Please use the following link to obtain a copy of the book: https://doi.org/10.1201/b15467.





Vlad Atanasiu

EXPERT BYTES

Computer Expertise in Forensic Documents

Players, Needs, Resources, and Image

CRC Press
Taylor & Francis Group

Comparison of millions of signatures in minutes, expertise in hundreds of languages, linking disparate intelligence — these are some of the promises of computing to forensic document examination. For this vision to happen, both computer scientists and forensic experts have to understand their partner's professional field almost like anthropologists would do. In this book Vlad Atanasiu provides a set of keys to a cultural encounter by unveiling the sociotechnological landscape in which the scientists will evolve: players, needs, resources, and pitfalls. While maintaining a global perspective on forensic documents, the focus is on writing in all its forms. On the computing side, the book is rooted in the field of document image analysis.

Author bio p. 198

CRC Press
Taylor & Francis Group
an **informa** business

9 781466 591905 90000

computer science / forensics / ISBN 987-1-4665-9190-5 / K20536

6000 Broken Sound Parkway, NW, Suite 300, Boca Raton, FL 334871; 711 3rd Ave, New York, NY 10017; 2 Park Sq., Milton Park, Abingdon, Oxon OX14 4RN, U.K. • www.crcpress.com

Contents

	Acknowledgments	ix
One	**Perspective**	
	Book framework	3
Two	**Object**	
	Forensic document expertise	11
Three	**Agenda**	
	Computational research	25
Four	**Technology**	
	Know-how + Tools	45
Five	**Context**	
	Expertise performance	55
	Policy landscape	75
Six	**Closure**	
	Why computerize?	87
Seven	**Fact sheets**	
	An ecosystem	93
	Basic forensic fields	94
	Common topics	95
	Diverse goals + aspects	97
	Example curricula	101
	Foremost organizations	103
	Gender demographics	105
	Health trends	107
	Information sources	108
	Just standards	111
	Key datasets + databases	113
	Landmark software	121
	Main descriptors	121
	Noticeable hardware	129
	Open research geography	130
Eight	**Links**	
	References	137
	Index	181
	Credits	197

****
Vlad Atanasiu, Expert Bytes: Computer Expertise in Forensic Documents - Players, Needs, Resources and Pitfalls, Boca Raton, FL: CRC Press, 2013
****

> ADDENDA & ERRATA

Last updated on 13 December 2015

Addenda

Matley 2008 reference [p. vii] — Marcel B. Matley points out the following regarding the reference to one of his articles: "You cite my review of Shrihari's paper on handwriting of twins at the beginning. Unfortunately, you must have had an improper copy since what you attribute to me is not what my critical review says. I do not deny computers can be of value, I only demonstrate how his particular study shows that his specific program and research is much mistaken and off base by a long ways." [email to the author, 2013.10.28]

What I found interesting in Matley's point of view is that it can be seen as an epitome perspective on computational expertise of forensic documents (FD), beyond characterizing a specific article, the one by Srihari. As Matley made it clear, it is not computers *per se* that he is battling, but "misbegotten computerized research." He writes: "After 20 or 30 years of tax and other funds being dumped into the bottomless pit of misbegotten computerized research, we have nothing of practical use to show for it." [same reference]

Computational paleography reference [p. 3n1] — A review of the current state of issues and perspectives in computational paleography is offered in Tal et al. 2013. — See new references at the end of the addenda.

On the writing style of law scholars [p. 5] — Of which it was famously said: "I am the last one to suppose that a piece about the law could be made to read like a juicy sex novel or a detective story, but I can not see why it has to resemble a cross between a nineteenth century sermon and a treatise on higher education." [Rodell 1936: 41]

User-centered design [p. 30] — Kansei engineering, a Japanese design philosophy aiming to elicit positive affects from users in respect to products, is particularly suited to issues of user-centered design.

Know-how + Tools [p. 45] — In complement of technical references, a recommended reading is NAS 2010 for its critical perspective on biometrics, the close relative of FD, which is envisioned as a system engineering problem, one that is not infallible, and is part of a social, cultural, and legal context.

Sampling [p. 47, replaces the second paragraph] — *Sampling* — Let's not forget that the choice of sampling criteria depends on what one wants to do with the samples once collected. A nation-wide sampling cast like an opinion poll utilizing census criteria (age, gender, education, income, and so forth) is an appropriate model for a software employed at country level (the United States

in one study), but less so for a local usage if the demographics differ from the national average (and apart from the fact that we do not know what the sample size appropriate for studying handwriting should be).[1] Indeed, numerous FD questions require specialized datasets, which can shed light on the individuation power of Latin script compared to Arabic or Chinese, distinguish between the standardized writing styles of calligraphers or monitor that of people suffering from Alzheimer's, penetrate the intriguing peculiarities of Basque terrorist handwriting[2] or the impact tactile screens on the ability to distinguish signatures (notoriously bad for anyone having had to sign a FedEx receipt).[3] Research of the past 15 years has also shown the importance of considering the production conditions when assessing writing individuality. For example, its value is lower for simulated writing collections when compared to purely genuine samples[4] or twins' handwriting compared to that of nontwins', an astute experiment since long favored by psychologists and FDEs.[5]

Given the thorny issue of sampling, researchers might want to look for alternatives. A class of such methods, called non-probabilistic because they don't rely on random sampling, are known from the area of survey design. One of them in particular, "purposeful sampling," is interesting for the topic of handwriting.[6] Contrary to probabilistic sampling, purposeful sampling is a qualitative approach that seeks to understand the mechanism behind a particular distribution, in order to provide explanations instead of merely quantitative descriptions. The model mechanism can then be used to simulate distributions, thus bypassing sampling (in the right conditions and to a certain degree, since even a model needs to be grounded in data). Single case studies are another option, with a long history in medicine, where FDEs can look for methodological inspiration when confronted with such questions as documenting the transformation of an individual's handwriting in time.[7]

1. Srihari et al. 2002: 858–859 propose to adopt the size of opinion polls, around 1,100 subjects, but Sacks 2003a: 916–917 points to medical experiments commonly with as few as 10–20 patients and epidemiological studies carried out on tens of thousands of individuals; NAS 2010: 30–31 discusses the global world scale. — **2.** Bisotti 2011. — **3.** Harralson 2013: 62–64. — **4.** E.g., Found and Rogers 2007. — **5.** References in Harrison and Seiger 2003, Huber and Headrick 1999: 196–198, study in, e.g., Srihari et al. 2008, critique in Page et al. 2011: 15; compare with sampling issues in the FBI mtDNA database in Kaestle et al. 2006. — **6.** On purposeful sampling see Leyden 2013: 113–128, Patton 2002: 45–46, 230–247; on the long history of sample vs model based inference see Brewer and Gregoire 2009; a practical introduction to sampling is Henry 1990; for reference works see Pfeffermann and Rao 2009, Gideon 2012. Usually sampling methods are quantitative, produce descriptions, are used for measuring prevalence of phenomena, and allow generalization of results to larger populations; model methods are qualitative, explanatory, generate mechanisms, and few of them are statistically representative. In terms of terminology the approaches are variously called in statistics and the social sciences random, probabilistic, nonparametric, design based, as opposed to nonrandom, nonprobabilistic, parametric, model or prediction based, to which the medical literature adds the terms homeothetic and extensive (general and group), respectively ideographic and intensive (individual). — **7.** Although concerned with clinical psychology, scientists from many other fields can benefit from the excellent overview of the history, issues, and methods of single case experiments compared to group and randomized studies is Barlow et al. 2009: 1–29 [history], 31–59 [issues], 46–52 [possibility of generalization from single cases].

Descriptors [p. 48] — Psychophysical experiments represent a possible way forward to understand what features are used by the human visual system for comparing script styles. [Features that allow the distinction between characters, an important aspect for understanding reading mechanisms, have been studied in psychology (references in Sanocki 2012: 135–136).]

Handwriting individuality [p. 55–57] — The US National Academy of Sciences calls for the development of a "science of human individual distinctiveness" and sees it as "essential to effective and appropriate use of biometric recognition," given the complexity and social impact of such systems. [NAS 2010: 4] On the dogma of biometric recognition and anthropometry see NAS 2010: 23–52, respectively Gould 1996.

Dataset representativeness [p. 61] — That "population sampling issues that affect performance in practice may not be fully appreciated," has also been highlighted by the US National Academy of Sciences and contrasted with the "very sophisticated statistical methods are used for the signal analysis and pattern recognition aspects of biometric technology." [NAS 2010: 17]

DNA profiling [p. 72n4] — See Zimmer 2013 on multiple genomes (mosaicism and chimerism).

On interdisciplinarity [p. 78n2] — In fact, the issue of interdisciplinarity arose from the very inception of computing. In the words of Grace Hopper, in the 1940s one of the first programmers and central figure to the invention of software compilers and the COBOL programming language, "We had to learn their vocabularies in order to be able to run their problems. I learned languages of oceanography, of this whole business of minesweeping, of detonators, of proximity fuses, of biomedical stuff." [Beyer 2009: 55]

HG4FDE, Anonymous [p. 124] — The software is now called "Masquerade", and has its website at http://www.nitesrl.com/products/masquerade.

References

BARLOW David H., Matthew K. Nock, Michel Hersen (2009) — *Single Case Experimental Design: Strategies for Studying Behavior Change*, Boston, MA: Pearson Education.

BEYER Kurt W. (2009) — *Grace Hopper and the Invention of the Information Age*, Cambridge, MA: The MIT Press.

BREWER Ken, Timothy G. Gregoire (2009) — *Introduction to Survey Sampling*, in Pfeffermann and Rao 2009 (29A): 9–38.

GIDEON Lior, ed. (1996) — *Handbook of Survey Methodology for the Social Sciences*, New York: Springer.

GOULD Stephen Jay (1996) — *The Mismeasure of Man*, New York: W.W. Norton.

HENRY Gary T. (1990) — *Practical Sampling*, Newbury Park, CA: Sage.

LAYDER Derek (2013) — *Doing Excellent Small-Scale Research*, Thousand Oaks, CA: Sage.

NAGAMACHI Mitsuo (2011) — *Kansei/Affective Engineering*, Boca Raton, FL: CRC Press.

NAGAMACHI Mitsuo, Anitawati Mohd Lokman (2011) — *Innovations of Kansei Engineering*, Boca Raton, FL: CRC Press.

NAS (National Academy of Sciences, Whither Biometrics Committee) (2010) — *Biometric Recognition: Challenges and Opportunities*, Washington, DC: The National Academies Press.

PATTON Michael Quinn (2002) — *Qualitative Research & Evaluation Methods*, Thousand Oaks, CA: Sage, 3rd ed.

PFEFFERMANN Danny, C. R. Rao, eds. (2009) — *Sample Surveys: Design, Methods and Applications* (vol 29A), *Inference and Analysis* (vol 29B), Amsterdam: North-Holland.

RODELL Fred (1936) — "Goodbye to Law Reviews," *Virginia Law Review*, 23: 38–45.

SANOCKI Thomas, Mary C. Dyson (2012) — "Letter Processing and Font Information During Reading: Beyond Distinctiveness, Where Vision Meets Design," *Attention, Perception, & Psychophysics*, 74: 132–145.

SCHÜTTE Simon (2005) — *Engineering Emotional Values in Product Design: Kansei Engineering in Development*, PhD thesis, Linköping: Department of Mechanical Engineering, Linköping University.

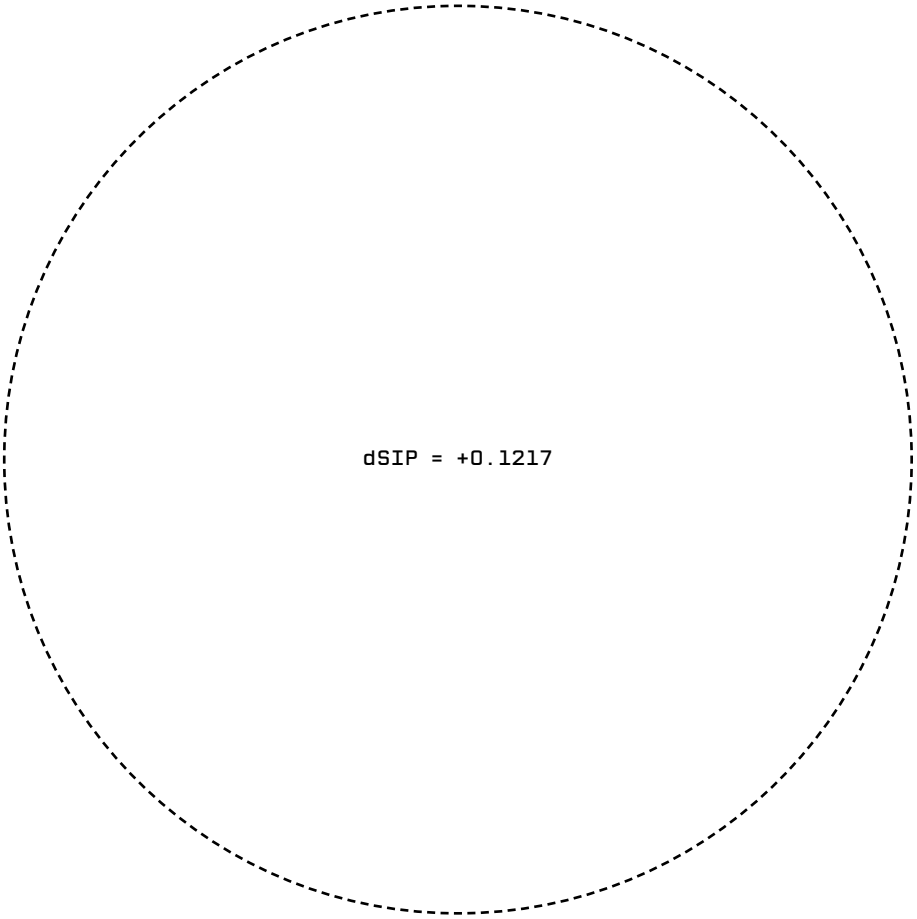
TAL Hassner, Malte Rehbein, Peter A. Stokes, Lior Wolf (2013) — "Computation and Palaeography: Potentials and Limits (Dagstuhl Perspectives Workshop 12382)," *Dagstuhl Manifestos*, 2(1): 14–35.

ZIMMER Carl (2013) — "DNA Double Take," *The New York Times*, online, accessed 2013.10.03, http://www.nytimes.com/2013/09/17/science/dna-double-take.html.

Errata

P. 99 — For “2. forgery ○ (class)“ read “2. forgery ● (class)“.

Publication
the Sixth


dSIP = +0.1217

International Journal on Document Analysis and Recognition (IJDAR) (2022) 25:129–160
https://doi.org/10.1007/s10032-021-00386-0

ORIGINAL PAPER

Personalizing image enhancement for critical visual tasks: improved legibility of papyri using color processing and visual illusions

Vlad Atanasiu[1,2] · **Isabelle Marthot-Santaniello**[2]

Received: 28 October 2020 / Revised: 16 May 2021 / Accepted: 26 August 2021 / Published online: 27 December 2021
© The Author(s) 2021, corrected publication 2022

Abstract
This article develops theoretical, algorithmic, perceptual, and interaction aspects of script legibility enhancement in the visible light spectrum for the purpose of scholarly editing of papyri texts. Novel legibility enhancement algorithms based on color processing and visual illusions are compared to classic methods in a user experience experiment. (1) The proposed methods outperformed the comparison methods. (2) Users exhibited a broad behavioral spectrum, under the influence of factors such as personality and social conditioning, tasks and application domains, expertise level and image quality, and affordances of software, hardware, and interfaces. No single enhancement method satisfied all factor configurations. Therefore, it is suggested to offer users a broad choice of methods to facilitate personalization, contextualization, and complementarity. (3) A distinction is made between casual and critical vision on the basis of signal ambiguity and error consequences. The criteria of a paradigm for enhancing images for critical applications comprise: interpreting images skeptically; approaching enhancement as a system problem; considering all image structures as potential information; and making uncertainty and alternative interpretations explicit, both visually and numerically.

Keywords Script legibility · Image enhancement · Color processing · Perceptual image processing · Image quality · Papyrology

It's obvious—any fool can see it.
Homer, *The Iliad,* 7.464 [77, p. 227]

1 Introduction

Of the images below, which would you consider more legible,

A. or B. ?

If you selected B, you behaved like an engineer, confident in the benefits of a high signal-to-noise ratio. It may, therefore, come as a surprise that, in the evaluation reported in this article, three out of eight scholars deciphering these ancient Greek papyri preferred item A, because it is the original document, and originals should not be altered. This is not unreasonable: to question what exactly one is looking at is sensible. We call this attitude *critical vision*, and place it at the core of image enhancement. Our psychophysical and sociological mini-experiment also illustrates how enhancement quality depends on more factors than algorithms alone. These two principles frame the work being presented here.

1.1 Objectives

In narrow terms, the scope of this article is the legibility enhancement in the visible light spectrum of papyri documents for scholarly text editing. The broader objectives derive from the challenges, approaches, and lessons of this work. On the technical level, the aim is to demonstrate the utility of color processing and visual illusions in legibility enhancement. To emphasize the power of these approaches,

✉ Vlad Atanasiu
atanasiu@alum.mit.edu

Isabelle Marthot-Santaniello
i.marthot-santaniello@unibas.ch

[1] Department of Informatics, University of Fribourg, Fribourg, Switzerland

[2] Department of Ancient Studies, University of Basel, Basel, Switzerland

the proposed methods will rely on a minimal algorithmic apparatus. On the theoretical level, a paradigm of legibility enhancement for computer-aided vision will be outlined. The central tenet of this paradigm is that critical vision tasks require enhancement technologies capable of making uncertainty explicit and adopting a systemic approach, that encompasses users, tasks, data, methods, tools, and interactions. The methodological objective is to develop methods and paradigms on an empirical, interdisciplinary basis.

1.2 Relevance

Papyri make up an extensive but underexploited trove of time capsules. The vast majority of these artifacts are accessible only as images taken in visible light and are often characterized by poor legibility due to degraded physical documents, non-calibrated imaging, and image compression. Digital enhancement promises to reduce the time and uncertainty associated with reading difficult texts from difficult reproductions. This work may be useful in application domains other than papyrology as well, as image enhancement is a ubiquitous stage in image processing. For example, there are similarities between deciphering papyri and examining medical, forensic, and military imagery, with respect to the uncertainty surrounding their interpretation.

1.3 Contributions

The present article adds to the limited existing work on papyri enhancement in the visible light spectrum. It promotes the use of color processing and computationally induced visual illusions. It works towards remedying the lack of a holistic approach to legibility enhancement. Making uncertainty explicit is also novel in this domain.

1.4 Methodology

The theoretical argument made in this article, the evaluation of the proposed methods, and the design of the ensuing software rely on an in-depth exploratory statistical analysis of data obtained from a user experiment. However, mathematics, algorithms, software, and hardware are never neutral when used by humans; this is especially so when the machine and the human form a closely coupled system, as in the case of legibility enhancement. Therefore, this article presents various points of view regarding how psychological and sociological factors are co-determinant in the design of legibility enhancement methods. The user-centric and context-aware approaches promoted herein are complementary to purely data-driven methods [162].

1.5 Applications

The principal application is an aid to human vision for use in transcribing papyri. The developed legibility software is also a tool for publishing better papyri images in print and online. Machine vision, particularly document binarization and text recognition, is the third potential application or the proposed methods, although human-readable texts are not necessarily legible to machines.

1.6 Organization

Section 2, "Problematic", introduces papyrology and scholarly text editing and then defines legibility enhancement through the concept of critical vision. Section 3, "Related work", reviews the computational methods of image enhancement used in papyrology and for similar tasks in other domains. Section 4, "Methods", presents the novel legibility enhancement methods. Section 5, "Experiment", describes a user experiment to empirically understand critical legibility enhancement. Section 7, "Paradigm", discusses a type of image enhancement that makes uncertainty explicit and summarizes the criteria of a computer-aided critical vision system. Section 8, "Implementation", shows how the paradigm and empirical findings were translated in software.

The image evaluation dataset is part of the article's online Supplementary Material. The software resulting from this research is freely available online [16].

2 Problematic

2.1 Relevance

Papyrus was the quintessential portable writing surface in the Mediterranean area for five thousand years, until the end of the first millennium CE, when paper technology spread. As such, papyrus is a crucial carrier of substantial knowledge about the Egyptian, Greek, Roman, Byzantine, and other civilizations of antiquity. Figures 1 and 2 present surviving samples of two foundational texts of Western mathematics and literature: Euclid's *Elements* and Homer's *Iliad*. The number of unpublished papyri is currently estimated to be approximately one to one and a half million, which "will keep papyrologists busy for centuries at least" [182, pp. 644–645].

2.2 Data

Cyperus papyrus L. is an aquatic plant growing on the banks of the Nile River. By slicing, assemblage, and gluing via its own sap, a fine, flexible, and smooth papyrus

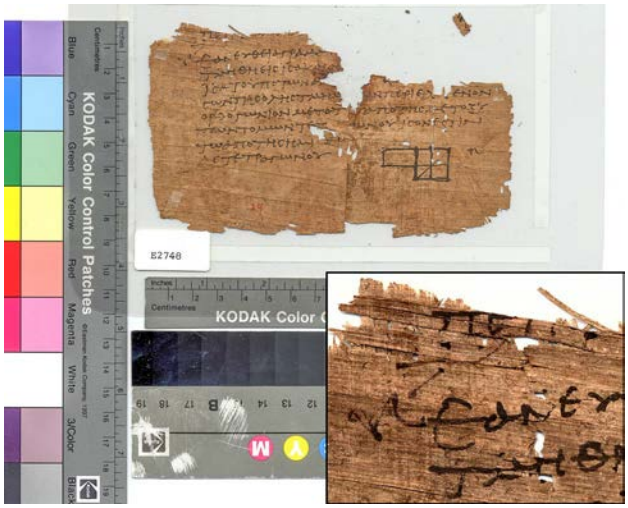


Fig. 1 Euclid, *Elements*, Book II, Proposition 5; papyrus of the 3rd–4th century CE from Oxyrhynchus, Egypt. (Credit: P.Oxy. I 29, University of Pennsylvania Museum of Archaeology and Anthropology, CC-BY-2.5.)

can be manufactured, that is suitable for writing [34]. The orthogonal fiber pattern and surface roughness were once commercial criteria in the papyrus trade; today, they offer valuable hints to historians on prices, tastes, and document origins. However, the closeness in spatial frequency between papyrus texture and script complicates computational graphonomics. Reading performance is also affected, given the higher perceptual sensitivity to horizontal and vertical gratings [26, pp. 270–271]. The spectral characteristics of papyri satisfy historical and computational aims, such as the recovery of ink traces by means of multispectral imaging. Further characteristic, inconvenient to scholars and scientist alike, include edge irregularity, holes, document fragmentation, and the presence of artifacts surrounding the papyrus, such as color charts, rulers, and shelf marks (Fig. 1).

2.3 Task

Papyrology is the study of papyri, although it also encompasses inscribed potsherds, wax tablets, metal foils, and other materials [17, p. xvii]. The aim of papyrology is to elucidate all aspects of the past, from the material to the spiritual. Its main source of information is the text content. Text editing is the process of transforming a poorly legible text in bitmap format into legible electronic text. This is not a purely mechanical task, as considerable paleographical, linguistic, and historical knowledge is required to construct a space of possible readings and then select one or more as the most probable. It involves more than "a pair of sharp eyes and a certain amount of common sense" as any person struggling with an unfamiliar script and text can attest, and it has been aptly compared to the work of a detective or a puzzle solver [161, pp. 197, 199]. For papyrologists, every word counts, because a single misread character can, literally speaking, change a queen into a fishmonger or teleport a city across space and time. Therefore, variant readings are common and indexed, since 1922, in a dedicated publication [161, pp. 212–213].

The task under consideration can thus be defined as *critical vision* in (a) the objective sense that its outcomes are of *significant importance*, and (b) the subjective sense of an observer exercising *skepticism* in the interpretation of signals. Here, the observers are the text editors, who are circumspect about the reality and meaning of what is seen in images, and their readers, who doubt the editors' interpretation. The opposite of critical vision is *casual vision*, which is characterized by lower task criticality and levels of observer criticism. In psychological terms the latter is a form of fast, reflexive, and easy cognition (System 1), and the former of slow, analytical, and deliberative cognition (System 2) [83].

The two vision types call for different image processing paradigms. Casual vision is best served by binary images, designed to minimize entropy and produce trustworthy, unambiguous texts. Critical vision requires that ambiguity be maintained. In this case, no image structure is noise; every one is potential information. They provide context to develop concurrent readings and are valuable as archaeological layers of the document's history. The text on a crumpled paper, to illustrate the argument, wouldn't be identifiable as a draft if the creases would be removed for the sake of noise reduction.

2.4 Strategies

It is not unusual for papyrologists to take part in archaeological excavations, and, like archaeologists, to be sensitive to the context of discoveries and use various techniques to analyze data. A quote from experimental participant FRG illustrates the roles of *contextualization* and *diversification*:

> "I found it important to look at the *original*, or a normalized version of the original (*vividness*), sometimes with the *stretchlim* or *lsv* version of it for the bulk of the transcription work. To understand the state of preservation of the papyrus, the *histeq* or *adapthisteq* files have been very useful. The contrast is very good, and so I always use this type of image at the beginning of a transcription and [for] comparison with the original [. . .]. The *negvividness* and *neglsv* images have been useful to detect differences between ink and holes, so I've used them occasionally. *Retinex* I expect to be useful, but not with the "Sammeltafeln" [different contents written on the same papyrus] we have in the Freiburg collection; while it is very good with large papyri, I would

P6 – 90

Fig. 2 Overview and detail of the original papyri images used in the evaluation. For higher resolution images please refer to the article's online Supplementary Material. **Top:** Notice the degradation of the physical documents, the differences in image quality, and the difficulty in comparing details within a limited display space. **Bottom:** Details scaled to equalize character height, demonstrating the difficulty associated with deciphering such documents. (Credits: (1) Columbia University Library, CC BY-NC 3.0; (2, 4, 5) University of Michigan Library, Public Domain Mark 1.0; (3) Yale University Library, Public Domain Mark 1.0; (6) Courtesy of The Egypt Exploration Society and the University of Oxford Imaging Papyri Project; (7, 8, 9) Istituto Papirologico Vitelli, by permission.)

> not use it for the transcription work throughout, only from time to time. *Locallapfilt* I have never found useful."

The user describes the creation of a coherent, rich, and justifiable interpretation based on multiple perspectives. Implicitly, he indicates that the legibility enhancement system must support the enhancement of a variety of image features, and that methods may be chosen for their good performance on a specific document and task. The strategy fits well with models of visual search [104] and information seeking [102]. In-depth user studies and theoretical models of papyrological reading strategies implemented in expert systems are discussed in [56, 155, 176, 177].

2.5 Duration

The text editing of papyri is a delicate and typically lengthy process, comprising document manipulation during reading, forming hypotheses regarding the content to achieve comprehension, and consulting auxiliary resources such as dictionaries. In addition to legibility, other factors affecting the amount of time required for this work include content difficulty, reader expertise, and work behavior. For example, it took the experienced participant FRG five to six hours split between two sessions to transcribe the papyrus used for training during evaluation, which was equivalent in length and legibility to the test documents presented in Fig. 2. The trainee OC, however, worked on and off for four to five weeks to transcribe a similar papyrus. AW also emphasized how, as time passed, he repeatedly modified his textual interpretation (or "belief", to use Reverend Bayes' terminology).

One implication is that dataset size and participant number is limited by necessity in papyrological legibility evaluations. A further consideration is the feasibility and appropriateness of experiments conducted in controlled laboratory settings. Not only is their organization complicated by their length, but the cognitive process of activity followed by rest phases is also impractical to reproduce.

Under these constraints, the "homework" approach is more practical and realistic.

Fig. 3 Enhancement with (left to right) DStretch options rgb and lbk, and the novel methods vividness and lsv (detail of papyrus 4 of Fig. 2)

3 Related work

3.1 Papyrological practice

Three approaches to script enhancement characterize common papyrological practice: interactive image manipulation, spectral decorrelation, and imaging techniques [35, 142].

3.1.1 Interactivity

Image editors such as Adobe Photoshop allow easy and quick enhancement of papyri for a wide range of users. The flexibility of interactive image editing promotes solutions adapted to specific document qualities and user interests. Manipulations beyond the most basic tasks, however, require greater expertise, and are time-consuming.

3.1.2 Decorrelation

In the early 2000s, Jon Harman created DStretch, a software for enhancing pre-historical rock art from the American Southwest. This has since become a commonly used tool among papyrologists [71, 142]. DStretch decorrelates spectral bands by using either eigenvectors or principal components [8, 66]. The process can be performed in various color spaces, to highlight different features. An image is obtained with pixels that are ideally well-separated in terms of color values, and represented as pseudocolors. The method yields excellent results for images with a Gaussian distribution, which is characteristic of inscriptions such as on walls, potsherds, or stones. DStretch is also considered user-friendly by papyrologists and can be used in a field-work context, as a smartphone application. One limitation of the decorrelation stretch method is its poorer performance for images with a non-Gaussian distribution, as in the case of papyri reproductions. Moreover, the features emerging from decorrelation are not necessarily meaningful. Figure 3 shows how DStretch induces both achromatic contrast and chromatic noise, in comparison with two of our novel methods.

3.1.3 Imaging

Most papyri images are obtained through color or monochrome photography; these are the images of principal interest in this article. However, the specific needs of paleography have also lead to the use or pioneering of many sophisticated techniques. Ultraviolet and infrared imaging, which can reveal ink traces in recycled and degraded documents invisible to unaided human vision, are used to analyze individual papyri [5, 60, 181]. With the advent of commercial narrow-band hyperspectral scanners, large-scale digitization has also become possible [41, 93, 94]. Reflectance Transformation Imaging retrieves the three-dimensional texture of the papyrus surface by illuminating it from a multitude of angles, thus providing a level of depth information that is inaccessible through classic photography [142]. X-ray and terahertz-based imaging have been used for the non-destructive reading of subsurface inscriptions in papyri and mummy cartonnage [64, 105]. These methods have been coupled with computed tomography methods for the virtual unrolling of the Herculaneum papyri, carbonized during the famous eruption of Mount Vesuvius in 79 CE [7, 112].

3.2 Computer science research

The authors have found very few articles that specifically examine the enhancement of papyri legibility from visible light images. Sparavigna [175] applies an edge detector based on the magnitude of dipole moments to increase the contrast of shape outlines, and thereby further enhancing the effect of a similar edge detection neural mechanism already present in the human visual system, known as the Mach bands [26, pp. 362–363]. The anisotropic filtering of the image in the frequency domain was tested as a preprocessing step to suppress papyri fibers and gaps; unfortunately, this also removed character segments collinear with the fibers.

The best papyri enhancements are obtained using multispectral imaging, as noted above in the papyrological practice overview. Moreover, in conjunction with multispectral images, color processing has also been used for papyri legibility enhancement specifically and for document processing more generally, as will be shown in the next section. We will here widen the survey to encompass other methods, topics, and applications of potential interest to this article, within the field of document processing and beyond, and conclude by discussing the applicability of machine learning.

Because noise—whether biological, mechanical, electrical, or digital—is a fundamental issue in communication, substantial research has aimed at improving document legibility by suppressing a great variety of what are usually considered manifestations of *noise*, including: ink bleed-through [46], see-through [48], foxing [163], shadows [95], termite bites [151], cross-outs [30], photocopied low-contrast carbon copies [36], low resolution raster images [141], and background texture interference [133] (for a history of image

P6 – 91

denoising, see [111]). The type of visual *media* can also dictate the typology of enhancement methods (e.g., methods for pictures and for movies differ in whether the time dimension is available as a source of contextual information for optimizing image processing) [24]. In applied science contexts, such as in the photography and video equipment industry, there is interest in developing enhancement methods predicated on an understanding of the *nature of noise*, e.g., optical, mechanical, and electronic noise sources in cameras [139]. Advances in image quality measurement [87, 192] have benefited from research into *visual perception* and *neuroscience* [19, 188], as well as models of *scene statistics* [189]. The role of *tasks* in image enhancement is of particular interest in cultural heritage applications [180]. The *systemic* and *critical* approaches to document analysis advocated in this article have been the subject of exemplary research over more than a half-century in two domains, each with specific aims, constraints, and solutions. First, the legibility of flight documentation and instrumentation plays a critical role in *aviation* performance and safety; here, optimization has been approached mainly through psychophysical experimentation [55]. Second, research on the enhancement of *medical* radiographic images stands out in terms of the extent to which the impact on clinical diagnosis of technologies and perception have been investigated, including the role of visual illusions [23, 158].

In addition to methods, prominent areas of focus in this field are datasets, benchmarks, and ground truthing [117]. Visual confirmation of the attenuation of conspicuous artifacts is a typical means of comparing methods, supplemented with numerical characterization if feasible (i.e. given the availability of reference images or appropriateness of reference-free image quality measurements). User evaluations are rare, and vary from a few participants for historical documents [12, 28] to thousands in online campaigns for industrial applications [63, 144]. Legibility has been systematically studied since the early 20th century in experimental psychology (notably for the design of traffic signage and car plates, flight documentation and instrumentation, and typography for visually impaired [14, 114]), in works on teaching handwriting [159], as well as in relation to optical character recognition [68] and document image quality [4]. Document and image quality differ from writing legibility (e.g., a noise-free, but unfamiliar writing style may be difficult to read), although they conjointly affect the processing of visual information.

While the use of machine learning for image enhancement is well established [178], its application to document image enhancement, in particular historical, is still rare [72, 74, 115, 130, 131]. Notwithstanding results comparable to the state of the art, the challenges it faces are considerable. Historical data is scarce and heterogeneous, and ground truthing is time-consuming, which hinders the generalizability of the learned models [54]. In respect to the task of critical vision, an implementable machine learning methodology has yet to be devised that can deal with images that are ambiguous, and their interpretations multiple, variable, or unknown. For scholarly applications, the resulting enhancements have to be, furthermore, trustworthy.

4 Methods

4.1 Justifi ation

The proposed methods are rooted in color science and visual psychology and thus stand in contrast to the mathematics- and data-driven methods typically utilized in the document analysis domain. This section presents the rationale for and benefits of these methods.

Color image processing is a interdisciplinary field, owing to the multifarious aspects of color, as diverse as electromagnetic, perceptual, genetic, and linguistic [62, 97]. The perception of shape as the prominent aspect of documents has resulted in color receiving moderate attention in document image analysis. However, if the image resolution is low, or the text and background interfere with each other, as often occurs with papyri, color processing becomes an excellent alternative to spatial and frequency domain processing. Application examples of color-based document processing include historical document image enhancement [133], denoising, and binarization [185], page object segmentation [9], as well text extraction from typographic documents [173], web images [86], and cartographic maps [184]. Figure 4 illustrates the outcomes of the proposed and comparison methods, highlighting interactions between color and shape.

One far-reaching tenet of color research [84, 183] is that the technological, cognitive, and perceptual modalities through which visual reality is apprehended are not necessarily consistent with one another. In the negative polarity method, for example, physically and numerically identical gray-level differences are perceptually distinct. Because such effects, experienced as paradoxical, are the core of some proposed methods, this fact is emphasized through the use of the term *visual illusion* (see definitions, typologies, theories, history, and samples in [147, 164, 165]). Effects appropriate for enhancement were chosen among color illusions, since geometrical illusions are difficult to elicit without degrading legibility, and that illusions of stereopsis, apparent motion, and temporal illusions are tiring to use for extended periods of work.

The use of color and illusions shifts the burden of image enhancement from the computer to the human visual system. This perceptual image processing approach makes possible to create algorithms that are both mathematically

Fig. 4 Enhancement results on papyri PSI XIII 1298 (15a) r 1 (top) and P.Oxy.XXII 2309 (bottom). From left to right: the original; five standard enhancement methods; and the new methods proposed in this article (italics). Note the different attenuation of the smudge (top row) and papyrus texture (lower row). The image labels refer to the eponymous methods described in Sects. 4, "Methods", and 5.4 "Algorithms", and are abbreviations of (from left to right) original image, stretch limits (*stretchlim*), histogram equalization (*histeq*), adaptive histogram equalization (*adapthisteq*), local Laplacian filters (*locallapfilt*), retina and cortex (*retinex*), lightness, saturation, and value (*lsv*), vividness (*vividness*), negative vividness (*negvividness*), and negative lsv (*neglsv*). Retinex and vividness are established names in the literature, while stretchlim, histeq, adapthisteq, and locallapfilt carry over the function names given in the MATLAB programming language, which was the development environment in the work presented herein. All ninety evaluation images are available in the article's online Supplementary Material.

Fig. 5 Color values at the same location in an image with sRGB (left) and Adobe RGB (1998) (right) embedded color profiles (detail of Fig. 7).

straightforward, and considered by the experimental participants useful. Researchers also have noticed that simplicity is beneficial to legibility enhancement quality [179, pp. 404, 416], and the likelihood of implementation of methods in software and their use by the end-users [100, p. 920].

4.2 Chroma contrasting by gamut expansion

In Fig. 5, the picture on the right appears more saturated, and the text is more clearly separated from the background. Yet the color values of both image files are identical. The perceptual difference results from the files being tagged with different color profiles, meaning that the same values represent different locations in the color space. The embedded profile of the left image is sRGB IEC 61966-2.1, a standard designed for displays with a limited amount of reproducible color, as were typical in the 1990s. The space defined by these colors is called a color gamut (Fig. 6). The

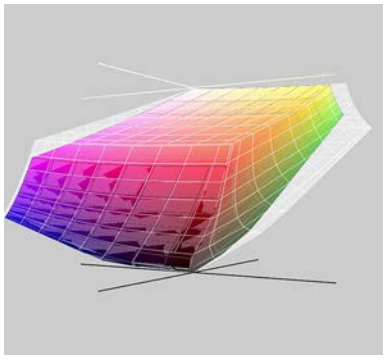


Fig. 6 The sRGB color gamut volume represented within the Adobe RGB (1998) gamut (gray hull) in the CIELAB color space. The vertical dimension represents lightness, L^*, while the horizontal axes correspond to the red-green, a^*, and the blue-yellow, b^*, opponent colors.

larger Adobe RGB (1998) gamut was designed to meet the needs of photographers and high-fidelity reproduction. Humans perceive far more colors than are present in either of these gamuts.

Numbers alone do not define colors; a coordinate system is also necessary. When a color space is replaced by a wider one, the gamut expansion increases the detectability of differences without affecting the values. The elegance of this legibility enhancement method resides in not requiring any numerical manipulation from software developers.

The underlying psychophysics is based on stronger chroma increasing the perceived level of luminance (*brightness*), which is known as the Helmholtz-Kohlrausch effect [45, 52]. This effect is leveraged as an unintended application of gamut mapping to improve the legibility of papyri.

A color space assignment and perceptual effects occur whenever an image file is read and visualized. Therefore being aware of their impact on legibility is important. In the case of papyri, the issue is compounded by so few reproductions having embedded color spaces. The most widespread space, sRGB, is assumed to apply in such situations.

The conversion of numerical values to visible color is done by the color management system of the image viewers, operating systems, and displays. For several reasons, transforming the numerical values to the destination color space may be desirable, instead of solely specifying it. For example, applications might disregard color space profiles, and thus deprive users of the enhancement effect. Nonetheless, the transforms among color spaces (*gamut mapping*) are user- and application-specific. The "*perceptual intent*" shifts out-of-gamut (OOG) color values to in-gamut values, thereby preserving perceptual color appearance; by contrast, the "*colorimetric intent*" strives to preserve numerical fidelity, usually by clipping outlier OOG values to the destination's gamut boundaries. This is a key topic in color management research [153, 166]. The legibility enhancement method suggested here can be made more robust by using standardized color conversion profiles, such as developed by the International Color Consortium (ICC). The relevant code in the MATLAB programming language is as follows:

```
source_profile = 'sRGB Profile.icc';
destination_profile = 'AdobeRGB1998.icc';
C = makecform('icc', destination_profile, source_profile);
I = imread('input_image.tif');
J = applycform(I, C);
imwrite(J,'output_image.tif', 'ColorSpace', 'rgb')
```

The results shown in Fig. 7 exhibit the change in chroma and color distribution. On the top is the original image with sRGB profile and its color distribution, and on the bottom the image after conversion to Adobe RGB (1998).

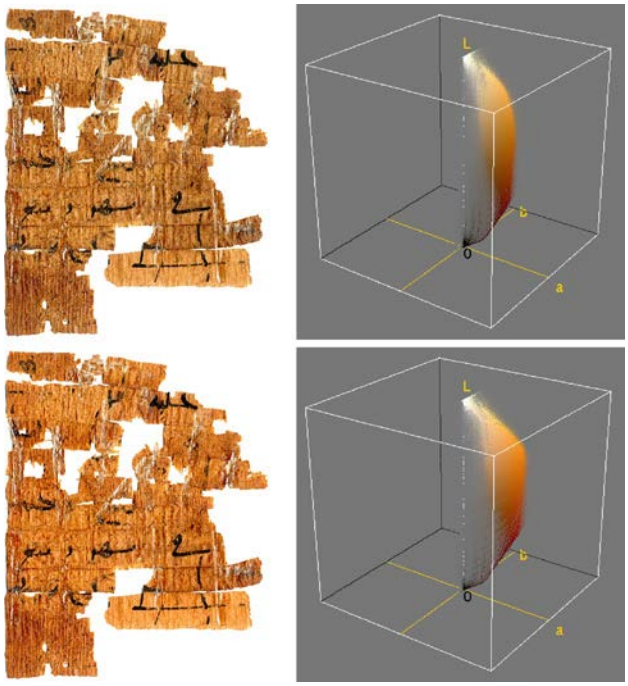
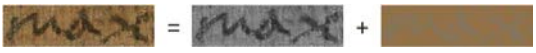


Fig. 7 Color distribution of a papyrus image with native sRGB color profile (top), and following conversion to Adobe RGB (1998) (bottom). (Credits of papyrus image: Sorbonne Université, Institut de Papyrologie, P. Sorbonne Arabe 201 1500a.)

P6 – 94

4.3 Lightness contrasting by stretching

Lightness refers to perceived reflectance, and is an achromatic component of color, together with brightness [157, pp. 2766–2767]. It is also a principal source of information about shapes for the human visual system, as can be ascertained via image decomposition into achromatic and chromatic colors:



Hence, increasing the lightness contrast substantially improves legibility. The enhancement formula (*stretchlim*) is

$$L^{*\prime} = [L^{*} - \min(L^{*})]/[\max(L^{*}) - \min(L^{*})], \quad (1)$$

where L^{*} is the lightness in the CIELAB color space, bound to the [0, 100] range and corresponding to the approximate number of perceptually noticeable lightness levels under optimal conditions [51, pp. 24, 202]. CIELAB was developed by the Commission Internationale de l'Éclairage (CIE) as a perceptually uniform color space allowing for the quantitative description and manipulation of color along fundamental phenomenological dimensions, such as lightness, chroma, and hue. Its perceptual proprieties often make

Fig. 8 Legibility varies with foreground/background and background/surround polarity.

CIELAB a better candidate for color processing than the Red, Green, and Blue (RGB) color space of common digital images. CIELAB is not without its shortcomings and more accurate color appearance models exist and are developed [51, 167]; however, for applications where faithful color reproduction is not the main goal, CIELAB remains adequate. The conversion between RGB and CIELAB is parametrized by a triplet that specifies the illuminant of the scene reproduced in the image [40]. D65 (noon daylight, 6504 K) provided in practice the greatest improvement to papyri legibility.

4.4 Lightness contrasting by negative polarity

The characters in Fig. 8, with CIELAB lightness values of 0 (black) or 100 (white), are set on squares of value 50 (gray). Despite identical numerical difference between figure and ground, the perceived contrast is stronger for light-on-gray configurations (bottom) and weaker for dark-on-gray (top); dark surrounds (left), and diminishing the stimuli (right), increase the effect. This occurs despite the perceptual linearity of CIELAB lightness and the use of calibrated displays.

This contrast illusion, known as *lightness induction*, is modulated by factors including asymmetrical polarity gain, nonlinear focus/surround contrast gain, and spatial frequency [26, 156]. It belongs to the broader category of *color constancy*, the neural and ecological basis of which has been a matter of debate since antiquity [2, 65]. Furthermore, the *irradiation effect*, well known to astronomers, makes light figures in dark surrounds (e.g. stars in the night sky) appear larger [143, 191].

Taken together, these effects let us posit that reversing the polarity of the image (i.e., so that the script appears lighter), will improve legibility. The empirical testing of this hypothesis as applied to papyri images is reported in Sect. 5, "Experiment". The formula for negative polarity (*neg*), $L^{*\prime}$, is

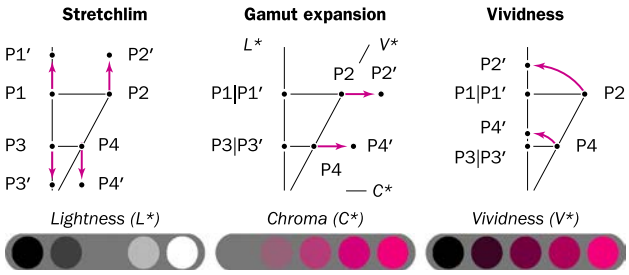


Fig. 9 Comparison of three enhancement methods described in this article. The vertical diagram axis represents the CIELAB lightness, L^*, the horizontal axis the chroma, C^*, and the diagonal the vividness, V^*. The variation of color along these axes is illustrated in the bottom part of the figure. For each enhancement type, the diagrams show the location of four exemplary points in a vertical section of the CIELAB color space before and after enhancement. Thus, stretchlim displaces all points towards the extremities of the lightness axis; gamut expansion displaces the points along the chroma axis (except for those situated directly on the lightness axis); finally, vividness applies a rotation transform resulting in all points being collinear with the lightness axis

$$L^{*\prime} = 100 - L^{*}, \tag{2}$$

where L^* is the image lightness in the CIELAB color space. The method is appropriate for papyri, a document type of medium lightness inscribed with dark ink, and typically imaged on a light background (see Fig. 2).

P6 – 95

4.5 Selective contrasting by vividness colorization

Chroma is a dimension of the CIELAB color space that is orthogonal to lightness and hue (Fig. 9). Chroma modification effects a change between achromatic and chromatic color. Another dimension is *vividness*, which defines a concomitant change in lightness and chroma [22]. Vividness can be used as a model of how a dark ink film applied to the papyrus reduces both the amount of reflected light, thereby changing the lightness, and acts as an achromatic filter. Thus, vividness helps to distinguish ink from papyrus better than lightness. They can be substituted one for the other to improve legibility. Enhancement based on vividness (vividness) preserves the papyrus appearance, because the chroma and hue values remain unchanged (Fig. 3). The process differs from stretchlim and gamut expansion in that the transform is nonlinear and selective: that is, more vivid colors are more strongly emphasized, while achromatic color values do not change. The enhanced lightness, $L^{*\prime}$, is given by

$$L^{*\prime} = (L^{*\,2} + a^{*\,2} + c^{*\,2})^{-1/2}, \tag{3}$$

where L^* is the lightness, and a^* and b^* are the chromatic components in CIELAB ($L^{*\prime}$ values exceeding 100 are

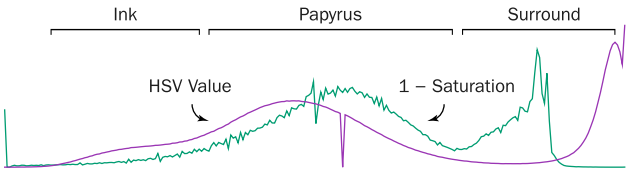


Fig. 10 The figure depicts an example of the color profile of a papyrus in the HSV color space (the "value" channel and the complement of the "saturation" channel), along with the object classes dominating specific bands of the spectrum (ink pixels, non-inked papyrus surface, and the imaging area surrounding the papyrus).

clipped). The mathematical expression of vividness corresponds to the l^2-norm of the CIELAB image values.

Because the vividness method modifies the image in a pixel-wise fashion, it increases the distinguishability of ink and papyrus even in the presence of objects such as color charts, which complicate the use of more sophisticated methods such as principal component analysis [27].

4.6 Background attenuation by difference of saturation and value

The hue, saturation, and value (HSV) color space is not a perceptual color space like CIELAB, rather, it is a simple mathematical transformation of the RGB color space [170]. Nonetheless, in HSV the distributions of saturation and value in papyri images roughly overlap (Fig. 10). On the basis of a simple difference between them, it is possible to attenuate the non-inked papyrus surface and, thus, enhance text legibility. Because the HSV value only approximates lightness, we fuse the difference of saturation and value with CIELAB lightness (e.g., by taking the mean) and accordingly obtain an image with sharper edges (Fig. 11).

The algorithm (*lsv*) represents a blind separation of a mixture of distributions. It consists in replacing the CIELAB lightness, L^*, by pixel-wise operations with $L^{*\prime}$, as follows:

$$L^{*\prime} = 100\ {}_0\lfloor 1 - ({}_0\lfloor 100 - L^* \rceil^1 + {}_0\lfloor V + S - 1 \rceil^1)/2 \rceil^1, \quad (4)$$

$$V = \max(R, G, B), \quad (5)$$

$$S = [V - \min(R, G, B)]/V, \quad (6)$$

where $L^{*\prime}$ is the enhanced lightness; V and S are the value and saturation image channels in the HSV color space, derived from the R, G, and B values of the RGB color space; and ${}_0\lfloor \cdot \rceil^1$ denotes normalization to the [0, 1] range.

Fig. 11 Papyrus before (left) and after (right) enhancement with the lsv method (detail of papyrus 8 in Fig. 2).

4.7 Hue contrasting by hue shift

4.7.1 Rationale

The spectral sensitivity of the human visual system is uneven: it peaks in the green band in bright light, and in blue-green in dim light; it is lowest in the red band [26, pp: 28-29, 118-119, 208-209]. As papyri have a brownish tint, a shift towards hues of higher acuity would be expected to improve legibility. The images in Fig. 12 result from applying this rationale. The first is the original, the second has all chromatic information discarded, and the remainder have the means of their CIELAB hue shifted to 246°, 162°, and 24°, which correspond to the loci of the primary colors [50, p. 343].

In addition for stronger acuity, a host of perceptual effects combine to make blue a compelling target for a hue shift. The increase in perceived lightness effected by an increase in chroma, as described above, is lowest for yellow and stronger for blue [52]. Moreover, the visual field for blue is broader than that for both green and red, thereby allowing for better comparison between focus and surround [26, pp. 108–109]. A blue text background has also been found to improve performance in analytic tasks [121]. Blue is furthermore robust to the most common colorblindness types [135].

Hue shift can be profitably combined with lightness polarity reversal: (*a*) lightness induction reinforces hue contrasting; (*b*) the asymmetry of the color space geometry reduces the dynamic range of chroma in the blue direction with respect to yellow for lightness levels above the medium [127, pp. 133–169].

Achromatic vision can resolve finer details than chromatic vision [128]. Because high spatial frequency is a characteristic of both script and papyri, using only the lightness channel for the decipherment task might make sense. However, the addition of color allows for ink to be better distinguished from stains, shadows, and other entities. Color vision in general improves image segmentation [29, 69].

4.7.2 Implementation

The shift of the papyrus towards blue (*blue* method) is accomplished via the following equation:

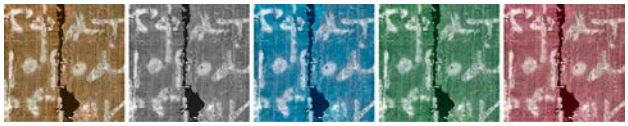


Fig. 12 From left to right: original, lightness channel L^*, and hue-shifted papyri; all have negative lightness polarity (detail of papyrus 4 in Fig. 2).

Fig. 13 Hue shift (left) vs sign change (right), plus negative polarity (detail of papyrus 4 in Fig. 2).

$$h_i' = h_i - h_p + h_b, \tag{7}$$

where h_i and h_i' are the input and output hue in the cylindrical CIELAB color space, with $h = \tan^{-1}(b^*/a^*)$ [51, p. 204]; while h_p is the centroid of the papyrus hue values and $h_b = 246°$ is the locus of blue in the hue dimension. An initial observation is that this is an ideal method, given the non-homogeneities in CIELAB. Second, the centroid of the papyrus hue is non-trivial to determine in the presence of other scene objects, such as color charts. Third, the location of maximum blue sensitivity varies across individuals, genders, ages, visual ecologies, and cultures, among other factors [49].

A simpler solution is available. This consists in changing the sign of the values in the a^* and b^* CIELAB channels:

$$a^{*\prime} = -a^* \quad \text{and} \quad b^{*\prime} = -b^* . \tag{8}$$

This is made possible because the brown hue of papyri has blue as its opponent color. It is this method, applied in conjunction with negative lightness, vividness, and gamut expansion, that was employed in the user evaluation. A comparison of the hue shift and sign change methods is shown in Fig. 13.

4.8 Dynamic range increase with CIELAB retinex

The retinex theory (from retina and cortex) provides an explanation for the constancy of color perception across a wide range of illumination conditions. Essentially, it expresses lightness as the product of ratios of spatial neighborhoods. As an image enhancement technique, it reveals details in dark and light areas. Many variants have been proposed since Land's and McCann's original work, and the out- comes differ according to image quality, content, and intent [107, 119]. The process is typically performed on each color channel individually, or on a single intensity image derived from the three RGB values, as a substitute

Fig. 14 Clockwise from top left: **a** Original image, with three faint annotations indicated by an index. **b** Enhancement by retinex in the RGB color space. **c** Retinex enhancement in the CIELAB color space. **d** Blue-shifted and negative of the SE image (detail of papyrus 8 in Fig. 2).

of lightness. Pursuing our thinking within the framework of color processing, we find that performing retinex in a perceptual color space (CIELAB) is beneficial to legibility enhancement. It results in a concomitant increase in contrast in several perceptual dimensions discussed in previous sections: lightness, saturation, vividness, and hue. Consequently, significant features may become visible. In Fig. 14, the presence of annotations in a different ink is striking in the CIELAB retinex process, and further details are revealed by the blue negative. The result, in terms of appearance, insights, and pitfalls, is analogous to watching the *Iliad* unfold in Technicolor rather than in black and white. The *cinematic metaphor of image processing* sums up the rewards and implications of color processing and visual illusions in support of critical vision.

4.8.1 Methodology

Conversion from RGB to CIELAB in the sRGB color space. Multiscale retinex with color restoration (MSRCR) [81, 140, 160] at three scales with Gaussian kernels of standard deviation 15, 80, 250, and a saturation of the 2.5% lowest and highest pixel values applied to (1) the RGB (Fig. 14b) and (2) the V^* (vividness), a^* (bluish–yellowish), and b^* (greenish–reddish) dimensions (Fig. 14c). Output normalization of (2) to the [0, 100], [−128, 127], and [−128, 127] ranges, respectively, and back-conversion to

RGB. (3) Blue negative applied after CIELAB retinex, with $V^{*\prime} = 100 - V^*$, $a^{*\prime} = -a^*$, $b^{*\prime} = -b^*$ (Fig. 14d).

4.9 Cross-spectral colorization

Although the proposed methods are designed for visible light images, their use may be extended to multispectral images.

4.9.1 Background

Multispectral images consist of a set of single-channel images taken in various bands of the light spectrum; hence, each one is perceptually achromatic. Image fusion, which is a common additional processing step, is based on component analysis, wavelet transform, and other dimensionality reduction techniques [12, 70, 80, 110, 123, 174].

The visualization of multispectral images and images from bands outside the visible spectrum (VIS) may be improved through the use of color processing and by integrating multiple images into a single image. The problem, however, is that non-VIS images have no natural color correspondence. Moreover, the number of dimensions of multispectral images (dozens or even hundreds) usually surpasses that which is perceptible by the trichromatic human visual system.

Many single- and multi-channel "pseudocoloring" solutions have accordingly been developed [39, 187, 195], ranging from analog optical [168] to digital frequency domain coloring [3] to the ubiquitous "colormaps" [1]. However, the perceptual intricacies that must be considered even when color mapping single grayscale images [98, 154] indicate considerable scope for additional research in this area.

In the document processing domain, the pseudocoloring of multispectral and non-VIS images has been used to increase the contrast between text and substrate. For example, George *et al.* [61] used the first three components of the principal and independent component analysis, respectively, of 31 spectral bands of an ostracon as input in the RGB color space. Easton *et al.* [47] inserted the ("invisible") ultraviolet-dominant grayscale image (long-wave ultraviolet light illumination) into the green and blue dimensions of the (visible) RGB color space, as well as the infrared-dominant grayscale image (tungsten illumination) into the red RGB dimension, thereby facilitating the legibility of the Archimedes parchment palimpsest. Gargano *et al.* [60] applied an overlay technique while working with papyri, which consisted in etching the binarized infrared image (or, alternatively, the red VIS channel) containing the document text into the visible light color image.

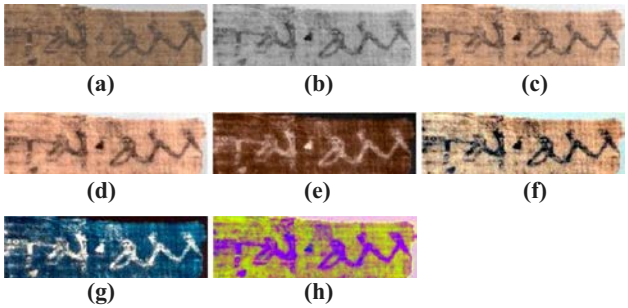


Fig. 15 Example of legibility enhancement by cross-spectral colorization. **a** Papyrus imaged in visible light; **b** infrared image; **c** the infrared, colorized with the chromatic channels of the visible light image and enhanced with the gamut expansion and stretchlim methods; **d** vividness enhancement of the image (**c**); **e** negative of image (**d**); **f** MSRCR-RGB retinex of image (**c**); **g** negative blue-shift of image (**f**); **h** and the decorrelation stretch of the blue, red, and infrared bands.

4.9.2 Method

Our solution does not rely on binarization, which presents problems when applied to papyri [149] and may introduce artifacts and suppress critical contextual information during the interpretation of difficult texts. The rationale is that a single-channel image from the non-visible spectrum (non-VIS), or a fusion of several bands, contains features of interest that are not or poorly identifiable in the visible spectrum. Conversely, the VIS image will contain contextual information, supporting the interpretation of the image. Because the "end-user" of the visualization process is the human visual system, we use the perceptual CIELAB color space rather than the physical RGB space.

The procedure consists in transforming the VIS image from RGB to CIELAB, inserting the non-VIS image into the lightness dimension L^*, and then converting back to RGB. We thereby obtain an image where the details (rendered as lightness) come from outside the visible spectrum, while the context (rendered as chroma and saturation) comes from within it. The use of color allows for visual discrimination between the two types of information, while the lesser spatial acuity of chromatic perception relative to achromatic perception [154] confers saliency to the script with respect to the background. Since we now have a color image, we may apply the methods presented above for further enhancement. Figure 15 illustrates the improved legibility resulting from the cross-spectral colorization process.

4.10 Method integration

The proposed methods lead to better results when used concurrently. A papyri legibility enhancement workflow is presented in Algorithm 1. The vividness method is taken as an

example, with gamut expansion, dynamic range stretching, negative polarity, and blue shift also included.

Algorithm 1. Papyrus legibility enhancement

▷ input
read RGB image values
▷ color gamut expansion (Section 4.2)
convert color profile using ICC profiles
· source color space: sRGB IEC 61966-2.1
· target color space: Adobe RGB (1998)
· source render intent: perceptual
· target render intent: perceptual
▷ use a perceptual color space for enhancement (§ 4.3)
convert color space from RGB to CIELAB
· source color space: sRGB IEC 61966-2.1
· whitepoint: D65
▷ vividness enhancement (§ 4.5)
replace lightness by vividness (Eq. 3)
$L^{*\prime} = (L^{*2} + a^{*2} + b^{*2})^{-1/2}$
▷ dynamic range increase (§ 4.3)
stretch dynamic range of lightness to bounds (Eq. 1)
$L^{*\prime} = [L^* - \min(L^*)] \,/\, \{\max[L^* - \min(L^*)]\}$
▷ negative polarity (§ 4.4)
reverse lightness polarity (Eq. 2)
$L^{*\prime} = 100 - L^*$
▷ blue shift (§ 4.7)
change sign of values in chromatic channels (Eq. 8)
$a^{*\prime} = -a^* \quad \text{and} \quad b^{*\prime} = -b^*$
▷ back-conversion to primary colors space
convert color space from CIELAB to RGB
· source color space: sRGB IEC 61966-2.1
· whitepoint: D65
▷ manage out-of-gamut values
clip values to the [0, 1] range
$I' = \min(0, \max(1, I))$
▷ output
save image in TIFF or JPEG format
▷ make color profile explicit
embed ICC color profile in image file
· color space: sRGB IEC 61966-2.1

5 Experiment

5.1 Objectives

An experiment was carried out with papyrologists as participants in pursuit of four objectives: (1) Test the hypothesis that the proposed methods outperform existing ones. (2) Use the results to support the decision as to which methods to implement. (3) Explore the data to discover further potentially relevant facts and operationalize the findings. (4) Design the experiment to gain insight into the realistic usage conditions of the legibility enhancement software.

5.2 Participants

The experiment participants were two female and five male graduate students, and one male faculty member participating in the Eucor Tri-national Papyrology Webinar 2019-2020 held by the Universities of Basel (Switzerland), Freiburg (Germany), and Strasbourg (France). All were trained in Ancient Greek. Participation in the experiment counted towards their grades. The sample is more representative than suggested by its small size, if it is considered relative to the total number of papyrologists (the International Association of Papyrologists had 484 members in May 2020), and its international heterogeneity.

5.3 Stimuli

The stimuli consisted of a training and an evaluation set of papyri images. The training set comprised eight images of unpublished papyri from the University of Freiburg Library, whereas the evaluation set comprised the nine publicly available papyri color images from the ICDAR 2019 Competition on Document Image Binarization (DIBCO) [149] (Fig. 2). The quality of the original images in the training set was similar to that of the noisiest evaluation images. The DIBCO images represent typical qualities of papyri documents and reproductions, and the authors participated in their selection process. Each original method was processed with ten of the methods described in the previous section, thus bringing the total of images to 80 and 90 in the training and evaluation set, respectively (the evaluation images are available in the online Supplementary Material).

P6 – 99

5.4 Algorithms

Four novel enhancement algorithms (*vividness*, *lsv*, *negvividness*, and *neglsv*) and five comparison algorithms (*stretchlim*, *histeq*, *adapthisteq*, *locallapfilt*, and *retinex*) were used. Given the variability of the lightness dynamic range in the test images, and to facilitate a bias-free comparison of methods, all images had their CIELAB lightness normalized in a preprocessing step, while their gamut was also expanded as described above. All methods were implemented in the MATLAB (R2020b) programming environment, except for retinex, which was produced with the ImageJ software package. The choice of comparison methods was based on their status as classical enhancement methods (stretchlim, histeq, adapthisteq), having outstanding edge-preserving performance (locallapfilt), and being a prime choice for achieving color contrast equalization (retinex). — Input images had no embedded color space information, except for P.Oxy. XXII 2309, which was in sRGB. All methods converted the input images to the Adobe RGB (1998) color space with D65 white point and saved them to sRGB to expand the color

gamut. Lightness was obtained from CIELAB. — The novel algorithms were those described in Sect. 4, "Methods". *Negvividness* and *neglsv* apply negative lightness to *vividness* and *lsv*. Due to late-breaking research hue shift and *CIELAB retinex* were not included in the experiment. — *Stretchlim* [118] expands the dynamic range of an image's CIELAB lightness to its bounds, [0, 100]. — *Histeq* [79, 118] is a global enhancement method that equalizes the spread of the histogram values of an intensity image over the available bit-depth dynamic range via linearization of its cumulative density function, thereby improving its contrast. — *Adapthisteq* (contrast-limited adaptive histogram equalization, CLAHE) [118, 198] is a local enhancement method similar to histeq, in that it equalizes the image histogram; however, it does this on small regions (here, 8-by-8 pixels). Accordingly, contrast for local structures is increased. — *Locallapfilt* [118, 134] is a state-of-the-art edge-preserving image enhancing method using pyramids of locally adapted Laplacian filters. Depending on parametrization, it can be used to either smooth or sharpen. Smoothing was used here, with parameters $\sigma = 0.4$, $\alpha = 2$, and $\beta = 1$, derived empirically for their appropriateness to papyri images.—*Retinex* images were produced with the ImageJ implementation (RGB color space, uniform level, scale 240, scale division 3, and dynamic 1.20) [73].

P6 – 100

5.5 Procedure

Each participant transcribed the text of a different image from the training set as part of their webinar examination. Afterward, they were asked to rate the processed images according to how useful they were for text transcription, using the following system: X: "*I used only one image for the transcription, this one*"; A: "*Very useful: it was the primary image used for transcription*"; B: "*I sometimes used this image*"; N: "*I didn't use this image*". (The results were similar to those obtained from the evaluation set.) This experimental step in real use conditions acquainted participants with the opportunities provided by the various processing methods for scholarly text editing.

During the evaluation step, all participants worked on all images of the evaluation set (Fig. 2). The first task involved rating the images for a hypothetical transcription on the same scale as used for the training set. The second task involved the ranking of the same images by utility. Given the effort and time invested in a scholarly transcription of ancient papyri, requiring participants to provide a transcription of all evaluation images it would have been impractical.

Several further questions were asked of the participants regarding how they organized the images and interacted with the computer to compare the versions (e.g., side by side, overlaid, or printed); whether they zoomed into the images; whether they used a desktop, laptop, tablet, or other device; how many displays were used and what sizes they were; what the operating system and image display application were used; and whether they could provide any other information about using the processed images for the transcription task.

5.6 Setting

The experimental instructions were sent by email to the participants, who downloaded the images to their computers and eventually sent back the completed questionnaires. Given the virtual nature of the papyrology webinar, the participants never physically convened in the same location as the experimenters, thus precluding testing in controlled laboratory conditions. This constraint, however, turned out to be a great opportunity. The homework "field study" enabled *in situ* simulation of the use of then future legibility enhancement software in realistic conditions in a variety of dimensions, and thus led to findings that could not have been obtained in a laboratory setting. In addition to leaving the choice of evaluation environment and time to participants, the setting also maintained their familiarity with the software and hardware employed. Data organization on personal computers, human-computer interaction patterns developed in coexistence with specific operating systems and applications, as well as the display size and number are only some of the factors that would have biased the results if participants had to work in the unfamiliar laboratory settings.

To more accurately simulate real-life conditions, the image filenames were also not anonymized. The identity of these images is indeed known to real users, since they select methods by name in the enhancement software interface. Furthermore, the methods employed in this evaluation produce images that are sufficiently peculiar to make them identifiable at a glance (such as the negative polarity method), which defeats the purpose of anonymization. The differentiated practices that may develop from this knowledge are part of normal user behavior and represent the insights that the evaluation aims to develop. If anonymization was employed, these real behaviors would not have been captured, meaning that the evaluation would have simulated a non-existent condition. An illuminating example of this rationale concerns the special treatment given by the experimental participants to the original image, which was ranked as the most reliable image for the transcription task despite broad acknowledgment of the better script-background contrast in the enhanced images.

Similar considerations to ours have led to remote evaluations in two other legibility enhancement evaluations of historical documents [12, 28]. Large-scale image and video quality assessments with 8100 and 985 participants from multiple countries, rating 1162 and 3000 images respectively, found the results of laboratory and unconstrained online evaluations to be equivalent [63, 144]. Methodologies

for crowdsourcing image quality assessment have also been actively investigated [78, 152], while the "ecological validity" of laboratory and field experimentation has been debated for over half a century in fields such as psychology, sociology, and economics [57, 76, 90, 124].

5.7 Frame

The experiment measured the preferences for legibility enhancement methods in the context of specific documents and available hardware and software, developed in real-life conditions by participants as they gained expertise in using these methods for scholarly text transcription. This statement is necessary to clarify how legibility enhancement was framed by the experimental design. It is not (as would be possible with a different design) transcription errors or duration of task completion that are measured to judge the utility of methods for enhancing legibility, but rather their utility as manifested in their rating and ranking by papyrologists.

6 Results

Table 1 presents the ratings and rankings of the image processing methods used in the experiment. Providing this data *in extenso* is essential if we are to make visible the statistical features quantified, summarized, and discussed in the next section. For example, while user behavior heterogeneity is apparent from Table 1, it is obfuscated in Table 3 through statistical data aggregation.

6.1 How good are the proposed methods?

A visual inspection of Table 1 suggests that, as a group, the proposed methods ranked overall higher than the comparison methods. This impression is substantiated by the distribution of ratings of the two method sets according to their utility: the proposed methods are prevalently of primary use, whereas the comparison methods are mostly of no use (Table 2).

6.2 How do the methods differ?

The methods' effects on legibility were examined by combining the participants' feedback with a visual inspection of the enhanced images (Fig. 4) and understanding of the algorithms. The original image was typically low-contrast, and some portions of the script were faded or missing. *Stretchlim* (i.e., normalization of lightness) was often insufficient. *Histeq* created high local contrast variation, because it is based on global contrast improvement. *Adapthisteq* strongly enhanced the papyrus structure, which interferes with the script. *Locallapfilt* had very low contrast, and its smoothness was too strong; parametrization for individual images becomes thus necessary, although this is difficult to do automatically and inconvenient when done manually. *Retinex* performed best out of all methods at revealing the script, creating very sharp edges. However, reading was also made difficult because of interference from the papyrus structure, and there was also a strong vignetting effect that sometimes completely obscured parts of the image. *LSV* had excellent smoothing properties, but often damaged script structures. *Vividness* gives the most balanced results, although the contrast could be improved.

6.3 Which methods should be implemented in software?

To determine the relative utility of the individual methods, their ratings were aggregated across all participants and documents in an overall ranking (Table 3). This ranking was obtained with two different ranking methods, after which the outputs were aggregated and optimized. The top-1 enhancement method is vividness, followed by the negative method. The least useful were locallapfilt and histeq.

6.3.1 Methodology

While ranking has been extensively investigated in statistics [11], we found methods with more useful properties for the task at hand in the domains of operations research, marketing, sports, and voting theory. These methods are non-parametric, and thus facilitate the handling of data with unknown distributions. They take categorical and ordinal data as input, which corresponds to the experimental ratings and rankings. They also differ in terms of their perspectives, with the mean-based methods being competitive and rewarding the individual performance of the enhancement methods, while the median-based method is consensus-based and satisfies the majority of users. These methods are briefly described below; the code was published in [15].

The *Centroids* ranking is based on matrix analysis [108, pp. 127–129]. The n-by-1 vector $\boldsymbol{r}$ of the aggregated rating scores of n items with individual scores s (or ranks k, for which $s = n + 1 - k$) is defined as follows:

$$\boldsymbol{r} = \text{round}[(\boldsymbol{S} - \boldsymbol{S}^{\mathbf{T}})\boldsymbol{e}/n], \tag{9}$$

where $\boldsymbol{S}$ is an n-by-n matrix, with $\boldsymbol{S}_{ij} = \sum(s_i)$ for $i \neq j$ and $\boldsymbol{S}_{ij} = 0$ for $i = j$ (diagonal of zeros), $\boldsymbol{S}^{\mathbf{T}}$ is the transpose of $\boldsymbol{S}$, and $\boldsymbol{e}$ is an n-by-1 vector of ones. (Note the row-wise averaging, from which the method derives its name.) Using Table 3A as an example, the items correspond to the enhancement methods; their corresponding ranks are obtained from Table 1, after which application of the above equations results in the values listed in the rating column

Table 1 Ranking and rating of legibility enhancement methods.

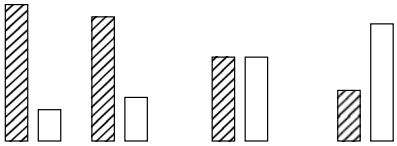
	AW ♂ 2 displays	FG ♀ 3 displays	FRG ♂ 1 display	LG ♂ 2 displays	OC ♂ 2 displays	OR ♀ 1 display	SR ♂ 2 displays	SN ♂ 1 display
columbia.apis. p367.f.0.600	A adapthisteq negisv retinex B negnorm norm stretchlim lsv *original* histeq N locallapfilt	A *original* adapthisteq histeq locallapfilt lsv B negisv negnorm norm stretchlim retinex	A adapthisteq B norm negisv *original* N lsv stretchlim histeq retinex negnorm locallapfilt	A negisv B lsv negnorm norm stretchlim *original* locallapfilt N retinex adapthisteq histeq	A norm negisv N *original* retinex adapthisteq histeq locallapfilt lsv negnorm stretchlim	A *original* B retinex negisv negnorm N adapthisteq histeq locallapfilt lsv norm stretchlim	A negnorm B locallapfilt lsv negisv norm stretchlim *original* N adapthisteq histeq retinex	A lsv B negisv negnorm locallapfilt *original* norm stretchlim adapthisteq N retinex histeq
P.Corn. Inv. MSS. A 101. XIII	A retinex negnorm *original* adapthisteq stretchlim B histeq negisv norm N locallapfilt lsv	A *original* adapthisteq histeq locallapfilt lsv B negisv negnorm norm stretchlim retinex	A *original* B negnorm norm retinex histeq N stretchlim lsv negisv adapthisteq locallapfilt	X negnorm N *original* retinex adapthisteq histeq locallapfilt lsv negisv norm stretchlim	X stretchlim N *original* retinex adapthisteq histeq locallapfilt lsv negisv negnorm norm	X *original* N retinex adapthisteq histeq locallapfilt lsv negisv negnorm norm stretchlim	A adapthisteq B *original* histeq negnorm locallapfilt lsv norm retinex stretchlim N negisv	A negnorm B *original* lsv stretchlim locallapfilt norm negisv adapthisteq histeq retinex
P.CtYBR inv. 69	A retinex norm *original* B negnorm adapthisteq stretchlim N histeq lsv negisv locallapfilt	A *original* adapthisteq histeq locallapfilt lsv B negisv negnorm norm stretchlim retinex	A norm B *original* histeq retinex negisv N stretchlim adapthisteq lsv negnorm locallapfilt	A negisv B negnorm lsv *original* norm stretchlim N retinex adapthisteq histeq locallapfilt	A norm locallapfilt N *original* retinex adapthisteq histeq lsv negisv negnorm stretchlim	A *original* B negisv retinex N adapthisteq histeq locallapfilt lsv negnorm norm stretchlim	A negnorm B *original* stretchlim lsv norm locallapfilt negisv adapthisteq N retinex histeq	A negnorm B *original* locallapfilt lsv norm stretchlim N negisv adapthisteq retinex histeq
P.Mich.inv.1318v	A retinex adapthisteq negisv norm *original* B histeq negnorm stretchlim N lsv locallapfilt	A *original* adapthisteq histeq locallapfilt lsv B negisv negnorm norm stretchlim retinex	A norm B stretchlim histeq *original* negnorm retinex N lsv negisv adapthisteq locallapfilt	A negnorm B negisv norm *original* N retinex lsv adapthisteq stretchlim histeq locallapfilt	A norm lsv N *original* retinex adapthisteq histeq locallapfilt negisv negnorm stretchlim	A *original* B retinex N adapthisteq histeq locallapfilt lsv negisv negnorm norm stretchlim	B negisv lsv norm negnorm adapthisteq stretchlim retinex N *original* locallapfilt histeq	A lsv B negnorm negisv norm locallapfilt *original* N retinex stretchlim adapthisteq histeq
P.Mich.inv.2755	A retinex norm negisv B adapthisteq stretchlim *original* N histeq lsv negnorm locallapfilt	A *original* adapthisteq histeq locallapfilt lsv B negisv negnorm norm stretchlim retinex	A norm B histeq negisv stretchlim retinex N lsv *original* negnorm adapthisteq locallapfilt	B retinex lsv norm stretchlim *original* N negisv negnorm adapthisteq locallapfilt histeq	X norm N *original* retinex adapthisteq histeq locallapfilt lsv negisv negnorm stretchlim	X *original* N retinex adapthisteq histeq locallapfilt lsv negisv negnorm norm stretchlim	A negnorm B negisv norm adapthisteq lsv stretchlim *original* retinex N histeq locallapfilt	A negnorm B *original* negisv lsv norm stretchlim adapthisteq N locallapfilt histeq retinex
P.Oxy.XXII 2309	A retinex norm adapthisteq negnorm *original* B stretchlim negisv N histeq locallapfilt lsv	A *original* adapthisteq histeq locallapfilt lsv B negisv negnorm norm stretchlim retinex	A norm B stretchlim negnorm retinex adapthisteq N histeq locallapfilt lsv negisv *original*	A *original* B negnorm norm stretchlim N negisv retinex adapthisteq locallapfilt histeq lsv	A norm negisv N *original* retinex adapthisteq histeq locallapfilt lsv negnorm stretchlim	X *original* N retinex adapthisteq histeq locallapfilt lsv negisv negnorm norm stretchlim	A negnorm B norm stretchlim *original* locallapfilt lsv negisv adapthisteq retinex N histeq	A negnorm B *original* stretchlim norm retinex lsv N negisv locallapfilt adapthisteq histeq
PSI XII 1274 r	A retinex adapthisteq negisv norm *original* B stretchlim lsv negnorm histeq N locallapfilt	A *original* adapthisteq histeq locallapfilt lsv B negisv negnorm norm stretchlim retinex	A norm B negisv retinex histeq N stretchlim *original* lsv adapthisteq negnorm locallapfilt	A negisv B negnorm norm *original* lsv stretchlim N retinex locallapfilt adapthisteq histeq	X norm N *original* retinex adapthisteq histeq locallapfilt lsv negisv negnorm stretchlim	A *original* B retinex negisv negnorm N adapthisteq histeq locallapfilt lsv norm stretchlim	A lsv B stretchlim locallapfilt negnorm norm adapthisteq negisv retinex N *original* histeq	A lsv B locallapfilt *original* norm stretchlim negnorm negisv N retinex adapthisteq histeq
PSI XIII 1298 (15a) r 1	A retinex negisv *original* B lsv negnorm adapthisteq norm N histeq stretchlim locallapfilt	A *original* adapthisteq histeq locallapfilt lsv B negisv negnorm norm stretchlim retinex	A lsv B *original* retinex histeq norm negisv N negnorm stretchlim adapthisteq locallapfilt	B negisv adapthisteq lsv stretchlim negnorm norm N *original* retinex histeq locallapfilt	X norm N *original* retinex adapthisteq histeq locallapfilt lsv negisv negnorm stretchlim	X *original* N retinex adapthisteq histeq locallapfilt lsv negisv negnorm norm stretchlim	A negisv B lsv norm negnorm stretchlim adapthisteq retinex N *original* histeq locallapfilt	A lsv B *original* negnorm negisv norm stretchlim N retinex locallapfilt adapthisteq histeq
PSI XIV 1376 r	A norm negisv lsv adapthisteq B negnorm stretchlim retinex *original* N histeq locallapfilt	A *original* adapthisteq histeq locallapfilt lsv B negisv negnorm norm stretchlim retinex	A lsv negisv norm histeq retinex N adapthisteq negnorm stretchlim *original* locallapfilt	B negisv negnorm lsv *original* N adapthisteq histeq locallapfilt norm retinex stretchlim	A norm adapthisteq negisv N *original* retinex histeq locallapfilt lsv negnorm stretchlim	A *original* B retinex negisv negnorm locallapfilt N adapthisteq histeq lsv norm stretchlim	A lsv B norm negnorm stretchlim negisv locallapfilt *original* adapthisteq retinex N histeq	A negnorm B negisv *original* lsv norm stretchlim adapthisteq locallapfilt N histeq retinex

Legend: "X" denotes methods used to the exclusion of all others, "A" those that were of primary use, "B" those that were of secondary use, and "N", on gray background, those that were of no use. Brackets indicate tied rankings

Table 2 Utility of enhancement methods

Methods	*Utility category counts*			
	Exclusive X	*Primary A*	*Secondary B*	*No use N*
original	4	22	**28**	18
Proposed methods				
lsv	0	10	30	**32**
vividness	3	16	**41**	12
neglsv	0	13	**38**	21
negvividness	1	12	**36**	23
Total proposed	*4*	*51*	***145***	*88*
Comparison methods				
stretchlim	1	1	**42**	28
histeq	0	0	22	**50**
adapthisteq	0	9	25	**38**
locallapfilt	0	1	23	**48**
retinex	0	8	**33**	31
Total comparison	*1*	*19*	*145*	***195***

Utility category percentages, column-wise



Proposed methods	**80**	**73**	**50**	31
Comparisons	20	27	**50**	**69**

Table 3 Overall ranking of enhancement methods by utility ratings.

A. Ranking by Centroids method

Rank	*Method*	*Rating*	*Spread* (1)
1	*original*	31.1	
2	**vividness**	29.1	
3	**neglsv**	11.1	
4	**negvividness**	10.1	
5	**lsv**	−2.9	
6	retinex	−3.9	
7	stretchlim	−5.9	
8	adapthisteq	−9.9	
9	locallapfilt	−27.9	
10	histeq	−30.9	

(1) Graphical representation of the ratings distribution.

B. Ranking by Majority Judgement method

Rank	*Method*	*Rating*	*Spread*	*Category*
1	*original*	2.37		Secondary use +
2	**vividness**	2.26		Secondary use +
3	**neglsv**	1.71		Secondary use −
4	**negvividness**	1.68		Secondary use −
5	stretchlim	1.61		Secondary use −
6	retinex	1.57		Secondary use −
7	**lsv**	1.56		Secondary use −
8	adapthisteq	1.48		No use
9	locallapfilt	1.33		No use
10	histeq	1.31		No use

C. Aggregation of rankings by ROD method

Rank	*Method*	*Rating*	*Spread*	*Optimization*
1	*original*	∞		*original*
2	**vividness**	73.908		**vividness**
3	**neglsv**	2.091		**neglsv**
4	**negvividness**	1.779		**negvividness**
5	stretchlim	0.483		stretchlim
6	retinex	0.475		retinex
7	**lsv**	0.460		**lsv**
8	adapthisteq	0.204		adapthisteq
9	locallapfilt	0.008		locallapfilt
10	histeq	0.000		histeq

of Table 3A. Their ranking is obtained by sorting the rating vector $\boldsymbol{r}$ in descending order of magnitude.

The *Ratio of Offense and Defense* (ROD) [108, pp. 176–197] takes as input the same score matrix $\boldsymbol{S}$ used by the Centroids method, from which it derives the ratings vector $\boldsymbol{r}$ through the following equation:

$$\boldsymbol{r} = \boldsymbol{o}/\boldsymbol{d}, \tag{10}$$

where the methods' "offensive" strength is $\boldsymbol{o} = \boldsymbol{R}\boldsymbol{e}$, while their "defensive" strength is $\boldsymbol{d} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{e}$. (The terminology comes from the world of sports, where this ranking method originates.)

In *Majority Judgment* each elector grades each candidate (instead of choosing a single one), and rates and ranks are obtained from mean scores, resulting in the elections of those candidates that are the most acceptable to the largest number of electors, i.e. consensual [18]. The algorithm begins by iteratively removing the median values $\{m_1, m_2, ..., m_p\}$ from the ratings $\{r_1, r_2, ..., r_p\}$ of each candidate $\{c_1, c_2, ..., c_p\}$, and concatenating them lexicographically. For example, the ratings vector $\{7, 9, 9, 11, 11\}$ is reordered by medians into $\{9, 9, 11, 7, 11\}$, and concatenated to generate the numerical value 9.09110711. The resulting scalars $\{v_1, v_2, ..., v_n\}$, called "majority values", are then sorted in descending order of magnitude to yield the candidate rankings $\{k_1, k_2, ..., k_p\}$, where $v(k_1) \succ v(k_2) \succ ... \succ v(k_p)$. This procedure was applied to the enhancement method utility ratings {X, A, B, N} of Table 1 in order to generate the rankings and ratings presented in Table 3B. While ranking is an ordinal representation and rating is numeric, verbal qualifiers have their own benefits. For example, the utility of the ranked enhancement methods is easily recognizable from

P6 – 103

Table 4 Intraclass ranking of enhancement methods by utility ratings (optimized aggregation by ROD method; ties in brackets)

A. Within–user ranking (column-wise aggregation of ratings from Table 1; ties in brackets)

AW	FG	FRG	LG	OC	OR	SR	SN
retinex	*original*	**vividness**	**negvividness**	**vividness**	*original*	**negvividness**	**negvividness**
original	**lsv**	retinex	**negIsv**	**negIsv**	retinex	**lsv**	**lsv**
vividness	**vividness**	histeq	*original*	stretchlim	**negIsv**	**negIsv**	*original*
adapthisteq	**negIsv**	**negIsv**	**vividness**	**lsv**	**negvividness**	adapthisteq	**vividness**
negIsv	**negvividness**	*original*	**lsv**	adapthisteq	locallapfilt	**vividness**	stretchlim
negvividness	stretchlim	**lsv**	stretchlim	locallapfilt	**lsv**	stretchlim	**negIsv**
stretchlim	histeq	**negvividness**	adapthisteq	*original*	**vividness**	retinex	locallapfilt
lsv	adapthisteq	stretchlim	locallapfilt	**negvividness**	stretchlim	*original*	adapthisteq
histeq	locallapfilt	adapthisteq	retinex	histeq	histeq	locallapfilt	retinex
locallapfilt	retinex	locallapfilt	histeq	retinex	adapthisteq	histeq	histeq

B. Within–document ranking (row-wise aggregation of ratings from Table 1; ties in brackets)

columbia.apis. p367.f.0.600	P.Corn.Inv. MSS.A 101.XIII	P.CtYBR inv.69	P.Mich. inv.1318v	P.Mich. inv.2755	P.Oxy.XXII 2309	PSIXII 1274r	PSIXIII 1298 (15a) r1	PSIXIV 1376r
negIsv	*original*	*original*	**vividness**	**vividness**	*original*	**vividness**	*original*	**negIsv**
original	**negvividness**	**vividness**	*original*	*original*	**vividness**	**negIsv**	**lsv**	**lsv**
vividness	stretchlim	**negvividness**	**negvividness**	stretchlim	**negvividness**	*original*	**negIsv**	*original*
negvividness	retinex	**negIsv**	**negIsv**	**negIsv**	stretchlim	**negvividness**	**vividness**	**vividness**
lsv	adapthisteq	stretchlim	retinex	retinex	retinex	**lsv**	**negvividness**	**negvividness**
adapthisteq	**vividness**	locallapfilt	**lsv**	**negvividness**	**negIsv**	retinex	retinex	adapthisteq
stretchlim	histeq	retinex	stretchlim	**lsv**	adapthisteq	stretchlim	stretchlim	retinex
locallapfilt	**lsv**	**lsv**	adapthisteq	adapthisteq	**lsv**	adapthisteq	adapthisteq	stretchlim
retinex	**negIsv**	adapthisteq	histeq	histeq	locallapfilt	histeq	histeq	locallapfilt
histeq	locallapfilt	histeq	locallapfilt	locallapfilt	histeq	locallapfilt	locallapfilt	histeq

the category column of Table 3B, which uses the spelled-out names of the four utility classes {Exclusive, Primary, Secondary, No use} corresponding to the ratings. The plus and minus signs following the class names indicate whether a method belongs to the upper or lower range of a utility class.

Empirical evidence suggests that aggregated opinions may outperform individual ones, an approach formalized by *ensemble methods* [103]. When comparing the results of the Centroids and Majority Judgment methods, differences can be seen in ranking and rating, reflecting the methods' differing outlooks. To obtain a balanced result, the two rating vectors are aggregated, first by deriving the matrices $\boldsymbol{R}$ of pairwise positive differences from each of the n (in our case $n = 2$) rating vectors $\boldsymbol{r}$:

$$d_{p(ij)} = [\boldsymbol{r}_{p(i)} - \boldsymbol{r}_{p(j)}], \tag{11}$$

$$\boldsymbol{R}_{p(ij)} = \begin{cases} d_{p(ij)}, & \text{for } d_{p(ij)} > 0, \\ 0, & \text{for } d_{p(ij)} \leq 0, \end{cases} \tag{12}$$

where $p = 1, 2, ..., n$. The matrices are then normalized and averaged:

$$\overline{\boldsymbol{R}} = \sum [\boldsymbol{R}_p / \sum (\boldsymbol{R}_p \boldsymbol{e})]/n, \tag{13}$$

where $\boldsymbol{e}$ is an n-by-1 vector of ones. Finally, the aggregated scores vector $\boldsymbol{r}$ is obtained using the ROD method: $\boldsymbol{r} = \overline{\boldsymbol{R}}\boldsymbol{e}/\overline{\boldsymbol{R}}^{\mathbf{T}}\boldsymbol{e}$. This procedure produced the ratings shown in Table 3C. The approach is akin to graph analysis, as commonly used in, e. g. website ranking methods, since negative differentials between pairwise scores are set to zero, meaning that the matrix $\overline{\boldsymbol{R}}$ reflects the dominance of each enhancement method.

A further refinement, named *Kemenization*, allows for the aggregated ranking to be optimized [88, pp. 9–23; 108, pp. 175–176]. It consists in switching item k_i with item k_{i+1} in the aggregated vector r if item k_{i+1} occurs more frequently among the ranks i of the input vectors $\boldsymbol{r}_p$. From the optimization column of Table 3C, we can observe that the procedure results in no modification of the aggregated ranking.

6.4 Are the evaluated methods suffici t?

The fact that multispectral images typically provide superior legibility indicates that additional enhancement methods are desirable and sometimes indispensable (e.g., when reading carbonized papyri [125, 171]). Nevertheless, participants attempting further image manipulation, such as in Photoshop, could not usefully improve legibility: "The enhanced pictures were most often so good that this didn't help as much as I hoped" (SN). Also, one method, i.e. vividness, was substantially better than the others (Table 3). Finally, over-reliance on the original image was apparent: "I like to work on the original version, and I find it's not always necessary to look at the others" (OR); "I mainly worked with the 'base' image (vividness or original). This represented 95% of my work. . . . The other . . . I briefly tried to use them, but quickly abandoned." (OC). This final comment was reflected in the numerous cases in which methods were deemed not useful (42% of $n = 720$ in Table 2; swaths of gray in Table 1). Provided that the DIBCO dataset is representative of papyri, the conclusion is that the amount and type of novel and existing methods made available are satisfactory on average.

Table 5 Agreement of ratings of enhancement methods, measured with Kendall's coefficient of concordance, *W*

A. Inter-class agreement

Between users	0.376
Between documents	0.698

B. Within-user agreement

1.	FG	1
2.	OR	0.727
3.	AW	0.616
4.	SN	0.607
5.	OC	0.500
6.	LG	0.474
7.	SR	0.453
8.	FRG	0.386

C. Within-document agreement

1.	P.Oxy.XXII 2309	0.376
2.	PSI XIII 1298 (15a) r 1	0.286
3.	P.Corn. Inv. MSS. A 101. XIII	0.283
4.	P.Mich.inv.2755	0.280
5.	columbia.apis.p367.f.0.600	0.280
6.	PSI XIV 1376 r	0.246
7.	P.CtYBR inv. 69	0.243
8.	P.Mich.inv.1318v	0.236
9.	PSI XII 1274 r	0.220

6.5 Enhancement methods are complementary and their utility is context-dependent

Table 1 reveals a substantive variability of method ranking and ratings within participants, between participants, and between documents. Retinex, for example, is the first choice for participant AW, but the last for OC (Table 4). The statistical analysis of agreement and spread quantitatively confirms the striking variability (Table 5 and Fig. 16). Various distributions are present: unimodal, bimodal, and uniform (Fig. 16a). Clearly, *no single method is adequate for every participant-document pairing*. The implication here is that an optimal text transcription can be obtained only by optimizing individual user/document/method configurations, and not by the method that performs best on average; in other words; the choice of enhancement method is best left to the user.

A second insight from Table 1 is provided by participant FG, who stated that all methods were equally useful. Other comments concur: "the images . . . definitely complemented each other" [FG]. Therefore, not only is no single method is the best, but *there is an explicit need for the concurrent use of multiple methods to optimize text transcription*.

The above precept is even more pertinent as image quality varies locally (e.g., AW ranked the method utility differently for the recto and verso of the same papyrus), and local information is critical to legibility. These findings further support the conclusions drawn from the overall method ranking. Rather than implementing a universal solution for legibility enhancement, *multiple methods should be offered to users*.

These results are unsurprising. The DStretch software, for example, offers more than fifty parameterizable methods. A recent comparison of the performance of eight optical, X-ray, and terahertz-based imaging approaches for recovering text within Egyptian mummy cartonnage concluded that it is only by "carefully selecting, optimizing and combining" them that success may be achieved [64, p. 1]. Similarly, the organizers of the Document Image Binarization competition explicitly state that "no binarization algorithm is good for all kinds of document images" [43, 113]. In industries dealing with color imaging and reproduction, the impossibility of a universal algorithm for color conversion (gamut mapping) is also acknowledged. Various strategies are therefore utilized to select appropriate methods, on the basis of (for example) the user's subjective intention (maintaining saturation or overall color appearance [114, pp. 3–5, 107–109, 194]), the linguistic dimension (preserving the names given to colors [127, pp. 218–219]), or the semantic content (e.g., skin color [127, p. 216]).

6.5.1 Comparison

P6 – 105

Our findings appear to differ from those obtained by two recent legibility enhancement evaluations of historical documents, which explicitly indicate the homogeneity of agreement between participants. Brenner and Sablatnig [28] produced an annotated dataset of 250 multispectral images from 20 paper and parchment manuscripts, with selected regions of interest graded for their legibility by 20 scholars; the settings were both laboratory and remote, and the interclass correlation coefficient (ICC) [96] was used for statistical analysis. Arsene *et al., Arsene-2018* had seven participants evaluate 15 pseudocolor multispectral fusion methods for their utility in enhancing the legibility of one parchment palimpsest page; a laboratory setting and ICC analysis were used.

There are evident similarities and differences in terms of intent and experimental setup between the above-mentioned works and the present research. We do however consider that the main reason for the behavioral *heterogeneity* observed in our study stems principally from three factors. First, all three evaluations contend with small sample sizes and unconstrained settings, which may amplify variance (while large datasets may hide diversity). Second, we adopted the following methodological "mantras": do not reject outliers, but question their significance (cf. the case of the original image); examine the raw data (Table 1) before aggregating it (Tables 2, 3, 4, and 5); do not assume normality for the data

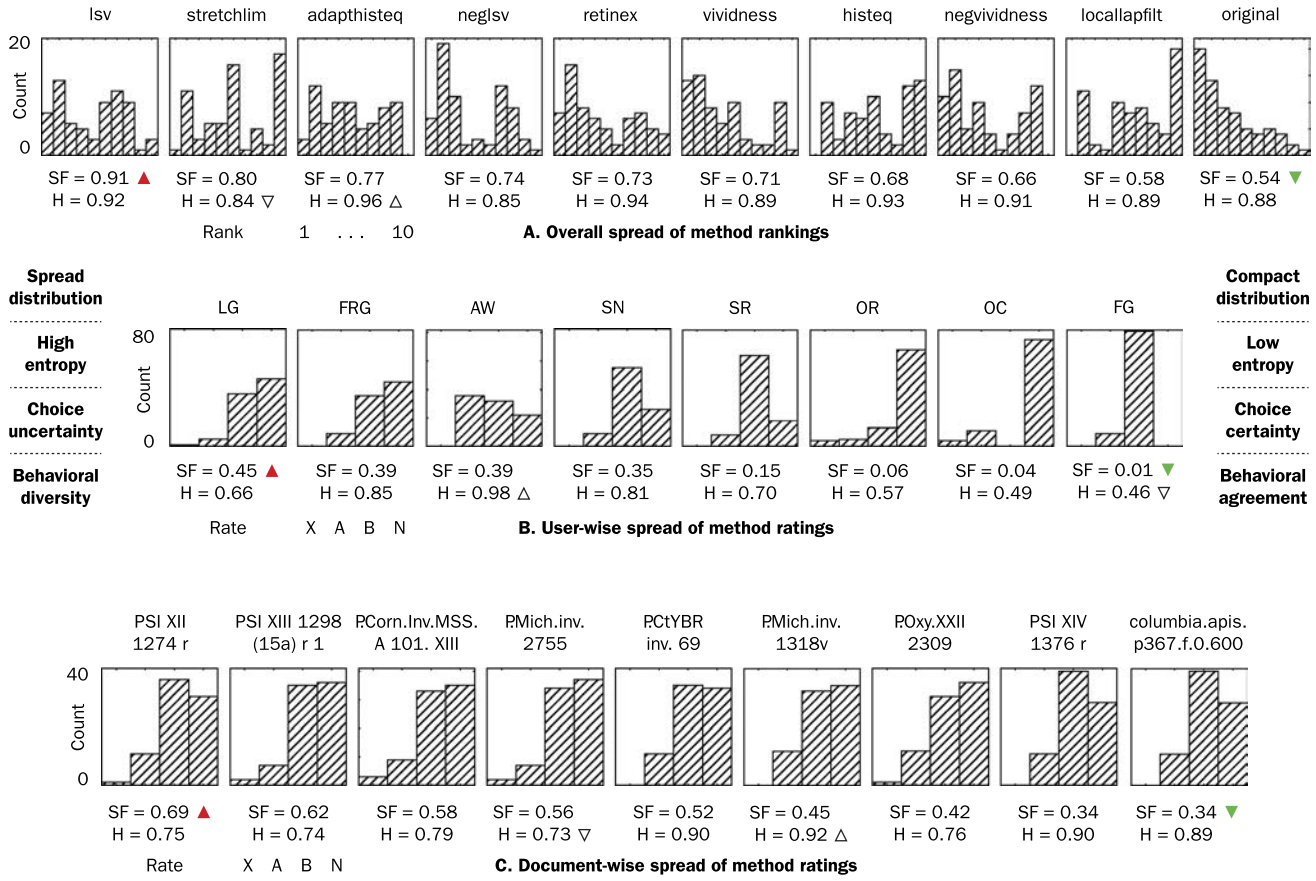


Fig. 16 Spread of enhancement method rankings and ratings, measured with spatial flatness *SF* and Shannon entropy *H*

distribution (while ICC does). Third, the reported behavioral differences may reflect the evaluators' differing outlooks on which experimental aspects to operationalize. The focus is on one hand on personalization and contextualization, and on the other hand on generalization. The choice depends on the situation and affordability: enhancement for critical vision benefits from local optimization, while global optimization is more efficient for casual vision.

6.5.2 Methodology

Kendall's coefficient of concordance W is non-parametric and distribution-independent statistic of agreement between multiple rankings, appropriate for the multimodal nature the analyzed data [89, pp. 94–116]. The results in Table 4 were obtained by adjusting for ties and by correcting for small sample sizes, followed by a χ^2 test of significance for large number of ties. The hypothesis H_0 that the between-user and between-document agreements are due to chance is rejected in favor of the alternative, with $p = 0.001$ and $p = 6.124\text{e-}9$, respectively, at an $\alpha = 0.1$ significance level. The coefficient of concordance, W, for m ranked items and n rankings is defined as

$$W = 12S/[m^2(n^3 - n) + 2 - m\sum(t^3 - t)], \tag{14}$$

where S is the sum of squares of rank deviations from the mean, and t is the length of runs of tied ranks. The formula for the χ^2 value is:

$$\chi^2 = 12S/[m(n^2 + n) + 2 - \sum(t^3 - t)/(n - 1)]. \tag{15}$$

Spatial flatness SF is the complement of the spectral flatness [31, 197], which is itself the ratio of the geometric and arithmetic mean of the magnitude of the Fourier transform $\mathcal{F}$ of a signal x of length n:

$$SF = 1 - n(\prod|\mathcal{F}(x)|)^{1/n}/\sum|\mathcal{F}(x)|. \tag{16}$$

Spatial flatness varies between 0 for an impulse and 1 for a uniform distribution. It is used in signal processing and is implemented as one of the descriptors of the MPEG-7 multimedia content description standard, with applications including classification into noise and tonal sounds [82, pp. 29–32]. We applied this method to characterize histograms. It is a powerful, non-parametric, descriptive statistic that, in a single scalar, conveys information on (1) the *coverage*

P6 – 106

of a range of possible values by actual values in a vector (e.g., to how many different ranks experimental participants assigned a method), (2) the *variance* of the distribution's amplitude (e.g., whether one rank is preferred over the others), and (3) the *homogeneity* of distribution (e.g., whether participants exhibit consistent or eclectic behavior). Such a rich description is made possible by the properties of the frequency domain, where regularities in signals are much easier to investigate than in the spatial domain. For the sake of comparison, Fig. 16 gives the Shannon entropy, H, of the ranking histograms. The "peakiest" histogram in Fig. 16a has the lowest SF value, but the lowest H value is that of a histogram wherein the values are fairly evenly distributed.

6.6 From facts to action: an operationalization pattern

The findings on rating variability derived from Table 1 were operationalized via the following design pattern: measurement ⊳ interpretation ⊳ implications ⊳ implementation.

The variability of ratings within columns represents the *within-participant* variability and is indicative of two behavioral attitudes: *eclecticism* in the use of enhancement methods and *adaptability* to document particularities. The prevalence of high eclecticism in the demographic may translate to a decision to accommodate this behavior by simplifying the access of users to the various enhancement methods, e.g., by tiling the interface windows. Conversely, low eclecticism may lead to methods organized in depth, via several menu levels. High adaptability may be considered an incentive to invest in the development of image quality measurement techniques, in order to predict the enhancement methods best suited to a given document, and to rank them dynamically in the software interface. The measurements of agreement in Table 4B show a wide distribution range between eclectic and conservative participants (FG has a uniform stimulus responses), and adaptable and rigid behaviors. The conclusion is that both *depth* and *flat interface designs* are needed, as well as that developing image quality models is desirable.

The principal factors in the level of *within-document rating agreement* in Table 5C are the document and image quality, script legibility, and user experience. This information could aid in evaluating the significance of enhancement method performance. In relating document rankings to their images in Fig. 2, no correlations with either factor are found. For example, moderate agreement characterizes both low and high document or image quality (PSI XIV 1376 r vs columbia.apis.p367.f.0.600), whereas low agreement is obtained for all script cursivity levels (PSI XII 1274 r vs P.Mich.inv.1318v). No conclusions are drawn from these data.

Comparison of *between-user* and *between-document rating agreement* may reveal the dominant factor driving the observed behavioral heterogeneity in the rating of the enhancement methods, owning either to user variability or to document quality variability. The results would have bearing on whether to invest more in *human-computer interaction* issues or *enhancement algorithms*. The values in Table 5A show a greater agreement in between-document ratings than between-user ratings, and thus support the focus on interaction facilities, more than on enhancement algorithms. This empirical finding is a surprising conclusion to an article devoted to algorithms, and thus merits further exploration.

6.7 Approaching legibility enhancement as a system

The preceding findings detailed the *high heterogeneity of user behavior* and the *large number of factors* affecting which legibility enhancement methods users choose to utilize. The following model of the legibility enhancement process aims to help better understand these phenomena.

The SYSTEM takes *bitmap* text as input and outputs *electronic text*. During this process the *entropy* of the text is lowered (i.e., a cryptic text has been read). The system is composed of *interacting elements* and their *properties*. The principal aspects are: (1) the TASK of critical reading, (2) the IMAGES with various degrees of content *affordance*, (3) the TOOL creating an ergonomic environment (such as image viewing software and displays), (4) the METHODS of legibility enhancement that are variously useful for the users, and (5) the USERS themselves, along their level of expertise in interacting with the system components. The BEHAVIOR of users emerges from the processes within the system, and is characterized by an idiosyncratic eclecticism in how they use the methods and the tools, as well as by contextual *adaptability* to the specificities of documents and their digital reproductions. LEGIBILITY is a property not only of images alone, but also of user *expertise* and tool *ergonomy*. Likewise, the *utility* of a legibility enhancement method is a property of the interaction among all elements of the system; therefore, its measurement is relative to the overall context.

P6 – 107

6.8 Contextual system optimization is preferable

Image interpretation in scholarly transcription, and even more so in medical, forensic, and military contexts, can have significant repercussions on people's lives (e.g., the role of the Rosetta Stone transcription in the Egyptian tourism industry). Optimal performance on a case-by-case basis, rather than considered statistically, is thus critical. This can be achieved only if the "variables" of the legibility system (i.e., methods and tools) are attuned to the "constants" (i.e., user, document, and reproduction). In short, choosing the *globally* optimal method for all combinations of the system states will yield suboptimal performance in most cases; only

when several analysts express themselves on the same document can their opinions be fused into a unique solution that might surpass individual solutions. To use an analogy, biometric security systems are optimized for individual users, whereas physical keys are universally functional for any user; evidently, the average of biometric signals from multiple users will result in very poor authentication performance.

The evidence in respect to legibility enhancement thus favors the *personalization* of methods and tools to specific users or demographics, the *contextualization* to the task, and the adaptation to individual documents and reproductions.

One simple implementation of a legibility enhancement system optimized to local conditions would be to offer a spectrum of methods and then *leave the choice to the users*. More sophisticated techniques are the ranking of methods according to their suitability for the document in question, and adaptation to the behavior of a given user. A fully-fledged context-aware critical visual enhancement system (e.g., autonomous robots) would have at its disposal a multitude of enhancement methods to choose from and combine in the function of the observed distortion types, scene content, environmental conditions, technological resource application domains, and user demographics and individualities.

6.8.1 Related work

P6 – 108

Digital photo editing is a research area where adaptation to individual and demographic preferences has been proposed as a way to develop image enhancement processes beyond the stage of universal filters [85]. Typical features of interest are the lightness transfer function (S-curve), temperature, and tint, while more complex edits may be modeled using deep learning [91].Taking user intent and image content into account are some of the challenges of this application type.

Demographics and individual adaptation have long been of interest for medical applications of analog and computational image enhancement. For example, macular degeneration is a common ophthalmologic pathology manifested by a blurred spot in the visual field. The detrimental impact of this condition, including the handicap it presents for reading, may be mitigated by bandpass filtering based on psychophysical models of low vision [137], a technique that could be implemented in an MPEG image decoder for real-time applications [92]. The progressive yellowing of the eye's crystalline lens with aging is another example of the demographic dimension of image enhancement. It leads to a darker field of vision, a condition for which it has been proposed to use lightness enhancement by means of non-linear tone mapping and high-pass filtering (sharpening) [126]. The several types and degrees of colorblindness (affecting about eight percent of males and one percent of females of European ancestry [120, 135, pp. 222–223]) have spurred the development of a further class of demographic-specific image enhancement techniques intended to simulate color vision deficiency for the normal-sighted [116, 186], as well as enhancing the color perception of the colorblind [53, 150, 196].

One challenge associated with medical-related image enhancement concerns the wide variety and individuality of pathologies that need to be accounted for in software [59]. Furthermore, the suboptimal ergonomics of the hardware in which this software embedded is often a reason why they fail to be used [136]. Consequently, a number of creative design solutions have been proposed, such as the integrated visual and audio representations of images [67]. An additional persistent problem remains the difficulty of evaluating the effectiveness of the experimental enhancement methods outside the laboratory, as part of the daily life of the visually impaired [138]. This aspect has considerable significance, given the far greater disparity for pathological applications than for non-pathological ones between laboratory and real-life environments (in a similar manner as we suggest to be the case for paleographic transcription).

6.9 The primacy of the "original" image

Overall, papyrologists ranked original images as the most useful for transcription. Participant SN is unambiguous: "I'm convinced that the original picture is the most important reference point in discussing a text." Why should this be so, and what might it mean for software development?

In many domains—papyrological, medical, or journalistic—the more removed the information from its source, the less trustworthy it is considered: (cf. witness vs hearsay evidence, primary vs secondary documentary sources). The dogma of the "original" is a socio-professional specificity inculcated early on in papyrological education. As sensible this is in the context of physical documents examined visually, digital reproductions are, however, inherently "manipulated" images, and an image at the exit point of an imaging system is no more faithful than, say, an image calibrated during post-production. Hence, suitable documentation should clarify alterations by the enhancement methods, and thus preserve the confidence of users in that they understand the source of features observed in images.

The "low toner effect" might also have contributed. Psychologists found that, counterintuitively, a job application letter of poor print quality was more trustworthy than a pristine print. The proposed explanation was that low legibility demands more cognitive effort, which is interpreted as a reflection of the high value of the letter content [132, pp. 149–151]. The effect was reproduced with mathematical exercises more often correctly solved when in an uncommon faded typeface [10].

It would however be misleading to rely solely on the overall ranking of enhancement methods (Table 3) and thereby

conclude that the original is the top-performing image under all conditions. For example, between-user variability is substantial: whereas OR consistently rates the original as being of "exclusive use" or "primary use", OC rejects it as "not useful" (Table 2). The software should thus be able to adapt to user preferences, as well as document specificities.

6.10 The unexpected usefulness of negative images

The trivial negative method is noteworthy not only for its excellent performance as a top-2 enhancement method after vividness (Table 2), but also for its *discriminatory qualities*. It helps to ascertain whether a specific pixel cluster might be ink ("I used the negative image to understand whether I was imagining an ink trace or it was real." [OC]), and to distinguish between ink and holes in the papyrus, a recurrent issue in reading papyri reproductions. One further point of practical relevance regarding this method is that it is applicable to *monochrome images*. Some papyri survive or are remotely accessible only in this format; other (such as the charred papyri) have little chromatic information.

The finding that negative polarity is considered by users to improve legibility is of interest because it is contrary to the prevailing *psychophysical evidence*, which consistently associates better legibility with positive contrast (i.e., dark script on light background) [26, pp. 206–207; 32; 99; 109 145]. The possible cause of the disparity is that these studies typically use binary black and white stimuli, rather than ternary black, gray, and white stimuli, as in the present experiment.

Negative polarity is known to be preferable only for special conditions, such as the degeneration of visual acuity in aging [26, pp. 244–243] or disease [109]. Also, aeronautical displays with negative polarity help to maintain the adaptation of the visual system to a dark environment during night flight [148]. Closer to papyrology is the similar case of readers repeatedly switching between the high luminance of a computer screen and the low luminance of paper documents [20]. These exceptions have a context of *low vision* in common, whether permanent or transitory, that demands increased visual and cognitive efforts. In other words, they are akin to critical reading tasks, such as the decipherment of ancient documents, characterized by high visual and semantic entropy.

In terms of *visual ecology*, positive contrast has dominated the history of writing, albeit mostly for practical reasons such as the cost of colored substrate. Negative polarity texts are often prestige objects, for example, the Rosetta inscription that features a light script with crystalline sparkle on a dark stone [122]; the word "God" calligraphed in gold on a green eight-meter-high panel in the Aya Sofia mosque of Istanbul [25, p. 504]; and the 12th-century "Blue Sutra" of Japan, in gold on blue paper [58, pp. 36, 68–69]. These examples are also sources of inspiration for how contrast—and thus legibility—might be further improved, using displays with special material properties, such as high-reflectance or fluorescence.

6.11 Smooth human–computer interaction is critical

Participant AW reported: "Even with the split-screen option, and using two monitors, I felt like my devices lacked the optimal Graphical User Interface to maximize the usefulness of those pictures. Comparing the same zoomed part of text through several images proved, for example, to be quite costly in terms of time and clicks, and I feel this disturbed the workflow." Five of eight participants juxtaposed two or more image variants during the transcription work; three used two displays, while one used three displays. Two participants printed the images (as more hard copies can be spread out on a table than on a computer display).

These explicit and implicit statements demonstrate how interaction with the images can impact the advantages derived from image processing. Indeed, without good interaction, there is little enhancement as far as the user is concerned. What, then, is *good interaction*? Observing the participants is informative in this regard: good interaction involves a wide visual field, the synchronization of manipulations across multiple images, the understanding of how different enhancement methods have varying utility for transcription, and a state of mental flow during task completion [42].

6.12 Gender may affect legibility enhancement

One female participant in the experiment had perfect agreement in her method ratings, while the second female had the second highest agreement score (Table 5B). If they are not statistical outliers, then the gender disparity in behavior could have implications, given the increased proportion of females in papyrology (202 [57%] out of 356 participants at the International Congress of Papyrology in 2019 were female, compared to 3 [5%] out of 57 at the congress in 1937 [6, 37]). For example, females have been found to have greater chromatic discrimination ability, whereas males are more sensitive to lightness variation [190]. Therefore, the enhancement method through chromatic contrasting might improve legibility more for females, while the negative polarity method could be more appropriate for males. A larger sample size would be needed to further the investigation of the topic.

However, the significance of this section is broader. In addition to gender, there are numerous other perceptual variability factors (age-related, genetic, cultural, and climatic, among others) [49, pp. 149–244] that indicate the utility

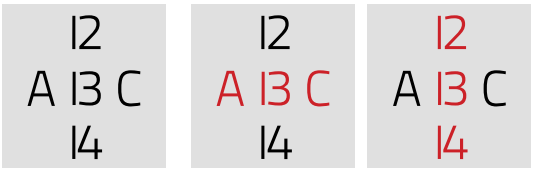


Fig. 17 The shapes in the center of the squares are perfectly legible, but ambiguous: do they represent the letter "B", the figure "13", or something else? Coloring helps differentiate between alternatives. Such shapes are said to be "multistable stimuli" [101], and have been the source of well known illusions, e. g. Arcimboldo's face made of books [164, p. 131]

of demographically tailored and even personalized image enhancements (in a similar fashion to how corrective glasses are fitted individually) rather than those calibrated to the statistical mean of the general population.

7 Paradigm

A paradigm of legibility enhancement for critical applications is provided herein, representing a step towards developing computer-assisted critical vision systems. This paradigm helps defining the goals, methods, and limits of legibility enhancement, and is rooted in the problem analysis and experiment carried out in the present research. The preceding sections have circumscribed the notion of critical vision, established that potential information should not be suppressed, and advocated a systems approach in the design of enhancement software. A further element will be discussed below: uncertainty. This element will help operationalize the paradigm and link the field of image enhancement to the vast interdisciplinary research on uncertainty.

P6
–
110

7.1 Requirements

Because vision is purposeful, we might sometimes stare at things without seeing them when there is no reason to seek them out ("inattentional blindness" [172]). Two converse problems concern our tendency to believe what we see and see what we believe ("confirmation bias" [169, pp. 61–75]). Generally speaking, the ideal enhancement system would inform us about objective uncertainty in images and guard against subjective cognitive biases during their interpretation. Uncertainty is important and its presence should be made explicit. The system should therefore aid in its detection and analysis, represent it visually and numerically, and preserve it for verification. *Uncertainty management should be at the core of the enhancement system and its components, from the algorithms to the interface.*

7.2 Examples

One illustrative analogy is the use of white material for pottery reconstruction in archeology to distinguish between missing and original parts. An example of script polysemy and the role of context in reading is shown in Fig. 17, wherein the central shapes can be interpreted as either letters or numbers depending on the reading direction. Mundane cases where an individual seeks to mentally solve visual ambiguities include faded maps or letters and frightening shadows at night, while imagery-based clinical diagnosis and military decisions represent domains in which uncertainty is both pervasive and critical. For the facial identification of suspects, police organizations have developed mechanical and computational aids for using witness recollections to produce faces composited from a catalog of elements, such as eyes, nose, and mouth, assembled by likelihood into a coherent whole [44].

7.3 Rationale

A paradigm called *conservative preprocessing* was proposed by Chen, Lopresti, and Nagy in a rare work on the theory of image enhancement [38]. It consists in preserving the original by means of reversible transformations and the use of derivative representations as a proxy. Other researchers have noted that applying the concept of noise suppression outside the casual reading task might be misplaced: for example, bleed-through removal may unwittingly obliterate the only trace of text that is inaccessible because the page is glued to the support, was not scanned, or belongs to a missing second or third page [180]. Such invasive enhancement can thus destroy historical, forensic, and conservation-related information in documents and reproductions [60, 129]. Problems of this kind are common in digital library projects. Technical factors can also make enhancement problematic; for example, because band-pass filtering is known to frequently suppress script parts [133, p. 445], while neural networks can "hallucinate" character-like shapes [21, p. 1221], should the user not be made aware of the possibility of erroneous readings?

7.4 Conceptualization

Let us here define uncertainty in the context of image analysis and consider possible solutions. As discussed above, image enhancement is the outcome of interacting stimuli, observers, tasks, contexts, and technologies. Multiple concurrent enhancements therefore appear to be better adapted to the high combinatorial variability potential of such a SYSTEM.

To clarify the uncertainty reduction PROCESS, we draw on cognitive psychology (in particular, decision-making in critical contexts, such as those encountered by medical emergency services, firefighters, police, and military personnel [169]). The process consists of an *interpretation* stage, in which possible stimuli of interest are detected and separated from the surrounds (building "situational awareness"), various options as to their identification are formulated (integration into coherent "gestalts"), and their likelihood is estimated (multidimensional weighting and/or ranking); this is followed by a *decision* stage, in which a choice of interpretations is made. For our proposed method, it is noteworthy that (1) *alternatives* to the visual stimulus are produced, (2) the visualization is *mental*, rather than physical, and (3) it involves mental *manipulations* of the alternative images.

Establishing a taxonomy of common FACTORS leading to uncertainty helps us to find appropriate remedies. The inability to detect visual targets (*lack* of information) and vaguely defined targets (*fuzzy* information) are types of uncertainty that arise early in the interpretation process and that may be countered by more and sharper imagery; that is, alternative views. Decisions regarding *implausible* interpretations may be solved by considering the target's compatibility with its visual context, which shows the utility of retaining context and even using alternative backgrounds. The uncertainty developing from the need to choose between *contradictory* interpretations demands an increased capability to compare alternatives. Entities need not necessarily be noisy to make interpretation difficult (viz. readily distinguishable characters in unknown scripts); to deal with *novelties*, information exogenous to the image and observer must be supplied. Finally, an *overabundance* of potential interpretations may become overwhelming and result in "decision inertia" [169, pp. 37–41]. Information visualization theory suggests for such cases filtering the amount of information, through selection or ranking [168].

In terms of LIMITATIONS, he number of items processable at a given time in human memory is however finite and can degrade under time pressure and stress; similarly, the visual field size is also limited and increasingly blurred as it moves away from the fovea. Thus, methods designed to *increase* the number of alternatives must be accompanied by mechanisms to *limit* visible alternatives and *manage* all available stock.

7.5 Operationalization

We propose to make uncertainty materially explicit through the physical rendering and cueing of mental interpretations of images (as a literal application of the notion that "seeing is believing"). When the alternatives are known, highlighting or inpainting of objects may be used to find fits between potential targets and a given context, as in a puzzle game where the pieces are materializations of mental images. For unknown alternatives, solutions may be elicited by purposive or random alterations at the global or local level, statically or as a movie, similar to the shimmering reflection on an undulating water surface, or searching for optimal shapes by sketching, or tilting a hologram to observe it from different perspectives (our deliberate use of metaphors is a method for problem solving, as well as for communication [75, 106]). Machine learning may help shape prediction and synthesis, while human-machine interaction techniques are used to guide the image exploration. The result is much more than an image enhancement algorithm: it is a *computer-assisted critical vision system*. Current research on such systems includes applications for, e.g., medical radiographic diagnosis [158, pp. 359–414].

This method operates as illustrated using Fig. 17, in which two images were derived from the original; each of these make explicit a different interpretation through the perceptual differentiation of distinct semantic classes. If semantic grouping became unfeasible, the random alternation of individual shapes between two colors would be the next best option (also revealing a less obvious solution when only the three vertical bars are red, in that they may represent the upper-case letter "I", the lower-case letter "l", or the figure "1"). The inpainting technique presented in this example consists in removing the gaps between adjacent central shapes to strengthen their identity as the letter "B". In the real case of papyri, the most common character allographs of interest may be statically inpainted or flickered. This is a complex and possibly contentious intervention in the image structure; thus, in our implementation, we avoided segmentation by opting to present variants of the whole scene.

P6 – 111

7.6 Frameworks

The presented method for making uncertainty explicit in view of helping decision-making may be understood in information theoretical terms as finding a balance between maximal (= quandary) and minimal choice (= no or unknown alternative) by comparing image alternatives with different entropies for different entities within the images. The heterogeneity of user behavior observed in our experiment suggests that the optimum is contextually determined by a combination of factors.

Bayesian theory offers an additional useful theoretical framework, in that establishing the identity of the visual target is dependent on both the prior assumptions of the observed shape having the hypothesized identity and the likelihood of observing the hypothetical entity in the given context. In this sense, the explicit presentation of multiple

alternatives for comparison is a means of updating and optimizing one's beliefs about the identity of visual stimuli.

7.7 Criteria

Our work yielded a set of core criteria defining our paradigm of legibility enhancement for critical applications:

1. *Critical vision*: Legibility enhancement for critical applications is a matter of critical vision. It consists in exerting skepticism about the interpretation of images.
2. *Potential information*: Any image structure should be viewed as potential information that should not be suppressed in critical vision applications; interpretations arise from connections made between image parts.
3. *Uncertainty explicitation*: Support for critical vision is realized by making alternative image interpretations materially explicit. Visual target identity depends on both the shape and its context.
4. *Systems approach*: Legibility enhancement for critical applications is a system problem. Optimization depends on tasks, data, users, tools, and their interactions.

8 Implementation

P6 – 112

The research findings were implemented in the open-source software Hierax [16]. To account for the heterogeneity of user behavior and the complementarity of enhancements, a range of methods were offered: our novel methods, and adapthisteq and retinex. Since interaction with the images was found to affect the usefulness of enhancements, the software includes an image viewer. Our design methodology conforms to elements of the so-called information visualization "mantra" for graphical user interfaces: overview, filter, detail, and context [168]. One interface tab presents an overview of one image enhanced using the selected methods (Fig. 18); through the spatial juxtaposition of patterns, the resulting mosaic facilitates understanding of the characteristics of different enhancement methods. A second tab allows for detailed side-by-side comparison, with zooming, panning, and rotation being synchronized; the diptych view is intended for the interpretation of details in close coupling of vision and motor control. Rapid switching between images via the keyboard is a temporal image analysis modality that leverages attentional capture through flicker, introduced to help detect potential features of interest [193]. The mosaic, diptych, and flicker image analysis techniques implement the paradigm of uncertainty explicitation at the whole scene level. These facilities were appreciated by the paleographers who beta-tested the software. Hierax has already been employed in a real-world setting, to enhance the 70 papyri of the University of Basel Library for a forthcoming online catalog.

9 Conclusions

This work aimed to improve the legibility of ancient papyri for text editions. Novel enhancement methods were developed on the basis of color processing and visual illusions. A user experiment demonstrated that these methods outperform classical and more complex enhancement methods. Future work could examine whether the new methods could also improve reading by machines.

The experiment also yielded unexpected findings. Critical for the software, user behavior was found to be *heterogeneous* in type, and not only variable in intensity. Thus, better performance might be achieved by *personalization*, *contextualization*, and *complementarity*, instead of searching for an overall optimum that may be spurious or counterproductive. Experiments with more users and images should be conducted to strengthen, nuance, and perhaps diversify the findings.

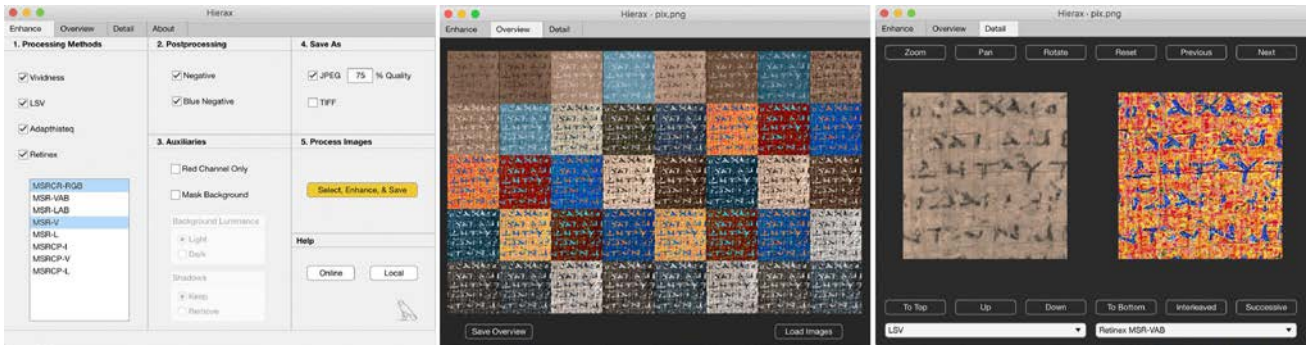


Fig. 18 The graphical user interface of the Hierax software for papyri legibility enhancement, which implements the methods described in this article. Method selection and parametrization (left); overview of all processed images (center); image comparison and interaction panel (right)

The findings also lead to a paradigm of legibility enhancement for critical applications, in view of a future computer-aided critical vision system.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10032-021-00386-0.

Acknowledgements This article is an output of the project n° PZ00P1_174149, *Reuniting fragments, identifying scribes and characterizing scripts: the digital paleography of Greek and Coptic papyri*, funded by the Swiss National Science Foundation. The experimental participants are heartily thanked for their valuable efforts. Gorge Nagy, Andreas Fischer, Irvin Schick, and Bernice Rogowitz, as well as the three anonymous reviewers, offered much appreciated critique. The authors wish to acknowledge the proofreaders' help in improving the manuscript readability: Jacqueline M. Huwyler, Thomas Hardy of Science Journal Editors, and K. F. of Proofed. — Dedicated by V. A. to Aries Arditi and Ruth Rosenholtz, for the vision from the lighthouse.

Author Contributions V. A. conceived the algorithms, programmed the software, designed the experiment, analyzed the results, and wrote the article. I. M.-S., project leader, defined, planned, and supervised the research (data collection, guidance, tests, experiment coordination, critical revision of the manuscript).

Funding Information Open Access funding provided by Université de Fribourg.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Abidi, B.R., Zheng, Y., Gribok, A.V., Abidi, M.A.: Improving weapon detection in single energy X-ray images through pseudocoloring. IEEE Trans. Syst. Man Cybern. C (Appl. Rev.) **36**(6), 784–796 (2006)
2. Adelson, E.H.: Lightness perception and lightness illusions. In: Gazzaniga, M.S. (ed.) The New Cognitive Neurosciences, 2nd edn., pp. 339–351. MIT Press, Cambridge, MA (2000)
3. Afruz, J., Wilson, V., Umbaugh, S.E.: Frequency domain pseudocolor to enhance ultrasound images. Comput. Inform. Sci. **3**(4), 24–34 (2010)
4. Alaei, A., Raveaux, R., Conte, D., Stantic, B.: "Quality" vs. "Readability" in document images: statistical analysis of human perception. In: Proceedings of the 13th IAPR International Workshop on Document Analysis Systems (DAS), 24–27 April 2018, Vienna, Austria, pp. 363–368 (2018)
5. Alexopoulou, A.A., Kaminari, A.A., Panagopoulos, A., Pöhlmann, E.: Multispectral documentation and image processing analysis of the papyrus of tomb II at Daphne. Greece. J. Archaeol. Sci. **40**(2), 1242–1249 (2013)
6. Allberry, C.R.C., et al.: Actes du Ve Congrès International de papyrologie, Oxford, 30 Août – 3 Septembre 1937. Fondation égyptologique Reine Élisabeth, Bruxelles (1938)
7. Allegra, D., Ciliberto, E., Ciliberto, P., Milotta, F.L.M., Petrillo, G., Stanco, F., Trombatore, C.: Virtual unrolling using X-ray computed tomography. In: Proceedings of the 23rd European Signal Processing Conference (EUSIPCO), 31 August–4 September 2015, Nice, France, pp. 2864–2868 (2015)
8. Alley, R.E.: Algorithm Theoretical Basis Document for Decorrelation Stretch, ASTER-AST06. Jet Propulsion Laboratory, Pasadena, CA (1996). http://www.dstretch.com/DecorrelationStretch.pdf
9. Altamura, O., Berardi, M., Ceci, M., Malerba, D., Varlaro, A.: Using colour information to understand censorship cards of film archives. Intl. J. Doc. Anal. Recogn. **9**(1), 281–297 (2007)
10. Alter, A.L., Oppenheimer, D.M., Epley, N., Eyre, R.N.: Overcoming intuition: metacognitive difficulty activates analytic reasoning. J. Exp. Psychol. **136**(4), 569–576 (2007)
11. Alvo, M., Yu, P.H.L.H.: Statistical Methods for Ranking Data. Springer, New York, NY (2014)
12. Arsene, C.T.C., Church, S., Dickinson, M.: High performance software in multidimensional reduction methods for image processing with application to ancient manuscripts. Manuscipt Cultures **11**, 73–96 (2018)
13. Asano, Y., Fairchild, M.D., Blondé, L.: Individual colorimetric observer model. PLoS ONE **11**(2), e0145671 (2016)
14. Atanasiu, V.: Ugraphia: The Utopia of a Perfectly Legible Script. /gʁafematik/ Collection of extended abstracts of the Conference Graphemics in the 21st century: From graphemes to knowledge, 14–15 June 2018, Brest, France, pp. 113–140 (2018)
15. Atanasiu, V.: Non-parametric rating and ranking functions. MATLAB Central File Exchange (2020). https://www.mathworks.com/matlabcentral/fileexchange/78384
16. Atanasiu, V., Marthot-Santaniello, I.: Hierax legibility enhancement software (2020). https://hierax.ch
17. Bagnall, R.S.: The Oxford Handbook of Papyrology. Oxford Univ. Press, Oxford (2009)
18. Balinski, M., Rida, L.: Majority Judgment Measuring, Ranking, and Electing. MIT Press, Cambridge, MA (2010)
19. Barten, P.G.J.: Contrast Sensitivity of the Human Eye and Its Effects on Image Quality. SPIE Optical Engineering Press, Bellingham, WA (1999)
20. Bauer, D., Cavonius, C.: Improving the legibility of visual display units through contrast reversal. In: Grandjean, E., Vigliani, E. (eds.) Ergonomic aspects of visual display terminals, pp. 137–142. Taylor & Francis, London (1980)
21. Belthangady, C., Royer, L.A.: Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. Nat. Methods **16**, 1215–1225 (2019)
22. Berns, R.S.: Extending CIELAB: Vividness, V^*_{ab}, Depth, D^*_{ab}, and Clarity, T^*_{ab}. Color Res. Appl. **39**(4), 322–330 (2014)
23. Beutel, J., Van Metter, R.L., Kundel, H.L. (eds.): Handbook of Medical Imaging, vol. 1. SPIE, Bellingham, WA (2000)
24. Bhowmik, S., Sarkar, R., Nasipuri, M., Doermann, D.: Text and non-text separation in offline document images: a survey. Intl. J. Doc. Anal. Recogn. **21**(1–2), 1–20 (2018)
25. Blair, S.: Islamic Calligraphy. Edinburgh Univ. Press, Edinburgh (2006)
26. Boff, K.R., Lincoln, J.E. (eds.): Engineering Data Compendium: Human Perception and Performance, vol. 1. Harry Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, OH (1988)

P6 – 113

27. Bouwmans, T., Javed, S., Zhang, H., Lin, Z., Otazo, R.: On the applications of robust PCA in image and video processing. Proc. IEEE **106**(8), 1427–1457 (2018)
28. Brenner, S., Sablatnig, R.: Subjective assessments of legibility in ancient manuscript images: the SALAMI dataset. In: Proceedings of the ICPR International Workshops and Challenges, 10–15 January 2021, Virtual Event, part VII, pp. 68–82 (2021)
29. Breuil, C., Jennings, B.J., Barthelmé, S., Guyader, N., Kingdom, F.A.A.: Color improves edge classification in human vision. PLoS Comput. Biol. **15**(10), e1007398 (2019)
30. Brink, A., van der Klauw, H., Schomaker, L.: Automatic removal of crossed-out handwritten text and the effect on writer verification and identification. In: Proceedings of the SPIE–IS&T Electronic Imaging Conference, Document Recognition and Retrieval XV, 29–31 January 2008, San Jose, CA, USA, vol. 6815: 68150A (2008)
31. Broersen, P.M.T.: Automatic Autocorrelation and Spectral Analysis. Springer, London (2006)
32. Buchner, A., Mayr, S., Brandt, M.: The advantage of positive text-background polarity is due to high display luminance. Ergonomics **52**(7), 882–886 (2009)
33. Buckle, C.E., Udawatta, V., Straus, C.M.: Now you see it, now you don't: visual illusions in radiology. Radiographics **33**(7), 2087–2102 (2013)
34. Bülow-Jacobsen, A.: Writing materials in the ancient world. In: Bagnall, R.S. (ed.) The Oxford Handbook of Papyrology, pp. 3–29. Oxford Univ. Press, Oxford (2009)
35. Bülow-Jacobsen, A.: Photography of Papyri and Ostraca. In: Caputo, C., Lougovaya, J. (eds.) Using Ostraca in the Ancient World, pp. 59–83. De Gruyter, New Discoveries and Methodologies, Berlin (2021)
36. Cannon, M., Hochberg, J., Kelly, P.: Quality assessment and restoration of typewritten document images. Int. J. Doc. Anal. Recogn. **2**(2–3), 80–88 (1999)
37. Capasso, M., Davoli, P.: Abstracts of the 29th International Congress of Papyrology, 28 July–8 August 2019, Lecce, Italy. Università del Salento, Lecce (2019)
38. Chen, J., Lopresti, D., Nagy, G.: Conservative preprocessing of document images. Intl. J. Doc. Anal. Recogn. **19**(4), 321–333 (2016)
39. Christen, M., Vitacco, D.A., Huber, L., Harboe, J., Fabrikant, Brugger, P.: Colorful brains: 14 years of display practice in functional neuroimaging. Neuroimage **73**(6), 30–39 (2013)
40. CIE: Colorimetry, TR 15:2004. CIE, Vienna (2004)
41. Ciortan, I., Deborah, H., George, S., Hardeberg, J.Y.: Color and hyperspectral image segmentation for historical documents. In: Proceedings of the 2015 Digital Heritage Conference, 28 September–2 October 2015, Granada, Spain, pp. 199–206 (2015)
42. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. HarperPerennial, New York, NY (2008)
43. DIB: Document Image Binarization (2021). https://dib.cin.ufpe.br
44. Davies, G., Valentine, T.: Facial Composites: Forensic Utility and Psychological Research. In: Lindsay, R.C.L., Ross, D.F., Read, J.D., Toglia, M.P. (eds.) Handbook of Eyewitness Psychology: Volume 2, Memory for People, pp. 59–82. Psychology Press, New York, NY (2014)
45. Donofrio, R.L.: Review paper: the Helmholtz–Kohlrausch effect. J. Soc. Inform. Disp. **19**(10), 658–664 (2011)
46. Drira, F., LeBourgeois, F.: Mean-Shift segmentation and PDE-based nonlinear diffusion: toward a common variational framework for foreground/background document image segmentation. Int. J. Doc. Anal. Recogn. **20**(3), 201–222 (2017)
47. Easton, R.L., Knox, K.T., Christens-Barry, W.A.: Multispectral imaging of the Archimedes palimpsest. In: Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop (IEEE-AIPR'03), 15–17 October 2003, Washington, DC, USA, pp. 111–116 (2003)
48. Elhedda, W., Mehri, M., Mahjoub, M.A.: Hyperkernel-based intuitionistic fuzzy c-means for denoising color archival document images. Intl. J. Doc. Anal. Recogn. **23**(3), 161–181 (2020)
49. Elliot, A.J., Fairchild, M.D., Franklin, A. (eds.): Handbook of Color Psychology. Cambridge Univ. Press, Cambridge (2015)
50. Fairchild, M.D.: Refinement of the RLAB color space. Color Color Res. Appl. **21**(5), 338–346 (1996)
51. Fairchild, M.D.: Color Appearance Models, 3rd edn. Wiley, Chichester (2013)
52. Fairchild, M.D., Pirrotta, E.: Predicting the lightness of chromatic object colors using CIELAB. Color. Res. Appl. **16**(6), 385–393 (1991)
53. Farup, I.: Individualised Halo-Free Gradient-Domain Colour Image Daltonisation. J. Imag. **6**(116), 1–10 (2020)
54. Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., James, S.: Machine learning for cultural heritage: a survey. Pattern Recogn. Lett. **133**(5), 102–108 (2020)
55. Fitts, P.M., (ed.), (1947). Psychological Research on Equipment Design, Report no. 19. Army Air Forces, Washington D.C
56. de la Flor, G., Luff, P., Jirotka, M., Pybus, J., Kirkham, R., Carusi, A.: The case of the disappearing ox: seeing through digital images to an analysis of ancient texts. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2010), 10–15 April 2010, Atlanta, GA, USA, pp. 473–482 (2010)
57. Fréchette, G.R., Schotter, A.: Handbook of Experimental Economic Methodology. Oxford University Press, Oxford (2015)
58. Fu, S., Lowry, G.D., Yonemura, A.: From Concept to Context: Approaches to Asian and Islamic Calligraphy. Smithsonian Institution, Washington, DC (1986)
59. Gao, X.W., Loomes, M.: A new approach to image enhancement for the visually impaired. In: Proceedings of the IS&T International Symposium on Electronic Imaging 2016, Color Imaging XXI: Displaying, Processing, Hardcopy, and Applications, 14–18 February 2016, San Francisco, CA, USA, COLOR-325: 1–7 (2016)
60. Gargano, M., Bertani, D., Greco, M., Cupitt, J., Gadia, D., Rizzi, A.: A perceptual approach to the fusion of visible and NIR images in the examination of ancient documents. J. Cult. Heritage **16**(4), 518–525 (2015)
61. George, S., Grecicosei, A.M., Waaler, E., Hardeberg, J.Y.: Spectral image analysis and visualisation of the khirbet qeiyafa ostracon. In: Proceedings of the 6th International Conference on Image and Signal Processing (ICISP 2014), 30 June–2 July 2014, Cherbourg, France, 272–279 (2014)
62. Gevers, T., Gijsenij, A., van de Weijer, J., Geusebroek, J.-M.: Color in Computer Vision: Fundamentals and Applications. Wiley, Chichester (2012)
63. Ghadiyaram, D., Bovik, A.C.: Massive online crowdsourced study of subjective and objective picture quality. IEEE Trans. Image Process. **25**(1), 372–387 (2016)
64. Gibson, A., Piquette, K.E., Bergmann, U., Christens-Barry, W., Davis, G., Endrizzi, M., Fan, S., et al.: An assessment of multimodal imaging of subsurface text in mummy cartonnage using surrogate papyrus phantoms. Herit. Sci. **6**, 7 (2018)
65. Gilchrist, A.: Seeing Black and White. Oxford Univ. Press, Oxford (2006)
66. Gillespie, A.R., Kahle, A.B., Walker, R.E.: Color enhancement of highly correlated images: I. Decorrelation and HSI contrast stretches. Remote Sens. Environ. **20**, 209–235 (1986)
67. Gonzalez, A., Benavente, R., Penacchio, O., Vazquez-Corral, J., Vanrell, M., Parraga, C.A.: Coloresia: An Interactive Colour

P6 – 114

Perception Device for the Visually Impaired. In: Sappa, A.D., Vitrià, J. (eds.) Multimodal Interaction in Image and Video Applications, pp. 47–66. Springer, Berlin (2013)
68. Govindaraju, V., Srihari, S.N.: Image quality and readability. In: Proceedings of the International Conference on Image Processing (ICIP), 23–26 October 1995, Washington, DC, USA, vol. 3, pp. 324–327 (1995)
69. Hansen, T., Gegenfurtner, K.R.: Independence of color and luminance edges in natural scenes. Vis. Neurosci. **26**, 35–49 (2009)
70. Hariharan, H., Koschan, A., Abidi, B., Gribok, A., Abidi, M.: Fusion of visible and infrared images using empirical mode decomposition to improve face recognition. In: Proceedings of the International Conference on Image Processing, 8–11 October 2006, Atlanta, GA, USA, pp. 2049–2052 (2006)
71. Harman, J.: Using decorrelation stretch to enhance rock art images. DStretch: Rock Art Digital Enhancement (2008). http://www.dstretch.com/AlgorithmDescription.html
72. He, S., Schomaker, L.: DeepOtsu: Document enhancement and binarization using iterative deep learning. Pattern Recogn. **91**(7), 379–390 (2019)
73. Hernández, F.-J.: Retinex for ImageJ (2010). https://web.archive.org/ web/ 20130 32305 3256/; http://www.dentistry.bham.ac.uk/landinig/software/retinex/retinex.html
74. Hidalgo, H.L., España, S., Castro, M.H., Pérez, H.A.: Enhancement and cleaning of handwritten data by using neural networks. In: Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA), 7–9 June 2005, Estoril, Portugal, part I, pp. 376–383 (2005)
75. Hofstadter, D.: Fluid Analogies Research Group: Fluid Concepts and Creative Analogies. BasicBooks, New York, NY (1995)
76. Holleman, G.A., Hooge, I.T.C., Kemner, C., Hessels, R.S.: The real-world approach and its problems: a critique of the term ecological validity. Front. Psychol. **11**, 721 (2020)
77. Homer: The Iliad, trans. Robert Fagles. New York, NY: Penguin (1990)
78. Hoßfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K., Tran-Gia, P.: Best practices for QoE crowdtesting: QoE assessment with crowdsourcing. IEEE Trans. Multimedia **16**(2), 541–558 (2014)
79. Hummel, R.: Image enhancement by histogram transformation. Comput. Graphics Image Process. **6**, 184–195 (1977)
80. James, A.P., Dasarathy, B.V.: Medical image fusion: a survey of the state of the art. Inform. Fusion **19**, 4–19 (2014)
81. Jobson, D.J., Rahman, Z., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Trans. Image Process. **6**(7), 965–976 (1997)
82. Hyoung-Gook, K., Moreau, N., Sikora, T.: MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval. Wiley, Chichester (2005)
83. Kahneman, D.: Thinking, Fast and Slow. Penguin, London (2011)
84. Kambara, A.: Effects of experiencing visual illusions and susceptibility to biases in one's social judgments. Percept. Mot. Skills **7**(4), 1–6 (2017)
85. Kapoor, A., Caicedo, J.C., Lischinski, D., Kang, S.B.: Collaborative personalization of image enhancement. Int. J. Comput. Vis. **108**(1–2), 148–164 (2014)
86. Karatzas, D., Antonacopoulos, A.: Colour text segmentation in web images based on human perception. Image Vis. Comput. **25**, 564–577 (2007)
87. Keelan, B.W.: Handbook of Image Quality: Characterization and Prediction. Marcel Dekker, New York, NY (2002)
88. Kemeny, J.G., Snell, J.L.: Mathematical Models in the Social Sciences. MIT Press, Cambridge, MA (1972)
89. Kendall, M.G.: Rank Correlation Methods, 4th edn. McGraw-Hill, New York, NY (1970)
90. Kihlstrom, J.F.: Ecological validity and "ecological validity." Perspect. Psychol. Sci. **16**(2), 466–471 (2021)
91. Kim, H.U., Koh, Y.J., Kim, C.S.: PieNet: Personalized Image Enhancement Network. In: Proceedings of the Computer Vision (ECCV 2020), 23–28 August 2020, Glasgow, UK, pp. 374–390 (2020)
92. Kim, J., Vora, A., Peli, E.: MPEG-based image enhancement for the visually impaired. Opt. Eng. **43**(6), 1318–1328 (2004)
93. Kim, S.J., Deng, F., Brown, M.S.: Visual enhancement of old documents with hyperspectral imaging. Pattern Recogn. **44**, 1461–1469 (2011)
94. Kim, S.J., Zhuo, S., Deng, F., Fu, C.-W., Brown, M.S.: Interactive visualization of hyperspectral images of historical documents. IEEE Trans. Visual. Comput. Graph. **16**(6), 1441–1448 (2010)
95. Kligler, N., Katz, S., Tal, A.: Document enhancement using visibility detection. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18–23 June 2018, Salt Lake City, UT, USA, pp. 2374–2382 (2018)
96. Koo, T.K., Li, M.Y.: A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropractic Med. **15**(2), 155–163 (2016)
97. Koschan, A., Abidi, M.: Digital Image Processing. Wiley, Chichester (2008)
98. Kovesi, P.: Good colour maps. How to design them (2015). arXiv:1509.03700
99. Kowalik, G., Nielek, R.: Senior programmers: characteristics of elderly users from stack overflow. In: Proceedings of the 8th Internationl Conference on Social Informatics (SocInfo 2016), 11–14 November 2016, Bellevue, WA, USA, part II, pp. 87–96 (2016)
100. Krupinski, E.A.: Practical applications of perceptual research. In: Beutel, J., Van Metter, R.L., Kundel, H.L. (eds.) Handbook of Medical Imaging, Bellingham, vol. 1, pp. 895–929. SPIE, WA (2009)
101. Kruse, P., Stadler, M. (eds.): Ambiguity in Mind and Nature: Multistable Cognitive Phenomena. Springer, Berlin (1995)
102. Kuhlthau, C.C.: Seeking Meaning: A Process Approach to Library and Information Services. Libraries Unlimited, Westport, CT (2004)
103. Kuncheva, L.I.: Combining pattern classifiers: methods and algorithms, 2nd edn. Wiley, New Jersey, NJ (2014)
104. Kundel, H.L.: Visual search in medical images. In: Beutel, J., Van Metter, R.L., Kundel, H.L. (eds.) Handbook of Medical Imaging, vol. 1, pp. 837–858. SPIE, Bellingham, WA (2000)
105. Labaune, J., Jackson, J.B., Pagès-Camagna, S., Menu, M., Mourou, G.A.: Terahertz investigation of Egyptian artifacts. In: Proceedings of the 35th International Conference on Infrared, Millimeter, and Terahertz Waves, 5–10 September 2010, Rome, Italy, pp. 1–3 (2010)
106. Lakoff, G., Johnson. M.: Metaphors We Live By. Univ. of Chicago Press, Chicago, IL (1980)
107. Land, E.H.: The retinex theory of color vision. Sci. Am. **237**(6), 108–128 (1977)
108. Langville, A.N., Meyer, C.D.: Who's #1? The Science of Rating and Ranking. Princeton Univ. Press, Princeton, NJ (2012)
109. Legge, G.E., Rubin, G.S., Schleske, M.M.: Contrast-polarity effects in low-vision reading. In: Woo, G.C. (ed.) Low Vision: Principles and Applications, pp. 288–307. Springer, New York, NY (1987)
110. Lettner, M., Sablatnig, R.: Spatial and spectral based segmentation of text in multispectral images of ancient documents. In: Proceedings of the 10th International Conference on Document Analysis and Recognition, 26–19 July 2009, Barcelona, Spain, pp. 813–817 (2009)

111. Li, X.: Image denoising: past, present, and future. In: Gunturk, B.K., Li, X. (eds.) Image restoration: fundamentals and advances, pp. 1–19. CRC Press, Boca Raton, FL (2013)
112. Lin, Y., Seales, W.B.: Opaque document imaging: building images of inaccessible texts. In: Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV' 05), 17–21 October 2005, Beijing, China, vol. 1, pp. 662–669
113. Lins, R.D., Almeida, M.M., Bernardino, R.B., Jesus, D., Oliveira, J.M.: Assessing binarization techniques for document images. In: Proceedings of the ACM Document Engineering Conference (DocEng' 17), 4–7 September 2017, Valletta, Malta, pp. 183–192 (2017)
114. Lund, O.: Knowledge construction in typography: the case of legibility research and the legibility of sans serif typefaces. PhD thesis, Univ. of Reading, Reading (1999)
115. Lombardi, F., Marinai, S.: Deep learning for historical document analysis and recognition: a survey. J. Imag. **6**(10), 110 (2020)
116. Machado, G.M., Oliveira, M.M., Fernandes, L.A.F.: A physiologically-based model for simulation of color vision deficiency. IEEE Trans. Visual. Comput. Graph. **15**(6), 1291–1298 (2009)
117. Marinai, S., Karatzas, D.: Report from the AND 2009 working group on noisy text datasets. Intl. J. Doc. Anal. Recogn. **14**(2), 113–116 (2011)
118. MathWorks: Contrast adjustment. Matlab: Image Processing Toolbox (2020). https://www.mathworks.com/help/images/contrastadjustment.html
119. McCann, J.J.: Retinex at 50: color theory and spatial algorithms: a review. J. Elect. Imag. **26**(3), 031204 (2017)
120. McIntyre, D.: Colour blindness: Causes and Effects. Dalton Publishing, Chester (2002)
121. Mehta, R., Zhu, R.: Blue or red? Exploring the effect of color on cognitive task performances. Science **323**(5918), 1226–1229 (2009)
122. Miller, E., Lee, N.J., Uprichard, K., Daniels, V.: The examination and conservation of the Rosetta stone at the British Museum. In: Roy, A., Smith, P. (eds.) Tradition and Innovation: Advances in Conservation, pp. 128–132. The Intl. Institute for Conservation of Historic and Artistic Works, London (2000)
123. Mindermann, S.: Hyperspectral imaging for readability enhancement of historic manuscripts. M.A. thesis, Technical Univ. München, München (2018)
124. Mitchell, G.: Revisiting truth or triviality: the external validity of research in the psychological laboratory. Perspect. Psychol. Sci. **7**(2), 109–117 (2012)
125. Mocella, V., Brun, E., Ferrero, C., Delattre, D.: Revealing letters in rolled Herculaneum papyri by X-ray phase-contrast imaging. Nat. Commun. **6**, 5895 (2015)
126. Moriyama, D., Azetsu, T., Ueda, Ch., Suetake, N., Uchino, E.: Image enhancement with lightness correction and image sharpening based on characteristics of vision for elderly persons. Opt. Rev. **27**(4), 352–360 (2020)
127. Morovič, J.: Color Gamut Mapping. Wiley, Chichester (2008)
128. Mullen, K.T.: The contrast sensitivity of human colour vision red-green and blue-yellow chromatic gratings. J. Physiol. **359**, 381–400 (1985)
129. Mussell, J.: The Nineteenth-Century Press in the Digital Age. Palgrave Macmillan, London (2012)
130. Neji, H., Nogueras-Iso, J., Lacasta, J., Ben Halima, M., Alimi, A.: Adversarial autoencoders for denoising digitized historical documents: the use case of incunabula. In: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR 2019), 20–25 September 2019, Sydney, Australia, pp. 31–34 (2019)
131. Nguyen, C.K., Nguyen, C.T., Hotta, S., Nakagawa, M.: A character attention generative adversarial network for degraded historical document restoration. In: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR 2019), 20–25 September 2019, Sydney, Australia, pp. 420–425 (2019)
132. Oppenheimer, D.M.: Consequences of erudite vernacular utilized irrespective of necessity: problems with using long words needlessly. Appl. Cogn. Psychol. **20**, 139–156 (2006)
133. Pan, X.-B., Brady, M., Bowman, A.K., Crowther, C., Tomlin, R.S.O.: Enhancement and feature extraction for images of incised and ink texts. Image Vis. Comput. **22**, 443–451 (2004)
134. Paris, S., Hasinoff, S.W., Kautz, J.: Local laplacian filters: edge-aware image processing with a laplacian pyramid. Commun. ACM **58**(3), 81–91 (2015)
135. Parry, N.R.A.: Color vision deficiencies. In: Elliot, A.J., Fairchild, M.D., Franklin, A. (eds.) Handbook of Color Psychology, pp. 216–242. Cambridge Univ. Press, Cambridge (2015)
136. Peli, E.: Limitations of image enhancement for the visually impaired. Optom. Vis. Sci. **69**(1), 15–24 (1992)
137. Peli, E., Lee, E., Trempe, C.L., Buzney, S.: Image enhancement for the visually impaired: the effects of enhancement on face recognition. J. Opt. Soc. Am. A **11**(7), 1929–1939 (1994)
138. Peli, E., Woods, R.L.: Image enhancement for impaired vision: the challenge of evaluation. Int. J. Artif. Intell. Tools **18**(3), 415–438 (2009)
139. Perry, S.: Image and video noise: an industry perspective. In: Bertalmío, M. (ed.) Denoising of photographic images and video: fundamentals, open challenges and new trends, pp. 217–230. Springer, Cham (2018)
140. Petro, A.B., Sbert, C., Morel, J.-M.: Multiscale Retinex. Image Process. On Line **4**, 71–88 (2014)
141. Pham, T.-A., Delalandre, M.: Post-processing coding artefacts for JPEG documents. Int. J. Doc. Anal. Recogn. **20**(3), 189–200 (2017)
142. Piquette, K.E.: Revealing the material world of ancient writing: digital techniques and theoretical considerations. In: Hoogendijk, F.A.J., van Gompel, S.M.T. (eds.) The Materiality of Texts from Ancient Egypt: New Approaches to the Study of Textual Material from the Early Pharaonic to the Late Antique Period, pp. 94–118. Brill, Leiden (2018)
143. Plateau, J.: Mémoire sur l'irradiation. Hayez, Bruxelles (1839)
144. Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., Kuo, C.-C.J.: Image database TID2013: Peculiarities, results and perspectives. Image Commun. **30**(1), 57–77 (2015)
145. Pons, C., Mazade, R., Jin, J., Dul, M.W., Zaidi, Q., Alonso, J.M.: Neuronal mechanisms underlying differences in spatial resolution between darks and lights in human vision. J. Vis. **17**(14), 5–24 (2017)
146. Popowicz, A., Smolka, B.: Overview of Grayscale Image Colorization Techniques. In: Celebi, M.E., Lecca, M., Smolka, B. (eds.) Color Image and Video Enhancement, pp. 345–370. Springer, Cham (2015)
147. Da Pos, O., Zambianchi, E.: Visual Illusions and Effects: A Collection. Guerini, Milano (1996)
148. Poston, A.M.: A literature review of cockpit lighting. Technical memorandum 10-64, US Army Human Engineering Laboratory, Aberdeen Proving Ground, MD (1974)
149. Pratikakis, I., Zagoris, K., Karagiannis, X., Tsochatzidis, L., Mondal, T., Marthot-Santaniello, I.: ICDAR 2019 Competition on Document Image Binarization (DIBCO 2019). In: Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR), 20–25 September 2019, Sydney, Australia, pp. 1547–1556 (2019)
150. Rasche, K., Geist, R., Westall, J.: Detail preserving reproduction of color images for monochromats and dichromats. IEEE Comput. Graph. Appl. **26**(3), 22–30 (2005)

P6 – 116

151. Raj, V., Arunkumar, C.: Content restoration of degraded termite bitten document images. Int. J. Eng. Res. Appl. **4**(5), 151–155 (2014)
152. Ribeiro, F., Florencio, D., Nascimento, V.: Crowdsourcing subjective image quality evaluation. In: Proceedings of the 18th IEEE Interantional Conference on Image Processing (ICIP 2011), 11–14 September 2011, Brussels, Belgium, pp. 3097–3100 (2011)
153. Rodney, A.: Color Management for Photographers. Focal Point, Oxford (2005)
154. Rogowitz, B.E., Treinish, L.A.: Data visualization: the end of the rainbow. IEEE Spectrum **12**, 52–59 (1998)
155. Roued-Cunliffe, H.: A decision support system for the reading of ancient documents. Ph.D. thesis, Faculty of Classics, Univ. of Oxford, Oxford (2011)
156. Rudd, M.E.: Lightness computation by the human visual system. J. Elect. Imag. **26**(3), 031209 (2017)
157. Rudd, M.E., Popa, D.: Stevens's brightness law, contrast gain control, and edge integration in achromatic color perception: a unified model. J. Opt. Soc. Am. **24**(9), 2766–2782 (2007)
158. Samei, E., Krupinski, E.A. (eds.): The Handbook of Medical Image Perception and Techniques, 2nd edn. Cambridge Univ. Press, Cambridge (2019)
159. Sassoon, R.: Handwriting of the Twentieth Century. Intellect, Bristol (2007)
160. Schivre, G.: Multiscale Retinex. MATLAB Central File Exchange (2020). https://www.mathworks.com/matlabcentral/fileexchange/71386-multiscale-retinex
161. Schubert, P.: Editing a Papyrus. In: Bagnall, R.S. (ed.) The Oxford Handbook of Papyrology, pp. 197–215. Oxford Univ. Press, Oxford (2009)
162. Schwartz-Shea, P., Yanow, D.: Interpretive research design: concepts and processes. Routledge, New York, NY (2012)
163. Seuret, M., Chen, K., Eichenberger, N., Liwicki, M., Ingold, R.: Gradient-domain degradations for improving historical documents images layout analysis. In: Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR), 23–26 August 2015, Nancy, France, pp. 1006–1010 (2015)
164. Shapiro, A., Todorović, D. (eds.): The Oxford Compendium of Visual Illusions. Oxford Univ. Press, Oxford (2017)
165. Shapiro, A.G., Hedjar, L.: Color illusion as a spatial binding problem. Curr. Opin. Behav. Sci. **30**, 149–155 (2019)
166. Sharma, A.: Understanding color management, 2nd edn. Wiley, Hoboken, NJ (2018)
167. Sharma, G., Rodríguez-Pardo, C.E.: The dark side of CIELAB. In: Proceedings of the SPIE–IS&T Electronic Imaging, Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications, vol. 8292: 82920D (2012)
168. Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: Proceedings of the 1996 IEEE Symposium on Visual Languages, 3–6 September 1996, Boulder, CO, USA, pp. 336–343 (1996)
169. Shortland, N.D., Alison, L.J., Moran, J.M. (2019). Conflict: How Soldiers Make Impossible Decisions. Oxford Univ. Press, Oxford (2019)
170. Smith, A.R.: Color gamut transform pairs. In: SIGGRAPH '78: Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques, 23–25 August 1978, Atlanta, GA, USA, pp. 12–19 (1978)
171. Sider, D.: The Library of the Villa Dei Papiri at Herculaneum. Getty Publications, Los Angeles, CA (2005)
172. Simons, D.J., Chabris, C.F.: Gorillas in our midst: Sustained inattentional blindness for dynamic events. Perception **28**(9), 1059–1074 (1999)
173. Sobottka, K., Kronenberg, H., Perroud, T., Bunke, H.: Text extraction from colored book and journal covers. Int. J. Doc. Anal. Recogn. **2**(4), 163–176 (2000)
174. Sonnad, S.: A survey on fusion of multispectral and panchromatic images for high spatial and spectral information. In: Proceedings of the International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 23–25 March 2016, Chennai, India, pp. 177–180 (2016)
175. Sparavigna, A.C.: Digital restoration of ancient papyri (2009). arXiv:0903.5045
176. Tarte, S.: Papyrological investigations: transferring perception and interpretation into the digital world. Liter. Ling. Comput. **26**(2), 233–247 (2011)
177. Terras, M.M.: Image to interpretation: an intelligent system to aid historians in reading the Vindolanda texts. Oxford Univ. Press, Oxford (2006)
178. Tian, C., Fei, L., Zheng, W., Xu, Y., Zuo, W., Lin, C.-W.: Deep learning on image denoising: an overview. Neural Netw. **131**(11), 251–275 (2020)
179. Tonazzini, A.: Color space transformations for analysis and enhancement of ancient degraded manuscripts. Pattern Recognit. Image Anal. **20**(3), 404–417 (2010)
180. Tonazzini, A., Gerace, I., Martinelli, F.: Document image restoration and analysis as separation of mixtures of patterns: from linear to nonlinear models. In: Gunturk, B.K., Li, X. (eds.) Image Restoration: Fundamentals and Advances, pp. 285–310. CRC Press, Boca Raton, FL (2012)
181. Tournié, A., Fleischer, K., Bukreeva, I., Palermo, F., Perino, M., Cedola, A., Andraud, C., Ranocchia, G.: Ancient Greek text concealed on the back of unrolled papyrus revealed through short-wave-infrared hyperspectral imaging. Sci. Adv. **5**(10), eaav8936 (2019)
182. Van Minnen, P.: The future of papyrology. In: Bagnall, R.S. (ed.) The oxford handbook of papyrology, pp. 644–660. Oxford Univ. Press, Oxford (2009)
183. VanderPlas, S., Hofmann, H.: Signs of the sine illusion: why we need to care. J. Comput. Graph. Stat. **24**(4), 1170–1190 (2015)
184. Velázquez, A., Levachkine, S.: Text/graphics separation and recognition in raster-scanned color cartographic maps. In: Proceedings of the International Workshop on Graphics Recognition, GREC 2003: Graphics Recognition. Recent Advances and Perspectives, Lecture Notes in Computer Science, vol 3088, pp. 63–74 (2004)
185. Ventzas, D., Ntogas, N., Ventza, M.M.: Digital restoration by denoising and binarization of historical manuscripts images. In: Ventzas, D. (ed.) Advanced Image Acquisition, Processing Techniques and Applications, pp. 73–108. Rijeka, InTech (2012)
186. Viénot, F., Brettel, H., Mollon, J.D.: Digital video colourmaps for checking the legibility of displays by dichromats. Color. Res. Appl. **24**(4), 243–252 (1999)
187. Vilaseca, M., Pujol, J., Arjona, M., Martínez-Verdú, F.M.: Color visualization system for near-infrared multispectral images. J. Imag. Sci. Technol. **49**(3), 246–255 (2005)
188. Wang, Z., Bovik, A.C.: Modern Image Quality Assessment. Morgan & Claypool, San Rafael, CA (2006)
189. Wang, Z., Simoncelli, E.P.: Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In: Proceedings of the SPIE-IS&T Electronic Imaging Conference, Human Vision and Electronic Imaging X, 18 March 2005, San Jose, CA, USA, pp. 149–159 (2005)
190. Webster, M.A.: Individual differences in color vision. In: Elliot, A.J., Fairchild, M.D., Franklin, A. (eds.) Handbook of Color Psychology, pp. 197–215. Cambridge Univ. Press, Cambridge (2015)
191. Westheimer, G.: Irradiation, border location and the shifted-chessboard pattern. Perception **36**(4), 483–494 (2007)

P6 – 117

192. Wu, H.R., Rao, K.R. (eds.): Digital Video Image Quality and Perceptual Coding. CRC Press, Boca Raton, FL (2006)
193. Yemelyanov, K.M., Lin, Sh.Sch., Luis, W.Q., Pugh, E.N. Jr, Engheta, N.: Bio-inspired display of polarization information using selected visual cues. In: Proceedings of the SPIE Optical Science and Technology, 5158, Polarization Science and Remote Sensing, 3–8 August 2003, San Diego, California, USA, pp. 71–84 (2003)
194. Yu, F.T.S., Tai, A., Chen, H.: Spatial filtered pseudocolor holographic imaging. J. Opt. **9**(5), 269–273 (1978)
195. Zheng, Y.: An overview of night vision colorization techniques using multispectral images: From color fusion to color mapping. In: International Conference Audio Language Image Processing, 16–18 July 2012, Shanghai, China, pp. 134–143 (2012)
196. Zhu, Zh., Toyoura, M., Go, K., Kashiwagi, K., Fujishiro, I., Wong, T.T., Mao, X.: Personalized Image Recoloring for Color Vision Deficiency Compensation. IEEE Transactions on Multimedia, in press (2021)
197. Ziyaee, T.: Unsupervised denoising via Wiener entropy masking in the STFT domain. In: Proceedings of the 2014 IEEE Military Communications Conference, 6–8 October 2014, Baltimore, MD, USA , pp. 467–472 (2014)
198. Zuiderveld, K.: Contrast limited adaptive histograph equalization. In: Graphic Gems IV, pp. 474–485. Academic Press Professional, San Diego, CA (1994)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

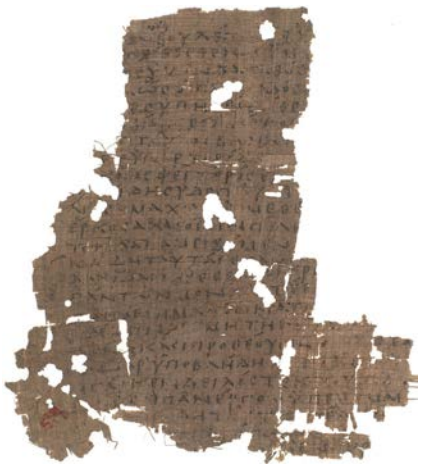
P6
–
118

Supplementary Material

Presented herein are the ninety images used in the evaluation of papyri legibility enhancement experiment. The methods are described in the article, Sections 4 Methods and 5.1 Algorithms.

TSVP →

Vlad Atanasiu, Isabelle Marthot-Santaniello (2021) — "Personalizing image enhancement for critical visual tasks: improved legibility of papyri using color processing and visual illusions", International Journal on Document Analysis and Recognition, https://doi.org/10.1007/s10032-021-00386-0.



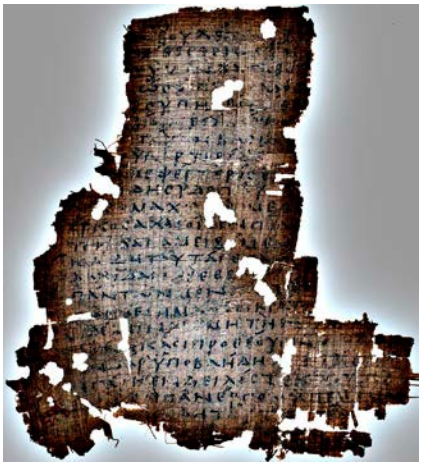
original

stretchlim

histeq

adapthisteq

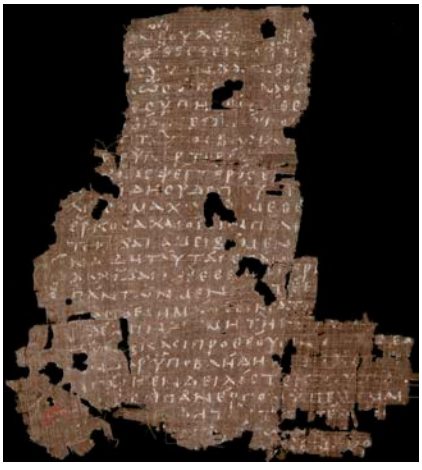
locallapfilt



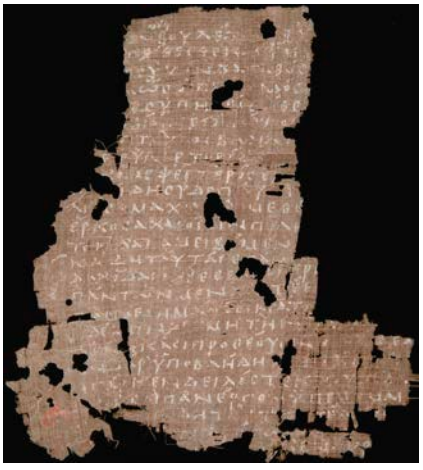
retinex

lsv

vividness



neglsv



negvividness

columbia.apis.p367.f.0.600

Credit of original papyrus reproduction: Columbia University Library, CC BY-NC 3.0

original

stretchlim

histeq

adapthisteq

locallapfilt

retinex

lsv

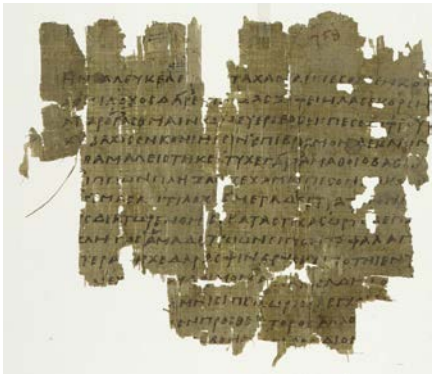
vividness

neglsv

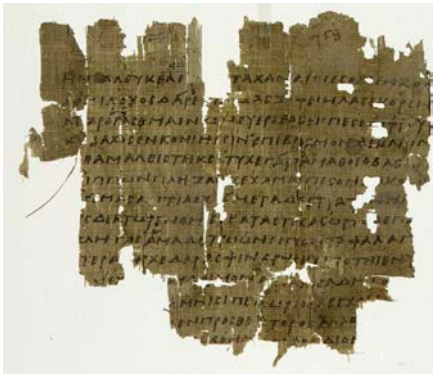
negvividness

P.Corn. Inv. MSS.A 101. XIII

Credit of original papyrus reproduction: University of Michigan Library, Public Domain Mark 1.0

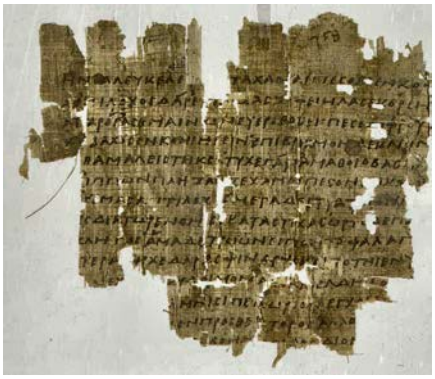


original

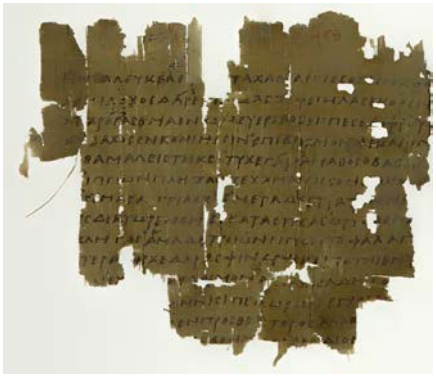


stretchlim

histeq



adapthisteq



locallapfilt

retinex

lsv

vividness

neglsv

negvividness

P.CtYBR inv. 69

Credit of original papyrus reproduction: Yale University Library, Public Domain Mark 1.0

original

stretchlim

histeq

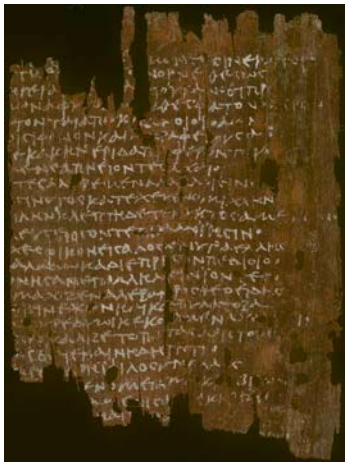
adapthisteq

locallapfilt

retinex

lsv

vividness



neglsv

negvividness

P.Mich.inv.1318v

Credit of original papyrus reproduction: University of Michigan Library, Public Domain Mark 1.0

original

stretchlim

histeq

adapthisteq

locallapfilt

retinex

lsv

vividness

neglsv

negvividness

P.Mich.inv.2755

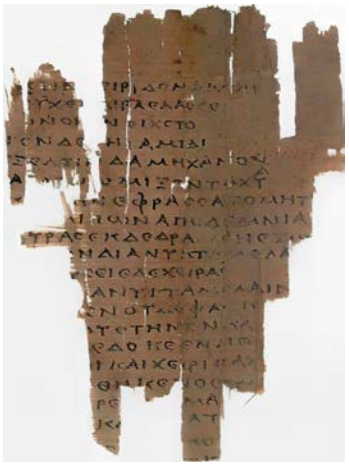
Credit of original papyrus reproduction: University of Michigan Library, Public Domain Mark 1.0

original

stretchlim

histeq

adapthisteq

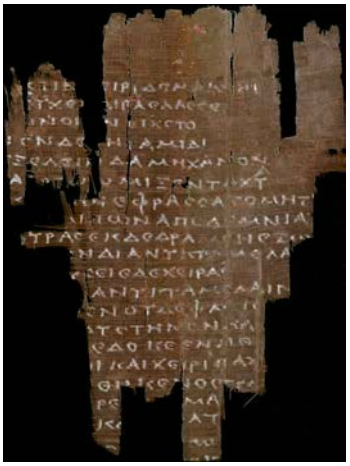


locallapfilt

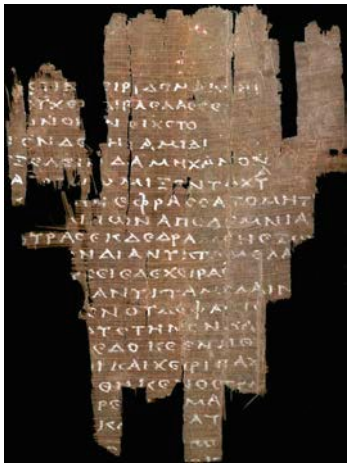
retinex

lsv

vividness



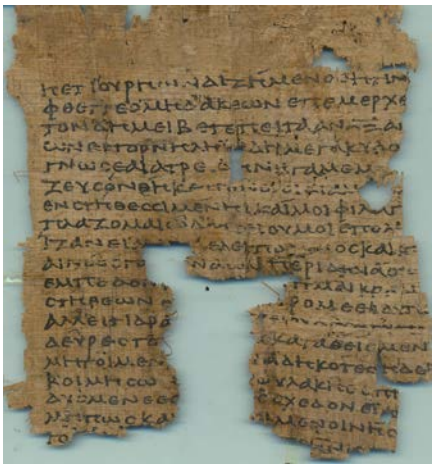
neglsv



negvividness

P.Oxy.XXII 2309

Credit of original papyrus reproduction: Courtesy of The Egypt Exploration Society and the University of Oxford Imaging Papyri Project

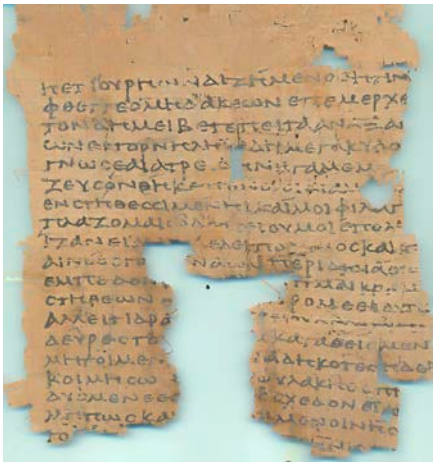


original

stretchlim

histeq

adapthisteq

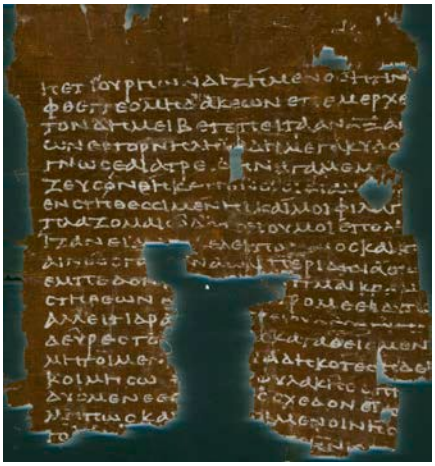


locallapfilt

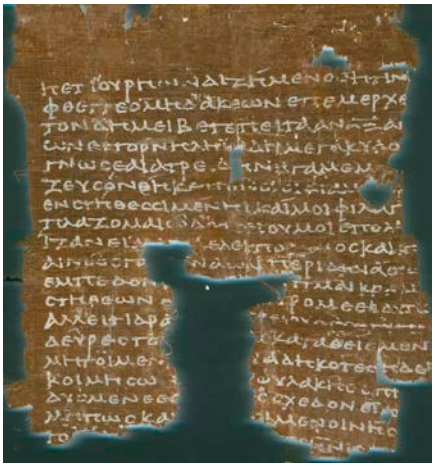
retinex

lsv

vividness



neglsv

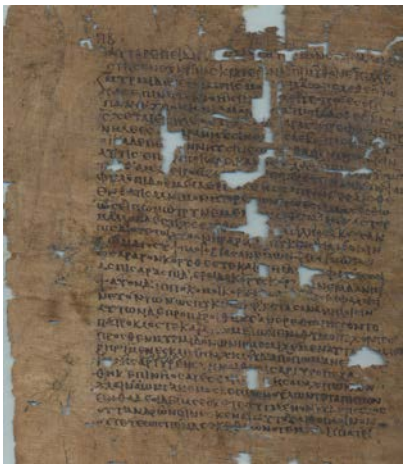


negvividness

P6 – 126

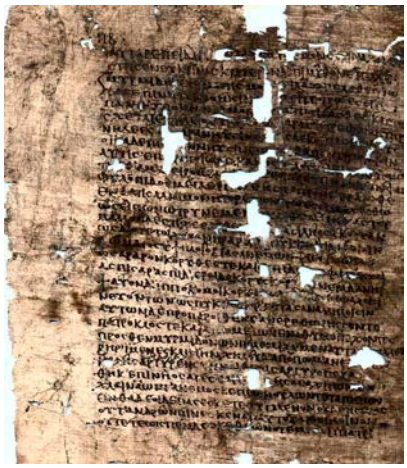
PSI XII 1274 r

Credit of original papyrus reproduction: Istituto Papirologico Vitelli, by permission



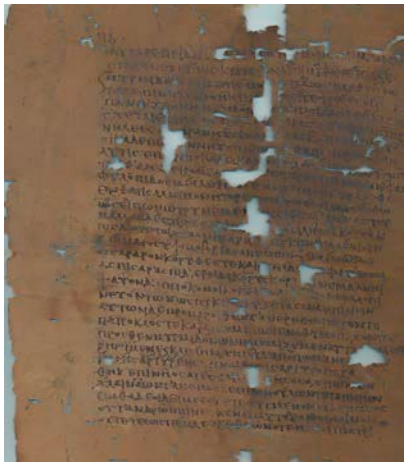
original

stretchlim

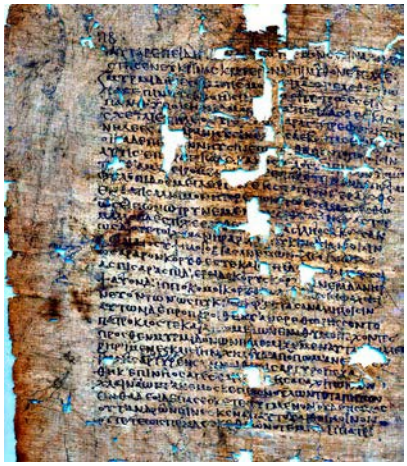


histeq

adapthisteq



locallapfilt

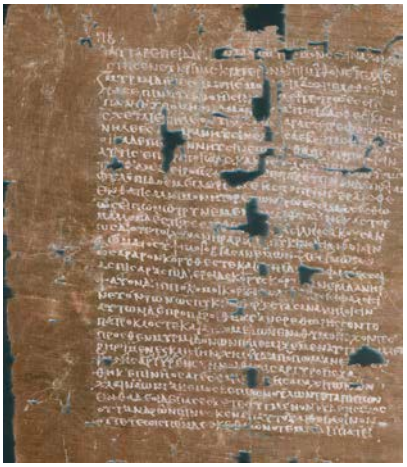


retinex

lsv

vividness

neglsv



negvividness

PSI XIII 1298 (15a) r 1

Credit of original papyrus reproduction: Istituto Papirologico Vitelli, by permission

original

stretchlim

histeq

adapthisteq

locallapfilt

retinex

lsv

vividness

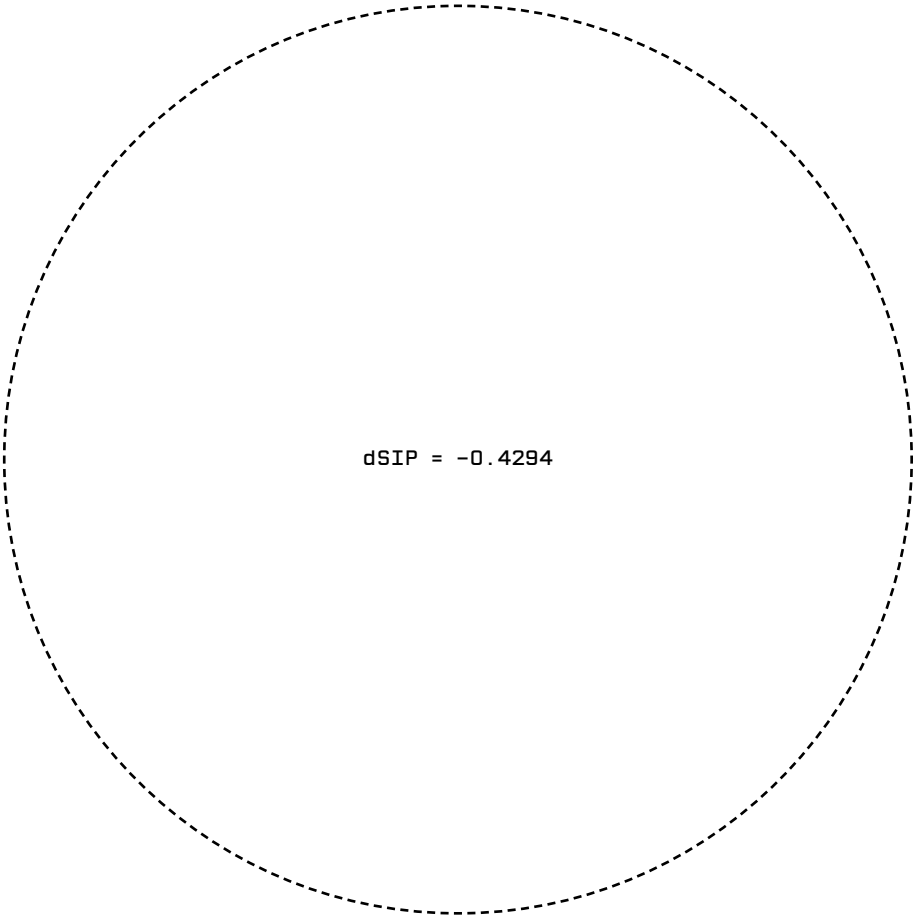
neglsv

negvividness

PSI XIV 1376 r

Credit of original papyrus reproduction: Istituto Papirologico Vitelli, by permission

Publication
the Seventh


dSIP = -0.4294

Heliyon 9 (2023) e20698

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

On the utility of Colour in shape analysis: An introduction to Colour science via palaeographical case studies

Vlad Atanasiu [a,*], Peter Fornaro [b]

[a] Department of Informatics, University of Fribourg, Boulevard de Pérolles 90, 1700, Fribourg, Switzerland
[b] Digital Humanities Lab, University of Basel, Spalenberg 65, 4051, Basel, Switzerland

ARTICLE INFO

Keywords:
Colour science
Colour processing
Colour perception
Image processing
Image enhancement
Palaeography

ABSTRACT

In this article, we explore the use of colour for the analysis of shapes in digital images. We argue that colour can provide unique information that is not available from shape alone, and that familiarity with the interdisciplinary field of colour science is essential for unlocking the potential of colour. Within this perspective, we offer an illustrated overview of the colour-related aspects of image management and processing, perceptual psychology, and cultural studies, using for exemplary purposes case studies focused on computational palaeography. We also discuss the changing roles of colour in society and the sciences, and provide technical solutions for using digital colour effectively, highlighting the impact of human factors. The article concludes with an annotated bibliography. This work is a primer, and its intended readership are scholars and computer scientists unfamiliar with colour science.

Whiteout, n. A white correction fluid used to cover textual errors.

— Oxford English Dictionary

1. Introduction

With the passing of time, and without restoration, the gold-plated engraving on the marble on the left side of Fig. 1 may come to resemble the one on the right, pictured in grayscale. This example of natural camouflage and anthropic decamouflage illustrates the role of colour in alternatively enhancing and diminishing the legibility of letter shapes. Moreover, it is the special phenomenological quality of light reflected by gold — warm and immaterial, transcending shape, texture, and colour — that constitutes one of the essential aesthetic appeals of this type of baroque epigraphy. These effects can be experienced only in the full ambient physical reality comprising oneself, the inscription, and its material and spiritual surroundings. Thus, colours can also tell different stories about shapes than their outlines do.

In nature and visual media, phenomena pertaining to the interaction between colour and shape are pervasive. At a time of mass digitisation, digital imagery and computational methods supply novel and powerful solutions to unlock the potentialities of colour, while also making the mastery of colour science a necessity for scientific work, unlike what was customary during the analogue era,

* Corresponding author.
E-mail addresses: atanasiu@alum.mit.edu (V. Atanasiu), peter.fornaro@unibas.ch (P. Fornaro).

https://doi.org/10.1016/j.heliyon.2023.e20698
Received 5 June 2023; Received in revised form 3 October 2023; Accepted 4 October 2023
Available online 10 October 2023
2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Fig. 1. Colour (a) and grayscale detail (b) of the foundation inscription on the Cantonal, City and University Library Zürich, Switzerland. — Credits: Vlad Atanasiu.

when not everybody was yet a photographer and a photoshopper. However, the interdisciplinary nature of colour science can make it difficult to be fully aware how a great variety of colour-related factors affect the capture, processing, display, and interpretation of shapes.

Informed by latent academic needs within a changing technological environment, the purpose of this article is to reach a readership new to colour science, in order to raise awareness of its benefits for shape analysis through an overview of the major aspects of theoretical and practical importance.*

This article is addressed to scholars and computer vision scientists working in the fields of archaeology and cultural heritage who are unfamiliar with colour science, but may find that there are numerous benefits to considering colour and shape conjointly when analysing artefacts such as epigraphic writing, wall paintings, aerial photography, and three-dimensional digital models. In fact, it has been hypothesised that human colour vision first evolved to improve the species' ability to detect and identify objects; in light of this fundamental function of colour, colour science becomes naturally relevant across a range of fields.

P7 – 132

For illustrative purposes, this article focuses on the branch of historical sciences that is palaeography, as considered in its broad sense independent of writing substrate and instrument, and thus including such disciplines as epigraphy and papyrology. Palaeography is essentially the study of shapes, not colour; this is an advantage in the context of the present article, since it makes debating the merits of colour more challenging. It should, however, be relatively intuitive to transfer what is learned from the present investigation of palaeography to the study of shapes in other disciplines.

Considering the many interacting sub-divisions of colour science, this article is organized serially according to how much influence can be exerted by users on the various factors affecting the use of colour in shape analysis. This organisation also coincides with a progression from basic to advanced topics that need to be understood. The motivation for using colour in shape analysis is subject to the persuasion power of colour science advocates with respect to the expected benefits (Section 2); colour science aspects such as physics, imaging, management, models, and displays can be approached as prescriptive givens, while users may also contribute to their development (Section 3); computer science aspects are the most open to creative developments (Section 4); psychological factors may be beneficially leveraged and influenced (Section 5); finally, social aspects are the least influenceable, but it is necessary to be aware of them in order to adopt appropriate attitudes (Section 6). Each topic is introduced by framing it in terms of colour utility in shape analysis, and illustrated by one or more case studies, supported by visual explanations. A synthetic overview of the main lessons is provided in the conclusions (Section 7). The extensive and annotated bibliography provides avenues for readers to pursue their interests beyond this article.

There is no prerequisite to understand the following exposition, other than a willingness to engage in transdisciplinarity. The text draws mainly on an eclectic mix of visual psychology and image processing, with abundant cultural and historical references.

* One recurring example for the need of colour science literacy can be gathered from articles on image processing, were sometimes the RGB and HSV colour spaces are used, while perceptually uniform spaces, such as CIELAB, would be more appropriate, for example, to convert an image to grayscale or manipulate the chromatic channels. A further (notorious) example is the use of the 'rainbow' colour map, which introduces spurious perceptual classes in otherwise monotonic linear data, and misleading image interpretation by showing shapes where there are none [36]. That ignorance of such trivial colour facts persists among scientists even decades after the issues were stated and remedies developed, indicates that colour literacy remains of actuality. (On colourmap misuses among scientists see Ref. [175]; on image processing for papyrology see [178: 94]; on the wanting state of colour literacy among visual artists, designers, and architects see Ref. [27]; the Color Literacy Project is an advocacy group from the sciences, arts, and the industry focused on promoting colour education in the fields of sciences, technology, engineering, arts, and mathematics (STEAM) [176]; the International Colour Day takes place annually on March 21; it was established by the International Colour Association (AIC) to promote colour science awareness [177].).

To create the right expectations about the nature of this work, we note here that it mixes elements of a *manifesto* programmatically (the exhortation to use colour), an *essay* stylistically (not a mechanistic proof), a *primer* content-wise (instructions for how to do things), a *survey* structurally (the catalogue of case studies), and *infographics* methodologically (visual illustrations). This multifaceted approach, similar to a *guided tour* showcasing the 'Wunderkammer' of colour, distinguishes it from typical research, review, and position papers. This approach was adopted to achieve our educational goals of 'raising awareness' and 'introducing colour science' in an accessible and comprehensive manner.

One strength of this article is that it introduces readers to a broadly relevant and timely shape analysis methodology that goes beyond the mainstream. In comparison to general-purpose handbooks and topic-focused articles on colour science, this publication offers a synthetic overview of a highly heterogenous field, containing both theoretical fundamentals and practical application examples. It deliberately emphasises multidisciplinary factors in colour-based shape analysis (given their objective ubiquity in this field) and the challenges of cross-disciplinary research. In doing so, it contributes to knowledge transfer between colour science, computer science, and the humanities. Last but not least, the article presents original content, including an overview of the changing scientific attitudes on colour, chromatic thinking models in sciences, and chromatic metaphors of shape, along with a comparison between 'manual' photo editing and image processing via programming (with remarkable results), and custom-made visualisations of concepts.

The limitations of this article derive from its format and the topic itself. As a synthesis article, it can offer only a glimpse of a vast subject of which much remains to be discovered. As for colour, its psychophysical effects and its interactions with shape are numerous, difficult to control, and subject to objective imaging variation and reproduction limitations, as well as subjective variation between individuals and cultures. Furthermore, the information presented here is a snapshot of our evolving understanding of colour. All these aspects are discussed in dedicated sections on (for example) colour management and display, colour-blindness and subjective colours, gender differences, and cultural attitudes between chromophobia to chromomania. Nevertheless, colour represents a supplemental lever in the analysis of shape that beseeches to be used. Aside from its practical benefits, it is also a fascinating intellectual topic.

2. Motivations to use colour in script analysis

Colour plays an important functional role within the document machinery as a device for improving readability and navigation. For the reproduction, processing, expertise, and display of written artefacts, this information should, in consequence, be preserved and utilized advantageously. Some rationales spanning these diverse tasks are presented below.

One striking case for considering colour in shape analysis is that of polychromatic script surfaces (Fig. 2). Like butterflies, polychromatic scripts are examples of shapes that cannot be fully analysed or appreciated if colour is discarded. Alternatively, the script surface may be uniformly coloured, while multiple colours are employed for different symbols, words, or sentences, to make semantic distinctions and mark structural text entities. This technique is common in palaeography across cultures and periods. For example, the diacritics of the early Qurans were written in red, green, and yellow, while in one remarkable Maghrebine Quran, every other word changes colour; Persian Qurans, moreover, may contain interlinear Farsi translations distinguishable from the Arabic lines by being inked in red Ref. [[1]: pp. 115–157]. Gold may be used for the headings in Italian Renaissance manuscripts, or for the entire content of luxury medieval Qurans from North Africa and Buddhist sutras from Japan [[2]: 68–69, 110]. In contemporary Arabic-script calligraphy, polychromatic script surface characterises the signature style of Hassan Massoudy and his followers, who use writing instruments pigmented with multiple inks [3]. Today (here we transition from palaeography to 'neography', the study of contemporary script), coloured texts have become an experimental subgenre in modern literature, and colour notation a versatile technique from music and diagrammatics to mathematics and electronics — while colour highlighters, sometimes fluorescent, are commonplace stationery items, that the reader of these lines might be using at this very instant (literature: example: [4]; literature: history: [5,6]; music: [7]; diagrammatics: [8]; mathematics: [9]). In brief, the proper colour reproduction and computational processing of all such documents is indispensable for their analysis.

P7 – 133

The palaeographical utility of colour extends beyond lavish polychromatic documents, if we consider the informativeness of microchromatic variation on script surfaces. Because ink accumulates at the end of a stroke where the writing instrument is lifted, while distinctive furrows are left when two strokes cross, the study of chromatic variation may allow for the inference of ductus direction and sequence [10]. This colour-based information is useful more generally to expertise, for dating and localising scripts, and for establishing authorship. On banknotes, chromatic effects are exquisitely intricated with shapes and incorporated as security devices, including in holograms, the importance of colour being such that the overall note colour may even give the currency by metonymy its popular name, e.g., 'greenback' for 'dollar' [11]. Again, colour reproduction fidelity and colour processing are indispensable to make such operations possible and successful.

Through experience, epigraphists working on worn-out inscriptions are well aware that script exists only insofar as it can be

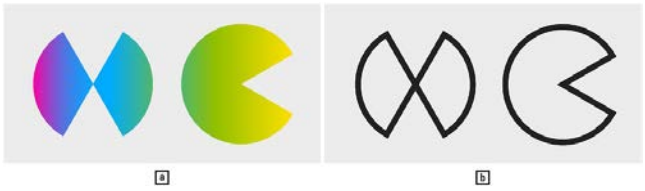


Fig. 2. Polychromatic script (a) and its monochromatic variant (b). — Credits: Letters H and C, one resembling a butterfly and the other a Pac-Man, form the colour and outline versions of the logo of the Swiss Association 'History and Computing' (https://ahc-ch.ch), designed by Vlad Atanasiu.

distinguished from the writing substrate, a segmentation task facilitated perceptually and computationally by the surplus of information introduced by colour. The legibility of flaking Pompeii inscriptions, gravestone epitaphs covered by lichens, and script intermixed with papyri fibres can be greatly improved through the use of colour-based computational enhancement techniques, as we shall see shortly.

Faithful colour reproduction and colour-aware image processing are also necessary for the documentation of the conservation state of documents. From art-historical and museographical points of view, extensive research must yet be conducted before the chromatic experience obtained from originals can be effectively conveyed through reproductions. The reflection of gold, the shimmering of texture, and the transparency of ink, for example, are notoriously difficult to capture, display, and measure with today's technologies, despite forming core aspects of the phenomenology of script, i.e., the emotions and connotations evoked by contemplating the written surface.

3. Fundamentals for the effective use of colour in shape analysis

The reproduction of documents in visible light (imaging) involves many technical aspects, such as lighting, camera optics and settings, and the use of reference rulers and colour patches, all of which need to be mastered to produce the kind of high-quality images necessary for fine palaeographical analysis. Since these techniques have been extensively studied in the pre-digital era, we will focus in this section on some aspects of digital colour management that are often overlooked by palaeographers and computer scientists unfamiliar with colour science.

Specifically, we demonstrate the benefits of colour fidelity using the case of document fragment matching; we illustrate how the digital image file format affects the human perception of shapes, as well as the quality of their computational processing; we point out that the judicious choice of the colour space in which digital images are represented facilitates their processing; and we explore the perhaps more obscure phenomenon of legibility varying with the colour gamut.

To show how these quite different topics fit into a single perspective and how they relate to the computer science, psychology, and sociology aspects described in the other sections, we will begin with an overview of colour science as a complex research domain.

P7
–
134

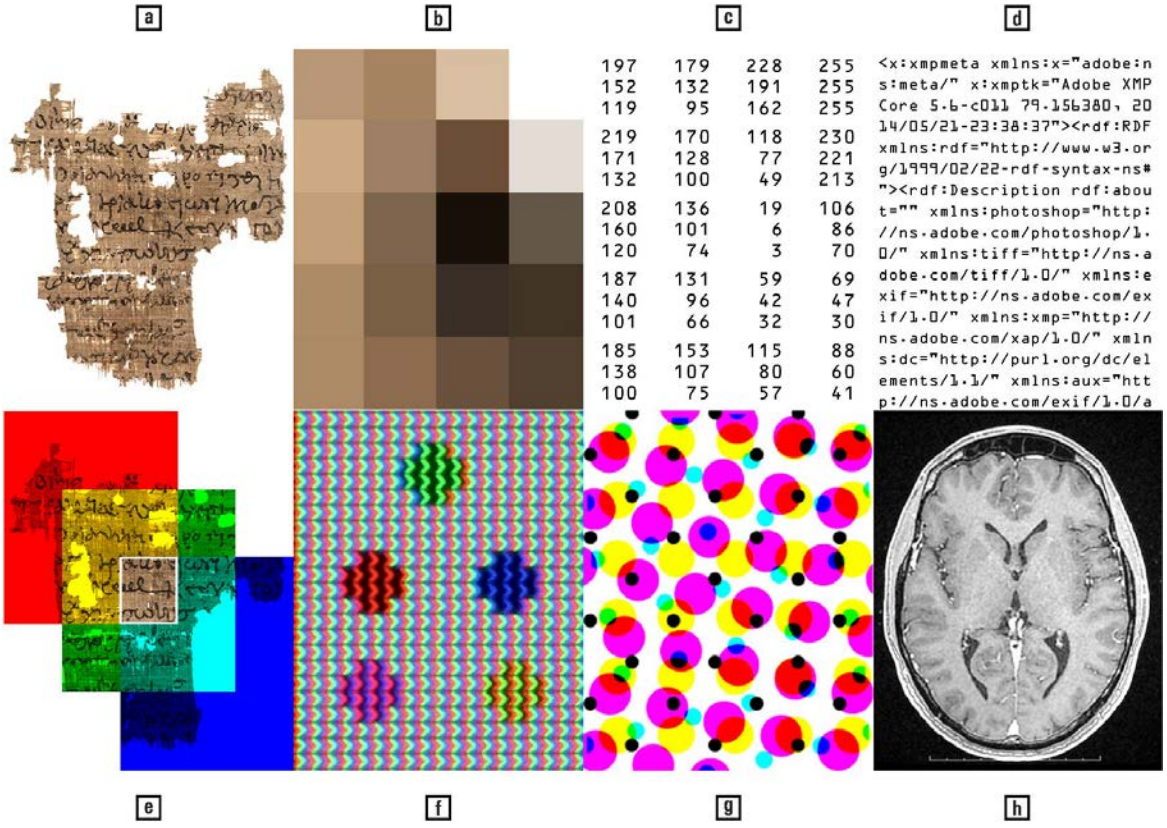


Fig. 3. What are colours and shapes in a digital image, such as that of a papyrus (a)? Depending on what manifestation we are looking at, the answer is: (b) square colour pixels, which appear when zooming into an image; (c) a number matrix, defining the colour values of the pixels; (d) programming language, found by opening a XMP image metadata file with a text editor; (e) red, green, and blue colour channels that produce a colour stimulus via superposition; (f) a raster pattern of red, green, and blue filtered light emitters of a computer monitor, which fuses into five coloured dots when seen from far away; (g) the cyan, yellow, magenta, and black ink dots of a printed image, which also fuse into a pinkish uniform surface when seen from afar; (h) neural activity within the brain, experienced as an 'image'. — Credits: a. P.Bas. inv. 47, Peter Fornaro, University of Basel Library; f, g. [26]; h. Vlad Atanasiu.

3.1. Understanding the complexity of colour science

This section aims to show the complexity of image formation, the multidisciplinarity of colour science, and the mental nature of images. It provides the conceptual framework within which colour-based shape analysis is performed.

Let us briefly recall the process by which the phenomena we perceive as colour and shape are formed, from the physical and material stage to imaging and digital images to the perception of visual stimuli. Illuminant bodies such as the sun, fireflies, and artificial lightbulbs emit electromagnetic waves of specific lengths, a narrow band of which can be perceived by the human visual system: this is 'visible light', in the range of approximately 390 nm–730 nm. These 'rays' of energy make contact with various materials — an ink-inscribed papyrus, for example — that modify the spectral power distribution of the reflected light by absorption, thereby giving rise to colours specific to objects under the given viewing conditions. The outgoing energy quanta — the photons — are captured by an imaging system, i.e., a light-sensitive device such as a photographic camera, which typically consists of optics, a photosensitive sensor, electronics, and software. The output is a digital file that can be further computationally processed, transmitted, stored, and then presented on a display or printed on paper. Those operations, which involve the preservation of colour consistency, constitute the so-called colour management (practical: [12]; technical: [13,14]; ICC profiles: [15]; gamma: [16]). The results of this process are, first, a stimulus for the physiology of the eye, and thereafter for the neural activity of the brain experienced as 'colour'. The interaction between colour and shape is interdependent on other perceptual features, such as transparency, shimmering, and texture, and other cognitive processes, such as memory, motricity, and language. These stimuli are interpreted according to individual, cultural, and social factors, and may prompt the viewer to take certain actions.

It is apparent that image formation is complex and involves many disciplines: physics, chemistry, optics, electronics, mathematics, computer science, psychology, neuroscience, philosophy, linguistics, history, and others. All these disciplines are part of colour science (compendia: [17,18,19]; data and formulae: [20]; history of colour science: [21]).

A further observation is that both colour and shape have radically different manifestations depending on their stage of formation. Fig. 3 presents some manifestations of a digital image: pixels, numbers, programming code, colour channels, liquid crystals, typographical dithering, and neural activity. By looking through a magnifier at a computer screen it becomes apparent how the shape of letters morph into rectangular dots, while the myriad hues dissolve into primary colours (red, green, and blue). In other words, what we call 'colour' and 'shape' exist only in our brains. These phenomena are illusions — sufficiently operational to keep us humans alive, but still illusions. They represent how we experience the world, not how it is. The perceptual interface theory, for example, explains this viewpoint by analogy with the icons on a computer desktop that facilitate interaction with the machine, but reveal little about its inner workings (psychology: [22]; philosophy: [[23]: 111–112]; nature of optical illusions: [[24]: xix–xxiii]; in Newton's words: 'For the [light] Rays to speak properly are not coloured.' [[25]: 83]).

P7 – 135

Fig. 4. (a, b) Two reproductions of the same papyrus varying in colour characteristics. As neither has an embedded colour profile, it is only the first author of this article's memory of the original physical document that suggests (a) to be closer to the observed reality than (b). — A physically single document with (c) strong colour variation and (d) fragments attributed by expertise to the same document. — Credits: a, b. P. Hamb. graec. 175 recto, Staats-und Universitätsbibliothek Hamburg, https://resolver.sub.uni-hamburg.de/kitodo/HANSh795, (CC BY-SA 4.0); c, d. Inv.Sorb. 2239abc +2242a r, Florent Jacques, Institut de Papyrologie, Sorbonne Université, Paris.

3.2. Colour management and data fidelity

According to the op art artist Victor Vasarely [[27]*: 990*]*, 'form and colour are one and the same'. But if colours are not reproduced with fidelity, then the same object may appear as two distinct ones. The issue of faithful reproduction is at the core of 'colour management', as discussed in this section.*

Both natural and anthropogenic factors can cause the decomposition of documents into fragments. While their defragmentation may be desirable for various reasons, this is difficult to achieve without automation. Reasons for this can include the high number of fragments (e.g., in the well-known cases of the shredded documents of the former East German State Security organisation and those of the US Embassy in Iran in 1979 [[28]: 23, 34], or the worldwide dispersion of fragments between collections, as in the case of papyri). While colour is one of the document features that can aid in defragmentation, it is useful only insofar as colours are reproduced with fidelity. Fig. 4 presents images of the same document, taken by staff from the same institution at different moments in time (a, b). If these two images were different fragments of the same original document, it would be impossible to reconstruct the original based solely on colour information, because the reproductions are so different. This demonstrates the importance of proper quality control while digitising, as defined in digitisation guidelines such as the US Federal Agencies Digital Guidelines Initiative [29], ISO, and the Dutch 'Metamorfoze' [30]. When considering that the same document can exhibit significant and sudden colour variations (Fig. 4 c, d), it is evident that reproduction fidelity can help to avoid false fragment joins based on colour similarity. A stern enjoinment in this respect comes from physicians speaking about how, despite the importance of proper colour calibration in medical imagery — where boundaries of lesions, tumours, and other regions of interest may be less clearly defined as in palaeography —, the issue is too often ignored by practicians: 'without reproducibility and accuracy of images, any attempt to measure colour or geometric properties is of little use' [[31]: 2].

While good colour management is indispensable, it also has its limits. As the same authors observe ('Taking two pictures of a wound with the same camera and settings, immediately after one another, normally results in two slightly different images'), the difficulty of determining shape boundaries is compounded by the fact that no two digital images of the same scene are objectively the same, due to environmental factors (e.g., change in illumination, object motion) and variations in imaging systems (e.g., different cameras, different settings, random electromagnetic noise) [[31]: 1, 2]. Image variability may be overcome in various ways. For example, electromagnetic noise in the camera can be reduced by averaging over multiple shots of the same scene, a technique known as 'stacking' commonly used in astrophysics [32]. Image variability may also be a desirable feature, as it reveals a host of information about the photographed scene and the imaging conditions [33].

P7 – 136

3.3. Image formats and shape distortions

While a number of sections in the present article explain how colour can improve shape processing and perception, this section illustrates the reverse case, i.e., how colour can negatively affect shape analysis.

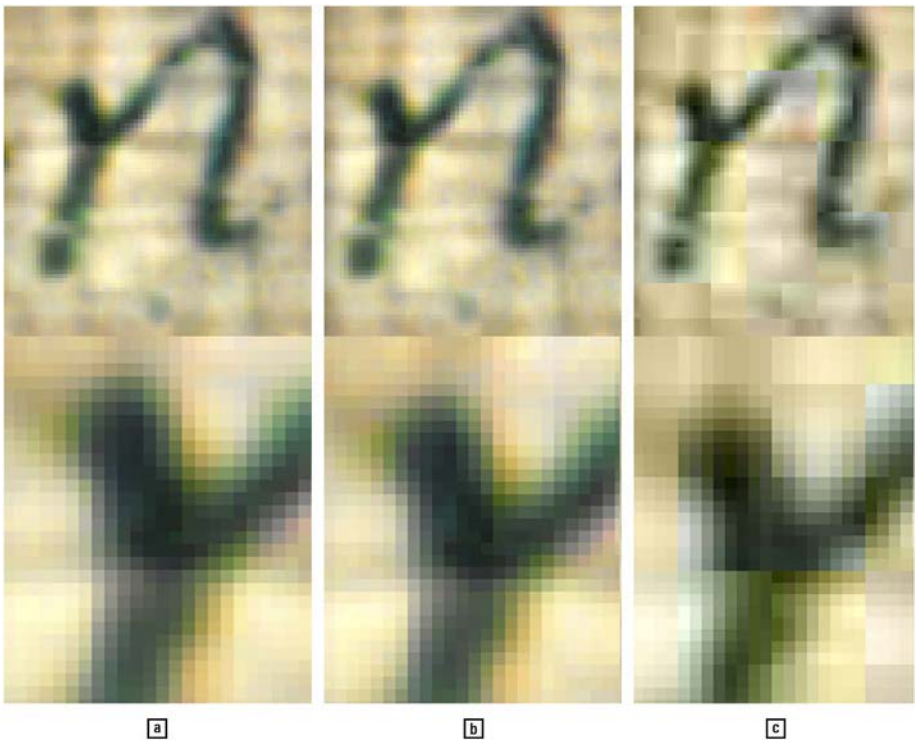


Fig. 5. (a) Original TIFF format image, (b) converted to JPEG with 100 % quality and (c) 10 % quality.

Fig. 5 presents details of a papyrus at two zoom levels, two file formats, and two compression levels. There are almost no visible differences between the first and second image, even when zooming in, and from far enough away, the third image also appears similar; however, substantial 'blocking' and shape distortion become apparent at close range [34]. Such artefacts are obviously problematic for fine palaeographical analysis, both human and computational.

The phenomenon is partially due to rationales involving colour: because the size of image files can be considerable (723 Mb for the entire image from which the detail in Fig. 9 a was extracted), they can take up significant storage space and be slow to transmit over computer networks. The JPEG format was thus developed to reduce file size while maintaining good perceptual image quality (86.4 Mb and 5.8 Mb in Fig. 9 b and 9.c, respectively), originally for application in videotext. Presently it is one of the most common digital image formats for cameras and on the Internet. Two of the methods implemented in the JPEG format to achieve such compression leverage specific characteristics of human visual perception: namely, the lower sensitivity to the blue colour band, and the lower sensitivity to high-frequency chromatic information (i.e., small spatial variations in saturation) as opposed to high-frequency achromatic information (i.e., details distinguished by lightness) [[35]: 23–28; [17]: 26–29; [36]]). By rounding a parameterizable number of values associated with these features (a process called quantisation and performed in the frequency domain on eight-by-eight blocks of pixels), the JPEG file size is reduced. Other common image file formats, such as PNG, GIF, and RAW, have other particularities.

While JPEG artefacts also occur for greyscale images, the reason why they occur in the first place is due to technical and perceptual solutions involving colour, for which this image format was optimised. This section attempts to point out these pitfalls for the benefit of the reader. Additionally, it showcases how colour management can affect shape quality through the choice of image formats, and further highlights that while the resulting degradations might be invisible to the human eye, they may nevertheless adversely affect computational image processing.

3.4. Colour spaces and image processing ease and validity

The way in which colour is described defines a 'colour space'. Because colour spaces are designed for specific applications, a large and increasing number of such spaces exist [37,38]. Here, we present some of the more prevalent types, along with their impact on colour and shape analysis, as exemplified by the colour distribution of a papyrus image (Fig. 5).

An RGB-based colour space defines colours in terms of the mix of the primary colours — red, green, and blue — that are necessary to obtain a given colour on most common displays. The display primaries are derived from the physiology of the human eye, which contains around seven million cells — 'cones' — that are individually sensitive to one of three overlapping ranges of light wavelengths (hence the term 'trichromatic' to characterise 'normal' human vision), and whose combined output is eventually perceived as colour [39]. Typical photographic cameras, file formats, and displays reproduce this architecture, which works as a biologically inspired means of capturing, storing, and transmitting colour. (Under low light conditions cones are insensitive, while around 120 million 'rod' cells become responsible for vision, and their activity results in achromatic perception.)

P7 – 137

L*a*b* (or CIELAB) is another colour space, based on psychophysical research dating back to the 1920s and carried out by the International Commission on Illumination (CIE), the leading standardisation organisation in colour matters, with the aim of characterising the perceptual qualities of colour. The orthogonal dimensions of L*a*b* are 'lightness' (L*, defining the black–grey–white achromatic axis), 'green–red' (a*), and 'blue–yellow' (b*; the two chromatic axes), and those of its cylindrical LChH version are 'lightness', 'chroma' (similar to 'saturation', and perpendicular to lightness), and 'hue' (circular around lightness) [[37]: 85–96]; these are not only more immediate and complete colour descriptors for humans than red, green, and blue, but also take the mathematical non-linearity of human colour perception into account. Additionally, using L*a*b* or LChH makes certain types of colour arithmetic more straightforward. For example, segmenting objects from their background according to their lightness level is simpler in L*a*b*, where lightness is one of the dimensions, than in RGB, where an approximation of lightness is given by the diagonal of the distribution cube. Moreover, the lightness of an image, which is useful in many computational applications, can be trivially obtained in the L*a*b* colour space by setting the a* and b* channels to zero.

The HSL colour space uses linear transformations of the display system R, G, and B values to encode colour as an approximation of hue, saturation, and lightness, and has gained popularity among early computer graphics researchers due to its intuitive colour representation; however, it has poor psychophysical validity. Colour differences, for example, are not perceptually uniform, as

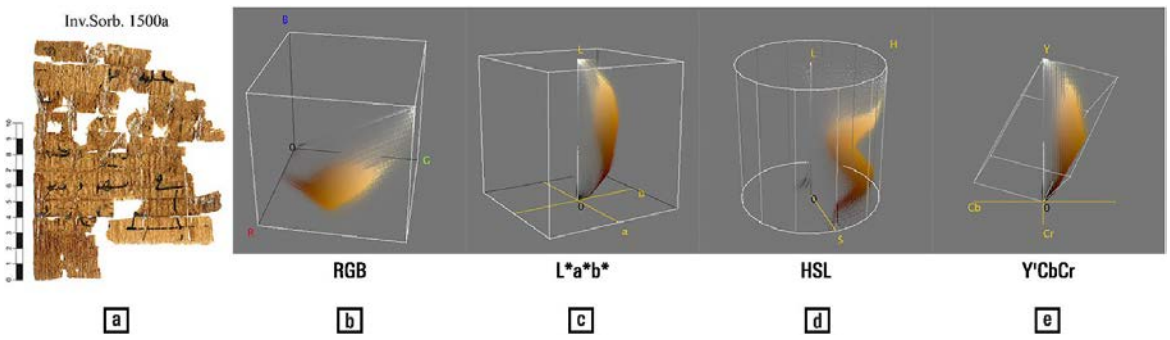


Fig. 6. (a) Papyrus image and (b–e) its colour distribution in various colour spaces. — Credits: papyrus: Inv.Sorb. 1500a, Florent Jacques, Institut de Papyrologie, Sorbonne Université, Paris; visualization: Colour Inspector 3D and ImageJ.

intended in the L*a*b* colour-space, which is obtained by non-linear mathematical transformations [[35]: 20]. Furthermore, colour distributions can acquire complex shapes, as presented in Fig. 6, and thus become difficult to manipulate mathematically.

The Y'CbCr colour space — where Y' stands for 'luma', and Cb and Cr for 'chroma', these descriptors being representative of lightness and chromaticity, respectively — was developed for television broadcasting and has been adopted for video production [40]. These are both applications in which signal transmission speed (or 'bandwidth') is critical, resulting in a reduction of the perceptually secondary chroma versus a full-quality luma. The sampling process is visible from the striped appearance of the distribution.

Note how the above colour spaces are limited to hue, saturation, brightness, while colour is characterized by many more effects: transparency, sheen, glitter, colour interactions, and so forth. These are considered in more sophisticated 'colour appearance models' (CAM) [37].

3.5. Colour gamut and shape discriminance

The volume and topology of the space of perceived colours determine shape discriminance.

Natural and artificial imaging and reproduction systems, such as eyes, cameras, and ink, are limited regarding the number of different colours to which they are sensitive, or that they can render; moreover, colourblind people are insensitive to some colours, while all of us perceive colours slightly differently, even if we are unaware of these phenomenological differences (in other words, roses are red for everybody in name alone) (on gender-based hue shift, see Refs. [41,42]). The amount of sensed or rendered colours is called a 'colour gamut', which defines a specific volume size and shape in a colour space. The larger the gamut, the more numerous, vivid, and pure will be the colours that can be sensed or rendered, with human vision having a vastly larger colour gamut than can be produced in print or on displays (Fig. 7). Colour gamuts also form part of the 'colour profiles' used to describe the overall colour characteristics of imaging and display devices, and allow conversion between them, using so-called Profile Connection Spaces [15]. Knowledge of the profile allows these devices to attribute the numerical values found in digital image files to specific colours of the gamut, and, as a consequence, in the physical world.

To maintain colour fidelity, the imaging gamut should be smaller or equal to the reproduction gamut, although this rule can sometimes be broken to produce specific effects. Fig. 8 illustrates an improvement in the legibility of the papyrus achieved by converting an sRGB gamut image to the Adobe RGB (1998) gamut. The phenomenon occurs due to an increase in saturation, which acts as a contrast enhancement in the colour domain at near-identical lightness distribution.

P7 – 138

3.6. Gamut mapping and cross-media colour constancy

We have seen that the colour gamut is a limited volume in the colour space and that different imaging and display systems have different gamut characteristics. Let us now consider what happens to images and shapes when they are transferred between distinct colour systems.

When the gamut of an image is larger than that of the target display, the gamut is shrunk to avoid the presence of colours that cannot be reproduced — this, and the converse operation of gamut expansion, is called 'gamut mapping' [43]. Intensive research has been carried out on how to best convert between gamuts for various applications. Fig. 9 illustrates shrinkage by (a) linear compression and (b) clamping out-of-bounds colours to the outer envelope of the target gamut. The first algorithm maintains the relative perceptual difference between colours, but reduces the overall saturation of the image, making it look less vivid and thus inadequate for applications where saturation is required (such as advertising). The second algorithm would preserve the colour appearance of the image if there were few out-of-bounds colours, but would drastically degrade its appearance if most colours were out of bounds.

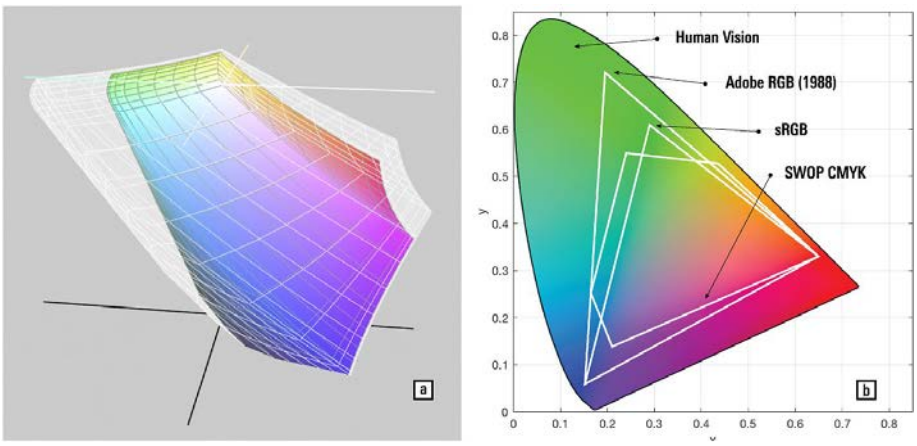


Fig. 7. (a) The coloured solid represents the sRGB colour gamut, the most common profile of digital images used by photographic cameras and exchanged over the Internet, within the wider Adobe RGB (1998) gamut (greyed solid), which is a richer gamut intended for higher photographic quality. This is a screenshot of ColoSynch, a colour profile selection tool included in Macintosh computers. (b) A cut through the CIE XYZ colour space showing the extent of four types of colour gamuts: human vision, Adobe RGB (1998), sRGB, and SWOP (Specifications for Web Offset Publications).

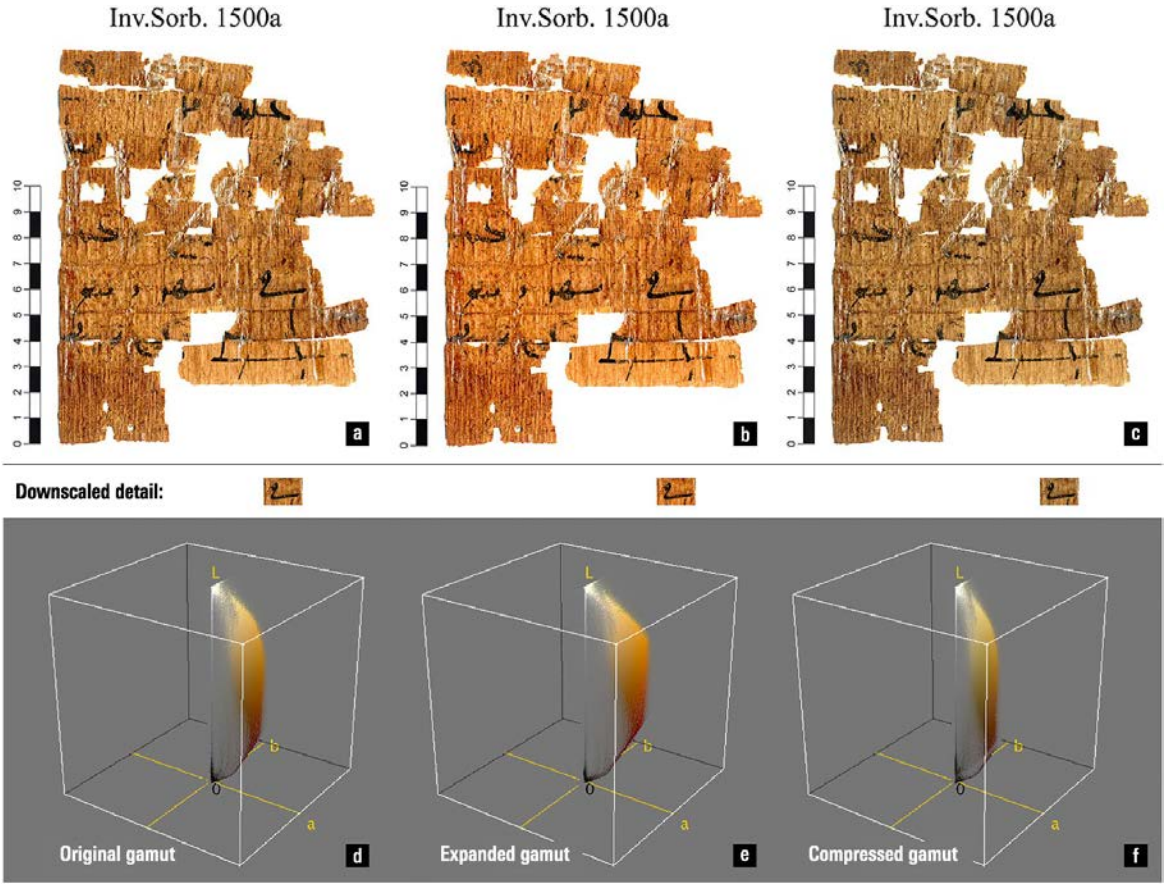


Fig. 8. A papyrus (a) imaged with the sRGB colour profile, (b) opened with the Adobe RGB (1998) colour profile, and (c) converted back to sRGB. Note the changes in saturation, colour contrast, and legibility. The image (b) has the largest gamut, strongest contrast, and best legibility; the effect is stronger for finer details, or when looking at the target from farther away. (d–f) The colour distribution of the papyrus images in the L*a*b* colour space. The shape in (f) has the largest extent, and the one in (g) the smallest. — To reproduce these images, follow the procedure outlined here [28]: The software used is Photoshop CS6; conversions use the Adobe ACE Conversion Engine and the Perceptual intent; files are saved to TIFF format, with embedded colour profile. (a) File > Open > select Assign profile: sRGB; Edit > Convert to Profile: RGB: sRGB; File > Save As; (b) File > Open > Assign profile: Adobe RGB (1988) > Convert to sRGB > Save; (c) File > Open > Assign profile: sRGB > Convert to Profile: Adobe RGB (1988) > Assign sRGB > Save. The colour distribution visualization is created with the Colour Inspector 3D plugin in the ImageJ software package. — Credits: a–c. image: Inv.Sorb. 1500a, Florent Jacques, Institut de Papyrologie, Sorbonne Université, Paris; image processing: Vlad Atanasiu.

P7 – 139

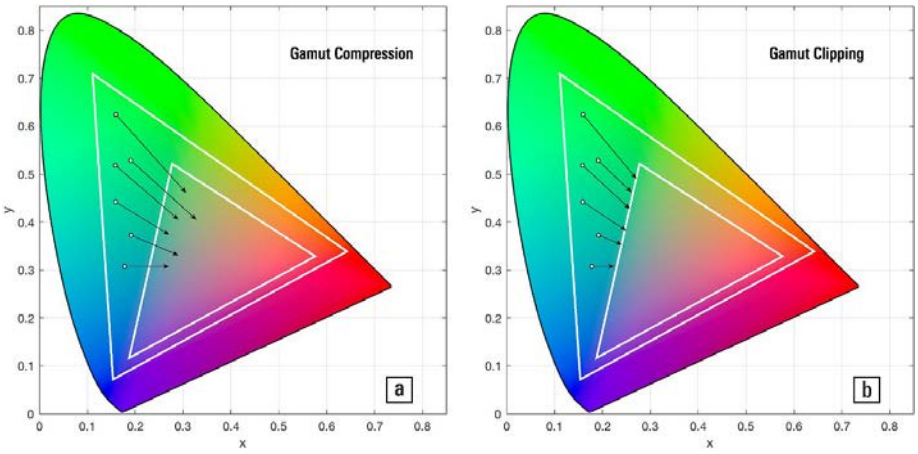


Fig. 9. Gamut mapping by (a) linear compression and (b) clipping (after [133: 34]).

A number of observations can be made at this point. First, the choice of gamut mapping algorithm, and implicitly its impact on shape integrity, is not a purely mathematical issue, but rather depends on both data characteristics and usage intent. Colour management applications typically allow users to select between methods for maintaining the fidelity of overall colour perception,

saturation, and numerical values.

Second, gamut mapping is often opaque, insofar as users can only select mappings by choosing from terms denoting general 'intents', which provide no information about how they are implemented algorithmically. These methods can and often do vary between software programmes, if not also within the same application. In Photoshop, for example, the difference between 'Assign Profile' and 'Convert to Profile' is not explicit; depending on which is used and how they are combined, the results can be inconsistent or perplexing.

Third, gamut mapping, with all the uncertainty and complexities involved, can take place without users being necessarily aware, by simply opening and saving a digital image file. For example, this occurs if the image has no colour profile specified, meaning that the software must attribute one by default in order to process the content (in such a case, the sRGB colour profile is typically assumed).

4. Methods and applications for computational colour-based shape analysis

The digital enhancement of images to improve legibility by humans and processing by machines, along with the segmentation of text from background, are routine and elementary stages of palaeographical work during which colour is often discarded and the processes are performed on grayscale images. This section illustrates how the promoted methodology of taking colour into account may yield improved results for these basic shape-related tasks by automated and manual means alike.

4.1. Colour-based image enhancement

Dropping colour information alters the ability to discriminate shapes – and vice-versa.

Two essential tasks carried out by members of the humanities disciplines in connection with documents are to read and publish them. Both of these tasks benefit from image legibility; thus, legibility enhancement is a valuable computational research area for the editing of texts. The challenge is to selectively enhance script rather than other structures while retaining sufficient non-script information to avoid misinterpretations and enable discoveries.

Fig. 10 illustrates the impact of various colour- and non-colour-based image enhancement techniques on legibility. The comparison

P7
–
140

Fig. 10. (a) Original papyrus image; (b) enhancement combining information across different colour spaces; (c) contrast-limited adaptive histogram equalisation (CLAHE) of the image's achromatic channel; (d) retinex-based enhancement. — Credits: P.Mich.inv. 1318v, University of Michigan Library; image processing by Vlad Atanasiu.

indicates that the addition of colour in the image processing procedure may lead to increased legibility (images b and d vs. image c), and that not all colour-based enhancement methods are equal (image b vs. d). Some of the other insights obtained from this specific figure also have general relevance. For example, typical papyrus images found in online databases often have low contrast and are thus difficult to read (a). Contrast-limited adaptive histogram equalisation (CLAHE) of the image's achromatic channel (i.e., lightness) visibly improves legibility, but also enhances the script and the papyrus texture in equal measure, creating a confusing interference between the two (c) [44]. By combining information across different colour spaces, better legibility may be achieved through selective script enhancement, background equalisation, and preservation of input information at low-intensity levels (b) [45]. Moreover, the fact that the technique applied here was developed by the first author specifically for papyri enhancement cautions against the uncritical use of generic enhancement methods. As a case in point, the illustrious retinex theory, developed by Edwin H. Land, the inventor of the Polaroid camera, makes sophisticated use of colour, and results in strong local contrast, although it also produces an undifferentiated enhancement of all papyrus structures, sometimes in addition to a darkening halo effect (d) (original theory: [46]; algorithm families: [47]; development examples: [48,49]). In conclusion, no single method is sufficient for all documents and applications.

4.2. *Multispectral imaging and revealing invisible shapes*

Light beyond the visible spectrum may reveal otherwise-hidden shapes.

The human visual system is sensitive only to three colour components, perceived as red, green and blue, which correspond to electromagnetic waves with peaks at 645 nm, 526 nm, and 444 nm, respectively. The full range of colours, with various wavelengths, can be generated through a linear combination of these signals. Due to the limitation of these three sensitivities, an infinite set of spectral power distributions exist that we perceive as the same colour. One example is the addition of red, green, and blue, on one hand, and cyan, magenta, and yellow, on the other hand, which both result in the perception of white. The phenomenon is called metamerism and characterises the human visual system, as well as all technical devices based on three colour components [[37]: 335–337; [39]]. While the confusion caused by metamerism may appear as an impediment to the survival of biological organism, its impact is mitigated by the low relative frequency of the phenomenon in natural scenes [50], and it forms, moreover, the basis of efficient colour vision. However, metamerism remains an impediment for the objective description of physical properties, such as ink. To overcome this limitation, capturing visual signals with more than three components is a state-of-the-art solution, known as multispectral imaging. By adding layers of non-visible wavelengths to an image, specific material properties of artefacts can also be taken into account. For example, certain chemical binders react strongly with ultraviolet light. Due to their material and chemical properties, the contrast between script and substrate can vary along the spectrum, resulting in better legibility in a specific band or mix of bands [51,52]. The science and art of the multispectral approach therefore consist in a combination of multispectral imaging, multispectral image processing, and multispectral image representation and interaction.

P7 – 141

Figure 11 reveals the differences in a papyrus inscribed with iron-gall ink, as it appears after the mixing of the wavelengths perceived as red, green, and blue (a) and in infrared, at 740 nm (b). Note that legibility can be improved simply by colouring the lightness component of the infrared image with the saturation and hue of the visible-light image, without the need for additional complex procedures, making the infrared image appear more familiar in terms of colour (c). Much better results are obtained by applying the mathematical procedure of decorrelation to the infrared lightness and visible-light red and blue bands, a process that effectively fuses pieces of a script from visible-light and infrared (d) [53–55]; colourisation of the result might also be preferred by some viewers (e).

One practical conclusion that can be drawn from this discussion on multispectral image processing concerns the power of combining existing methods to improve overall performance. In the following sections, the reader will discover more examples of this technique.

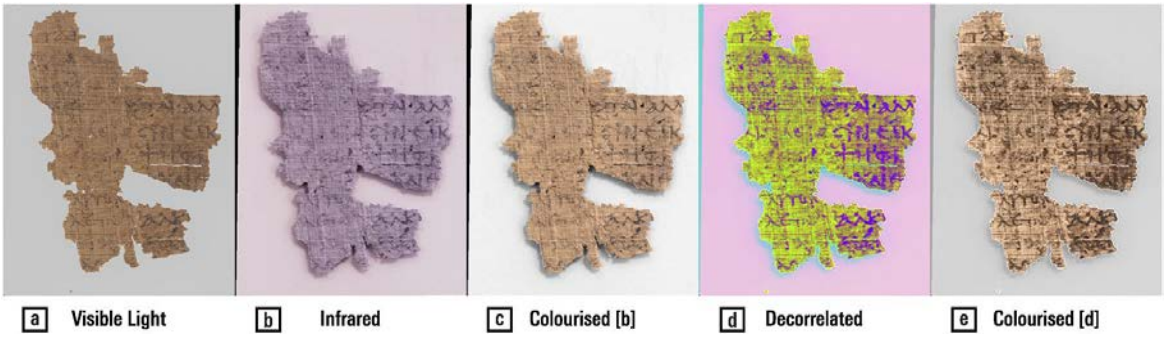


Fig. 11. Multispectral images of a papyrus inscribed with iron-gall ink: (a) visible light; (b) infrared; (c) colourised infrared lightness with visible light saturation and hue; (d) decorrelated red visible light, blue visible light, and infrared lightness; (e) colourisation of (d). — Credits: a. P.Bas. inv 27, University of Basel Library; b. Peter Fornaro; c–e. Vlad Atanasiu.

4.3. Colour-based shape segmentation

In black and white, it is difficult to distinguish shapes that may be trivial to separate in colour vision.

It is notoriously difficult to segment script from background in papyri documents, as their textured surface produces shadows of similar intensity to the script (see examples from a papyri binarization competition in Ref. [56]). However, better results can be obtained by performing the segmentation in the colour domain (Fig. 12), which provides additional useful information, i.e., saturation and hue [57]. The outcome may be further improved by integrating data from multiple features, such as colour and texture [58].

4.4. Colour-based background removal

Even apparently monochromatic shadows contain traces of colour and may thus benefit from colour processing.

To further showcase the utility of computational colour image processing, we present the task illustrated in Fig. 13, which consists in segmenting objects from the background and from each other (a). To this end, the image was, in one case, reduced to lightness and binarized using the popular Otsu method (b) [60]; in the second case, the shadows were removed by taking all colour information in the red, green, and blue channels into account, thereby producing a colour-invariant image [[61]: 50–51] (c) that is subsequently binarized (d).

Biological systems also have demonstrably more difficulty suppressing shadows in the lightness channel alone than in all channels concomitantly [62]. Consider that a colour-edge usually characterises a change in material, while a lightness-edge is ambiguous: it may be either a change in material or the edge produced by a shadow (the latter phenomenon is used for camouflage) [63,23]. In this case, too, colour improves a purely shape-oriented task, as it is the source of a different kind of information about our surroundings. The utility of colour in our experience of the world as stable under a wide range of changing illumination conditions — named 'lightness' and 'colour constancy' [64] — rather than a perpetual kaleidoscopic composition and decomposition of shapes has been invoked to explain the emergence of colour vision itself [[65,66]: 224–225].

4.5. Integrating human and machine abilities

Up to this point, we have championed the utility of automated methods for colour-based image processing — but what about interactive photo editing? Can the human outperform the machine? In which contexts and for which reasons?

The post-processing of historical images (e.g., for restoration, colouring, colour balance and superresolution) is a domain of predilection for photo-editing, sometimes with stunning results (see Ref. [67] for colouring famous black-and-white pictures, and [68] for superresolution images from NASA's Apollo programme). As Fig. 14 demonstrates, humans can achieve superior results on specific images and tasks, provided that they possess the right skills. The reasons behind this disparity include the flexible integration of various methods into workflows, know-how relative to their sequencing and parametrisation gained through lengthy experimentation, the specific enhancement of different image regions beyond the capabilities of existing locally adaptive methods, contextual semantic knowledge, application-specific reasoning, and the cognitive ability to perform effective image quality assessment.

P7 – 142

Some of these aspects, taken individually, have been thoroughly studied in image processing (compendium: [70]), and the desire for their integration has led to the development of methods fusion research [71]. Other aspects, such as binarization, remain challenging despite decades of research [45]. Faced with certain limits of automated image processing when confronted with real-world applications (document enhancement is one example), and realising their complementarity with the human aptitude for interpreting visual stimuli, researchers have progressively established the interdisciplinary field of perceptual image processing. Examples of research directions in this field include the coupling of image enhancement and image quality assessment [72], along with the

Fig. 12. (a) Colour image of a papyrus. (b) Edge detection using the Canny method for the lightness gradient [59]. (c) Applying the same method on the colour gradient. — Credits: a: Inv.Sorb. 2346, Florent Jacques, Institut de Papyrologie, Sorbonne Université, Paris.

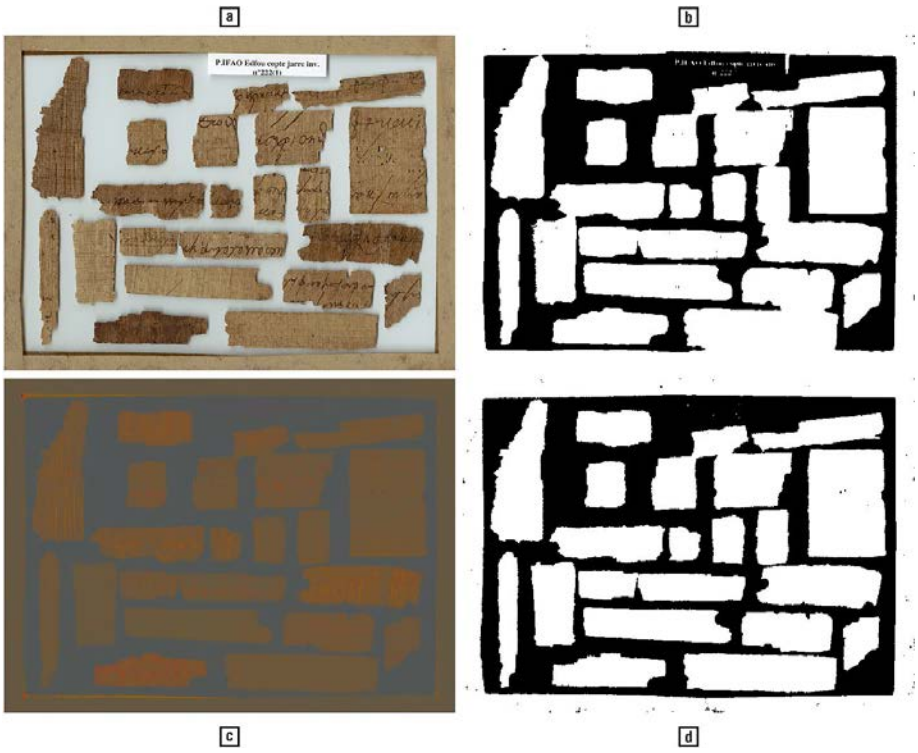


Fig. 13. (a) Colour image of papyri fragments; (b) binarization of image 'a' using the popular Otsu method; (c) colour-invariant version of image 'a'; (d) binarization of image 'c'. — Credits: Institut français d'archéologie orientale (Ifao), Cairo; d-scribes project, University of Basel; 'Edfou au VII^{e} siècle' project, Ifao.

P7
–
143

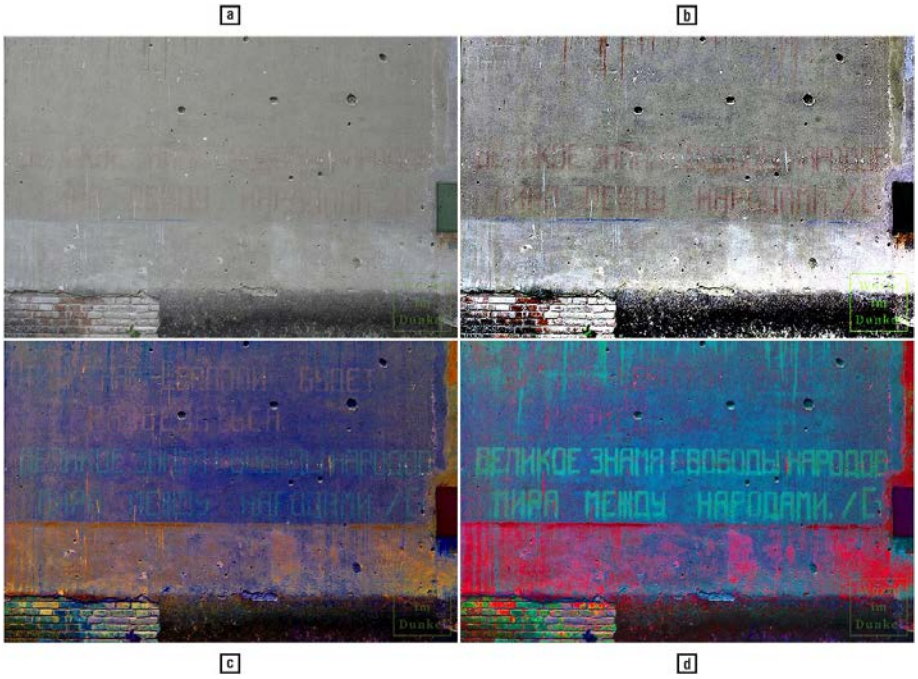


Fig. 14. (a) Unprocessed image; (b) image processed in ImageJ with one implementation of the retinex approach, which gave the best results in terms of script legibility among those tested by the authors; (c, d) image edited in Adobe Photoshop by an expert in digital inscription recovery [69]. The modified pictures are processed on the three colour channels of the input RGB image. Note that the top two lines of text are almost invisible in both the unprocessed and the retinex versions. — Credits: Russian inscription from 1945 on a mansion in Lower Austria, announcing the end of the Second World War; photographed and processed by Thomas Keplinger, https://www.worteimdunkel.at/?p=6713.

systematic exploration of leveraging perceptual phenomena for the purpose of image enhancement [73,74]. Human colour perception and computational processing is a further example, as discussed throughout this article.

It remains — and this is the 'lesson' of this section — that image processing is both science and art, especially when colour is involved.

5. Leveraging psychological factors for colour-based shape analysis

When using colour in the analysis of script, the palaeographer must contend with a host of effects that affect the perception of colour and shape in both beneficial and detrimental ways. The computer scientist, too, should be aware of such human factors insofar as the computational methods to be developed need to match the specificities of the human visual system. Moreover, synchronic and diachronic cultural variability in colour perception and naming and in colour-related technologies must also be accounted for in machine learning and automated shape classification. The case studies discussed in this section should make these points clear.

5.1. Utility of colour vision: past and present

A look at the prehistorical origins of colour vision illuminates contemporary phenomena related to the use and neglect of colour in shape analysis, particularly as affected by gender differences.

Exactly why animals developed colour vision is unclear. The extraction of information about the environment from a spectral analysis of light wave frequencies, amplitudes, and phases requires multiple photoreceptors, and the trichromatic architecture of 'opponent colours' provides an efficient and compact mechanism for this purpose [[75]: 14–18; [37]: 22]. Clear benefits for trichromats are a reduction in the ambiguity induced by shadows in respect to shape detection and identification (i.e., improve 'colour constancy'). But what do mantis shrimps do with up to fifteen types of photoreceptors, including sensitivity to ultraviolet and circularly polarized light [[76]: 35, 42–44, 235–237]? For the colour vision of primates there are various explanatory theories, mostly argued in terms of statistical properties of the visual environment, involving the detection of ripe fruits in thick foliage and avoidance of toxic plants, the primates' visual sensitivity peeking in the green-yellow band of light (Fig. 15) [[21]: 198]. This is perhaps a mechanism for the co-evolution of flora and fauna: the animals helping the spread of seeds, in a similar manner to flower-bearing plants encouraging pollination by insects attuned to the ultraviolet colour of their petals, to the point that their reliance for navigation and feeding may be even greater on colour vision than shape discriminance (overviews: [[77–79]: 112–119]; ecological factors: [80]; fruit theory: [81,82]; ultraviolet vision: [83]).

P7 – 144

As urban humans no longer need to pick berries for survival, they may afford a higher amount of colourblindness than rural dwellers. Urban humans, however, still benefit from colour vision for intellectual work. The pictures in Fig. 16 show the extensive usage of colour for bookmarking and highlighting documents, as practiced by law students at the University of Fribourg, Switzerland. These markings make the documents browsable for navigational purposes on the basis of information encoded in colour alone, just as riffling green leaves to look for red berries. They also question the notion that outlines are sufficient for object recognition [[75]: 8, 121–138].

Remarkably, the importance of colour for safety has persisted in the transition from the natural to the human-made environment, even becoming ubiquitous in conjunction with shape-based markers (text, symbols, and patterns) applied on objects in areas as diverse as pharmaceutics (distinguishing drugs in hospitals [84]), transportation (the yellow school bus), construction (green fluorescent jackets), and electronics (colour-coded specifications).

In general, the number of colours used in drawing tasks is higher for females than males [85], reflecting a broader spectrum of gender-related evolutionary characteristics of female chromatic perception, such as faster, richer, and more accurate colour naming ability [[86]: 320], better discrimination in the middle of the visible light spectrum than males [41], as well as higher rates of being genetically potential tetrachromats [[87]: 200, 206; [88]], and thus endowed with finer chromatic sensitivity.

By comparison, males have increased spatial awareness as opposed to vividness of object imagery [[89]: 648; [90]], more accurate depth vision [91], inter alia due to a narrower visual field [[92] (1): 106–107], and faster motion perception [41,93,94]. All these are advantages for males to catch prey and hide from ferocious beasts in the savannas of the hominids, but hardly of utility for deciphering the Rosetta Stone and other palaeographical artefacts (Fig. 17).

Fig. 15. (a) How many ripe berries will you eat at breakfast using monochromatic vision? Answer: perhaps one. (b) Using dichromatic vision? Also one. (c) And using trichromatic vision? At least three! — Credits: picture: Vlad Atanasiu; berries: Lower Austria.

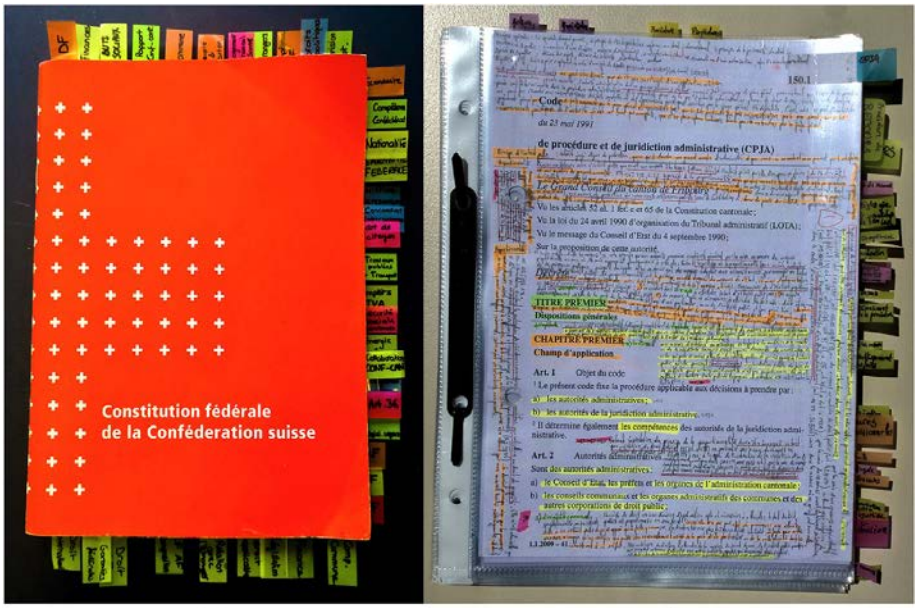


Fig. 16. Colour vision and academic achievement: (a) colour-coded bookmarking and (b) highlighting by university students. Note that the number of hues used is higher than the basic four opponent colours, one female being observed using a twenty-three-colours highlighter megapack. From an ergonomics perspective, there are a few remarkable points suggested by these images: some tasks and methods are apparently more effectively approached using colour (structuring a black-and-white document), some individuals and groups are more able to discriminate between closely related colours, and the benefits of using colour are greater than the effort of managing a large array of colouring tools. — Credits: Vlad Atanasiu.

Fig. 17. Motion perception is indispensable for detecting well-camouflaged and fast-moving targets; colour vision, however, is less helpful for tasks such as locating a snow leopard in its natural habitat. Picture (a) is in greyscale, and picture (c) is binarized, to mimic winter conditions). — Credits: Snow leopard (*Panthera uncia*), Basel Zoo, 2020; picture: Vlad Atanasiu.

Considering the social specialisation of males as hunters and females as foragers, in conjunction with the opportunities to easily hide and to observe open spaces afforded by primeval human habitats, and supported by archaeological and palaeontological data, psychophysical findings such as those listed above have formed the basis of the 'hunter–gatherer ecological hypothesis' of the evolution of sexual dimorphisms of human vision [91,95] (but woman are also participating in hunting [96], and males are also painting in colour). This conjecture should acquire a stronger significance for the reader if we revolve the historical perspective towards

P7
–
145

implications for the present. In other words, contemporary scientists may wish to consider the impact on their interests and work methods of perpetuating traits evolved in prehistoric cultural and environmental contexts. As a practical example, the interpretation of inscriptions with low legibility could be carried out collaboratively, to profit from the complementarity of visual abilities, while image enhancement methods could be developed with personalisation in mind, just as reading glasses are not made for the average viewer [32].

5.2. Colour as facilitator in shape analysis

The detection and localization of shapes in the visual field is an example of a task considerably facilitated by colour; for synaesthetes, this holds true even in the absence of physical colours.

During early childhood, humans display a preference for colour over shape when classifying stimuli [[97]: 150]. While this propensity diminishes with age, it does not disappear entirely. Typically, the gaze more easily detects objects with a stronger lightness contrast; at equal lightness, however, objects can appear to 'pop out' from their surroundings if they contrast in hue or saturation, a 'law' of Gestalt psychology called 'colour saliency' [98,99]. Fig. 18 illustrates this phenomenon, showing how shape identification and texture segmentation is influenced by colour.

The mental colouring of numbers, letters, and other shapes in the absence of physical colouring — and thus their faster classification when similarly coloured — occurs for synaesthetes, those people endowed with cross-modal perception, such as experiencing shapes as colours or sounds as flavours [101,100,102]. In the sonnet *Voyelles,* the French poet Arthur Rimbaud famously described vowels as being coloured, an example of sound-as-colour cross-modal perception: 'black A, white E, red I, green U, blue O vowels/one day I'll tell your ineffable birth' (it has been suggested that the order of vowels in this poem follows that in the evolution of colour naming; [103]).

5.3. Colour as complication in shape analysis

In horology, the term 'complication' designates 'any feature of a mechanical timepiece beyond the display of hours, minutes, and seconds', such as a calendar, or, on certain smartwatches, heart rate [104]. By analogy, we may say that colour has many 'complications': it is a thicket of interacting phenomena and surprising implications. Colourblindness and subjective colours are two of the more distinctive in terms of shape perception and cognition.

When persons with red-green colourblindness, one form of colour deficiency of varying intensity, look at Fig. 19 a, they read the number '3', while normally sighted people see an '8'; those who are totally colourblind see no number at all (plates of the so-called Ishihara colourblindness test: [[105]: 4]; history of colourblindess tests: [106]).

P7 – 146

The mechanism at work in the stimuli of the so-called Ishihara tests of colourblindness, is worth our attention, since it combines at least three distinct visual effects: chromatic masking, perceptual grouping, and edge smoothing. First, the red dots belonging to the digit '8', but not to the digit '3', result in the same green as the background when projected on the green–red colour dimension perceived by green–red colourblind persons. Second, there exists a cluster of red dots that are common to both '8' and '3', whose integration forms the shape '8'. Third, the random size and placement of dots ensures that the joints between '8' and '3' are inconspicuous (similar to how typographical stochastic halftoning makes prints appear smooth, by randomizing phases in the frequency domain representation of images [107,108).

At physiological level, the phenomenon illustrated here is due to the lack of red-sensitive pigments in the eye's photoreceptor cells [[109,110]: 35–37]. Depending on whether the red, green, or blue cones are affected, the condition is known as protanopia, deuteranopia, or tritanopia. Colourblindness affects about 8 % of males, and less than 1 % of women of North European ancestry, with lesser percentages in other demographics; for physiological reasons (such as the yellowing of the cornea), colour deficiencies also vary with age [[111]: 222–223; [40] (1): 388–390; [110]: 38].

The prevalence of males in computer sciences — the percentage of bachelor degrees awarded to males in computer science in the United States dropped from 85 % in 1966 to 63 % in 1984, and rebounded to 82 % in 2008 [[28]: 105–106] — has as consequence a higher average of colourblind individuals among this demographic than in the general population, a situation worth investigation for

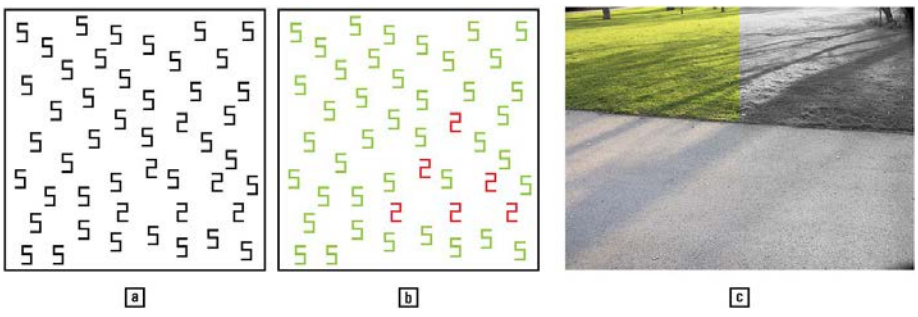


Fig. 18. (a) A mix of '2' and '5' black figures, with the '2's arranged in a triangle. (b) Differentiation between '2's and '5's; detection of the triangle becomes much easier for both humans and machines when the figures are coloured. (c) Colour also considerably facilitates the automatic segmentation of the image into 'grass' and 'pavement' areas. — Credits: a, b. after [100]; c. Vlad Atanasiu.

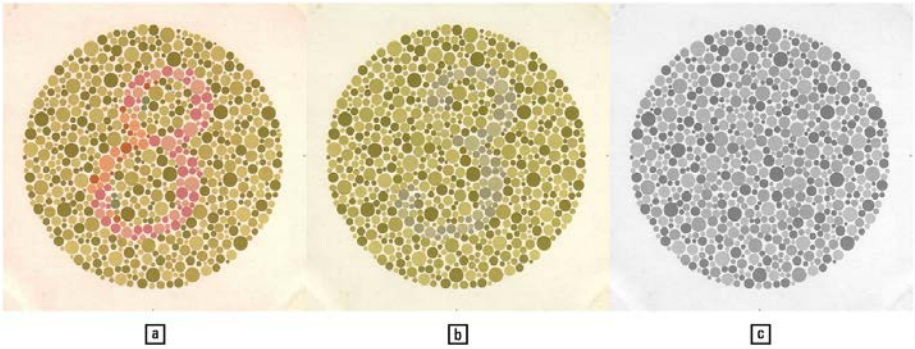


Fig. 19. Colours create shapes — (a) Normal-sighted persons read the figure '8' in this colourblindness test plate, red–green deficient persons read '3', and monochromats see no shape at all. (b, c) These two plates are image processed variants of plate (a), simulating for the benefit of normal-sighted persons the red–green deficiency and the monochromatic vision, respectively; specifically, a faint '3' should be seen in plate (b), and no digit in plate (c). The strength of the effect varies with the viewing distance. — Picture (b) is obtained by converting the RGB image to the L*a*b* colour space, and setting the values of the a* image channel, i.e., the green–red colour space axis, to zero, while picture (c) is obtained by applying the same procedure to the b*-channel, i.e., the blue–yellow channel, as well. The '3' in picture (b) is difficult to distinguish, which illustrates that the processes involved in colourblindness are more complex than can be simulated by even a refined colour appearance model such as L*a*b*. — Credits: [Ref. [105]: plate 2].

elucidating whether it is related in any way to a neglect of colour as a computational analysis feature in favour of shape. However, it can also be noted that colour-anomalous observers have the benefits of superior visual acuity and contrast sensitivity [112]. Likewise, it is yet to be determined whether the current trend in the feminisation of papyrology — about 56 % of participants at the 2019 International Congress of Papyrology were women [113] compared to only 5 % in 1937 [114] — will lead to a heightened interest in colour among scholars, and thus help advance the field.

The colourblindness 'deficiency' is a valuable ability for applications in steganalysis and military camouflage detection, as objects and patterns that are invisible to normal-sighted people might be detected by the colourblind [[115]; 86: 106]. There are implications for palaeographers, archaeologists, and computer scientists as well. Colourblind researchers may seek to use software packages that aid the analysis of coloured artefacts (e.g., Ref. [116]. Also, it is possible that the chromatic characteristics of artefacts change under the influence of natural or artificial factors, creating spurious shapes. It is therefore recommended that the inspection of colour images by normal-sighted persons includes the isolation of the green–red and blue–yellow channels to detect potential alterations.

The effects of colourblindness, such as revealed through the Ishihara tests, are a striking demonstration of shape identification being incontrovertibly contingent on colour. More generally, colour vision plays an important role in the construction and deconstruction of shape, as attested (with existential implications) by the use of camouflage among animals and humans [117,118,119] and a host of other perceptual processes, such as the determination of texture from colour [120]. The basic principle is that shape is determined by edges and edges by contrast, which can occur due to a difference in lightness or saturation or hue. If these examples show that colour can affect shape, the reverse is also true: shape can induce the perception of subjective colours [[92] (1): 368–369]. The detection of optical illusions [24,121] in digital imagery should thus provide computer scientists with an intellectually rewarding research field.

To substantiate the well-founded nature of our entreatment, let us consider the diversity of ramifications resulting from studying the subjective colours perceivable in Fig. 20. Illusory pastel-coloured patches emerge from this achromatic hatched stimulus if the eyes drift slowly over the striped texture. Hatching is a technique used in graphic arts to simulate grey levels by means of black and white, in engraving, cartography, technical drawing, and as 'guilloche' interlaces (used e.g., as security features for banknotes). Such was the level of mastery attained in hatching during the Renaissance that the humanist Erasmus considered colour as almost superfluous: 'Dürer … could express absolutely everything in monochrome, that is in black and white … And so alive is it to the eye that if you were to add colour you would spoil the effect.' [[122]: 399] Five hundred years later, the same reticence towards colour and the belief in the

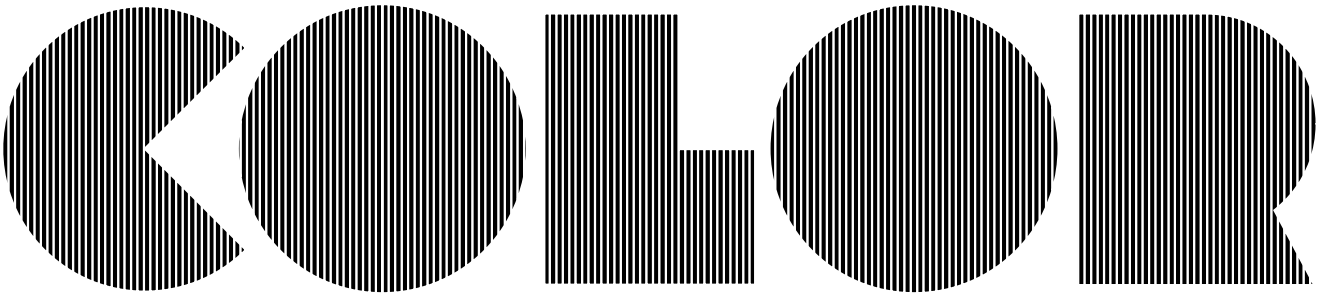


Fig. 20. By drifting the eyes slowly over the striped texture, the observer should perceive pastel-coloured patches emerge from the black and white hatching. — Credits: composition: Vlad Atanasiu; typeface: Lot, by Svetoslav Simov.

ability of achromatic stimuli to simulate it — or at least to simulate the emotions colour evokes — made many great photographers and film directors wary of adopting colour film. An explicit link between hatching and colour in Western cultures is provided by its use in heraldry, where, for example, vertical hatching stands for gules (red), horizontals for azure (blue), and left diagonals for vert (green) [123,124]. Elsewhere, black-and-white patterned pottery of the pre-Columbian Mimbres Native American culture has been found to produce the same subjective colours when spun as the test disks of modern psychologists studying these so-called Fechner colours or Bentham illusion [125].

5.4. Challenges to benefiting from colour in shape analysis

Colour science, with its great variety of interdependent branches, presents researchers with the concomitantly daunting and exciting challenge of becoming knowledgeable of domains outside their specializations. Two of these domains, namely historical and cultural, are discussed in the following in the context of shape analysis as impacted by the colours of writing substrates, by colour naming (the linguistic dimension of colour science), and by display hardware.

As the Enlightenment scientist, philosopher, and novelist Margaret Cavendish once remarked, 'no body is without colour' [[126]: 307]; likewise, the history of script cannot be decoupled from that of the writing substrate. It suffices to consider the three-dimensionality of the majestic Roman epigraphic styles (the shadows shifting within its furrows with each passing cloud and the diurnal movement of the sun) or reflect on how, in the bureaucracies of the tenth century Middle East, handwriting styles were developed that were specific to various paper sizes [127] and, during the late twentieth century, a cornucopia of pixelated digital typefaces emerged due to the low resolution of displays and printers [128]. It would however be naive to consider colour as the primary classification criterion for substrates, as the aforementioned examples already suggest. The chromatic contrast between script and background, however important to legibility (and, thus, to the very function of writing), was often secondary to other considerations regarding script. Looking at major types of modern paper, the difference between coated and uncoated paper is one of shininess, of script-outline sharpness (due to the extent of ink absorption by the substrate), and of the velvet feel of the surface; recycled paper is not about low saturation, but about environmental issues; acid-free paper is not about the loss of whiteness, but about long-term document conservation; handmade paper has more to do with social status and nostalgia than a furtive veil of sepia colour; and the quest for increasingly white paper, observed throughout the history of Western papermaking, is interpreted as the negation of colour by desaturation and the achievement of an achromatic surface to satisfy chromophobic socio-cultural tendencies related to ideologies, race, ethnicity, class, and gender (overview: [129]; colour, computing, and ideologies: [130]; ethnography: [131]; paper colour, gender, and race: [[132]: 125–156]; early colour science and race: [[133]: 292–293]).

P7 – 148

Human subjects research is often impacted by a lack of practical access to participants (due to e.g., financial, time, and setup complexity constraints), a limited absolute number of cases (e.g., medical conditions), and, at times, a scientific culture of complacency; this commonly results in samples of experimental participants that are too small and/or insufficiently representative, thus reducing the statistical power of the results [134]. Colour science is no exception: the foundationally important CIEXYZ 1931 'standard colorimetric observer' was derived by averaging data obtained from only seventeen Britons [[135]: 189], while a formula for simulating colourblindness for the normal-sighted commonly used in computer science was developed from studies of two (sic) French people [[136]: 2650]. Furthermore, standards such as CIEXYZ 1931 colour space and the CIE 1939 definition of colour are 'cultural developments', guided by the input of various individuals, commissions, and national interests in addition to scientific research [[135]: 159–190]. To take an example, the reliance on lightness, saturation, and hue in contemporary colour science for defining colour spaces and studying colour naming is 'a completely arbitrary one', in the own words of the physiologist Johannes von Kries who was instrumental in propagating this colour model at the end of the nineteenth century [[137]: 129–132]. However, its success and continued use, are due to it being conceptually sound in terms of the physics of light and materials, and the utility of its mathematical modelling, as well as the weight of the social status of physicians and engineers within the 1930s CIE, in addition to the historical context that favoured these three colour appearance modes in Western cultures and the emergence of modern colour science within it [138,139].

Just as important as the scientific validity in colour research is the fact that human colour perception is a variable and evolving phenomenon. In this sense, for example, the study of the colour vocabulary of the ancient Greeks reveals that they were attracted not so much by hue and saturation, as in the modern Western world, but by other colour aspects such as lightness and shimmer [140,141], while the Polynesian people of the Bellona Island seem little interested in colour as an ontological category, preferring object categorization by patterns and utility [142]. More generally, it has been observed that industrialization (i.e., the complex of the chemical colour industry, colour science, standardization, and marketing) favours consistent colour naming [[143]: 10788]. Certainly, it has also contributed to augment the colourfulness of the modern world, and thereby impacting the development of colour cognition: in the span of the twentieth century the acquisition by Western children of the ability to correctly use the names of the four basic opponent colours red, blue, green, and yellow, has decreased from about eight to about four years [[75]: 149–150]. Not only are children immersed in an environment that has become chromatically richer, but toys such Lego bricks and colour crayons allow colour-focused object manipulation and contribute to develop colour as an expressive modality closely integrated with shape. Increased urbanization and intercontinental migration engender the chromatic acculturation and evolution explaining such cases as the westward spread of words for the colour 'turqouise' from the Middle East (with mines in the Sinai Peninsula, Iran, and Afghanistan), where it is a quintessential colour of Egyptian, Persian, and Ottoman arts (as the name implies: the 'Turkish' colour); as well the more frequent use of the terms 'brown' and 'orange' by Japanese migrating from countryside to Tokyo and from Japan to Germany [[79]: 141–151].

The relationship between colour and shape is also ingrained in how humans not only name objects, but think through language. Consider the metonymic use of colour terms to speak about the visibility of shapes, as in 'pitch-black night' (invisible shapes), 'white

nights' (visible shapes), 'snow white-out' (shapeless), 'all cats are grey at night' (indistinct shapes), and 'blue hour' (ambiguous shapes) [144]. That the replacement of colour words with shape words is rarer, such as the hypothetical and poetical 'mellow pink' (marshmallows) and 'sharp blue' (ice), reveals an important aspect in the discussion of colour-based shape analysis: colour cognition has evolved in humans more recently than shape cognition [[145]: 162].

From an applications point of view (e.g., for interface design), linguistic colour categories can speed up colour differentiation, and thus potentially coloured-shape identification: light blue and dark blue targets have been classified faster by Russian speakers who have two distinct words for them, 'sinij' and 'goluboj', and no overall term for blue, than by English speakers, who possess only one basic colour term for the blue range [146,147].

Circumstances such as discussed above add to the argument that it would be misleading to apply contemporary scientific concepts of colour to the digital processing of documents from the past, or non-Western or non-English-speaking societies, or to employ a distinct linguistic, cultural, or professional group of people to label colours for machine learning when the resulting algorithm is to be used by specific, mixed, or global demographics [148]. (On colour naming, see the original study of Berlin and Kay [149], its update in Ref. [52], and the overview of [23]; on the debate on the universality versus relativity of colour cognition, see e.g., Refs. [150,151,152, 153,147,154].)

Limited cultural awareness and historical knowledge are not the only impediments to benefitting from colour for shape analysis: modern colour capture, description, and display technology promote certain aspects and concepts of colours over others. A flat, light-emitting, and rigid computer display is ill-fitted to suitably reproduce the colour 'oinops', Homer's word for the liquid glimmer of the ocean's moving surface that evokes wine in a drinking cup [[141]; [155]: 25–40; [156]]. If our technologists shared Homer's interests, our modern colour technologies might be very different to those we have today. Instead, photographic film employed for portraits was predominantly calibrated using 'white' models ('Shirleys', or, in cinema, 'China Girls'), meaning that other and especially darker skin tones appeared with poor shape detail and contrast with the background; this issue was not so much technical as the result of sociocultural inequalities [157,158,159].

6. Impact of changing social attitudes on colour use by scholars and scientists

Having arrived at the end of this article, let us step back from the extensive technical data provided herein and conclude with a few reflections on the changing roles of colour in society and the sciences, since the social and cultural environments are co-determinant factors in how colour is perceived as an aid in shape analysis.

A long-forgotten art-theoretical debate of the Italian Renaissance concerned the question of the primacy of shape ('disegno', drawing shape outlines) or colour ('colore', brushing to cover surfaces), which stood for meaning and affect respectively, and was decided in favour of shape [[130]: 29]. The argument has been rekindled repeatedly in the history of the arts, sciences, and technologies. Although it has been well established that Antique Greco-Roman sculptures were painted in vivid colours rather than white, museum-goers remain so conditioned to the canon of an 'ivory' art that they find it difficult to consider the original polychromy anything but 'garish' [160]. Following the invention of colour photography, its merits in respect to black and white photography were vehemently contested, with the latter retaining a more 'noble' status. Examples include the monochromatic work of Ansel Adams, Henri Cartier-Bresson, Richard Avedon and other great artists (in the words of Cartier-Bresson: 'black and white transports, abstracts, and is not "normal". Reality is a chaotic deluge […]: why then bother with colour?' [[161]: 155–156]); consider also the continued production of monochromatic cameras for art photography, such as the high-end digital Monochrom Leica (its marketing slogan: 'Discover more in black and white'). Reflecting recently on the status of colour studies, the eminent French historian of colour Michel Pastoureau still considers that 'when you are a historian of painting, focusing your academic work on the issue of colour is handicapping yourself, at least in France' [162] — as if the terms *history* and *colour* are antithetical, as if we tend to think of the past as a realm of Piranesian grayscale shadows, a new achromatic trick of human memory.

P7 – 149

These examples of chromophobia in Western societies across time and place also have their manifestations in the sciences, notwithstanding Goethe, who valued the results of his research on colour more highly than his entire literary output [[21]: 115–119]. In respect to the pace of scientific progress, topics of interest, and the scientists' attitudes, it has been noted that 'the quantitative measurement of light intensity was not commonplace until the 1930s […] long after quantification had become central to other fields of science', [[135]: 3], while until the end of the twentieth century, colour was secondary to the study of shape in psychology, with many psychophysical experiments still being conducted on greyscale stimuli [[24]: xx]. In medical radiography, black-and-white reproduction remains ubiquitous, despite the presence of alternatives (blue-tinted X-ray films have been commercially available since 1933 [[25]: 138]). The dominant 'Colourbook theory' considered colour as almost a frivolous act of filling in the space between edges after the all-important shape segmentation had been completed by more fundamental, and evolutionarily older, neural processes, causing one researcher to exclaim in frustration: 'Colours are surely not just ornamental baubles, installed in us by Providence to soften the harsh realities of life and give employment to painters, interior decorators and crayon manufacturers' [[23]: 371]. Colours were also seen as properties of objects, as opposed to the anti-realist perspective, according to which they are constructs of the brain, and thus not reducible to purely physical investigation. Because 'human factors' are inherently multifactorial, and difficult to measure, reproduce, and automate, this dichotomy was a source of dissension between psychologists and physicists during the early CIE standardisation work on colorimetry in the 1930s [[135]: 130–132]. In respect to the cognitive paradigms of another discipline, you can do mathematics with shapes, but hardly with colours (geometry and arithmetic are the oldest areas of mathematics, while the more limited-in-scope chromatic graph theory emerged only during the nineteenth century; interestingly, it did so as a solution domain for solving the shape-related mapmaking problem known as the Four-Colour Problem [1852]. This was contemporary with the fascinating work by Oliver Byrne, a British surveyor, on representing the demonstrations in Euclid's *Elements* by means of coloured diagrams [9];

the connexion of these events to cartography indicate a discipline that entertains a traditionally strong interest in using the interaction of shape and colour as representation modalities, with the refinement of shape-from-texture-and-colour techniques for simulating topography by relief shading being one example [163,66,164].).

The social upheavals of the 1960s were accompanied by a firework of colours, from 'flower power' politics to the psychedelic colour worlds produced by drug consumption, no doubt in conjunction with technologies such as colour photography, cinema, and television [130]. Even the Apollo missions to the moon generated interest in colour and shape perception, as astronauts reported the lunar exterior to be 'a grey surface modified by weak-to-strong overtones of green, brown, or blue', quite difficult to describe and very different from our concept of an achromatic environment, an otherworldly experience further enhanced by the extreme sharp details up to the horizon enabled by the absence of atmosphere [68].

Soon, the sciences took this 'colour turn' themselves: with the formulation of quantum chromodynamics in the 1970s, physicists began to use trichromatic additive and subtractive colours as a 'chromogrammatic' tool to think and communicate by words and colouring about the interaction between quarks and gluons (to the regrets of Richard Feynman, who had preferred 'wonderful Greek names' [[79]: 3]), while signal processing engineers had an established and growing chromatic naming tradition for noise types – attested since 1943, 'white noise' has since entered common language (tellingly, by relating colour to noise, colour is again construed as something undesirable) [[165–167]: 'white noise']. A veritable 'colourmania' took hold of scientists attracted to chromatic colourmaps, such as the 'rainbow' one, when colour displays and graphic cards became widely available. Colour also permeated popular science: weren't the kids of the 1980s fascinated by growing multicolour fractals on their home computers and squinting their eyes at colourful random-dot stereograms? In psychology, research shifted from the study of colour and shape in isolation to their integration. The theories of embodiment (i.e., the mind is shaped by the specificities of a given body), enaction (i.e., the mind is shaped by interaction with the environment), and ecology (i.e., perception and cognition are predictively driven by statistics of various aspects of the environment) converged to give rise to new theories of colour, such as adverbialism, which posits that colour is the manifestation of interactions between environment, body, and mind [[168,169]: 131–158]. The psychophysics of low-level vision is identifying an increasingly long list of phenomena attesting to the utility of colour per se and its interaction with shape and other representations of the world [19]. In computer science, too, colour has become since some time already routinely used in conjunction with other features; for example, researchers proposed twenty years ago the integration of object colour, shape, and semantics for the navigation of digital image libraries [170], an approach subsequently implemented in general public applications such as Google Image, which offers a search-by-colour feature. Such developments must be appreciated within the context of the broader trend towards integrative technologies, e.g., ensemble methods that fuse the classification results obtained by different methods [171], and multibiometrics, which integrates evidence across sensors and modalities [172].

P7 – 150

Summing up the trends described above, we observe an increase in the performance of humans in both biological-based and technological-assisted extraction of colour information from light, in the natural environment for survival (ripe fruits), and in the human-made environment for social advantage in terms of status (the imperial purple and the blue jeans) and information management capabilities (fluorescent highlighters and digital colourmaps).

The practice of script analysis by humans and machines, however, remains dominated by a focus on shape and an almost pathological rejection of colour, as exemplified by the systematic monochromatic reproduction of script samples whenever it is possible to discard the background, as well as by the binarization of images prior to computational analysis. Time and accidents, however, have ensured that real documents are seldom perfectly achromatic, and that even otherwise achromatic digital documents may contain unintended colour pixels. Besides the historical and forensic information that these colour artefacts represent, they also provide additional and useful information supporting the goals of preprocessing and script analysis. This article has presented selected case studies from along the chain of image formation, with the goal of convincing the reader of the utility of colour for shape analysis. Encouragingly, we already see numerous indicators to suggest increasing interest in colour among the disciplines involved in one way or another with script and other shapes. As a case in point, even typography, the 'black art', today wears a brighter and more colourful face, aided by a modern digital technology known as ' ' [71,173].

7. Conclusions

Considering, on the one hand, the lack of familiarity with colour science among scholars and computer scientists and, on the other hand, the potential of colour for shape analysis, this article has endeavoured to improve the state of colour science literacy by raising awareness of its benefits and contributing a 'nutshell' introduction to colour science, which is illustrated with case studies from the domain of script analysis and covers management, computational, psychological, and social colour aspects.

While each section title and introductory paragraph points toward specific rationales for colour-based shape analysis, we summarize here the main lessons:

- there exists a need to improve colour science literacy;
- colour has many potential benefits for analysing shapes;
- colour science is complex, constituted of many interacting disciplines;
- data fidelity in reproducing, formatting, converting, and displaying images, along with task-oriented solutions, are necessary for benefitting from colour;
- even simple computational methods are helpful in a variety of applications;

- psychological factors in colour processing ought not to be neglected and can be beneficially leveraged, often demanding only limited resources;
- sociocultural attitudes to colour unobtrusively influence its use.

Looking up from script towards the stars, astronomers give us a persuasive ultimate rationale for the utility of colour images, referring to the inference of star temperature from their colour: 'colour is knowledge' [[174]: 34].

Funding sources

This work was supported by the Swiss National Science Foundation, project n∘ PZ00P1_174149, 'Reuniting fragments, identifying scribes and characterizing scripts: the digital palaeography of Greek and Coptic papyri'.

Data availability statement

No data was used for the research described in this article.

CRediT authorship contribution statement

Vlad Atanasiu: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Peter Fornaro:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

To colourblind Irvin Schick, who collects Ottoman calligraphy because it is achromatic; to weak-synaesthete Bernice Rogowitz, who likes colours because she sees shapeless numbers in them.

References

Note: Asterisks, ' * ', mark particularly useful references.

[1] F. Déroche, A. with Berthier, M.-G. Guesdon, B. Guineau, F. Richard, A. Vernay-Nouri, J. Vezin, M.I. Waley, *Islamic Codicology: an Introduction to the Study of Manuscripts in Arabic Script*, Al-Furqan Islamic Heritage Foundation, London, 2006.
[2] Sh Fu, G.D. Lowry, A. Yonemura, From Concept to Context: Approaches to Asian and Islamic Calligraphy, Smithsonian Institution, Washington, DC, 1986.
[3] H. Massoudy, Le Chemin D'un Calligraphe, Phébus, Paris, 1991.
[4] B. Cendrars, *La prose du Transsibérien et de la petite Jehanne de France*, Beinecke Rare Book Library, New Haven, CN, 2009.
[5] J. Le Rider, Les couleurs et les mots, Presses Universitaires de France, Paris, 1997.
[6] É. Ripoll, Penser la couleur en littérature : Explorations romanesques des Lumières au réalisme, Garnier, Paris, 2018.
[7] Th Sauer, Notations 21, Mark Batty Publisher, New York, NY, 2009.
[8] M.-H. Caraës, N. Marchand-Zanartu, *Images de pensée*, Réunion des Musées Nationaux, Paris, 2011.
[9] O. Byrne, *The First Six Books of the Elements of Euclid in Which Coloured Diagrams and Symbols Are Used Instead of Letters for the Greater Ease of Learners*, William Pickering, London, 1847. https://archive.org/details/firstsixbooksofe00byrn.
[10] V. Atanasiu, Le Retroencrage: analyse du ductus des écritures d'après le dégradé du coloris de l'encre, Gazette du Livre Medieval 37 (2000) 34–42.
[11] R. L. van Renesse, Optical Document Security, Artech House, Boston, MA, 2005 (An inspiring book for understanding the great variety of colour and shape interactions from the very specialized perspective of the design of optical document security devices, such as banknotes and passports).
[12] A. Rodney, Color Management for Photographers: Hands on Techniques for Photoshop Users, Focal Press, Burlington, MA, 2005 (A practical 'how-to' guide to managing colour in digital images).
[13] E.J. Giorgianni, Th E. Madden, Digital Color Management: Encoding Solutions, second ed., John Wiley & Sons, Chichester, 2008 (On colour management in digital workflows).
[14] A. Sharma, Understanding Color Management, second ed., John Wiley & Sons, Hoboken, NJ, 2018 (A technical view on colour management).
[15] Ph Green (Ed.), Color Management: Understanding and Using ICC Profiles, John Wiley & Sons, Chichester, 2010 (On the use of standardised profiles for colour specification, interchange, conversion, printing, and displaying).
[16] Ch Poynton, The Rehabilitation of Gamma. Proceedings of the SPIE/IS&T Conference 3299, Human Vision and Electronic Imaging III, 26–30 January 1998, 1998, pp. 232–249. San Jose, CA, USA.
[17] R.S. Berns, *Principles of Color Technology*, fourth ed., John Wiley & Sons, Hoboken, NJ, 2019.
[18] E. Reinhard, E.A. Khan, A.O. Akyüz, G. Johnson, Color Imaging: Fundamentals and Applications, Wellesley, MA: A. K. Peters, 2008 (Covers the entire process of image formation, focusing on various aspects of imaging including high-dynamic range imaging).
[19] S.K. Shevell (Ed.), The Science of Color, second ed., Elsevier/Optical Society of America, Oxford, 2003 (An introduction to colour science).
[20] G. Wyszecki, W.S. Stiles, Color Science: Concepts and Methods, Quantitative Data and Formulae, second ed., John Wiley & Sons, Hoboken, NJ, 2000 (The reference for data and formulae on colour perception).
[21] R.A. Crone, *A History of Color: the Evolution of Theories of Lights and Color*, Kluwer, Dordrecht, 1999. (A handy reference about the history of colour science.).
[22] D.D. Hoffman, Conscious realism and the mind-body problem, Mind Matter 6 (1) (2008) 87–121.
[23] C.L. Hardin, *Color for Philosophers: Unweaving the Rainbow*. Indianapolis, Hackett Publishing Company, 1988 (Philosophical aspects of colour).

[24] A. Shapiro, D. Todorović (Eds.), The Oxford Compendium of Visual Illusions, Oxford University Press, Oxford, 2017 (A scientific encyclopaedia of visual illusions).
[25] T.S. Curry III, J.E. Dowdey, R.C. Murry Jr., *Christensen's Physics of Diagnostic Radiology*, fourth ed., Williams & Wilikins, Media, PA, 1990.
[26] Wikimedia Commons Contributors, Wiki Dell lcd.Jpg; Halftoningcolor.Svg. *Wikimedia Commons*, 2023. https://commons.wikimedia.org/wiki/File:Wiki_dell_lcd.jpg. (Accessed 4 September 2023). https://commons.wikimedia.org/wiki/File:Halftoningcolor.svg.
[27] Hirschler R., Csillag P., Manyé P., Neder M., How much colour science is not too much?, Color Ressearch & Application 43 (2018) 977–992, doi:10.1002/col.22275. (Critique of colour literacy among visual artists, designers, and architects; discusses misconceptions about colour, and presents the fundamentals for a colour science curriculum for non-scientists.).
[28] V. Atanasiu, *Expert Bytes: Computer Expertise in Forensic Documents: Players, Needs, Resources, and Pitfalls*, CRC Press, Boca Raton, FL, 2014.
[29] [FADGI] United States Government, Federal Agencies Digital Guidelines Initiative, 10.20, 2022, http://www.digitizationguidelines.gov. (Accessed 22 October 2022) (One of the major guidelines on digitisation for libraries, archives, and museums).
[30] [Metamorfoze] National Library of the Netherlands, Metamorfoze: National Programme for the Preservation of Paper Heritage, 2022. https://www.metamorfoze.nl. (Accessed 22 October 2022) (One of the major guidelines on digitisation for libraries, archives, and museums).
[31] S. Van Poucke, Y. Vander Haeghen, K. Vissers, Th Meert, Ph Jorens, Automatic colorimetric calibration of human wounds, BMC Med. Imag. 10 (7) (2010) 10. http://www.biomedcentral.com/1471-2342/10/7.
[32] N. Baba, S. Isobe, Y. Norimoto, M. Noguchi, Stellar speckle image reconstruction by the shift-and-add method, Appl. Opt. 24 (10) (1985) 1403–1405.
[33] R. Raskar, Computational photography: epsilon to coded photography, in: F. Nielsen (Ed.), Emerging Trends in Visual Computing: LIX Fall Colloquium, ETVC 2008, Palaiseau, France, November 18-20, 2008, Revised Selected and Invited Papers, Springer, Berlin, 2009, pp. 238–253, https://doi.org/10.1007/978-3-642-00826-9.
[34] M. Yuen, Coding artifacts and visual distortions, in: H.R. Wu, K.R. Rao (Eds.), Digital Video Image Quality and Perceptual Coding, CRC Press, Boca Raton, FL, 2005, pp. 87–122 (Collection of visual distortions in digital images and videos).
[35] W.B. Pennebaker, J.L. Mitchell, JPEG, Still Image Data Compression Standard, Van Nostrand Reinhold, New York, NY, 1993.
[36] B.E. Rogowitz, L.A. Treinish, Data visualization: the end of the rainbow, IEEE Spectrum 12 (1998) 52–59.
[37] M.D. Fairchild, Color Appearance Models, third ed., John Wiley & Sons, Chichester, 2013 (A fundamental reference on colour spaces).
[38] R.G. Kuehni, A. Schwarz, Color Ordered: A Survey of Color Order Systems from Antiquity to the Present, Oxford University Press, Oxford, 2008 (A compendium of colour description models).
[39] A. Stockman, D.H. Brainard, Fundamental of color vision I: color processing in the eye, in: A.J. Elliot, M.D. Fairchild, A. Franklin (Eds.), Handbook of Color Psychology, Cambridge University Press, Cambridge, 2015, pp. 27–69.
[40] Ch Poynton, Digital Video and HDTV: Algorithms and Interfaces, Morgan Kaufmann, San Francisco, CA, 2002 (Technical aspects of colour for video and television).
[41] I. Abramov, J. Gordon, O. Feldman, A. Chavarga, Sex & vision I: spatio-temporal resolution. Biology of sex differences, Biol. Sex Differ. 3 (2012), https://doi.org/10.1186/2042-6410-3-20 article 20.
[42] Abramov, I., Gordon, J., Feldman, O., and Chavarga, A. Sex and vision II: Color appearance of monochromatic lights, *Biol. Sex Differ.* 3 (2012) article 21, doi: 10.1186/2042-6410-3-21. **(On sexual differences in human vision.)**.
[43] J. Morovič, Color Gamut Mapping, John Wiley & Sons, Chichester, 2008 (A technical reference on converting between colour spaces).
[44] K. Zuiderveld, Contrast limited adaptive histogram equalization, in: P.S. Heckbert (Ed.), Graphic Gems IV, Academic Press, San Diego, CA, 1994, pp. 474–485.
[45] V. Atanasiu, I. Marthot-Santaniello, Personalizing image enhancement for critical visual tasks: improved legibility of papyri using color processing and visual illusions, Int. J. Doc. Anal. Recogn. 25 (2021) 129–160, doi:10.1007/s10032-021- 00386-0. (Advocates the use of personalized computational methods for image enhancement, and describes an open source software implementing such an approach.).
P7 – 152
[46] E.H. Land, The retinex theory of color vision, Sci. Am. 237 (6) (1977) 108–128 (A classic on colour constancy, by the inventor of the Polaroid photographic camera).
[47] J.J. McCann, Retinex at 50: color theory and spatial algorithms, a review, J. Electron. Imag. 26 (3) (2017), 031204, https://doi.org/10.1117/1.JEI.26.3.031204, 1–14.
[48] D.J. Jobson, Z.-ur Rahman, G.A. Woodell, A multiscale retinex for bridging the gap between color images and the human observation of scenes, IEEE Trans. Image Process. 6 (7) (1997) 965–976.
[49] Special section on retinex at 50, J. Electron. Imag. 26 (3) (2017). A collection of modern developments of Land's (1977) colour constancy model SPIE & IS&T.
[50] D.H. Foster, K. Amano, S.M.C. Nascimento, M.J. Foster, Frequency of metamerism in natural scenes, J. Optical Soc.Am. A, Optics, Image Sci. Vision 23 (10) (2006) 2359–2372.
[51] H. Hariharan, A. Koschan, B. Abidi, A. Gribok, M. Abidi, Fusion of visible and infrared images using empirical mode decomposition to improve face recognition, Proceedings of the IEEE 2006 International Conference on Image Processing (ICIP 2006), 8–11 October 2006, Atlanta, GA, USA, (2006) 2049–2052, https://doi.org/10.1109/ICIP.2006.312860.
[52] P. Kay, L. Maffi, Color appearance and the emergence and evolution of basic color lexicons, Am. Anthropol. 101 (4) (1999) 743–760.
[53] A.R. Gillespie, A.B. Kahle, R.E. Walker, Color enhancement of highly correlated images. I. Decorrelation and HSI contrast stretches, Rem. Sens. Environ. 20 (3) (1986) 209–235, doi:10.1016/0034-4257(86)90044-1.
[54] A.R. Gillespie, A.B. Kahle, R.E. Walker, A.R. Erratum, A.B. Kahle, R.E. Walker, Color enhancement of highly correlated images. I. Decorrelation and HSI contrast stretches, remote sens. Environ. 20: 209–235, 1987, Rem. Sens. Environ. 21 (3) (1986) 362. doi:10.1016/0034-4257(87)90021-6.
[55] A.R. Gillespie, A.B. Kahle, R.E. Walker, Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques, Rem. Sens. Environ. 22 (3) (1987) 343–365, doi:10.1016/0034-4257(87)90088-5.
[56] I. Pratikakis, K. Zagoris, X. Karagiannis, L. Tsochatzidis, T. Mondal, I. Marthot-Santaniello, ICDAR 2019 Competition on Document Image Binarization (DIBCO 2019), *Proceedings of the 2019 International. Conference on Document Analysis and Recognition (ICDAR), 20–25 September 2019*, 2019, pp. 1547–1556. Sydney, Australia.
[57] J. Van de Weijer, Th Gevers, A.W. Smeulders, Robust photometric invariant features from the color tensor, IEEE Trans. Image Process. 15 (1) (2006) 118–127, https://doi.org/10.1109/tip.2005.860343.
[58] J. Chen, T.N. Pappas, A. Mojsilovic, B.E. Rogowitz, Adaptive perceptual color-texture image segmentation, IEEE Trans. Image Process. 14 (10) (2005) 1524–1536.
[59] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. 8 (6) (1986) 678–698.
[60] Otsu Nobuyuki, Auto thresholding, IEEE Transactions on Systems, Man, and Cybernetics, SMC- 9 (1) (1979) 62–66.
[61] Th Gevers, A. Gijsenij, J. van de Weijer, J.-M. Geusebroek, Color in Computer Vision: Fundamentals and Applications, John Wiley & Sons, Chichester, 2012 (A handbook on computational colour processing).
[62] F.A.A. Kingdom, C. Beauce, L. Hunter, Colour vision brings clarity to shadows, Perception 33 (2004) 907–914, https://doi.org/10.1068/p526.
[63] C. Breuil, B.J. Jennings, S. Barthelme, N. Guyader, F.A.A. Kingdom, Color improves edge classification in human vision, PLoS Comput. Biol. 15 (10) (2019), e1007398, https://doi.org/10.1371/journal.pcbi.1007398.
[64] M. Ebner, Color Constancy, John Wiley & Sons, Chichester, 2007.. Why a colour can be recognised as the same colour under different lighting conditions (direct sunlight, in shadows, etc.), and how to approach this important technical issue from a computational perspective.
[65] C. von Campenhausen, Photoreceptors, lightness constancy and color vision, Naturwissenschaften 73 (1986) 674–675.
[66] P. Collier, D. Forrest, A. Pearson, The representation of topographic information on maps: the depiction of relief, Cartogr. J. 40 (1) (2003) 17–26, https://doi.org/10.1179/000870403235002033.
[67] D. Jones, M. Amaral, The Colour of Time: A New History of the World 1850–1960, Head of Zeus, London, 2019.
[68] A. Saunders, Apollo Remastered: the Ultimate Photographic Record, Black Dog & Leventhal, New York, NY, 2022.

[69] Th Keplinger, 05.11). 1945 – Der Krieg ist aus!. *Worte im Dunkel*, 2020. https://www.worteimdunkel.at/?p=6713. (Accessed 31 October 2022).
[70] J.C. Russ, F.B. Neal, The Image Processing Handbook, seventh ed., CRC Press, Boca Raton, FL, 2016 (An eminently comprehensive and practical compendium on image processing concepts and techniques).
[71] Th L'Excellent, *Penser la typographie en couleurs*. M.A. thesis, Graphic design and multimedia, Paris: École Nationale Supérieure des Arts Décoratifs, 2009. https://typo.thomaslexcellent.com. (Accessed 31 October 2020).
[72] D.M. Chandler, Seven challenges in image quality assessment: past, present, and future research, Int. Scholarly Res. Notices (ISRN) Signal Process. (2013), https://doi.org/10.1155/2013/905685 art. 905685.
[73] A. Beghdadi, M.-C. Larabi, A. Bouzerdoum, K.M. Iftekharuddin, A survey of perceptual image processing methods, Signal Process. Image Commun. 28 (2013) 811–831, https://doi.org/10.1016/j.image.2013.06.003.
[74] S. Le Moan, I. Farup, J. Blahová, Towards exploiting change blindness for image processing, J. Vis. Commun. Image Represent. 54 (2018) 31–38.
[75] J. Davidoff, *Cognition through Color*, Bradford/MIT Press, Cambridge, MA, 1991.
[76] M.F. Land, D.-E. Nilsson, Animal Eyes, second ed., Oxford University Press, Oxford, 2012 (An excellent companion to pursue the investigation of colour beyond human vision).
[77] G.H. Jacobs, Evolution of color vision and its reflections in contemporary mammals, in: A.J. Elliot, M.D. Fairchild, A. Franklin (Eds.), Handbook of Color Psychology, Cambridge University Press, Cambridge, 2015, pp. 110–130.
[78] Sh Kawamura, Ch Hiramatsu, A.D. Melin, C.M. Schaffner, F. Aureli, L.M. Fedigan, Polymorphic color vision in primates: evolutionary considerations, in: H. Hirai, H. Imai, Y. Go (Eds.), Post-genome Biology of Primates, Springer, Tokyo, 2012, pp. 93–120, https://doi.org/10.1007/978-4-431-54011-3_7.
[79] H. Zollinger, Color: A Multidisciplinary Approach, Zürich/Weinheim: Helvetica Chimica Acta/Wiley-VCH, 1999 (Lively, erudite, and interdisciplinary introduction to colour science, including expositions on colour chemistry, intercultural aspects, linguistics, and art).
[80] R.N. Shepard, The perceptual organization of colors: an adaptation to regularities of the terrestrial world? in: J. Barkow, L. Cosmides, J. Tooby (Eds.), The Adapted Mind: Evolutionary Psychology and the Generation of Culture Oxford University Press, Oxford, 1992, pp. 495–532 (The environmental hypothesis of gender-based visual dimorphisms).
[81] B.C. Regan, C. Julliot, B. Simmen, F. Viénot, P. Charles-Dominique, J.D. Mollon, Fruits, foliage and the evolution of primate colour vision, Phil. Trans.: Biol. Sci. 356 (1407) (2001) 229–283.
[82] P. Sumner, J.D. Mollon, Catarrhine photopigments are optimized for detecting targets against a foliage background, J. Exp. Biol. 203 (2000) 1963–1986.
[83] G.H. Jacobs, Ultraviolet vision in vertebrates, Am. Zool. 32 (4) (1992) 544–554 (Colour vision beyond human capabilities).
[84] R. Dolder, The importance of colour in the hospital pharmacy, in: J.N. Counsell (Ed.), Natural Colours for Food and Other Uses, Applied Science Publishers, London, 1981, pp. 83–89 (Colour as critical discriminant marker of solid and fluid drugs in clinical practice).
[85] L. Wright, F. Black, Monochrome males and colorful females: do gender and age influence the color and content of drawings? Sage Open 3 (4) (2013) 1–9, https://doi.org/10.1177/2158244013509254.
[86] E.A. Sjoberg, R.G. Wilner, A. D'Souza, G.G. Cole, The stroop task sex difference: evolved inhibition or color naming? Arch. Sex. Behav. 52 (2023) 315–323, https://doi.org/10.1007/s10508-022-02439-9.
[87] M.A. Webster, Individual differences in color vision, in: A.J. Elliot, M.D. Fairchild, A. Franklin (Eds.), Handbook of Color Psychology, Cambridge University Press, Cambridge, 2015, pp. 197–215.
[88] K.A. Jameson, A.D. Winkler, K. Goldfarb, Art, interpersonal comparisons of color experience, and potential tetrachromacy. Proceedings of the Human Vision and Electronic Imaging, HVEI 2016, in: IS&T International Symposium on Electronic Imaging, 2016, https://doi.org/10.2352/ISSN.2470-1173.2016.16HVEI-145. *February 14*–18, 2016, San Francisco, California, USA (12.
[89] O. Blajenkova, M. Kozhevnikov, The new object-spatial-verbal cognitive style model: theory and measurement, Appl. Cognit. Psychol. 23 (2009) 638–663, https://doi.org/10.1002/acp.1473.
[90] D. Voyer, S. Voyer, M.P. Bryden, Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables, Psychol. Bull. 117 (2) (1995) 250–270.
[91] G. Sanders, K. Sinclair, T. Walsh, Testing predictions from the hunter-gatherer hypothesis — 2: sex Differences in the visual processing of near and far space, Evol. Psychol. 5 (2007) 666–679.
[92] K.R. Boff, J.E. Lincoln (Eds.), *Engineering Data Compendium: Human Perception and Performance*, Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, OH, 1988 (An unequalled encyclopaedic resource on psychometric data, including effect description, conditions, charts, and references).
[93] S.O. Murray, M.-P. Schallmo, T. Kolodny, R. Millin, A. Kale, Ph Thomas, T.H. Rammsayer, S.J. Troche, R.A. Bernier, D. Tadin, Sex differences in visual motion processing, Curr. Biol. 28 (17) (2018) 2794–2799.e3, https://doi.org/10.1016/j.cub.2018.06.014.
[94] A. Shaqiri, M. Roinishvili, L. Grzeczkowski, E. Chkonia, K. Pilz, A. Mohr, Brand, M. Kunchulia, M.H. Herzog, Sex-related differences in vision are heterogeneous, 10, Sci. Rep. 8 (2018) 7521, https://doi.org/10.1038/s41598-018-25298-8 (Large-scale research sexual differences in human vision).
[95] I. Silverman, M. Eals, Sex differences in spatial abilities: evolutionary theory and data, in: J. Barkow, L. Cosmides, J. Tooby (Eds.), The Adapted Mind: Evolutionary Psychology and the Generation of Culture, Oxford University Press, Oxford, 1992, pp. 533–549 (The social hypothesis of gender-based visual dimorphisms).
[96] A. Anderson, S. Chilczuk, K. Nelson, R. Ruther, C. Wall-Scheffler, The Myth of Man the Hunter: women's contribution to the hunt across ethnographic contexts, PLoS One 18 (6) (2023), e0287101, https://doi.org/10.1371/journal.pone.0287101.
[97] M.H. Bornstein, Emergence and early development of color vision and color perception, in: A.J. Elliot, M.D. Fairchild, A. Franklin (Eds.), *Handbook of Color Psychology*, Cambridge University Press, Cambridge, 2015, pp. 149–179.
[98] Ch L. Folk, The role of color in the voluntary and involuntary guidance of selective attention, in: A.J. Elliot, M.D. Fairchild, A. Franklin (Eds.), Handbook of Color Psychology, Cambridge University Press, Cambridge, 2015, pp. 481–501.
[99] W. Metzger, Laws of Seeing, MIT Press, Cambridge, MA, 2006 (A classic of the early twentieth century school of visual Gestalt psychology).
[100] V.S. Ramachandran, E.M. Hubbard, Synaesthesia — a window into perception, thought and language, J. Conscious. Stud. 8 (12) (2001) 3–34.
[101] R.E. Cytowic, D.M. Eagleman, *Wednesday Is Indigo Blue. Discovering the Brain of Synesthesia*, MIT Press, Cambridge, MA, 2009 (On experiencing shapes as colour, colours as shapes, and other cross-modal effects).
[102] L.C. Robertson, N. Sagiv (Eds.), SYNESTHESIA: Perspectives from Cognitive Neuroscience, Oxford University Press, Oxford, 2005.
[103] V. Ginsburgh, S. Metzidakis, On rimbaud's "vowels," again: vowels or colors? Athens J. Philol. 6 (4) (2019) 225–234, https://doi.org/10.30958/ajp.6-4-1.
[104] Wikipedia Contributors, Complication (Horology). *Wikipedia*, 2023. https://en.wikipedia.org/wiki/Complication_(horology. (Accessed 4 September 2023).
[105] Sh Ishihara, The Series of Plates Designed as a Test for Colour-Blindness, Kanehara Shuppan, Tokyo, 1972.
[106] A. French, K. Rose, K. Thompson, E. Cornell, The Evolution of Colour Vision Testing. Australian Orthoptic Journal 40 (2) (2008) 7–15.
[107] P. Kovesi, Phase congruency detects corners and edges, Proceedings of the Seventh International Conference on Digital Image Computing: Techniques and Applications (2003) 309–318. DICTA 2003, 10–12 December, Sydney, Australia.
[108] D.L. Lau, G.R. Arce, Modern Digital Halftoning, second ed., CRC Press, Boca Raton, FL, 2001.
[109] A.J. Elliot, M.D. Fairchild, A. Franklin (Eds.), Handbook of Color Psychology, Cambridge University Press, Cambridge, 2015 (An in-depth reference on the fundamental aspects of human colour vision).
[110] D. McIntyre, Colour Blindness: Causes and Effects, Dalton Publishing, Chester, 2002 (Concise and comprehensive introduction to the various types of inability to perceiving colour).
[111] N.R.A. Parry, Color vision deficiencies, in: A.J. Elliot, M.D. Fairchild, A. Franklin (Eds.), Handbook of Color Psychology, Cambridge University Press, Cambridge, 2015, pp. 216–242.
[112] R. Doron, A. Sterkin, M. Fried, O. Yehezkel, M. Lev, M. Belkin, M. Lev, M. Belkin, M. Rosner, A.S. Solomon, Y. Mandel, U. Polat, Spatial visual function in anomalous trichromats: is less more? PLoS One 14 (1) (2019), e0209662 https://doi.org/10.1371/journal.pone.0209662, 12.

P7 – 153

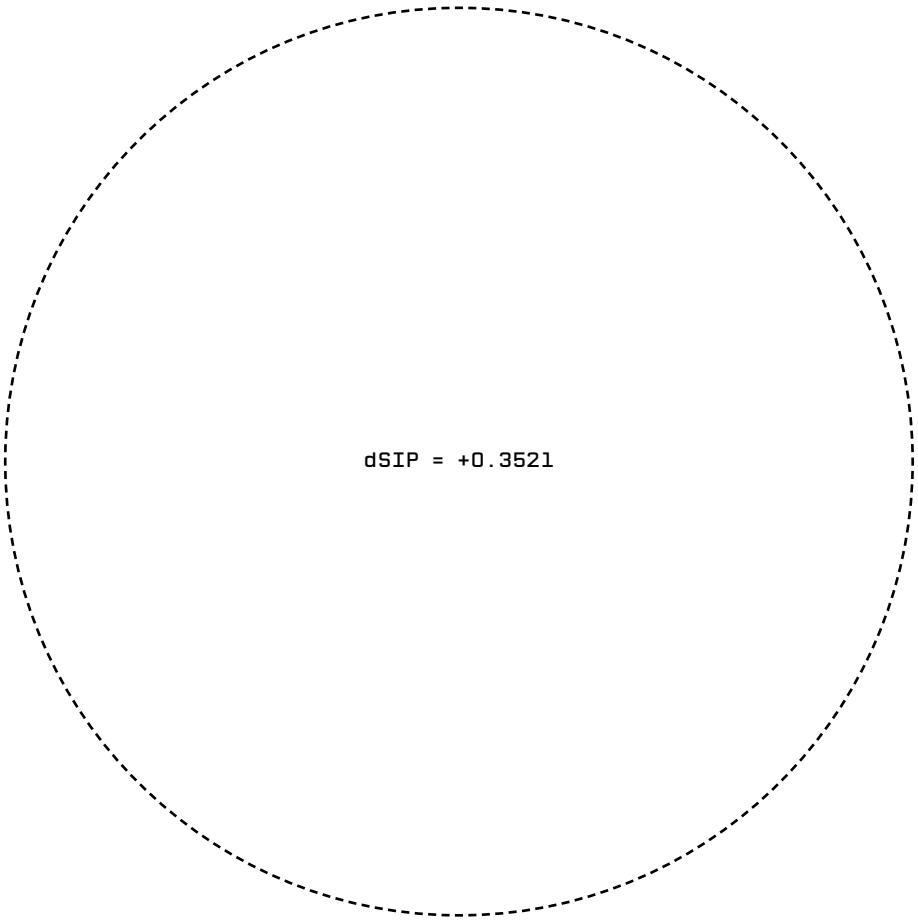
[113] M. Capasso, P. Davoli (Eds.), *Abstracts of the 29th International Congress of Papyrology, 28 July – 8 August 2019, Lecce, Italy*, University of Salento, Lecce, 2019.
[114] C.R.C. Allberry, et al., *Actes du V^e Congrès international de papyrologie, Oxford, 30 Août – 3 Septembre 1937*, Bruxelles: Fondation égyptologique Reine Élisabeth, 1938.
[115] M.J. Morgan, A. Adam, J.D. Mollon, Dichromats detect colour-camouflaged objects that are not detected by trichromats, Proc. Biol. Sci. 248 (1323) (1992) 291–295.
[116] I. Farup, Individualised halo-free gradient-domain colour image daltonisation, J. Imaging 6 (2020) 10, https://doi.org/10.3390/jimaging6110116, article 116.
[117] R.R. Behrens, *False Colors: Art, Design and Modern Camouflage*, Bobolink Books, Dysart, IA, 2002 (A historical account of the role of artists in the development of modern camouflage).
[118] G.D. Ruxton, W.L. Allen, Th N. Sherratt, M.P. Speed, Avoiding Attack. The Evolutionary Ecology of Crypsis, Aposematism, and Mimicry, second ed., Oxford University Press, Oxford, 2019.
[119] M. Stevens, S. Merilaita, Animal Camouflage: Mechanisms and Function, Cambridge University Press, Cambridge, 2011 (A fascinating survey of how animals appear to change shape using colour camouflage).
[120] F.A.A. Kingdom, Color brings relief to human vision, Nat. Neurosci. 6 (6) (2003) 641–644, https://doi.org/10.1038/nn1060 (One of several works by a scientist whose research focuses specifically on colour and shape interactions in human vision).
[121] N.J. Wade, Perception and Illusion: Historical Perspectives, Springer + Business Media, Dordrecht, 2005.
[122] D. Erasmus, in: J.K. Sowards (Ed.), De Pueris Instituendis, De Recta Pronuntiatione, Collected Works of Erasmus: Literary and Educational Writings, vol. 4, University of Toronto Press, Toronto, 1985.
[123] Wikipedia Contributors, Hatching (Heraldry). *Wikipedia*, 2023. https://en.wikipedia.org/wiki/Hatching_(heraldry. (Accessed 4 September 2023).
[124] Wikipedia Contributors, Schraffur [de]. *Wikipedia*. https://de.wikipedia.org/wiki/Schraffur, 2023. (Accessed 4 September 2023).
[125] Whittlesey S.M., Subjective color in Mimbres black-on-white pottery, KIVA: J. Southwestern Anthropology and History 80 (1) (2014) 45–70, doi:10.1179/0023194015z.00000000039. (Making colour with black and white.).
[126] C. Chamberlain, Color in a material world: Margaret cavendish against the early modern mechanists, Phil. Rev. 128 (3) (2019) 293–336, https://doi.org/10.1215/00318108-7537283.
[127] V. Atanasiu, Les réalités subjectives d'un paléographe arabe du X^e siècle, Gazette du Livre Medieval 43 (2004) 14–22.
[128] T. Omagari, Arcade Game Typography: the Art of Pixel Type, Thames & Hudson, London, 2019.
[129] M. Pastoureau, Black: the History of a Color, *Princeton*. Princeton, NJ, Princeton University Press, 2008 (One in a series of books by this author on the cultural history of individual colours; erudite and lively).
[130] C.L. Kane, Chromatic Algorithms: Synthetic Color, Computer Art, and Aesthetics after Code, The University of Chicago Press, Chicago, 2014 (Historical perspectives on modern colour technologies).
[131] M. Taussig, What Color Is the Sacred? University of Chicago Press, Chicago, 2009 (An anthropological perspective on the history of colours).
[132] J. Senchyne, The Intimacy of Paper in Early and Nineteenth-Century American Literature, University of Massachusetts Press, Amherst, MA, 2020.
[133] G. Taylor, Buying Whiteness: Race, Culture, and Identity from Columbus to Hip Hop, Palgrave Macmillan, New York, NY, 2005 (On the interweaving of racism and the early history of colour science in Enlightenment England).
[134] M. Brysbaert, How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables, J. Cognition 2 (1) (2019) 1–38, https://doi.org/10.5334/joc.72 (Major reference of practical importance for establishing the number of participants in human experiments).
[135] S.F. Johnston, A History of Light and Colour Measurement: Science in the Shadows, Institute of Physics Publishing, Bristol, 2001 (On the social construct of colour science and its scientific marginality).
[136] H. Brettel, F. Viénot, J.D. Mollon, Computerized simulation of color appearance for dichromats, J. Opt. Soc. Am. 14 (10) (1997) 2647–2655.
[137] R. Mausfeld, Attributes of color and elementaristic misconceptions of color representations, in: B. Saunders, J. van Brakel (Eds.), Theories, Technologies, Instrumentalities of Color: Anthropological and Historiographic Perspectives, University Press of America, Lanham, MD, 2002, pp. 127–146.
[138] L. Decock, The phenomenal color 'space' is not a space, in: B. Saunders, J. van Brakel (Eds.), *Theories, Technologies, Instrumentalities of Color: Anthropological and Historiographic Perspectives*, University Press of America, Lanham, MD, 2002, pp. 343–351.
[139] J. Van Brakel, Chromatic language games and thier congeners, in: B. Saunders, J. van Brakel (Eds.), Theories, Technologies, Instrumentalities of Color: Anthropological and Historiographic Perspectives, University Press of America, Lanham, MD, 2002, pp. 147–168.
[140] J. Gage, Color and Culture: Practice and Meaning from Antiquity to Abstraction, University of California Press, Berkeley, CA, 1993.
[141] M.M. Sassi, The Sea Was Never Blue: the Greek Colour Experience Was Made of Movement and Shimmer. Can We Ever Glimpse what They Saw when Gazing Out to Sea?, 2017 *aeon*, https://aeon.co/essays/can-we-hope-to-understand-how-the-greeks-saw-their-world, 2020.02.22.
[142] R. Kuschel, T. Monberg, 'We don't talk much about colour here': a study of colour semantics on Bellona Island, Man 9 (2) (1974) 213–242.
[143] E. Gibson, R. Futrell, J. Jara-Ettinger, K. Mahowald, L. Bergen, S. Ratnasingam, M. Gibson, S.T. Piantadosi, B.R. Conway, Color naming across languages reflects color use, Proc. Natl. Acad. Sci. USA 114 (40) (2017) 10785–10790, https://doi.org/10.1073/pnas.1619666114.
[144] F. Strik Lievers, Ch-R. Huang, J. Xiong, Linguistic synaesthesia, in: X. Wen, J.R. Taylor (Eds.), The Routledge Handbook of Cognitive Linguistics, Routledge, New York, NY, 2021. Routledge, https://www.routledgehandbooks.com/doi/10.4324/9781351034708-25.
[145] H. Zollinger, Biological aspects of color naming, in: I. Rentschler, B. Herzberger, D. Epstein (Eds.), Beauty and the Brain: Biological Aspects of Aesthetics, Birkhauser, Basel, 1988, pp. 149–164 (A clear and convincing exposition of the interaction between the neuropsysiology of colour cognition and ecological and cultural factors).
[146] G.V. Paramei, Singing the Russian blues: an argument for culturally basic color terms, Cross Cult. Res. 39 (2005) 10–38, https://doi.org/10.1177/1069397104267888.
[147] J. Winawer, N. Witthoft, M.C. Frank, L. Wu, A.R. Wade, L. Boroditsky, Russian blues reveal effects of language on color discrimination, Proc. Natl. Acad. Sci. USA 104 (19) (2007) 7780–7785, https://doi.org/10.1073/pnas.0701644104.
[148] Ch Xiang, AI Isn't Artificial or Intelligent: How AI Innovation Is Powered by Underpaid Workers in Foreign Countries. Motherboard: Tech by Vice, 2022, 12.06, https://www.vice.com/en/article/wxnaqz/ai-isnt-artificial-or-intelligent. (Accessed 6 December 2022).
[149] B. Berlin, P. Kay, *Basic Color Terms: Their Universality and Evolution*, University of California Press, Berkeley and Los Angeles, 1969 (A classic on colour naming across cultures).
[150] J.T. Abbott, T.L. Griffiths, Regier, Focal colors across languages are representative members of color categories, Proc. Natl. Acad. Sci. USA 40 (113) (2016) 11178–11183, https://doi.org/10.1073/pnas.1513298113.
[151] D.T. Lindsey, A.M. Brown, D.H. Brainard, C.L. Apicella, Hunter-gatherer color naming provides new insight into the evolution of color terms, Curr. Biol. 25 (2015) 1–6, https://doi.org/10.1016/j.cub.2015.08.006.
[152] D. Roberson, J. Davidoff, I.R.L. Davies, L.R. Shapiro, Color categories: evidence for the cultural relativity hypothesis? Cognit. Psychol. 50 (2005) 378–411, https://doi.org/10.1016/j.cogpsych.2004.10.001.
[153] B.A.C. Saunders, J. van Brakel, Are there nontrivial constraints on colour categorization? Behav. Brain Sci. 20 (1997) 167–228, https://doi.org/10.1017/S0140525X97001428.
[154] Ch Witzel, New insights into the evolution of color terms or an effect of saturation? i-Perception (2016) https://doi.org/10.1177/2041669516662040. September–October: 1–4.
[155] G. Deutscher, Through the Language Glass: Why the World Looks Different in Other Languages, Metropolitan Books, New York, NY, 2010.
[156] M.D. Fairchild, The colors of wine, Int. J. Wine Res. 10 (2006) 13–31, https://doi.org/10.2147/IJWR.S161891.
[157] Harding X., Keeping 'Insecure' Lit: HBO Cinematographer Ava Berkofsky on Properly Lighting Black Faces; MIC, June 9, 2017. https://www.mic.com/articles/184244/keeping-insecure-lit-hbo-cinematographer-ava-berkofsky-on-properly-lighting-black-faces. (Accessed 4 September 2023).

P7 – 154

[158] L. Roth, Looking at shirley, the ultimate norm: colour balance, image technologies, and cognitive equity, Can. J. Commun. 34 (1) (2009) 111–136, https://doi.org/10.22230/cjc.2009v34n1a2196.
[159] G. Yue, The China girl on the margins of film, 2015, pp. 96–116. October. 153 (Summer), https://doi.org/10.1162/OCTO_a_00228.
[160] V. Brinkmann, R. Dreyfus, U. Kock-Brinkmann (Eds.), *Gods in Color: Polychromy in the Ancient World*, Prestel, New York, NY, 2017.
[161] M. Guerrin, *Henri Cartier-Bresson et* Le Monde, Gallimard, Paris, 2008.
[162] C. Broué, M. Pastoureau, 2023, Michel Pastoureau, l'homme arc-en-ciel, Épisode 4/5: La couleur, cet unique et insaisissable objet; À voix nue, France Culture, radio interview, June 29, 2023, https://www.radiofrance.fr/franceculture/podcasts/a-voix-nue/la-couleur-cet-unique-et-insaisissable-objet-1890712 (An intellectual biography of an eminent historian of colour in the Western civilization).
[163] L. Ammann, P. Barla, X. Granier, G. Guennebaud, P. Reuter, Surface relief analysis for illustrative shading, Comput. Graph. Forum 31 (4) (2012) 1481–1490, https://doi.org/10.1111/j.1467-8659.2012.03144.x.
[164] L. Hurni, Cartographic Mountain Relief Presentation: 150 Years of Tradition and Progress at ETH Zurich. Proceedings of the 6th ICA Mountain Cartography Workshop: Mountain Mapping and Visualisation, February 11–15, 2008, Lenk, Switzerland, 2008, pp. 85–91 (Excellent examples from the domain of cartography for the use of colour to suggest three-dimensional shape).
[165] B. Bock von Wülfingen (Ed.), *Science in Color: Visualizing Achromatic Knowledge*, De Gruyter, Berlin, 2019 (A disciplinary cross-section of the use of colour in sciences).
[166] OED = *Oxford English dictionary*. https://www.oed.com.
[167] Wikipedia Contributors, Colors of Noise, Wikipedia, 2023. https://en.wikipedia.org/wiki/Colors_of_noise. (Accessed 4 September 2023).
[168] M. Chirimuuta, *Outside Color: Perceptual Science and the Puzzle of Color in Psychology*, MIT Press, Cambridge, MA, 2015.
[169] Chirimuuta M., Kingdom F.A.A., The uses of colour vision: ornamental, practical, and theoretical, Minds Mach. 25 (2015) 213–229, doi:10.1007/s11023-015-9364-z. (On the interaction of colour and shape; see also the other publications of author Kingdom.).
[170] A. Mojsilović, J. Gomes, B. Rogowitz, ISee: perceptual features for image library navigation, Proceedings of SPIE Conference on Human Vision and Electronic Imaging VII (4662) (2002) 266–277, https://doi.org/10.1117/12.469523.
[171] L.I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, second ed., John Wiley & Sons, Hoboken, NJ, 2014.
[172] A.A. Ross, K. Nandakumar, A.K. Jain, Handbook of Multibiometrics, Springer + Business Media, New York, 2006.
[173] M. Van Wageningen, Type and Color: How to Design and Use Multicolored Typefaces, Princeton Architectural Press, New York, NY, 2019.
[174] T.A. Rector, K. Arcand, M. Watzke, Coloring the Universe: an Insider's Look at Making Spectacular Images of Space, University of Alaska Press, Fairbanks, AK, 2015.
[175] F. Crameri, G.E. Shephard, P.J. Heron, The misuse of colour in science communication, Nat. Commun. 11 (2020) 10, https://doi.org/10.1038/s41467-020-19160-7, 5444.
[176] [CLP] Color literacy project. 2023. https://colourliteracy.org.gov. (Accessed 23 January 2023).
[177] Wikipedia Contributors, International Colour Day. *Wikipedia*, 2023. https://en.wikipedia.org/wiki/International_Colour_Day. (Accessed 4 September 2023).
[178] K.E. Piquette, Revealing the material world of ancient writing: digital techniques and theoretical considerations, in: F.A.J. Hoogendijk, van Steffie, M. T. Gompel (Eds.), The Materiality of Texts from Ancient Egypt: New Approaches to the Study of Textual Material from the Early Pharaonic to the Late Antique Period, Brill, Leiden, 2018, pp. 94–118.

Publication
the Eight

Please use the following link to obtain a copy of this book:
https://doi.org/10.36824/2024-atanasiu.



learn shapes multimodally
by touch and with eyes closed

characters maximizing distinctiveness
to improve recognition & avoid tampering

UGRAPHIA

THE PURSUIT OF PERFECT LEGIBILITY

VLAD ATANASIU

cyborg typeface designed for
both human and machine reading

Fluxus Editions

GRAPHOLINGUISTICS AND ITS APPLICATIONS

MEDIA STUDIES | TYPOGRAPHY

FALSE

true or false? messages written in the metamorphosis typeface change with reading distance, information being encoded in distinct spatial frequency bands

a history of legibility seen through the multifocal lens of cultural, social, aesthetic, technological, psychological, and informational constraints on progress in handwriting and typography

P8 – 160

is perfect legibility possible outside of thomas more's *utopia* (𐤕𐤃𐤋𐤓𐤉𐤀)? understanding the limits of legibility is of fundamental practical relevance to typeface design and optometry, in addition to being of theoretical interest for research in human communication. drawing on a wide range of material from paleography, typography, psychology, information theory, and cinema, and supplemented with an original experiment in script design, its argument reflected in idiosyncratically writing and layout styles, this cross-cultural history of legibility explores an interlocking complex of factors affecting progress in the long-term evolution of legibility ⇈ this work reveals the divergence between industrial ideals, scientific theories, and popular representations of legibility, how power games are played and aesthetic fashions develop through scripts. already loaded with heritage, scripts are being concomitantly constrained by character frequency variation between languages, visual perceptual effects, the kinematics of handwriting production, and the evolution of imaging, print and display technologies. there is limited progress found with respect to character structure, and a persistent flow and ebb in the legibility of handwriting ('bioscript'); however, there is ample optimization of artificially produced writing ('mechascript'), as well as a potential improvement in reading performance due to script proliferation and diversification

vlad atanasiu, phd scd, is researcher at the university of bern, switzerland, specialized in the history of writing and computational document processing & visualization

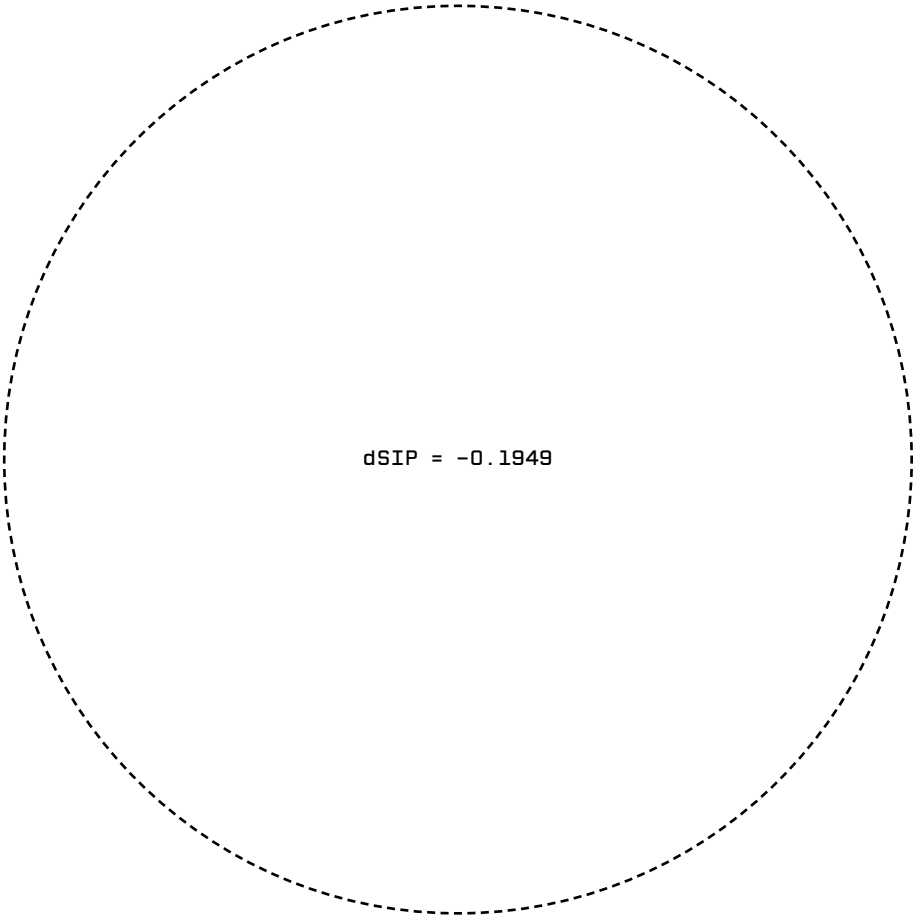
lege feliciter!

cover: this is not blindfolded justitia objectively evaluating the legibility of a typeface, but a child tracing sand-paper letters, demonstrating the montessori multi-sensory method of learning to write. picture commissioned by carl r. byoir (1886–1957), pioneer of public relations and early american promoter of montessori schooling © CANFIELD FISHER 1912: 86 ⇈ spine: book structure by layout: right bars: pages with symmetrical layout; middle bars: asymmetrical layout: left bars: near-empty © details on p. 2016

ISBN 9782487055001 90000 9 782487 055001

Table of contents. **1908 Aims** of this book: investigate if perfect legibility is possible. **1909 Rhetoric:** a scientific, literary, & graphical catalog of legibility constraints, illustrated with pictures, quotes, & typefaces, ranging from psychophysics to typographic practice to sociocultural pragmatics to theoretical models to cinematic representations to a typeface design experiment. **1910 Frame.** Epistemological perspectives & concept definitions. **1913 Empirics.** It is difficult to recognize the perfectly legible script when it is unclear how to measure legibility. **1915 Ideals.** The present typographical ideal of legibility is invisibility: script should not distract from reading. **1916 Opinions** on legibility. **1932 Theories.** Legibility is impermanent: subjective & contextual, fluctuating with social & material conditions. **1935 Consequences.** Does perfect legibility make us think differently ? **1937 Mythology.** Genesis, treatment, & ultima ratio of illegibility. **1943 Legacy.** Legibility is constrained by the legacy of character shapes. **1944 Overload.** Legibility is constrained by non-visual factors: linguistic, economical, political, etc. **1949 Power.** Legibility is the effectiveness of social control & command exerted by way of normative pattern production, while illegibility is a defense mechanism. **1953 Sausage.** A diet of only one script, however perfect, will eventually make you sick. **1954 Entropy.** Shape distinctiveness increases legibility. **1955 Statistics.** Differences in character frequency & combinatorics across languages, orthographies, & contents result in patterns with different legibility. **1956 Gestalt.** Contextual legibility optimization beyond local is difficult. **1959 Gray.** The textural pattern of paragraphs cannot be influenced by the script producer at content level. **1961 Contrast.** Color contrast between script & background, & between parts of script affects legibility. **1963 Perception.** Space perception is anisotropic, especially for those with dyslexia, one example among the many known & unknown perceptual phenomena affecting legibility. **1964 Kinematics.** Scripts are optimized for production facility, not solely visual appearance. **1967 Memory.** Legibility is a matter of perfecting mnemotechnics. **1978 Synthesis.** Mobiles (initial & para-usages), constraining factors, & attitudes (drives & effects) are components of the legibility machinery. **1980 Evolution.** Exposure to diverse scripts improves recognition performance. **1987 Acculturation.** Syncretisms keep legibility in dynamic equilibrium. **1990 Materiality.** Legibility varies with physical embodiment & environmental conditions. **1990 Aesthetics.** Script attractiveness motivates legibility performance. **1991 Personalization.** Legibility is optimal when tuned to a specific reader. **1991 Plasticity.** The tolerance range for deciphering bad writing is wide enough that most will care little about legibility until bad eyesight & old age catch up with us. **1993 Script Acts: Effective Imperfection.** What counts in written communication are outcomes, not character recognition alone, which may even benefit from imperfection. **1996 Undesirability.** There are instances in which less legibility is desirable. **1998 Spectrum.** There are degrees & kinds of legibility, according to applications. **1998 Evidence.** Legibility improves at a micromorphological level in artificially produced scripts, while the evolutionary improvement of character structure & handwriting is uncertain. **2007 Size.** There is an optimal script size range for every user & application. **2012 Polygraphy.** The presence of multiple scripts in the same document decreases legibility. **2015 Layout.** Layout ergonomy depends on information selection & manipulation ease. **2023 Time.** Legibility is degraded by the script decay over time. **2025 Representation.** Cinema represents legibility as a feat & a mystery to be pried open. **2036 Experiment.** Script design shows how scripts are compromises between many variables, with only one being legibility. **2040 Next.** The road ahead for legibility optimization. **2041 Exit.** Concluding remarks. **2048 References. 2066 Ontoindex. 2082 Fonts. 2084 Acknowledgments.**

Publication
the Ninth


dSIP = -0.1949

SoftwareX 14 (2021) 100684

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

Document Towers: A MATLAB software implementing a three-dimensional architectural paradigm for the visual exploration of digital documents and libraries

Vlad Atanasiu *, Rolf Ingold

University of Fribourg, Department of Informatics, Bd. de Pérolles 90, 1700 Fribourg, Switzerland

ARTICLE INFO

Article history:
Received 12 January 2021
Received in revised form 25 February 2021
Accepted 26 February 2021

Keywords:
Digital libraries
Document structure
Visualization
Information exploration
Serendipity
Metaphors

ABSTRACT

This article introduces the generic Document Towers paradigm, visualization, and software for visualizing the structure of paginated documents, based on the metaphor of DOCUMENTS-AS-ARCHITECTURE. The Document Towers visualizations resemble three-dimensional building models and represent the physical boundaries of logical (e.g., titles, images), semantic (e.g., topics, named entities), graphical (e.g., typefaces, colors), and other types of information with spatial extent as a stack of rooms and floors. The software takes as input user-supplied JSON-formatted coordinates and labels of document entities, or extracts them itself from ALTO and InDesign IDML files. The Document Towers paradigm and visualization enable information systems to support information behaviors other than goal-oriented searches. Visualization encourages exploration by generating panoramic overviews and fostering serendipitous insights, while the use of metaphors assists with comprehension of the representations through the application of a familiar cognitive model. Document Towers visualizations also provide access to types of information other than textual content, specifically by means of their physical structure, which corresponds to the material, logical, semantic, and contextual aspects of documents. Visualization renders documents transparent, making the invisible visible and facilitating analysis at a glance and without the need for physical manipulation. Keyword searches and other language-based interactions with documents must be clearly expressed and will return only answers to questions asked; by contrast, visual observation is well suited to fuzzy goals and uncovering unexpected aspects of the data.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

P9 – 165

Code metadata

Current code version	v2020.02.15
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-21-00009
Code Ocean compute capsule	–
Legal Code License	BSD-3-Clause
Code versioning system used	None
Software code languages, tools, and services used	MATLAB (R2020b)
Compilation requirements, operating environments & dependencies	Linux, Mac, Unix, Windows
If available Link to developer documentation/manual	Documentation included in software.
Support email for questions	atanasiu@alum.mit.edu

1. Motivation, principles and significance

"*Mathematics can only become truly interesting and original when it involves the operation of seeing something as something else.*"

[Reviel Netz]

* Corresponding author.
E-mail addresses: atanasiu@alum.mit.edu (Vlad Atanasiu), rolf.ingold@unifr.ch (Rolf Ingold).
URLs: http://waqwaq.info/ (Vlad Atanasiu), https://www3.unifr.ch/inf/de/all/people/16738/0a54b (Rolf Ingold).

https://doi.org/10.1016/j.softx.2021.100684
2352-7110/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Documents composed of pages are ubiquitous information storage formats, whether in digital (e.g., PDF and Word documents) or analog form (e.g., books, articles). Massive ongoing efforts are being made to realize the analog-to-digital conversion of historical document collections (e.g., Google Books), while digital-to-digital format conversions are also routine (e.g., Word to PDF). Thus, the relevance of developing diverse solutions that facilitate interactions with this type of information cannot be overstated. The Document Towers paradigm, visualization, and software introduced herein are one of these solutions.

While, typically, document content is accessed through the semantic analysis of text and graphical objects, this information is not always available (e.g., optical character recognition [OCR] is needed to make text in document images searchable) or may be expensive (e.g., high-quality handwriting recognition remains costly today). One solution, implemented in the Document Towers software, is to exploit the information contained in the document structure. For example, the first page of an article usually features a greater number and diversity of visually distinct entities than the subsequent pages; it also contains a greater variety of information levels, such as title, abstract, and authors. In other words, the structure of information is potentially indicative of its informativeness. It could in fact be argued that information *is* structure: letters and pixels create different texts and images depending on how they are arranged.

Natural language and numerical descriptions are other common paradigms for accessing information. The visual modality adopted by the Document Towers human–machine interface offers an alternative with several advantageous characteristics, including its support for exploration, serendipitous discoveries, and overview. These kinds of information behaviors differ from the targeted search implicit in the concept of the text input fields used almost ubiquitously in Internet search engines [1–3]. Visual search is thus well suited to cases in which the information sought cannot be clearly expressed in words, in which users employ different terminology and languages to describe the same concept, or in which the findings may be unexpected (as in quality control).

The Document Towers visualization represents documents as wireframe architectural models. The purpose of this cognitive model is to create a sense of familiarity for the user while comprehending and manipulating opaque physical objects and abstract digital structures [4,5]. There is a natural material correspondence between stacks of pages and buildings, and between libraries and cities; moreover, spatial and architectural metaphors are ingrained in software and hardware terminology (e.g., "information superhighway", "homepage", "desktop", "tunnel", "cloud", and the scientific field of "Information Architecture") [6], and have also been used in the past for experimental document representations. Examples include visualizing software structure as an urban landscape [7], extruding the nested structure of webpage objects [8], representing text columns in a three-dimensional space [9], rendering digital documents as look-alikes of physical books in a virtual reality library [10], and conceptualizing navigation between Internet domains as a walk through tunnels that connect various rooms [11]. Research into compact document overview are addressed in the literature, e.g., in the form of semantically highlighted thumbnails [12], or a dashboard of topic distribution in documents [13]. Illustrated surveys of similar experiments can be found in [14–16].

Fig. 1 illustrates how a three-dimensional architectural model is obtained by extruding the bounding boxes of various objects on a document page, thereby creating slabs that resemble rooms, walls, or pillars. Fig. 3 depicts how the page boundaries look like floors and the entire document like a building. Fig. 6 shows the similarity between a collection of documents and a cityscape.

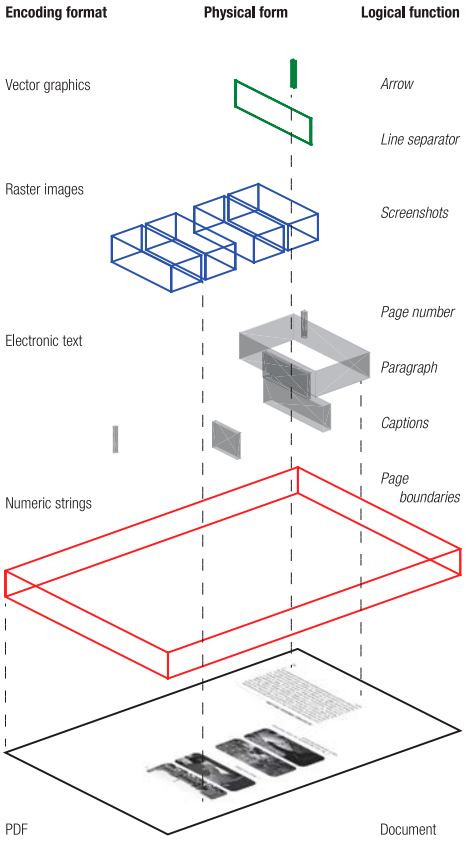


Fig. 1. Principle by which page entities are transformed into an architectural model.

Two effects of representing documents as wireframe architectural models contribute to the paradigm's effectiveness in eliciting insights. First, the models preserve the physical structure of paginated documents, which facilitates their analysis. Second, the simplification of shapes to slabs helps on one hand focus on the essential, and on the other hand spurs the imagination to search for meanings to sibylline structures, which is the goal of a visualization meant for exploration.

The Document Towers are first and foremost a visualization *paradigm* for paginated documents, while the Document Towers *software* implements this paradigm in an interactive visual *interface* (Fig. 2, left). Its purpose is to enable the exploration of document structures for practical applications, to aid research into the interpretation of document structures, and to promote the Document Towers paradigm for potential adoption in third-party applications. A distinct functionality of the software is its ability to read document object geometry and metadata from selected document formats, although the software scope in this case is not automatic analysis conducted by machines, but visual analysis performed by humans.

The separation between generic visualization and processing specific document formats is achieved by using as input to the software a custom JSON-formatted intermediary file for simple geometry and metadata description, which can be either generated from document files by the software or supplied by the users. To produce Document Towers visualizations all is needed are coordinates of the spatial extent of entities within documents, while the labels adding metadata to the entities are useful, but not indispensable. This makes the proposed method a generic information visualization concept for paginated documents, just like a histogram is a generic concept for representing a set of scalars.

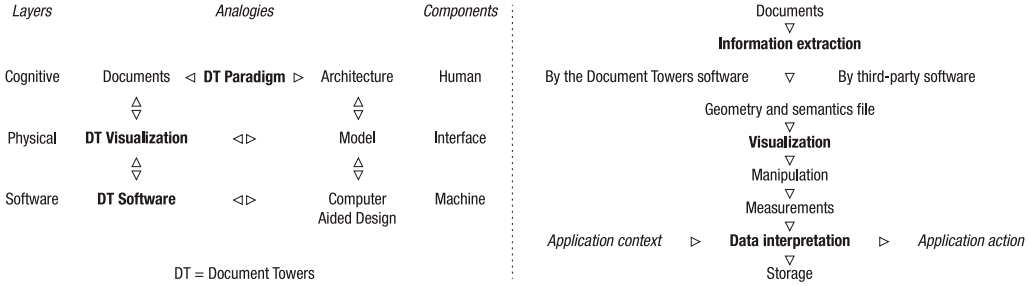


Fig. 2. Document Towers representation process (left) and workflow (right).

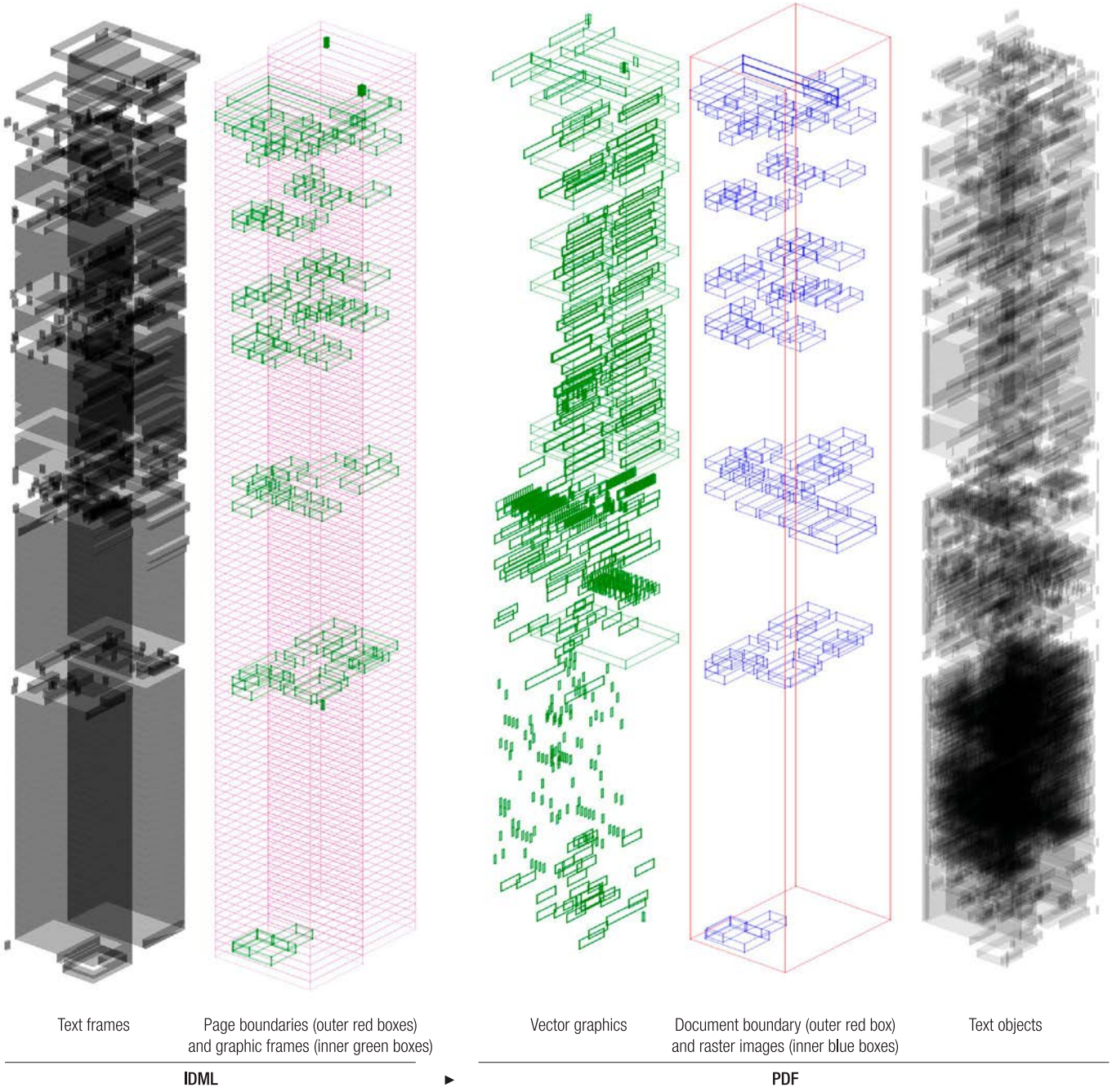


Fig. 3. Document Towers visualizations representing the distribution of text and graphics in the same document, generated from an IDML file (left) and a PDF file (right). (For a color version of this figure, the reader is referred to the web version of this article.)

P9 – 167

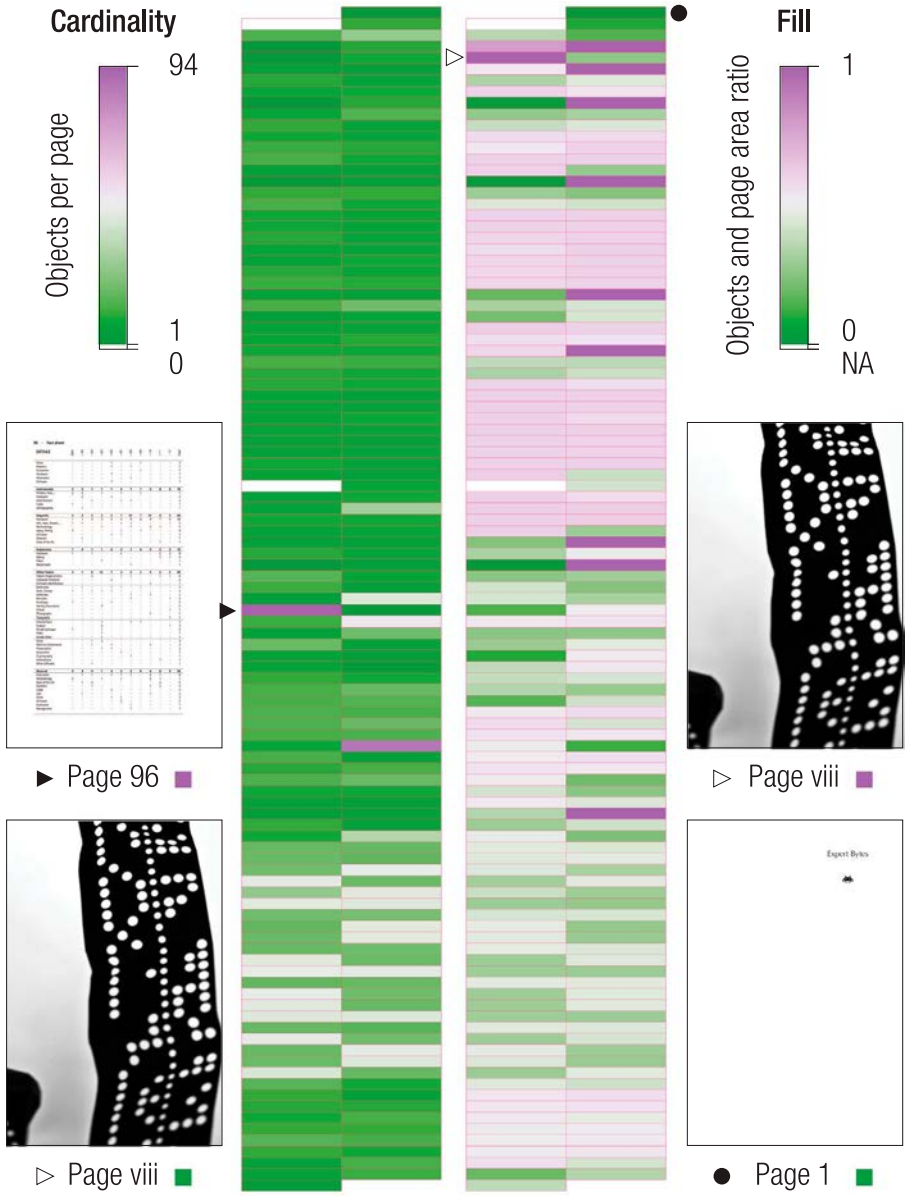


Fig. 4. Measures and Ribbons. Please refer to Section 2.6 "Measurements" for details. (For a color version of this figure, the reader is referred to the web version of this article.)

2. Software description

2.1. Software architecture and workflow

The software consists of a data capture and processing module, an interactive document structure viewer, and feature measurement functions. The user begins by using the software interface to select one or more electronic document files containing coordinates of document objects, and possibly also labels specifying the object classes (Fig. 2, right). The software extracts the object geometry and labels, then saves them to a file. This file can also be produced by a third-party software package in cases where it is necessary to process formats not supported by the Document Towers software (see the software documentation for specifications). In the next step, the user selects these geometry files, after which the software displays three-dimensional wireframes at the specified locations, and colors them according to the labels. The user can now interactively explore and interpret the visualization, which can subsequently be saved for later reuse.

2.2. Programming environment

The Document Towers software is written in the MATLAB programming language (R2020b). The software can be used in conjunction with MathWorks' MATLAB commercial application, or distributed for free as a compiled standalone desktop application running on top of the royalty-free MATLAB Runtime Compiler [17].

The visual quality of the 3D wireframes was first tested in Java (JavaFX), but early releases proved inadequate [18]. A JavaScript three.js library for 3D animation and using WebGL was found to be both graphically satisfactory and appealing due to its allowing for Web delivery of the visualization [19]. However, the exceedingly poor performance when importing geometry files larger than 1 MB made this solution impractical. For reference, the uncompressed ASCII file size for the geometry of a single Tower in Fig. 3 is 343 kB (4265 objects), while the collection of 89 Document Towers visualizations in Fig. 6 is 6.7 MB (85 203 objects).

Creating Document Towers visualizations in MATLAB resulted in none of the graphical drawbacks of JavaFX and was reasonably fast: a single Tower can be generated in four seconds on a MacBook Pro 2018 (3 GHz Intel Core i9). Use of the game engine Unity [20] and the design software AutoCAD [21] was contemplated, but MATLAB has the advantage of being a scientific software development environment that supports a broad range of applications, including image processing and statistics, both of which are useful for extending the document analysis power of the Document Towers software.

2.3. Extracted information

The information extracted from document files comprises the coordinates of the spatial extent of entities within the document, along with labels defining these entities (when available). For example, the IDML and ALTO files discussed in the next section provide labels for four spatially defined basic document entities: pages, text frames, raster images, and vector graphics. ALTO files typically contain a further set of labels, created ad hoc by the file producers, which introduce subcategories for the basic document entities (such as "title" for a text frame, "stamp" for a raster image, and color values for a vector font). The graphical interface of the Document Towers software can be used to filter entities for display. The Document Towers visualizations are agnostic to what they represent; their power depends on the richness of the information in various document formats, the possibility of automatically extracting it, and the user's ability to interpret it [22].

2.4. Data formats

The software is capable of reading the object coordinates and labels present in XML-formatted IDML (InDesign Markup Language) [23] and ALTO (Analyzed Layout and Text Object) [24] files. The former is employed by Adobe InDesign, the market-dominant typography software used by the publishing industry for complex layouts, while the latter is a standard for digitization projects.

Coordinates within pages, page numbers, object types, and labels, along with document metadata such as file names, are saved to a file in JSON (JavaScript Object Notation) format [25]. As an example, the first line in the file describing the Document Tower in Fig. 1 is `[0, 0, 0, 441, 0, 441, 666, 0, 666]`: here, the initial zero defines the object as a page, while the following numbers are the coordinate pairs of its bounding box. Sample data and detailed formatting specifications are provided with the software.

The geometry file is subsequently read by the visualization module. This hand-over mechanism allows the user to create document structure libraries, split long documents, merge documents into a single Document Tower, modify labels and metadata, or process data structure files generated using third-party software. For example, the information extracted from PDFs (Portable Document Format) presented in Figs. 3 and 6 is obtained through a proprietary application programming interface (API) for the Enlighter software of the Swiss company Sugarcube [26].

There may well be substantial differences in the information obtained from the various document formats; as a result, the visualized structures can be quite distinct from each other. Taking as an example an academic book written and produced in InDesign by the first author [27], Fig. 3 illustrates how the IDML-based visualization (left) emphasizes a recurrent irregular layout in the first part of the document, resulting from the presence of illustrations, while the Tower derived from the PDF of the same document (right) highlights the stark paragraph density of the reference section at the document's end. Many of the vector graphics visible in the PDF are rulers, which separate text from footnotes or table sections; in the IDML, these are defined as paragraph *parameters*, not as identifiable *objects*.

Fig. 5. The authors in front of a Document Towers visualization mural.

These differences raise questions, such as that of how much uncertainty related to document layout is present in a particular document description file. For example, a LaTeX `.tex` file on its own reveals little about the final appearance of the document, while a PDF file is expected to provide faithful graphical representation. Rather than considering the data format-dependency phenomenon a limitation, it may be useful to instead interpret differences as providing valuable insights into the specificities of various document formats, some of which may become visible through the Document Tower visualization paradigm. It is also useful to consider the complementarity of document formats, and investigate, via the Document Towers visualizations as thinking tools, how the different information provided by these formats might be used individually or fused. Such questions are relevant to document engineering, archival, and forensic applications, and the Document Towers paradigm provide a perspective distinct from others to address these questions.

2.5. Visualization and interaction

The main document structure representation paradigm implemented by the Document Towers visualization is that of three-dimensional wireframes that utilize the architectural metaphors of OBJECTS-AS-ROOMS, PAGES-AS-FLOORS, DOCUMENTS-AS-TOWERS, and LIBRARIES-AS-CITIES. By color-coding the facade of each floor according to specific semantic and quantitative criteria, so-called Ribbons are obtained, which are a more space-saving representation than the Document Towers visualization (Fig. 4).

Through a graphical interface, parameters such as colors and transparency can be set, the type of objects to display can be selected, and the projection changed between axonometric, elevation, and plan view; the user can also zoom, pan, and rotate the Document Towers visualizations in order to move from overview to details. Hyperlinks can be added, enabling a PDF version of the document to be opened at desired pages in a web browser; in this way, the Document Towers visualizations acquire a document navigation function similar to a table of contents.

2.6. Measurements

The software provides numerical statistics on object categories, as well as colorcoded information (Ribbons) on the number of objects per page (*cardinality*), and on the percentage of the

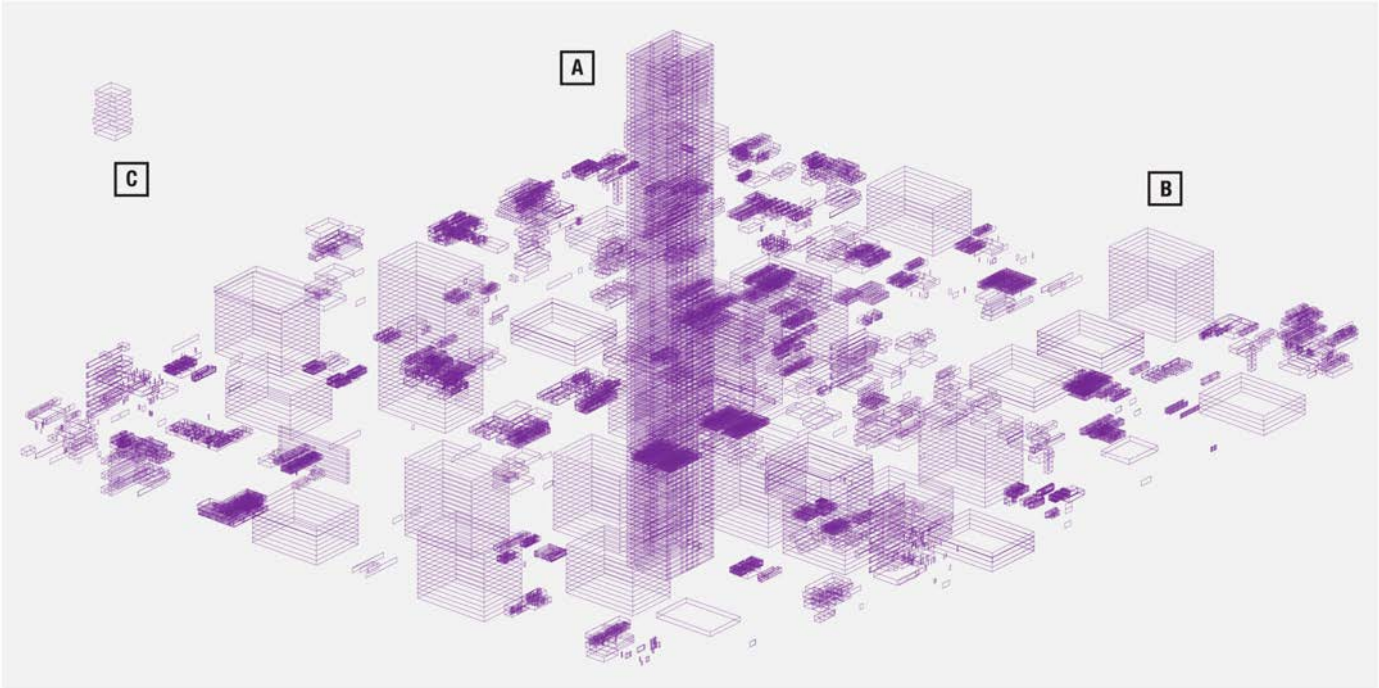


Fig. 6. A Document Towers City visualization having lead to three insights discussed in Section 3 "Illustrative example".

area of a page covered by objects (*fill*). Fig. 4 presents a front view of a Document Tower from the same PDF file as in Fig. 3, in which the facade was colorcoded on two measurements with reference to text, vector, and raster objects. The pages corresponding to the extreme values are shown. Page 96 is a table and has the highest number of objects per page as extracted by the PDF analysis algorithm, while page viii is a single raster image that covers the entirety of the page surface, in contrast to the almost empty page 1.

2.7. Scalability

The scalability of rendering documents as Document Towers visualizations is constrained by performance and physical factors. The former was discussed in Section 2.2 "Programming environment", while the latter derives from the limited surface and resolution available on typical computer displays, restricting the detail and overview capacity of representations. About 3000 article-length documents such as in Fig. 6 might be shown with reasonable legibility on the 3072 × 1920-pixels screen of a MacBook Pro laptop. To represent more documents at higher resolutions, it is suggested to make hard copies and display them on walls, a technique used by graphic designers for magazines, and architects and engineers for plans. This approach was also adopted by the authors to compare document structures, yielding additional benefits of persistent visualizations and more space to accommodate a larger group of viewers (Fig. 5).

3. Illustrative example

The Document Towers visualizations shown in Fig. 6 represent eighty-nine PDF documents in a folder ordered by filename on one of the authors' computers; here, the blue slabs represent the location and extent of raster images in the digital documents. The image exemplifies the following three benefits of visualizing document structures. All of these are serendipitous discoveries made by visualizing the document with the Document Tower software.

A. Misclassification: While this collection was supposed to contain only articles, the presence of a high-rising Tower reveals a monograph among them. This misclassification would most likely have gone unnoticed if not for its serendipitous discovery due to visualization.

B. Search quality: The regular Document Towers visualizations represent scanned documents, where each page is a single raster image, while the fragmented Document Towers visualizations represent native electronic documents, in which images cover only a portion of pages (if present at all). For a library wishing to offer its readers searchable digital documents, the implication is that scanned documents must be identified and the text and logical structure extracted, which is time-consuming and costly. Moreover, the text recognition rate from raster images rarely yields an electronic text identical to the one in the imaged documents, particularly in the case of noisy, historical, or handwritten documents, which leads to suboptimal textual search results. Visualization, by contrast, offers a quick and lightweight means to estimate document search quality before performing text recognition.

C. Forensics: The small Tower in the upper left corner, which draws the eye due to its outlying location, is not an individual document, but rather represents images located outside the visible frame of the PDF documents. Could it conceal a hidden message?

4. Impact

The Document Towers paradigm combines document structure with an architectural metaphor. This paradigm has a potentially broad range of both applications and users, and covers the entire document lifecycle.

Its main impact pertains to the information-seeking strategy it fosters: namely, exploration as opposed to targeted search. This is of particular interest in rich, unstructured, and undocumented environments, such as archives.

Document digitization and conversion, in particular, are applications likely to benefit from visual document structure representation, as this method is cheap to implement (e.g., can be

applied before optical character recognition) and provides rich insights through non-linguistic means by leveraging human pattern recognition capabilities (helpful for, e.g., outlier detection in quality control).

The Document Towers visualizations also make information in documents visible without the need to open them. As such, they could act as document navigation aids (e.g., next to the table of contents in e-book readers). They might further help designers to overview documents they are laying out.

Last but not least, Document Towers visualizations have a certain aesthetic appeal, which is a useful user experience factor [28].

5. Conclusions

The article introduced the Document Towers paradigm, visualization, and software, designed to facilitate the visual exploration of paginated document structures using an architectural cognitive model. Beyond the specifics of the visualization and software, the usefulness of the work lies in the paradigm it promotes: that of a technological information-seeking aid complementary to linguistic and numerical targeted search.

CRediT authorship contribution statement

Vlad Atanasiu: Conceptualization, Methodology, Software, Writing – original draft, Visualization. **Rolf Ingold:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

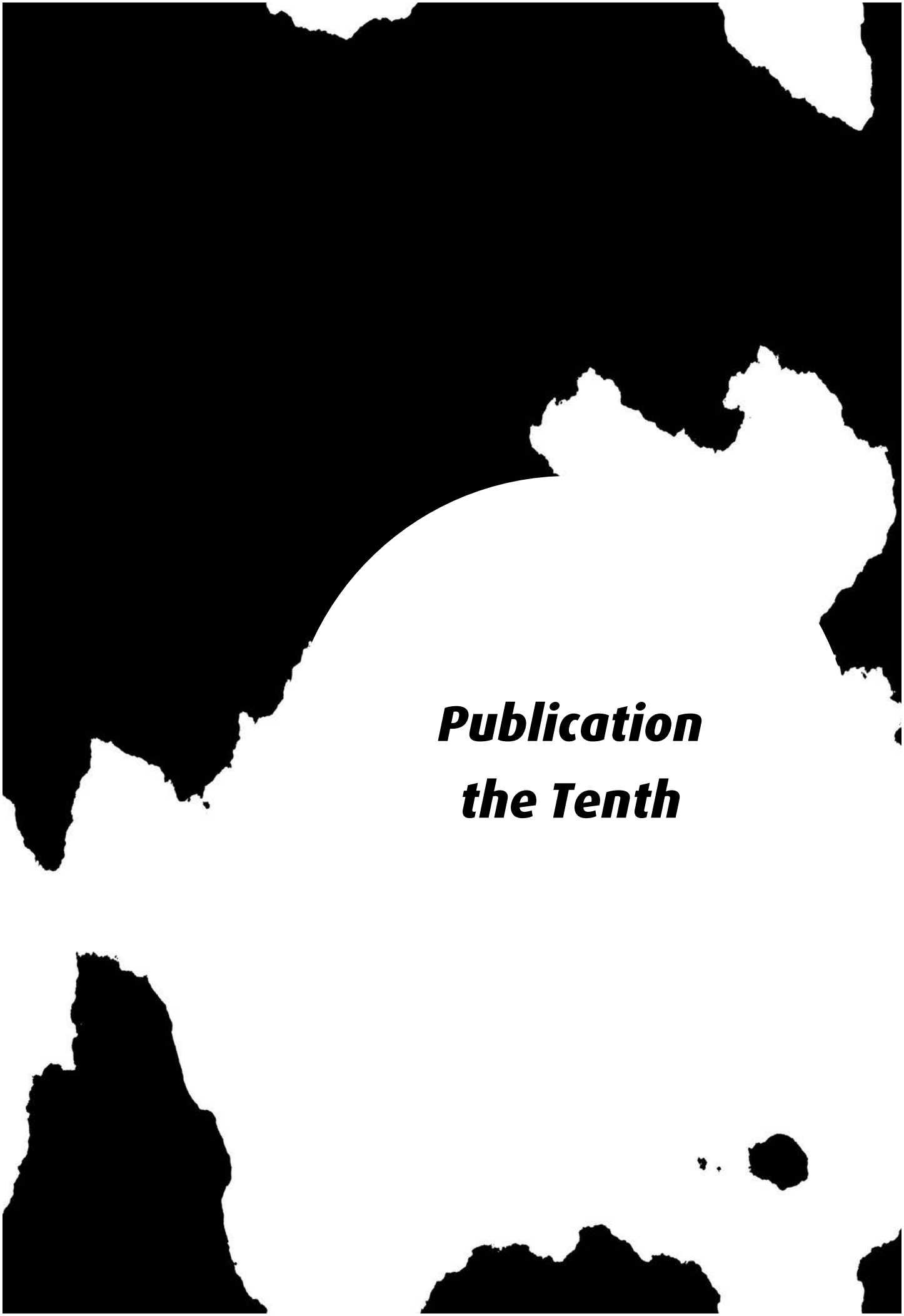
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

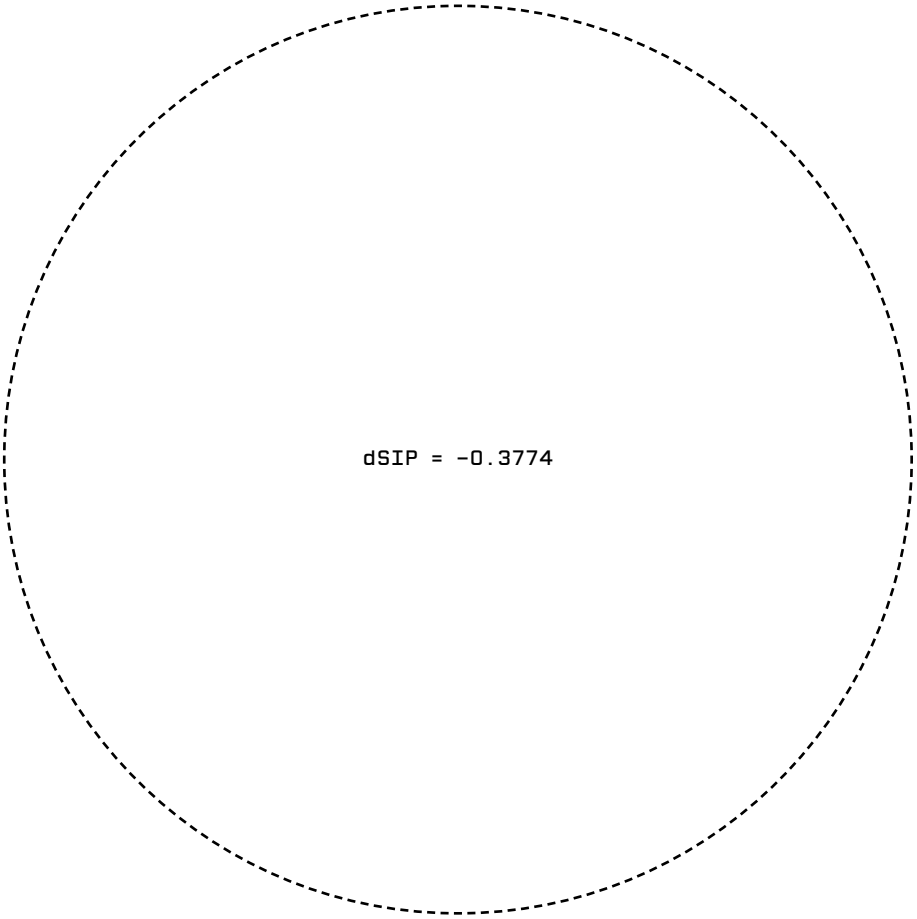
The first author expresses gratitude for the support and stimulating discussions with Andreas Fischer of the University of Fribourg, and Jean-Luc Bloeche and Maurizio Rigamonti of Sugarcube. Both authors acknowledge the improvement brought by the gracious comments of the anonymous reviewers. K. F. is thanked for proofreading. This research was supported by the Fund in Support of Innovation of the canton of Fribourg, Switzerland, under grant No. 2013.03.

References

[1] Bates MJ. What is browsing—really? A model drawing from behavioral science research. Inf Res 2007;12(4). URL http://www.informationr.net/ir/12-4/paper330.html.
[2] McCay-Peet L, Toms EG. Researching Serendipity in Digital Information Environments. San Rafael, CA: Morgan & Claypool; 2018.
[3] Fisher KE, Erdelez S, McKechnie LEF. Theories of Information Behavior. Medford, NJ: Information Today; 2006.
[4] Norman AD. Living with Complexity. Cambridge, MA: The MIT Press; 2010.
[5] Lakoff G, Johnson M. Metaphors We Live By. Chicago, IL: University of Chicago Press; 1980.
[6] Wertheim M. The Pearly Gates of Cyberspace: A History of Space from Dante to the Internet. W. W. Norton: New York, NY; 1999.
[7] Alam S, Boccuzzo S, Wettel R, Dugerdil P, Gall H, Lanza M. EvoSpaces: Multi-dimensional navigation spaces for software evolution. In: Lalanne D, Kohlas J, editors. Human machine interaction: Research results of the MMI program. Berlin: Spinger; 2009, p. 167–92.
[8] Mozilla developers. Firefox developer tools: 3D view [cited 2021.02.14]. 2021, URL https://developer.mozilla.org/en-US/docs/Tools/3D_View.
[9] Small DL. Rethinking the Book. (Ph.D. thesis), Cambridge, MA: MediaLab, Massachusetts Institute of Technology; 1999, No. 951.
[10] Almeida RAd, Cubaud P, Dupire J, Topol A. Interactions et métadonnées riches pour les bibliothèques numérisées. Doc Numér 2006;9(2):83–109. http://dx.doi.org/10.3166/dn.9.2.83-109.
[11] Selfridge P, Kirk T. COSPACE: Combining web browsing and dynamically generated, 3D, multiuser environments. Intelligence 1999;10(1):24–32.
[12] Stoffel A, Kinnemann H, Spretke D, Keim DA. Enhancing document structure analysis using visual analytics. In: Proceedings of the 2010 ACM Symposium on Applied Computing (SAC), 22–26 March 2010, Sierre, Switzerland. 2010, p. 8–12. http://dx.doi.org/10.1145/1774088.1774091.
[13] Humphreys A, Spencer C, Brown L, Loy M, Snyder R. Reimagining the digital monograph: Design thinking to build new tools for researchers. J Electron Publ 2018;21(1). http://dx.doi.org/10.3998/3336451.0021.102.
[14] Dodge M, Kitchin R. Atlas of Cyberspace. Harlow: Addison-Wesley; 2001.
[15] Card S. Information Visualization. In: Sears A, Jacko JA, editors. The Human–Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. 2nd ed.. New York, NY: Lawrence Erlbaum; 2014, p. 509–43.
[16] Fang X, Jacquemin C, Vernier F, Luo B. A survey of 3D document corpus visualization. Inf Technol J 2009;8(1):1–15.
[17] MathWorks. Matlab Compiler [cited 2021.01.03]. 2021, URL https://ch.mathworks.com/products/compiler.html.
[18] Oracle. JavaFX [cited 2021.01.03]. 2021, URL https://docs.oracle.com/javase/8/javase-clienttechnologies.htm.
[19] Threejs. JavaScript 3D library [cited 2021.01.03]. 2021, URL https://threejs.org.
[20] Unity. Unity: Real-Time Development Platform [cited 2021.01.03]. 2021, URL https://unity.com.
[21] AutoDesk. AutoCAD [cited 2021.01.03]. 2021, URL https://www.autodesk.com.
[22] Ingold R, Bloechle J-L, Rigamonti M. Reverse-engineering of PDF files. In: Chaudhuri BB, Parui SK, editors. Advances in Digital Document Processing and Retrieval. Singapore: World Scientific; 2014, p. 249–85.
[23] Systems A. IDML file format specification, version 8.0. San Jose, CA: Adobe Systems; 2006, URL https://www.adobe.com/content/dam/acom/en/devnet/indesign/sdk/cs6/idml/idml-specification.pdf.
[24] The Library of Congress. ALTO: Technical Metadata for Layout and Text Objects [cited 2021.01.03]. 2021, https://www.loc.gov/standards/alto/.
[25] JSONorg. Introducing JSON [cited 2021.01.03]. 2021, URL https://www.json.org/json-en.html.
[26] Sugarcube. Home page [cited 2021.01.03]. 2021, URL https://www.sugarcube.ch.
[27] Atanasiu V. Expert Bytes: Computer Expertise in Forensic Documents – Players, Needs, Resources and Pitfalls. Boca Raton, FL: CRC Press; 2014, p. 208, URL https://www.amazon.com/Expert-Bytes-Expertise-Documents-Resources-ebook/dp/B00OD404IM/.
[28] Turner P. A Psychology of User Experience: Involvement, Affect and Aesthetics. Cham: Springer; 2017.



Publication the Tenth


dSIP = -0.3774

Vlad Atanasiu

Improving the Quality Control and Exploration of Digital Documents Using Visualization and Serendipity

Abstract

This article contributes to the quality control of digitization workflows, and to the exploitation of historical document collections, by demonstrating the utility of serendipitous exploration (as opposed to targeted searches) through the use of Document Towers, a visual representation of the physical structure of documents as architectural models. The Document Towers are evaluated via a qualitative case study on historical Swiss newspapers, a thought experiment comparing them to alternative solutions, a diagrammatic visual and numeric assessment of exploratory tasks, and several quantitative and empirical usability measurements and psychometric surveys. The experiments confirmed that both serendipitous exploration and the Document Towers visualization are objectively well-suited for the quality control and exploration of digital documents. A significant disparity in subjective usability was observed between librarians and academics, who rated the visualization concept below and above average, respectively. Other findings included the redefinition of quality control as a tool of knowledge and dialogue among stakeholders in information systems; a generic diagrammatic instrument for evaluating the outcomes of explorations; demonstrating how the adoption of novel information technology may benefit from adaptation to individual psychologies and socio-professional contexts; and, unexpectedly, novel historical insights into Swiss map-making, newspaper history, and censorship.

P10 – 175

Introduction

> *"The greatest value of a picture is when it forces us to notice what we never expected to see" — John W. Tukey, statistician (Tukey 1977, p. vi)*

Goals and contributions — This article intends to advance the quality control of mass digitization workflows, specifically by empirically demonstrating the utility of serendipitous exploration (as opposed to targeted searches) for information systems. To this end, a series of experiments are carried out to evaluate the technology developed to enable the serendipitous information-seeking strategy, a paradigm named Document Towers. The Document

Towers are a literal interpretation of the concept of Information Architecture, in that they visually represent the physical structure of digital documents in the form of architectural models. The power of these visualizations lies in their potential for surprise, and hence, informativeness.

The contributions of this article extend beyond its initial goals to cover some unintended research outcomes. In the course of evaluating the Document Towers for quality control, their potential for historical research was discovered; thus, the case study on quality control was supplemented by one on the historical exploitation of digital archives. Furthermore, a link among quality control and historical research was identified, which lead to a redefinition of quality control as a task integrated with other dimensions of digital archives (such as the relationship between service providers and users in shaping technologies). Other unexpected contributions were novel historical findings, and the development of a formal diagrammatic method for the evaluation of explorations. Perhaps a more far-reaching discovery was that of substantial differences between librarians and academics with respect to the usability of information technologies with similar characteristics to those of the Document Towers; these findings may be ascribed to differences in individual psychological styles and socio-professional contexts. All these insights helped redefine the theoretical and practical perspectives on quality control with respect to the importance of serendipity as a tool of knowledge, as a dialogue among stakeholders, and as sensitive to social and technological contexts.

Relevance — The broad perspective according to which "the shape-shifting quality of mass digitization, and its social dynamics, alters the politics of cultural memory institutions" (Thylstrup 2019, pp. 3–4) is contingent on the usability of digital libraries and archives, which is itself commensurate with the quality of data as ensured by quality control. The role of quality control is also in an operational sense larger than any individual stage in a digitization workflow: it can further be instrumental in the identification of user demands and service opportunities, and can intervene in shaping the very architecture of information systems.

In terms of functionality, interaction with digital libraries and archives is dominated by targeted search strategies, which compel users to articulate clear ideas about what is being searched and to express themselves in ways that are comprehensible to the computerized information system. Reality, however, is less clear-cut, and information-seeking and historical research often incorporate imprecise and unknown information. As will be further explored in this article, novel and different strategies are thus required, such as panoramic overviews, serendipitous insights, open-ended exploration, affective human–technology bonds, and complementarity between automated

and interactive approaches.

The adoption of these technologies is, however, contingent on their subjective usability, independent of their objectively demonstrable qualities. It is therefore desirable that such technologies benefit all stakeholders in information systems (data providers, technology providers, and users). Notably, our empirical findings suggest that this is not always the case; hence, another significant aspect of the present article is that it highlights the role of personalized information technologies, along with the need to determine how to adapt socio-professional contexts to make them more receptive to technological progress.

Reading keys — As the above remarks make clear, one way to benefit from this text with multiple disciplinary perspectives is to generalize and transfer the knowledge derived from it. For example, the antiquarian investigation into Swiss map-making, along with the minutiae of Document Towers usability statistics and user psychometrics, have value beyond their own specific scopes; we encourage the reader to apply the methods of generalization and analogy while engaging with this article.

Readership — This article is of equal interest to librarians and archivists who design and supervise mass digitization projects, and to computer scientists and service providers who create software for the quality control and exploration of digital libraries and archives.

Background — The work here described builds on prior collaboration between the author, the case study participants, and their respective institutions (the Department of Informatics of HES-SO//FR and University of Fribourg and the Swiss National Library in Berne). It is intended as a pilot study to inform future, larger information systems projects.

Terminology — Technically, the case study deals with collections of digitized historical newspapers, called newspaper "archives" by the curating institution, which is a library. As many para-textual investigations of the documents were performed for this study (on material, technological, and institutional aspects) the approach is arguably closer to research in archives than in libraries, resulting in the use of the term "archives".

As for the term "document" — which, famously, can designate even a living antelope if it becomes part of a zoological collection, according to the French documentalist Suzanne Briet (Briet 1951, pp. 7–8) — the focus is on paginated documents (codices), although any information with a spatial extent can in principle be represented by Document Towers.

The term "serendipity" was originally coined in 1754 to describe two fictitious princes that were "making discoveries, by accident and sagacity, of things they were not in quest of" (Merton & Barber 2004, pp. 1–3; McCay-Peet & Toms 2018). A practical (albeit unhistorical) image to keep in mind while reading this article, with respect to the distinction between "exploration" and "search", is that of Marco Polo riding wide-eyed into the unknown Orient, and of Livingstone, wearing blinders, to keep him focused on finding the source of the Nile.

"Socio-professional" (a French loanword) designates social characteristics of professional environments, such as market pressures for technology companies and publishing pressures in the academic world.

Organization — The article begins by providing the background to the case study in terms of defining quality control for document digitization in general and at the Swiss National Library (Swiss National Library n.d.) in particular, then discusses the state of the art in digitization quality control (Section 2). Next, it introduces the Document Towers paradigm, explains how to use it, and compares it to existing solutions (Section 3). The evaluation experiments (Section 4) make up the bulk of the article. A discussion of the research outcomes concludes the article (Section 5).

Further reading — The role of exploration and serendipity in libraries, and the digital technologies supporting them has been surveyed by the author in (Atanasiu 2022a, Annex). For technical aspects related to the Document Towers, the rationale of its design paradigm, its technological and cultural background, and the utility of document structure representation in a multifarious range of applications, the reader may refer to the author's dedicated publications (Atanasiu 2022a; Atanasiu and Ingold 2021).

Background on Digitization Quality Control

Knowns — Quality control of document digitization consists principally in checking that all required document elements (e.g. titles, paragraphs, notes, styling, cross-references, tables, illustrations) have been correctly identified, that there are no missing, double, or faulty scanned pages, that the page order is correct, that tables are not wrongly segmented, that captions are linked to pictures, that pictures have the required resolution and color profile, and so on (Riley & Whitsel 2005, pp. 41–43; Almeida et al. 2009, pp. 150–151, 153; Chapman & Leonard 2013, p. 406). The workflow is subject to a number of strategic tasks, notably planning acquisition and allocation of resources (e.g. personnel, costs, time, know-how, software, hardware), monitoring digitization status and progress, and communicating between team members, with

the management, and with external service providers or institutional partners.

Unknowns — In addition to these well-defined tasks, there are quality-related aspects that cannot be foreseen and are therefore not included in the service provider specifications or captured by off-the-shelf software — these are discovered (or not) by chance during the quality control process. This case study will present a number of examples.

Automation and Manual Control — The conformance of file format, image aspect ratio, and colorimetric parameters are examples of digitization aspects whose control can be easily automated. Statistical sampling techniques further reduce the workload. Practitioners recognize, however, that manual intervention remains necessary and that achieving high quality results is a considerable time and finance drain (Riley & Whitsel 2005, p. 43; Almeida et al. 2009, p. 151). The workload can range from ten percent of the digitized batch (for controlling simple image parameters) to its entirety (such as for historical documents with pages that are brittle or stick together, or dictionaries, where character formatting is functionally meaningful (Chapman & Leonard 2013, p. 408)). In addition to the identification of individual errors, formal methods must also be implemented to evaluate errors in terms of type, quantity, and relevance. The experience gained from manual control steps should flow into the development of the automatic quality control. Digitization and conversion projects are burdensome endeavors, and it is best to solve issues early on; e.g. to avoid "legacy bugs", which are difficult and costly to eradicate once contracts are signed, or projects finished and programmers departed (Steffen 2016, pp. 2–3, 7).

Requirements — Decisions regarding the components of the quality control process are, as apparent from the above paragraphs, made primarily on technical and administrative grounds. Libraries and other document repositories do also involve the end-users in the planning process to improve the utility of public services and augment market revenues (Serenson 2000). User requirements, however, evolve over time, according to cultural, social, financial, and political interests (Buckland 1991, p. 55–67), sometimes with a consequential impact on quality control.

The Swiss National Library (SNL), in occurrence, has collected only a limited number of samples from the total output of Swiss newspapers since its founding in 1895, before switching to systematic collection during the mid-20th century, resulting today in substantial digital collation work and administrative overhead for various libraries in order to retroactively fill the gaps in newly designed online portals. The respective status of text and image parallels the ancient debate regarding the merits of orality and literacy

(Ong & Hartley 2012), in that images were scares in newspapers until fairly recently (e.g. the late adoption of photography by major newspapers, despite the early success of the illustrated press) and in need of study (cf. the history of maps in newspapers, which remains an underexplored domain of the otherwise vast fields of cartography and media studies). In this context, the SNL deemed it unnecessary to microfilm advertisements in newspapers until their historical interest was acknowledged in the 1980s, meaning that part of the current quality control effort consists in identifying missing advertisements.

Summary — The quality control of document digitization and conversion is resource-consuming, experience-based, not necessarily methodical or efficient, and at times even frustrating. Potential improvements are numerous: provide transparency regarding decisions made about technological solutions; allow searching for known issues, as well as exploration of unknown issues; support quantitative and qualitative issue evaluation; facilitate overview, planning, monitoring, and communication. Ideally, the solutions should be effective in on-site conditions, polyvalent with respect to users and tasks, and lightweight in terms of development costs, maintenance, and learning curve.

State of the Art

There are three different states of the art in document digitization quality control, depending on whether one focuses on the document repositories sector, the digitization and conversion industry, or the academic research world. Their salient characteristics and interactions, with a focus on newspaper digitization (Center for Research Libraries 2015; Dunning et al. 2012), are briefly described below.

Software — Several software have been developed for the quality control and monitoring of the physical-to-digital and digital-to-digital format conversion workflows; these are typically integrated in large enterprise-level systems, such as docWorks Validator by Germany's Content Conversion Specialists (CCS) (Content Conversion Specialists n.d.), LIMB Processing by France's i2S (i2s n.d.), and Kitodo Meta and Flow, an open-source project of German-speaking libraries funded by the German Research Foundation (DFG) (Kitodo n.d.). The main role of the modules with respect to quality control is to check the conformance of XML schemes and image quality aspects such as color profiles, compression level, and resolution. Analysis results are reported through infographics, and some corrective measures can be automated. The main drawbacks are that document structures remain invisible and the solutions are costly and complex.

Standards — In partnership with the industry, academia, and funding agen-

cies, libraries have developed various de facto document description standards, including ALTO (Analyzed Layout and Text Object), originating in the META-E project (2000–2003) co-funded by the European Commission, and maintained initially by CCS and now by the Library of Congress (Wikipedia contributors [n.d., a]), and METS (Metadata Encoding and Transmission Standard), also maintained by the Library of Congress (Wikipedia contributors [n.d., b]). Since the 2000s, the Australian National Library (Wikipedia contributors [n.d., e]) and the pan-European Europeana Newspapers project (Europeana Newspapers n.d.) have pioneered many aspects of newspaper digitization and conversion, including the adaptation of ALTO and METS, best practice metadata recommendations, and computational analysis and workflow support methods. An interesting aspect of the German pilot project as regards planning a national comprehensive newspaper digitization program (2013–2015) is that apart from providing an overview of the status quo and proposing recommendations, careful consideration is given to local organizational, cultural, and other particularities, in addition to striving for generic technical solutions (Staatsbibliothek zu Berlin 2017, pp. 5–6). As standards in professional digitization projects, the ALTO/METS formats are interesting for generating Document Towers, but are not common formats in which digital documents are created and exchanged (such as Microsoft Word for office documents, Adobe InDesign IDML in the publishing industry, PDF for interchange and archiving, EPUB for general readers, LaTex for scientific publications, and bitmap images for scanned documents). Formats are important, since they differ in terms of the type and quality of information they carry.

Research — Academic research has been involved in several of the software and related projects discussed above. For example, applied research carried out within the Europeana Newspapers project yielded in methods and software for workflow quality prediction (Clausner et al. 2016), document structure viewing (Mühlberger & Hackl 2015, pp. 56–59), recording image processing steps (Pletschacher & Antonacopoulos 2010), ground-truthing (Clausner et al. 2011), and OCR performance analysis (Clausner et al. 2015), as well as datasets for document structure ground-truthing (Clausner et al. 2016). The theoretical basis derives from long-standing and diverse fields, such as statistical process control in industrial and management settings (Oakland 2008), data quality in the field of "big data" (Sadiq 2013), outliers theory in statistics (Barnett & Lewis 1978), and misclassification research in pattern recognition (Atanasiu 2016, p. 2).

Vannevar Bush's Memex reading machine of the 1940s and Xerox's windows and graphical user interface of the 1970s are two examples that achieved notoriety while being representative of the research on document and library visualization carried out in the information visualization field (Bush 1945;

Card 2008; Fang et al. 2009). Several related techniques have been adapted to digitization quality control, such as the software–hardware hybrid prototype developed at the Conservatoire national des arts et métiers (CNAM), Paris, consisting in a zoomable juxtaposition of page thumbnails projected on a cabinet-sized half-sphere (Almeida et al. 2009, 2006). Nonetheless, much work remains to be done with respect to visual quality control solutions.

Proposed Solution: The Document Towers

> *"The good bookstore sells books, but its primary product, if you will, is the browsing experience" — Jeff Deutsch, bookseller (Deutsch 2022, p. 24)*

> *"It is probably wise to include a random element in a learning machine" — Alan Turing, computer science pioneer (Turing 1950, p. 459)*

Theory: Understanding the Document Towers

Paradigm — Document Towers were created by the author as a means of gaining an overview of the layout of digital documents without the need to browse them page by page, a laborious exercise. More generally, the Document Towers are a visual discovery technique for patterns and singularities in digital documents. The Document Towers represent the three-dimensional physical structure of documents: that is, the location, size, and shape of text paragraphs, bitmap images, vector graphics, page boundaries, colors, typefaces, keywords, and indeed any entity that can be attributed a spatial extent. The resulting visualization resembles an architectural model of a building, hence its name.

Rationale — The physical patterns revealed by the Document Towers are traces indicative of the document's logical structure, semantic content, and meta-information. This valuable equivalence of form and function is the principle that enables the user to become informed through observation, by interpreting what is seen rather than by querying that which might not even be there. The essence of this exploration is serendipity, and the Document Towers foster this type of information-seeking strategy. For this reason, they seemed an appropriate solution for tasks such as quality control and historical research, where serendipity — or surprise and high entropy, in information-theoretical terms — is expected and welcome. One goal of the present article is to test this hypothesis.

Implementation — The Document Towers paradigm has been implemented in a software program, called Crystal, because in a crystal labyrinth you cannot loose yourself, unlike in paper or digital documents, made of opaque or

intangible matter. Crystal takes as input a JSON-formatted file containing the coordinates and labels of entities extracted from documents, renders them as Document Towers, and provides interaction capabilities, such as filtering, zooming, panning, rotation, and various measurements. Crystal extracts the label and geometry data from IDML and ALTO files; to process PDF documents, an API for the Enlighter software by Sugarcube (Ingold et al. 2014; Sugarcube n.d.) is used, while for document images, OCR software such as ABBY FineReader is employed.

Comparison — There are a number of known issues with the use of visualization for the quality control of digitization processes (Almeida et al. 2009, pp. 151–151). The most common and consequential of these is the difficulty of concomitantly maintaining an informative degree of overview and detail on a sufficiently large set of documents (Cockburn et al. 2008). A typical quality control interface can provide great page-level detail, but an overview of the dataset is limited to a few thumbnails of the current document under examination (fig. 1). An increase in the number of thumbnails, usually accompanied by a reduction in their size, easily leads to crowding and loss of insight. A further issue is the physical manipulation of the data, especially in the case of highly dynamic visualizations that necessitate human–machine interfaces more sophisticated than a keyboard and mouse. By way of illustration, the use of a isometric joystick with six degrees of freedom to zoom in and out of the documents is notoriously difficult for first-time users, while scrolling the thumbnails in the viewport can be disorienting, just as when webpages are scrolled too fast (Igarashi & Hinckley 2000; Almeida et al. 2009, pp. 153–155). Among the most common software solutions are space-filling thumbnails (Cockburn et al. 2006) and dynamic and interactive visualizations, in addition to experimental techniques such as fish-eye view (Bederson et al. 2004), perspective view (Reinfurt & Wiesenberger 2014), rapid serial visual presentation (Back et al. 2006; Spence & Witkowski 2013), and augmented reality (Lee et al. 2013).

How do the Document Towers deal with the visualization problems and prospects described above? In short, through the three-dimensional stacking of pages, they preserve the natural structure of paginated documents, making it easier to identify (for example) misaligned, misplaced, missing, or spurious elements that are repetitive and appear at fixed locations, such as page numbers, chapter headings, or pages left blank. Three-dimensional stacking also increases spatial information density and the range between overview and detail. The simplification of data from detailed images to a wireframe of entity bounding boxes diminishes crowding. The color-coding of physical, logical, and semantic entities facilitates data classification, and thus expands the type of insights that can be obtained beyond what unmarked thumbnails can provide. Interaction issues are avoided so long as static views of document

collections and items are sufficient for their exploration. The interface does however have zooming capability and supports filtering via on/off criteria selection. Through techniques such as representing paper sheets as slabs and organizing the document structures in city-like grids, the Document Towers portray documents as buildings and libraries as cities. Familiarity with the target domain of the architectural and urban metaphor utilized by Document Towers contributes to make the abstract entities (document structures) more affordable and hence facilitates insight discovery. The use of a metaphor in itself, apart from its specific embodiment, can carry a number of benefits, notably being a cognitive stimulant if it can intrigue the user and thus lessen the tediousness of a task such as interactive quality control.

Automation — It is possible to combine visual analytics and computational pattern analysis within the Document Towers. For example, the classification of layouts based on raster images of the document pages is useful for interpreting scanned documents in cases where segmentation into objects is not available. The numeric values corresponding to the layout types can be color-coded and displayed on the external "walls" of the Document Towers. Even if object coordinates are present in the digital document files, their spatial distribution patterns may be displayed in a more compact and legible way as color-coded "Ribbons". Figs. 6 to 8 present layout characteristics of sample documents using the combined visual–numeric approach. For technical details on the measurements, see (Atanasiu and Ingold 2021) and (Atanasiu 2022b).

P10 – 184

Media — The Document Towers may be implemented in a variety of media, the usual form being that of an architectural wireframe model. These representations can further be augmented by color-coded Ribbons, which may be engraved on a physical medium to facilitate document navigation for users with low vision. Other embodiments, such as sound objects, could be useful in special applications (Atanasiu 2022a).

Art — Last but not least, the Document Towers are not devoid of a certain aesthetic appeal. They may be simply enjoyed "as is", or could alternatively be employed for communication purposes (e.g. as paintings or sculptures in the lobby of a library) to showcase their contents to a broader public audience in a new light.

Gallery: Visual Exploration of the Document Towers Universe

This pictorial section presents some of the dominant digital document exploration paradigms in response to which the Document Towers emerged, and illustrates the broad variety of their embodiments and applications.

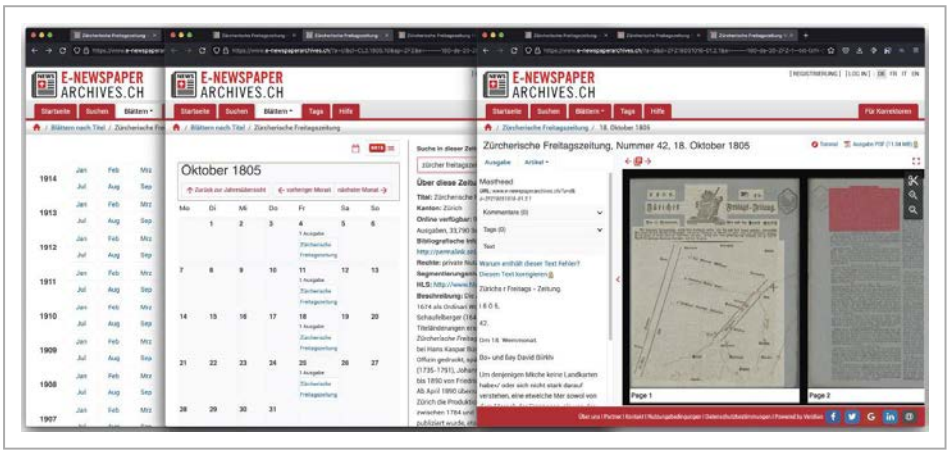


Figure 1. Logical navigation

Logical navigation — Time is an essential dimension of periodicals, such as newspapers; thus, it is intuitive to design the interface to a digital newspaper collection based on a temporal navigation principle. In this picture, we see screenshots from the e-newspaperarchives.ch website, from which we drew content for our case study. To access a specific issue, the users first scroll through the list of years and months, then the list of days, after which they slide individual pages within the viewing window, move the cursor over paragraphs and pictures to highlight available objects (in red on page 2), and finally click on them to read the text or the metadata (highlighted on page 1). While this navigation principle is logical, it is not particularly efficient for browsing a large document collection: it is abstract, requires too many manipulations, and offers only a limited area for viewing the actual document image. Note that this interface is very similar to that used by the library employees for the quality control of this dataset.

Logical structure — Figure 2 shows a screenshot of a typical viewer of the logical structure of PDF files (Apache PDFBox's PDFDebugger (Apache Software Foundation n.d.)). The information pertains to the document pictured in Fig. 1. The CropBox object (highlighted) contains the coordinates of the area visible to document readers. As the name of the software suggests, it is mostly useful to professionals searching for issues in digital files.

Figure 2. Logical structure

Content visibility — While physical libraries may be impressive works of art and architecture, as well as splendid places to work in or visit, their content can be as inscrutable as electronic documents, as most titles on the book spines are too far away to be legible. Science fiction has found a solution to this problem (fig. 3): information about the content of the works is communicated via color codes and dynamically displayed on the luminous spines of holographic books, as envisioned in "Star Wars" (Episode 2, "Attack of the Clones" (2002); (Fandom contributors n.d.)). As cover design is one of the factors considered when prospective readers select books in libraries and bookshops (Lador 1990; Kawaguchi & Suzuki 2018; Ooi 2019, pp. 81, 96), it is easy to imagine electronic book spines that personalize their appearance, enticing the reader to "Come hither!" in a manner reminiscent of the holograms that emerge like fairy-tale djinns from the giant outdoor advertisements in 2017 "Blade Runner 2049".

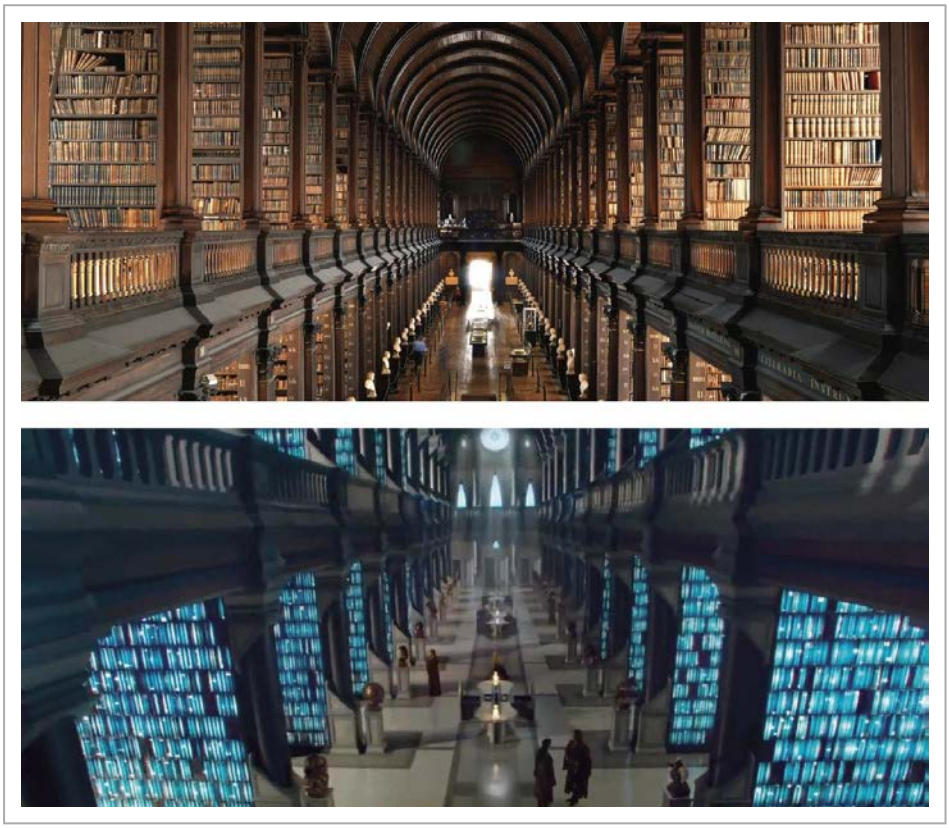


Figure 3. Content visibility: Trinity College Library, Dublin (top) and Jedi Temple Library, Coruscant (bottom). Images: Irish Welcome Tours and 20th Century Fox / Lucasfilm.

Spatial structure — Figure 4 shows a view of the user interface of the Crystal software that implements the Document Towers concept (Atanasiu & Ingold 2021).

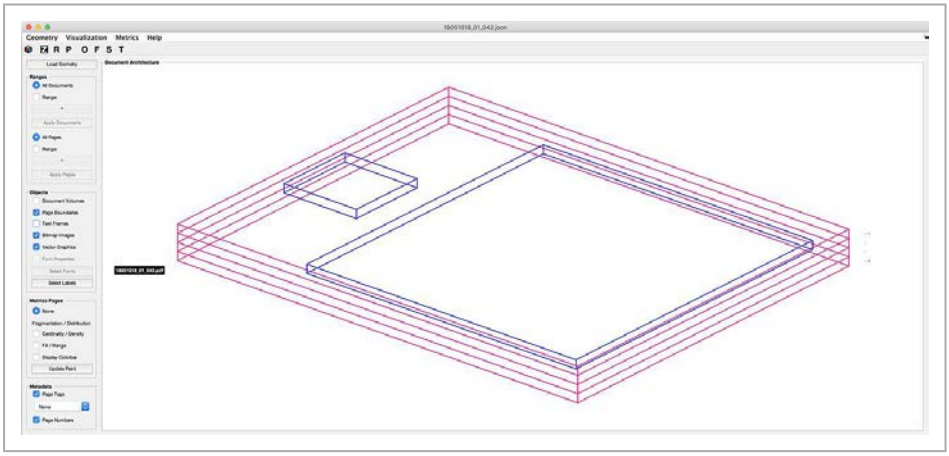


Figure 4. Spatial structure

It represents the spatial structure of the newspaper shown in Fig. 1 in the form of an isometric projection, as described by sets of coordinates in the document's ALTO file. There are four pages in this newspaper, the physical boundaries of which are rendered top-to-bottom by stacked red slabs. The two blue objects represent the two bitmap images on the first page. Also available for visualization through the interface menu are the bounding boxes of text paragraphs. The document model can be zoomed in and out, panned, and rotated. The software provides a schematic representation of the document layout in a single view.

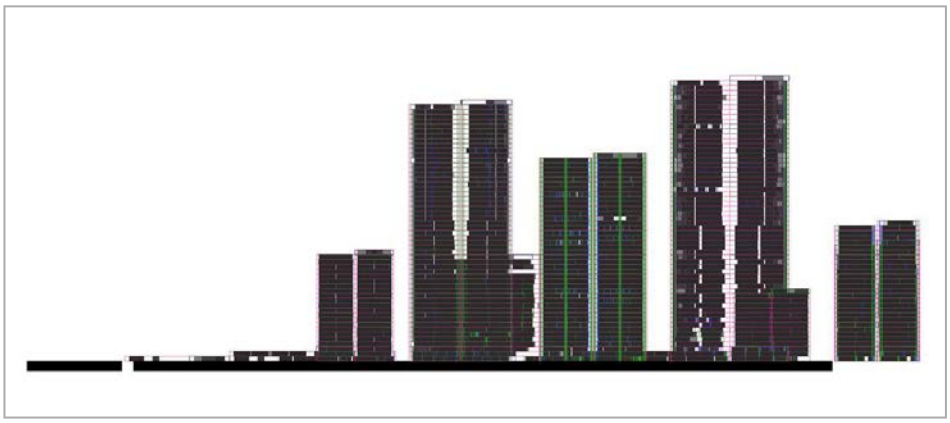


Figure 5. Urban metaphor

Urban metaphor — By exaggerating the width of a single page and the other objects on its surface, making them transparent or reducing them to a wireframe, and retaining the double-page structure of physical documents, objects that resemble architectural models of buildings and cities can be produced. This picture, entitled "Kendall Square" and derived from catalogs of the MIT Press, recalls an urban skyline. The Document Towers concept uses the document-as-a-building and the library-as-a-city metaphors to facilitate thinking about document collections (which in design terms is called "to increase affordability"; (Nielsen 1993)).

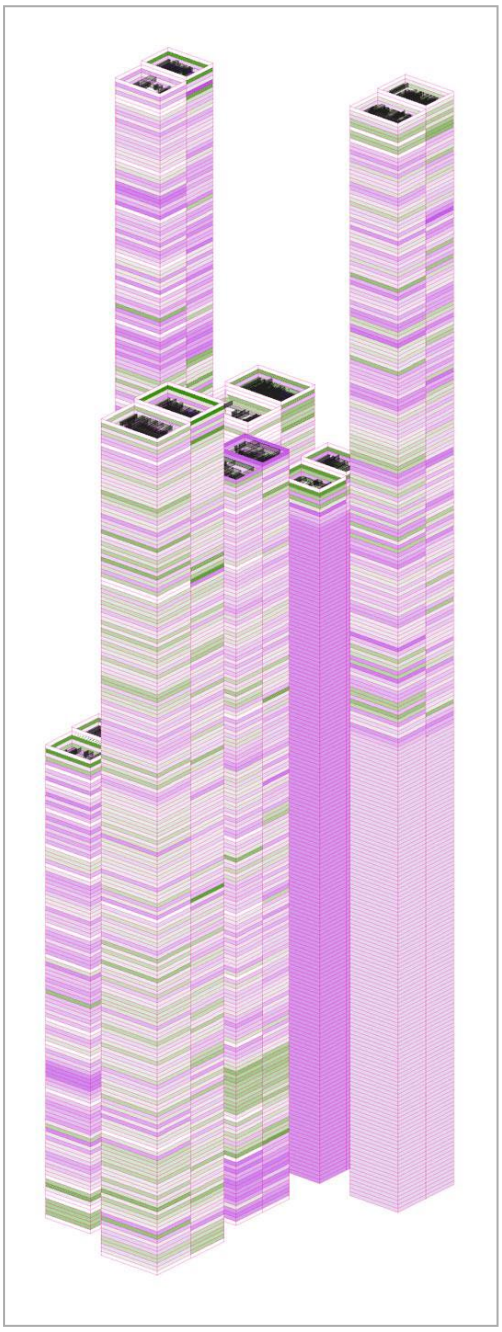


Figure 6. Architectural model

Architectural model — Figure 6 illustrates how variations of the Document Tower visualization technique may evoke various different architectural styles. For instance, these scholarly books produced by the École pratique des Hautes Études, Paris, recall the ceramic tiles covering the ziggurats of Babylon, the "Towers of Babel" so to speak. From a technical perspective, the colors encode the homogeneity of page layouts (with red indicating greater

homogeneity). (Documents courtesy of Laurent Garrigues)

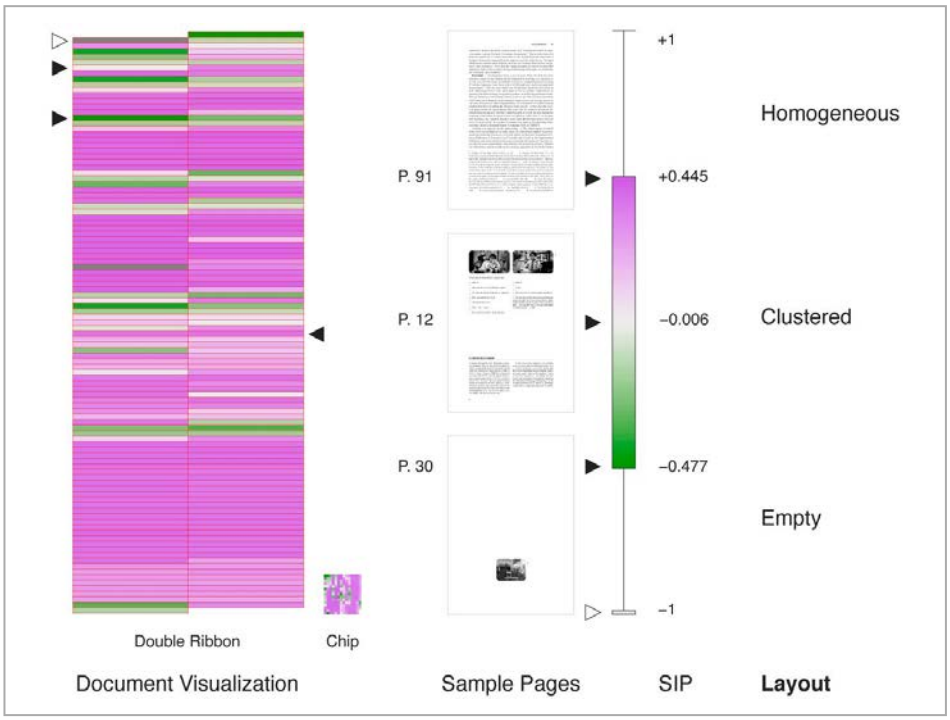


Figure 7. Semantic color-coding

Semantic color-coding — Figure 7 demonstrates how the computational pattern analysis of documents may be combined with visual analytics. In the first step, the type of page layout (varying between homogeneous, clustered, and empty) was characterized numerically using the structural information potential (SIP) algorithm (Atanasiu 2022b). In the second step, the resulting values were color-coded and rendered both as a "Double Ribbon" (for left and right-hand pages) and as a more compact "Chip" to facilitate further visual inspection. The document represented in Fig. 8, one of the author's books, was further analyzed in Document Tower form in (Atanasiu & Ingold 2021). From the visualization, we can see that chapters of text with homogeneous layout (in red) are interspersed with image sections characterized by a clustered layout (in white), while there is one mostly empty page between sections (in green). The red stretch at the end of the book (the lower part of the Ribbon) represents the visual uniformity of the bibliography and index. A correspondence exists between the visual and semantic structure of the pages: visually clustered pages usually contain semantically differentiated and hierarchized content (such as title, paragraphs, footnotes, figures, etc.), homogeneous pages tend to have a more limited semantic focus (signified by the absence of titles), and pages on which "ink pixels" are concentrated offer little information, being largely empty. Jan Tschichold, one of the

most influential Modernist typographers, referred to irregularly shaped layouts as "asymmetric typography" (Tschichold 1995). In information-theoretical terms, "asymmetric" pages are potentially more informative because the size range (and thus types) of objects extend over multiple scales, similarly to fractals. The same principle can be applied when interpreting Document Towers: in short, irregularities are perceptually salient. This is why the outsized map in the newspaper collection shown in Fig. 12 could be easily detected.

Figure 8. Tactile information

Tactile information — Information derived from processing digital documents can be displayed on the spines of physical documents. Here, the author's book (Atanasiu 2014) is augmented by a "Ribbon" representing the size of bitmap images on each page. To make this information accessible to persons with low vision, the Ribbon was engraved with a laser cutter on a wooden tablet.

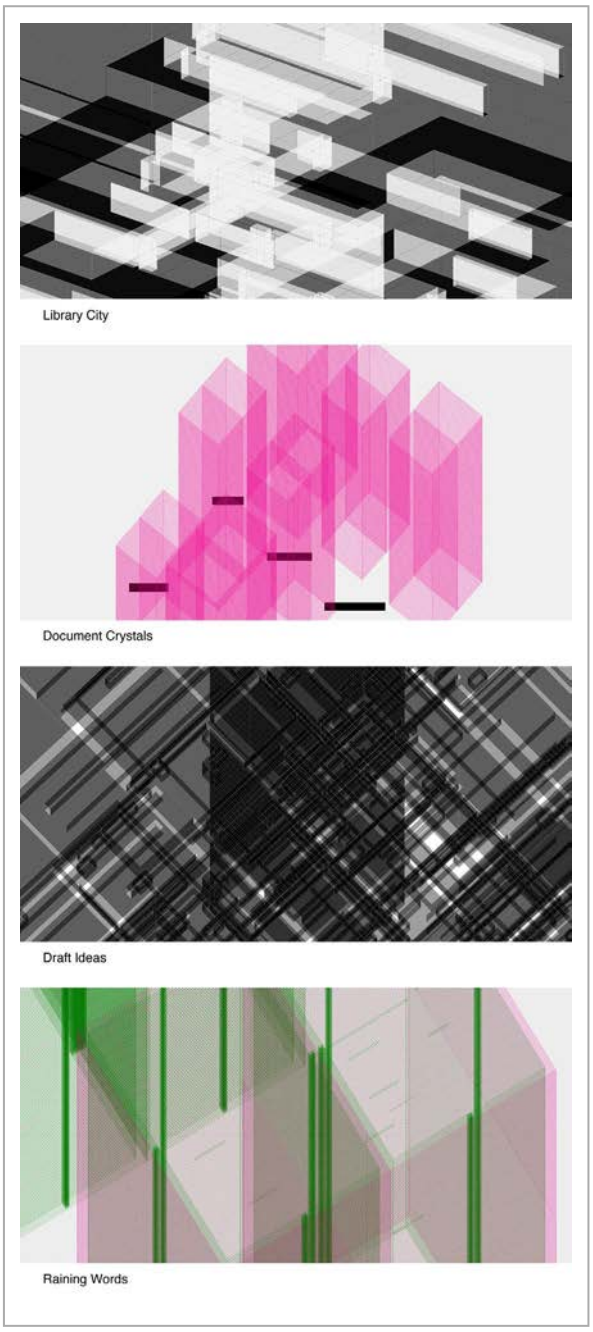


Figure 9. Poetic machines

Poetic machines — When zooming into Document Towers, very strange patterns often emerge. Such patterns tend, at least to this author, to evoke various poetical ideas through the juxtaposition of the abstract and the concrete. The ensuing synesthetic pleasure enhances one's motivation to explore these fascinating document structures at length. Apart from their oneiric qualities, the aesthetic dimensions of the Document Towers may also play a role in

improving their usability given the close connection between fascination and exploration.

Practice: How to Explore Documents in Three Steps Using the Document Towers

Figure 10 shows eighty-nine PDF documents represented as Document Towers placed on a regular grid, with each wireframe object standing for one raster image. Some notable patterns include the tall Document Tower vs the small ones, the regular vs the irregular, and the outliers; these correspond, respectively, to a book misclassified as an article, scanned vs natively digital documents, and images that fall outside the PDF viewport.

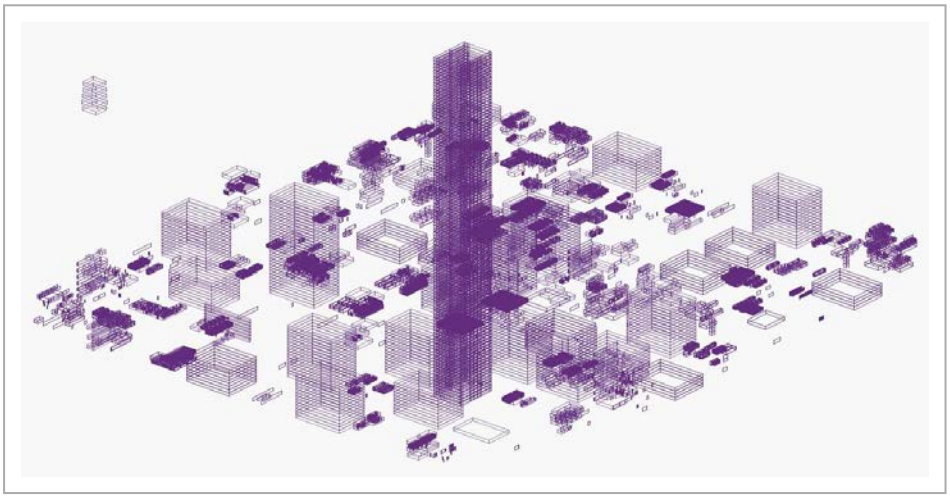


Figure 10. Eighty-nine documents as Document Towers

Exploration happens through a three-step workflow:

Observe — Observe the document visualization and identify patterns of interest. E.g. a tall Document Tower; regular vs irregular Document Towers (fig. 10).

Interpret — Determine the source generating the patterns. E.g. the tall Document Tower is a book misclassified among a set of articles; moreover, the regular Document Towers correspond to scanned documents, in which each page is a raster image the size of the physical document page, while the irregular Document Towers represent natively digital documents containing raster images located in various places throughout.

Act — Operationalize the findings. E.g. Remove the misclassified book from

among the articles and place it in the folder dedicated to books; perform optical character recognition on the scanned to document to make the text content searchable.

Patterns to look for are:

- Amounts (e.g. number of objects per page, pages per document, documents per visualized set).
- Size (e.g. page area covered by objects).
- Distribution (e.g. uniform, clustered, random, and regular patterns).
- Categories (e.g. the Document Towers may represent the bounding boxes of physical pages, raster images, vector graphics, or text blocks).
- Diversity (e.g. relative amounts of object categories).
- Outliers (e.g. object located outside the viewport of a PDF document). Absences (e.g. a stamp that is visible on the scanned page raster image, but not within the document object model).

Concepts to identify include:

- Visualization leverages the capabilities of the human visual system, and therefore differs from the numerical and linguistic analysis of documents.
- Depending on their cognitive abilities and preferences, different users may receive different benefits from visual document representations than from numerical and verbal representations.
- The spatial structure of documents can potentially provide information about their contents and other characteristics (Atanasiu 2022b).
- Exploring the spatial structure is quick (a whole document in one glance), economical (object coordinates are available in digital documents), and polyvalent (many potential applications).
- Exploration is about finding the unexpected, while searching involves finding the expected.

- The spatial structure of the same document may differ according to the digitization process and digital storage format employed (Atanasiu & Ingold 2021); moreover, its usability may vary with the user's skillset and mindset, as well as the application context.

In addition to document exploration, the Document Towers have an extensive and diverse range of applications, including document overview, navigation, quality control, misclassification detection, document design, representation for readers with low vision, and aesthetic appeal, among many others (Atanasiu 2022a).

Evaluation

Approaches — Usability is defined by the ISO standard 9241-11:2018(E) as having three facets, namely effectiveness, efficiency, and user satisfaction (ISO 9241-11:2018(E) 2018), which we address from the perspectives of different methods: qualitative, theoretical, diagrammatic, and quantitative. The choice to apply such an extensive approach was also made due to the difficulties associated with evaluating the outcomes of explorations and serendipitous activities.

The first experiment performed is a qualitative case study, comprising self-reports, interviews, behavioral observations, and historical research, the purpose of which is to evaluate the effectiveness of the Document Towers; here, the evaluation criteria are the number, type, and relevance of insights gained.

The second experiment is intended to evaluate the efficiency of the Document Towers, which we achieve by comparing its various qualities to those of alternative solutions using the method of thought experiments; this allows for the theoretical consideration of a large variety of use cases that would be impractical to test empirically.

The third experiment uses diagrammatics to facilitate an integrated evaluation of effectiveness and efficiency. Its specific contribution is to offer a solution to the problematic issue of evaluating exploration outcomes.

The fourth experiment empirically measures user satisfaction with the Document Towers, using quantitative analyses of one rating-scale-based usability satisfaction questionnaire, and furthering the interpretation of the results via two rating-scale questionnaire-based psychometric evaluation approaches.

Focus — The Document Towers are in their early stages of development, meaning that it is necessary to first evaluate the paradigm before the per-

formance of any specific implementation. The goal is now to find out what insights produce the Document Towers, with what quality, and for which applications. Therefore, it is not appropriate for an exploration endeavor to decide a priori on a specific evaluation task, as other tasks not yet imagined might be missed. This approach could be called "exploratory evaluation", in reference to the "exploratory data analysis" in statistics, first popularized by John W. Tukey, and defined in his own words as follows: "It is important to understand what you can do before you learn to measure how well you seem to have done it." (Tukey 1977, p. v).

Object — By representing the content of opaque documents, the Document Towers make the invisible visible; thus, from this point of view, their utility is self-evident, similar to the first X-ray photograph of Marie Curie's hand. The Document Towers also have polyvalent uses, so that only a case study specific to one application can be presented here. A survey of potential applications is published separately in [Atanasiu 2022a], as mentioned before.

Data — Since the evaluation tasks are quality control and historical research, any finding derived from the Document Towers is helpful, especially if significant. Therefore, the evaluation will not necessarily require large amounts of testing data.

Participants — To evaluate the Document Towers for digitization quality control and for supporting historical research in a library setting, participants with multiple skill sets are needed. That comprises digital technologies and humanities, and how such projects may be integrated in the specific missions, management, and resources of a library. Ideally, the participants should also know how to use the Document Towers to obtain insights, so that the evaluation can concentrate on the Document Towers' capacity to yield insight rather than their learning affordance. The combination of these requirements restricts the pool of respondents eligible for participation in the evaluation. While studies with small numbers of participants have low statistical power (Brysbaert 2019), and some usability evaluations involve massive numbers of users (e.g. N = 11,429 for a single study on recommendation serendipity carried out via smartphones by the Chinese IT giant Alibaba (Chen et al. 2020)), small-sample studies are an unavoidable reality in many fields and circumstances. Reasons for this may include the relative rarity of the occurrences under study (e.g. aviation accidents), the complexity of situations (e.g. in personalized medicine), ethical considerations (e.g. crimes), costs and duration (e.g. training machine learning algorithms), or issues concerning access (e.g. to archival records). The growing awareness of the relevance of such cases has given rise to a body of research focused on small sample size and single-case statistics and experimental design (Franklin et al. 2014; Morley 2017; Schoot & Miočević 2020). From this perspective, if the results are demon-

strably significant for the end users, tasks, or contexts, then a small number of participants is both sufficient and justified. Given the stated constraints on our subject matter, we accordingly adopt this approach as the basis for the four distinct evaluation methods used in this article ($N = 3$ in the first experiment, and $N = 21$ in the fourth experiment). As a case in point, while our first case study has just three participants, their use of the Document Towers was found to generate a broad range of insights about the test documents, which are further analyzed for their historical relevance to a depth that would not be possible for practical reasons with a large participant sample.

Tasks — As previously stated, the quality control of document digitization processes comprises well-defined issues, such as identifying missing pages, which can be dealt with by dedicated procedures and software, and the performance of which can be evaluated using standard information retrieval criteria such as precision and recall (Zhang 2008, p. 240). An essential, but technically much more difficult aspect of quality control is the myriad of serendipitous findings. To use an analogy, Columbus' journey would have been deemed a failure if evaluated strictly on the basis of the task he set out to complete (to reach India by sailing westwards from Spain). In the case of library documents, the problem is compelled by the findings' diversity, the fact that one finding might lead to another, the context-dependent nature of the data, and the subjectivity of interests among library users. A holistic, qualitative approach, as adopted in this article, seems appropriate for containing such a fluid problem.

Regarding the question of result replicability, it will become evident from the evaluation outcome that participants have seized the opportunity to use the Document Towers to arrive at idiosyncratic ends. While "qualitative research doesn't pretend to be replicable" (Marshall & Rossman 1995, p. 144), response variability is also a desirable and enriching feature of the task to be tested (i.e. exploration).

Incidentally, these characteristics of quality control parallel some of those of research in the domains of the humanities and cultural heritage, which are included among the core operational areas of national libraries. The evaluation will stress how the Document Towers establish a precious feedback loop between library personnel and library users and expand the definition of quality control from narrow factual issues (e.g. a missing page) to an instrument of knowledge (e.g. the missing page reveals censorship).

First Experiment: Qualitative Case Study

"If you can't measure it, you can't improve it."— Peter Drucker, apocryphal

> *"Incongruities do not, however, usually manifest themselves in the figures or reports executives receive and pay attention to. They are qualitative rather than quantitative."— Peter Drucker, management theoretician (Drucker 2015, p. 69)*

Method

Approach — Why does this experiment use qualitative evaluation? While quantitative, analytic, and objective evaluations prevail in computer science, qualitative evaluations, common in psychology and social sciences among other fields, can produce different insights, and describe, with flexibility and synthetically, contexts that are noisy, fluid, hyperconnected, and circumstantial. These are properties of the problems to be evaluated, namely quality control, historical research, serendipity, and exploration. This study also concerns itself less with software engineering, which naturally lends itself to quantitative evaluations, than with the complexity of human–computer interactions and interface design, the cognitive processes of historical research, and how para-textual document information is captured by and percolates through digital information systems. Moreover, the Document Towers integrate paradigmatically machine processing of documents and human knowledge about their meaning, thus naturally lending themselves to quantitative analyses, in line with the view that "our current fascination with information technologies must be balanced by careful attention to the more specifically human dimensions of librarianship" (Nauratil 1989, p. 104).

Scope — The elements of the Document Towers to evaluate are identified in a systematic fashion using the so-called "Nested Levels" model for visualization design, which distinguish between the domain situation at the top of the design parameters, the mid-layers of data and task abstraction and the visual encoding and interaction idiom, and the lower level of the algorithm (Munzner 2015, pp. 66–93). The evaluation object is the proposed paradigm, i.e. an architectural representation metaphor of document structures: first in terms of insights into documents and collections, then for its utility to the specific tasks of quality control and historical research, and possibly other tasks such as planning and communication within the library administration, or as an information system for the library users. The evaluation perspectives are that of the point of view of the librarian performing quality control and offering access to information, the computer scientist inventing digitization and information systems, and the historian using them.

Participants — The participants were Florian Steffen (FS), head of the Digitization Unit of the Swiss National Library, representing the quality control practitioner in this case study; Andreas Fischer (AF), computer science professor, Department of Informatics, HES-SO//FR and University of Fribourg,

bringing the technologist's perspective to the study; and the author (VA), informatics researcher and historian, standing for the patrons using the library and archive services. While the participants have engaged in ongoing professional collaborations, including on the Document Towers, it is not their sympathetic disposition that has led to the discovery of objective facts in the data through the use of the Document Towers. Indeed, prior to the evaluation, they were unaware of the particular quality control issues in the dataset to be tested, which were selected randomly among newly digitized documents, such as for example the nature and significance of one of the main findings: a remarkable map, a missing image, and unlabeled white space. One important dimension of the case study — the recognition of the utility of the Document Towers for historical research and its implications for quality control that lead to the inclusion of the author as study participant — was itself an unplanned outcome of observations made by FS and AF, along with collegial discussion by all participants. Although the design and interpretation of the case study were developed by VA, these results were made possible only by teamwork. This approach is akin to participatory research and action research, which actively involve designers and users to challenge the mentalities and behavior of both parties (Marshall and Rossman 1995, p. 4).

Data — The evaluation made use of documents with which the Library's Digitization Unit was engaged at the time of the case study, specifically a newly digitized batch of historical newspapers that had not yet entered the quality control process. This choice provides the evaluation with real-world data for evaluation, and may result in real-world actions by the Library if interesting findings emerge. The data consists in eighty-two items in nine annual batches of the major historical Swiss newspaper Züricher Freitags-Zeitung [ZFZ, Zurich's Friday Newspaper] (1674–1914; in 1814, it was the most read Swiss newspaper [Bollinger 2018]) sampled over the period 1766 to 1905, and one hundred items in one batch of the journal of the influential Christlicher Metallarbeiterverband der Schweiz [Swiss Christian Metallurgical Union] (1902–1999) from 1968 to 1970. The documents form a collection of more than one hundred newspaper titles made available by the Library and its partners via the website e-newspaperarchives.ch. Logical entities were extracted from the scanned documents by a subcontractor and returned as ALTO/METS and enriched PDF files; page, image, and headline boundaries were then represented for each batch as a City of Document Towers (fig. 2).

Procedure — The evaluations were carried out by VA individually and lasted approximately one and a half hours, followed by group discussion (VA self-administered the protocol and was subsequently questioned by AF and FS). VA summarized the results, with the exception of those concerning AF, which he presented; the final text was approved unanimously. Participants

had previous knowledge of the Document Towers visualization paradigm and software, Crystal. At the start of the evaluation, they were invited to experiment with a sample document (e.g. upload data, filter information, select views, zoom, explore, and interpret). Next, the participants were asked to visualize the ten datasets using Crystal and verbalize their intentions, actions, and experiences, as well as answer specific questions; discussion with the evaluator was encouraged. The questions were as follows: "Please describe what you see in the visualization; highlight what appears remarkable", "For what tasks and for which users do you think the visualization would be useful?", "How could the visualization be useful for quality control?", "Describe your experience, especially concerning interface ergonomy and visualization readability", and "What is your bottom-line impression of the visualization paradigm and software implementation?"

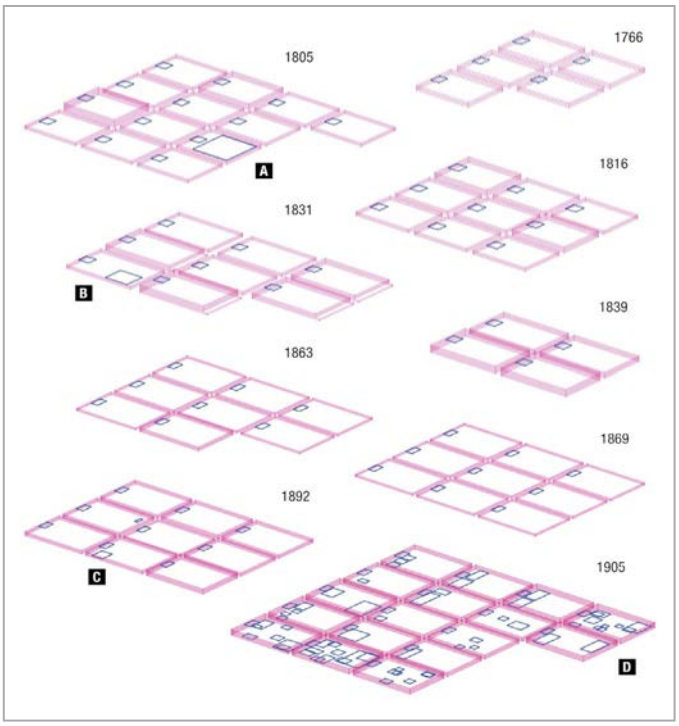


Figure 11. Objects evaluated

Figure 11 shows the Document Towers visualization of the 82 ZFZ issues used in the case study, (see sample pages in Fig. 12). Page boundaries are represented as red wire boxes, single issues as stacks, and issues are grouped by year. The blue boxes are items labeled as "Illustration" in the ALTO files of the digitized documents from which coordinates were obtained. — The newspaper logo is visible at the top of every issue; from 1863 onward, they show a different aspect ratio, suggesting a design change, confirmed by comparing front pages (B vs C). The larger page sizes in the 1831 batch are sup-

plements, and their identification through visualization facilitates planning and performing quality control. The conspicuous blue box of 1805 (A) stands for a map. The Document Towers provide visual proof that illustrations were quasi-absent from the ZFZ before the 20th century. — Comparing information available in the ALTO files as Document Towers and the document images reveals the intentional exclusion of tax and library stamps from digitization (no box around them in A, B, and C). This action costs money and time, as the objects have to be removed most likely manually, and is prone to errors, since it was discovered that not all stamps were excised. During interactions with the information system, both users and automated search engines are deprived of knowledge about the historically and possibly legally significant information carried by stamps.

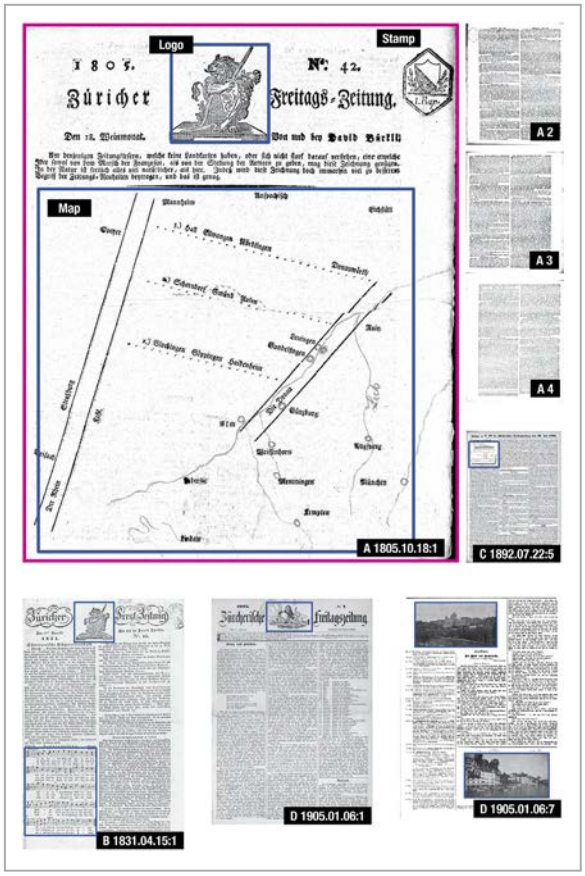


Figure 12. Actual ZFZ sample pages relevant to the visualization of fig. 11

Results

From this point onward we are discussing the primary findings derived from evaluating the Document Towers, where those findings subsequently led,

which concrete actions were taken by the participants as a result of the pilot evaluation, their significance, and the role played by the document visualizations.

Participant FS — FS found the Document Towers "extremely useful for quality control". An essential reason is that they allow identification of potential anomalies; here, FS pointed to how "monotonous" certain document runs represented as Document Towers appear, which he interpreted as a sign that there is in all likelihood no error to be expected. Another factor is that the Document Towers substantially facilitate "the identification of artifacts that can be subsequently searched [in other datasets], which is the ultimate quality control [method]". The discovery of overall patterns and trends also help in estimating the extent of, severity of, and costs incurred due to quality issues. Finally, the Document Towers spark novel ideas about quality control and, more generally, the use of digital documents and services within the wider context of the library's activities.

The Document Towers' importance was clear to FS, as they mitigate the great difficulty ("impossibility", in his words) of discovering features (e.g. when an image viewer is parametrized to fit images to the viewport, users are unaware that images with the same aspect ratio might be of different sizes) and interpreting data (e.g. coordinates are more easily assessed as a map than as a numerical list). FS noted with interest that the Document Towers prompt questioning of the data in many and important respects; among these were some not envisioned by the designer, such as using the Document Towers to train library staff on quality control and to "keep [patrons] longer on the library website".

In conclusion, FS, as head of the Digitization Unit, was (i) convinced of the Document Towers' utility for quality control and other tasks, and (ii) supported a common initiative for developing the paradigm through a future academic, industry, and government research and development project. Furthermore, spurred on by the aesthetic appeal of the Document Towers and Ribbons, the possibility of using them as part of forthcoming exhibitions at the Swiss National Library will be explored.

Participant AF — AF made some of the same central observations as FS and AV about the Document Towers, namely their discovery quality, complementarity to existing quality control instruments, polyvalence in terms of users and tasks, and attractiveness.

From a computer scientist's perspective, AF saw future potential of the Document Towers to support machine learning methods for document analysis. Such methods learn by example how to segment scanned pages into layout

elements, read printed or handwritten text, and detect logical structures, to name only a few applications. In order to achieve high accuracy, machine learning requires a considerable number of human-annotated learning samples, which should ideally cover all particularities of a document collection. AF viewed the Document Towers as an excellent means of obtaining an overview of the document collection and selecting learning samples accordingly. Another intriguing use of the Document Towers is to display the results of the document analysis methods for human inspection, aiding quality control and (in particular) the immediate identification of systematic errors repeated over large parts of a document collection, as these errors create distinct patterns. Finally, AF pointed out that the Document Towers may serve as a platform for visualizing search results, referring to it as "a city at night with lights indicating hits of a query" (e.g. search terms or whole document structures).

Participant VA — VA adopted the role of a historian and assessed how the Document Towers impact research into newspaper history. They focused on the map, the stamp, and the censorship gaps.

The map — An outlier image-block of singularly large size in the Document Towers batch instantly attracted attention (fig. 2: A). By consulting the relevant document-image, that of the front page of the 18 October 1805 ZFZ issue, the object was identified as a cartographic diagram of the principal localities and rivers of Napoleon's military campaign in Southern Germany during the War of the Third Coalition (1803–1806), which was discussed in the news conveyed by the document. Before the end of the 19th century illustrations were rare in this prominent Swiss newspaper — a well-known fact about the use of images in periodical mass media in general, but made explicit through the Document Towers visualization (occasional single-sheet news pamphlets or "Flugblätter", however, relied heavily on illustrations since their inception in the 16th century (Weisz 1933, p. 33; Neue Zürcher Zeitung 1980, pp. 135–138)). A subsequent visual check was performed on all currently online ZFZ issues from 1705 to 1815, as well as a keyword search ("carte" in French and "Karte"/"Carte" in pre-20th century German, i.e. "map") of the entire e-newspaperarchives.ch database. The first three maps to be found were from 1805, 1827, and 1859.

An additional literature survey, interviews with domain experts, and consultation of original documents in multiple libraries were conducted by VA (Weisz 1933; Gidalewitsch 1956; Bogel 1973; Neue Zürcher Zeitung 1980; Maissen 2005; Mussell 2012; Hafner 2015; Rickenbacher 2011, pp. 293–296, 299–301) resulting in the detection of one monochrome political map of Switzerland published in 1799 and reprinted on 15 August 1800 in Der Schweizer Bote of Lucerne as a single-page issue of the newspaper (Höhen-

er 2018). Another political map, of canton Säntis in 1798, is described in the literature as a newspaper map (Höhener 2018, p. 13) because it was commissioned by the Wochenblatt für den Kanton Säntis newspaper (1798 (1, 2): 6); however, the multicolor (i.e., more expensive) map was likely not part of the regular subscription, since it was sold by the publisher separately (1798 (9): 38). In conclusion, a simple document structure visualization led to a serendipitous discovery of the first known example of Swiss cartographic journalism specifically made for and incorporated into the newspaper layout, the first cartographic representation of military events in a Swiss newspaper, and the second attestation of cartographic content in a Swiss newspaper.

The deeper merit of the map is that it raises various questions about the nexus between military cartography and the mass media industry (an underexplored field (Monmonier 1989, p. ix; Rickenbacher 2011, p. 11; Novaes 2019, p. 3)), the marketing and technological road to the emergence of modern Switzerland (more than two and a half thousand maps were printed by the British press prior to 1800, while seemingly none were printed in Switzerland during the same period (Lehman 2011, p. 343)), and cartographic literacy (conceivably low, if we consider that the ZFZ was advertising the weekly sale in Zürich of military maps of Napoleon's campaigns for no less than the price of half of its yearly subscription rate, or the equivalent of two pounds of pork meat; such high production costs — or marketing shrewdness? — may explain why we see so few maps in Swiss newspapers of the era (maps sold: ZFZ 1805.11.08:4, ZFZ 1806.10.24:4; comparative costs: ZFZ 1801.12.18:1, Zürcherisches Wochen-Blatt 1809.05.08:3 (e-newspaperarchives.ch n.d.)).

The stamp — While viewing the page containing the map, VA observed a discrepancy between the newspaper image and the metadata extracted from it and visualized by the Document Towers: specifically, the image showed a stamp missing in the metadata (fig. 2: A). The precise nature and significance of the stamp and its omission was at first mystifying to the evaluation participants. What were we looking at? Perhaps a library mark? Was the omission an error made by the subcontractor, or did the Library digitization specifications direct the service providers to ignore stamps? The result of a rather convolved inquiry lead to three surprising discoveries:

- The item represents a tax stamp for newspapers introduced in 1801 during the Helvetic Republic (1798–1803), to be applied on empty paper sheets or after they were printed (ZFZ 1801.04.17:1) (e-newspaperarchives.ch n.d.; Stockar 2012). The tax was an instrument of financial censorship of the press [34: 283], as its cost was fully half the price of the actual newspaper (ZFZ 1801.12.18:1) (e-

newspaperarchives.ch n.d.; Guggenbühl 1996, p. 284), and derided by journalists as "saving reading time" and promoting among them the "potent brevity of [the illustrious Roman orator] Tacitus", as it resulted in fewer words being printed (Der helvetische Volksfreund 1801.05.02:167) (e-newspaperarchives.ch n.d.). The import of this humble stamp scheme should be judged with reference to its famous predecessor, the equally unpopular Stamp Act of 1765, created by the British parliament for its American colonies and regarded as one of the key events leading up to American independence (Wikipedia contributors [n.d., c]). The ability to derive such diverse and historically significant information underscores both the interest of the finding and the Document Towers visualization that was instrumental in its discovery.

- The Swiss National Library did not explicitly require the digitization of objects incorporated in documents after printing, such as library stamps or handwritten notes.

- The subcontractor randomly included and excluded such objects from ALTO/METS files, as attested by a library stamp in another newspaper (Zürcher Wochenchronik 1904.01.02:2) (e-newspaperarchives.ch n.d.) discovered by FS after being alerted of the general issue of omitting non-editorial objects.

The censorship gaps — A third set of results that emerged from the visualization experiment concerned the so-called "censorship gaps", empty spaces in printed publications denoting information that has been suppressed by official censorship (fig. 3) (Demm n.d.).

There is no means and no information in the typical OCR metadata to identify censorship gaps; thus, current digital information systems perpetuate the obfuscation of information effectuated in times long past for modern readers. This topic is not only historically and socially important (Mussell 2012, p. 90–95), but also difficult to quantify (Guggenbühl 1996, p. 15) and challenging for computer science: without additional knowledge, it is impossible to differentiate a genuine unprinted area from a censured one on a purely visual basis; some censorship-related gaps are half a line short and can be confounded with paragraph endings; other censorship techniques were used in addition to gaps, such as the blackening the relevant information with ink (Wikimedia Commons contributors n.d.), or the removal and rearranging of text in order to hide any visible hint of censorship, as apparently engaged in by the ZFZ. That VA was first made aware of the censorship issue through researching the stamp question, itself identified by physical juxta-

position with the map, again demonstrates the importance of serendipity for historical research in particular and information systems in general (Martin & Quan-Haase 2016, 2017). The Document Towers are an example of how exploration enabled by "panoramic vision" — as opposed to the "tunnel vision" enforced by keyword searches — can be implemented in a technology serving historians. They also demonstrate why exploration and serendipity are useful information behaviors that should not be neglected in the design of digital information systems.

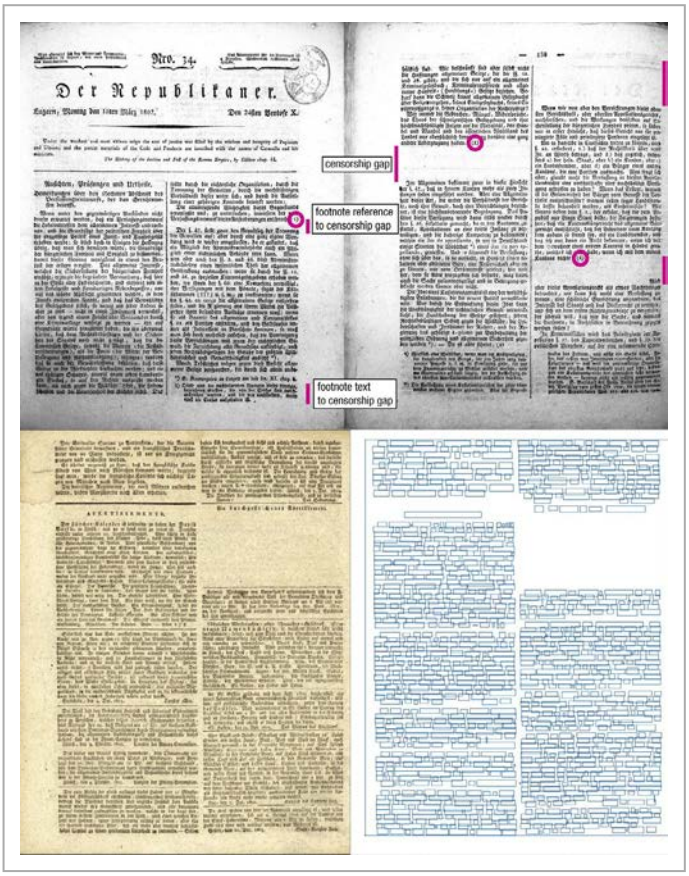


Fig. 13. Meaningful gaps

The top of Figure 13 shows censorship gaps in Der Republikaner of March 14, 1802, a pro-Helvetic Swiss newspaper and antagonist of the ZFZ (34: 140–151). The gaps have the typographical oddity of being marked by footnote numbers; these footnotes identify the "whitespace" as made by "wise" censorship (as sarcastically noted by the publisher, who promises to supply the missing data once censorship is abolished). The information void is made visible by the Republikaner and turned into a subversive political statement, just as their absence from the ZFZ reflects its pro-regime attitude (34: 199–206). (Bern University Library, Q7/BeM.His.Alt.SQ44 1802/

03. Photo: VA). The empty space within the second column of this ZFZ issue (1805.12.13: 8) at the bottom is explained by a caption as being a censored advertisement. The diagram on the far right locates text strings as extracted by OCR. Nothing in the metadata distinguishes the empty spaces that are margins from deleted paragraphs.

All participants — The quality control performed by participants with the Document Towers was visual analytics rather than linguistically or numerically mediated search. It highlighted how the use of a classical search interface — the textual input field — presupposes that users know what they are searching for and can formulate appropriate keywords; by contrast, a visual representation supports the discovery of unknown patterns, through visual scanning, instead of translating possibly vague intentions into words.

It was telling that the participants did not refer to the Document Towers' aptitude for browsing and navigating documents, two of the initial motivations behind the development of the Document Towers. Browsing is a basic document interaction, efficient for physical codices but (as has been empirically observed in the past) frustrating for digital documents, (e.g. the slow rendering of e-paper and the inability to navigate quickly between text, endnotes, bibliography, and indices) (Liebert 1994, p. 449–500; Signer and Norrie 2010). It is suggested that because what users are doing with Document Towers is precisely browsing and navigating, they do not need to be specific about it; the difference is that they do it with their eyes rather than with their hands.

In addition to verbal feedback, the participants' behavior was also looked into, and commonalities were found between the visualization medium and its communicative power. After initially using the software interface, paper hard copies prepared in advance were used for the remainder of the evaluation (fig. 14). The cause was the slowness of graphical rendering, repeated for each newspaper batch. Also important was that hard copies offered more physical display space than the laptop screens typically used by the local communities represented by the participants (i.e. library professionals, library patrons, and computer scientists).

In fact, no single computer screen provides as much space (and multiple screens are often unavailable for financial and technical reasons) as that obtained by the participants, who spread hard copies of different document representations across several tables, making it convenient to compare, organize, discuss, and collaborate on a broad dataset sample. Even the fact that the analysts did not need crowd around a screen and jostle for control of the computer mouse and keyboard contributed to a more relaxed, fluid, and integrative communication environment. Hard copies are furthermore per-

sistent, preventing the need for time-consuming regeneration of past digital representations.

Fig. 14. Participants AF and FS evaluating the Document Towers on screen and on paper

The evaluator pasted Document Towers hard copies on the office walls, which led to the discovery of an additional quality: without requiring any intervention (e.g. opening a computer, generating a view) or technical infrastructure (e.g. no energy consumption), the unobtrusive presence of hard-copy murals facilitate the memorization of information, prompt the re-thinking of insights, and provide openings for discussion with passing colleagues, who may be drawn in by the intriguing Document Towers.

In conclusion, the software interface and hard-copies are complementary modalities of interaction with the document representations. The relevance of these observations exceed the specifics of this case study, insofar as they have been repeatedly made in the past in the broader context of the debate regarding the advent of the so-called "paperless office" and the respective affordances of paper and electronic means of information presentation (Heath & Luff 2000; Sellen and Harper 2002).

Second Experiment: Thought Experiment

> *"Think!"— Thomas J. Watson, chairman and CEO of IBM (Wikipedia contributors [n.d., d])*

Under normal circumstances, no library would voluntary incur the expense and conservation hazards associated with moving two hundred years' worth of newspapers (a sizable physical volume) from the potentially off-site storage to a reading room in order to answer a single reader's question. Thinking alone is sufficient to prove this claim, and there is no need to test librarians' resolve in their function as guardians of information — as is apparent at least to those readers still accustomed dealing with physical documents, humans, and bureaucracies in a world of digital libraries. The author would like to demonstrate here that it is possible to apply similar thought experiments to compare the Document Towers to other document analytics techniques, since their salient characteristics are well understood.

"Thought experiments" are those for which testing does not require implementation, or that cannot be tested for various practical reasons. In this way, they are akin to simulations. Readers may be familiar with their potency from one of the seminal events in computer science history, Alan Turing's 1950 thought experiment for testing the intelligence of machines of the future (Turing 1950).

Our task is to find the first three maps published in the ZFZ newspaper; thus, it is a historically-oriented follow up to the quality control experiment. The search is performed physically on the original documents, linguistically through keywords, visually via thumbnails and Document Towers, and mixed linguistically–visually on images obtained by metadata filtering.

P10 – 209

Physical — Owning to considerations of conservation, costs, availability, administration, and (particularly) slow speed, consulting large amounts of physical documents is practically impossible for inconsequential questions such as ours. Physical documents nevertheless remain indispensable, as the digitization process filters and distorts the information of the original medium.

As an example, the map found in the digitized ZFZ shows hand-drawn rivers and locations (fig. 12: A1), while the rest of the map is created with typographic means, i.e. by arrangement of characters and lines (a German-Swiss invention of the late 18th century called "typometry" (Hoffmann-Feer 1969)). To determine whether the hand-drawing on the ZFZ map was part of the publication or subsequently added (by the publisher or a reader? for which purpose?) and evaluate its bearing on map production costs (were all the possibly 800 to 1000 exemplars of an issue hand-"illuminated", as the German text puts it? (Wiskeman 1959, p. 10) ZFZ 1805.11.08: 4 (e-newspaperarchives.ch n.d.)), the digitized document is not enough to assess material aspects of paper and ink. Additionally, as is well understood by historians, there is often only one digitized copy of an entire print run available, even though instances are not identical and thus carry different information (Mus-

sell 2012). When a physical exemplar of the 1805.10.18 ZFZ issue was located in Basel and examined, it contained no hand-drawings, suggesting the need for further research.

From a historical perspective the finding ascertains that newspaper cartography at the time was costly, which might explain why so few maps were published in Switzerland before the middle of the 19th century. From an information systems perspective, this case study exemplifies the informational richness of physical documents in comparison to their digitized avatars. The extreme simplification of documents to entity boundaries carried out by the Document Towers may constitute a reminder to users to remain aware of the differences between physical and digitized objects.

Keywords — Searching for maps using keywords is based on the assumption that maps have captions or are referenced in text (likely by the word "map"); this is the case for the ZFZ map (fig. 1). Notwithstanding, this requires a certain amount of user expertise, as they need to be able to read an unusual script style (Fraktur) to double-check the OCR transcription (containing about 7.5% errors on this specific page), be knowledgeable of historical spellings (under French cultural influence, "map" was written in pre-20th century German both as "Karte" and "Carte"), and master German, French, and Italian in order to be operational in the multilingual Swiss environment. Nevertheless, linguistic search excels relative to other search techniques in terms of the high quantity of information that can be searched.

Thumbnails — Searching maps by browsing newspaper page thumbnails potentially results in the highest quality, since what is being examined are rich, relatively unfiltered document images. As the volumes of thumbnails increase, however, the process becomes problematical. On e-newspaperarchives.ch, there are currently 33,790 pages of ZFZ to visualize for the period 1704 to 1914; this is not only slow to do, but also unreliable due to the resulting boredom (an important human factor that should not be neglected in information systems design). It took VA thirty-two hours to search the ZFZ from 1705 to 1815 for maps by looking at thumbnails, since thumbnails are generated only for one issue at a time and are too big to fit in the viewport in their entirety. One librarian was asked how to find the first Swiss newspaper map, and she suggested to manually search the other tens of old newspaper titles, which defeats the utility of automatic processing solutions for online documents.

Document Towers — All currently available ZFZ pages could be displayed as a mosaic on the considerably large surface of about 10 by 8 meters (10 by 9 yards) at 5 by 4 cm (2 by 1.5 in) per thumbnail. Document Towers make overview at a glance possible, with speed being their main benefit for the

map searching task. This is unsurprising, since Document Towers can be conceived as stacked thumbnails that are reduced to their bounding boxes.

Metadata — Non-textual objects are labeled "illustrations" in the ZFZ metadata of ALTO files. Although the public interface does not currently support search by metadata, a user of the future might ask for all "illustrations" to be displayed as thumbnails, thus obtaining a considerable improvement in speed and quality using this method. Of course, the location of the "illustrations" within the three-dimensional document is lost, and with it an additional hint to the object semantics that the Document Towers provide.

Conclusion — No single search method is sufficient or optimal: we were able to discover potentially the first attested military cartographic representation in a Swiss newspaper simply because illustrations were very rare for a long period of time, such that their pattern "pops out" from the surrounding text texture. Other pattern features or search methods would be required for identifying map frequency in later periods. Nevertheless, from a pragmatic perspective, it is the Document Towers among all approaches that lead to the discovery.

Third Experiment: Diagrammatic Evaluation

> *"The first things I found out were that all mathematical reasoning is diagrammatic and that all necessary reasoning is mathematical reasoning, no matter how simple it may be." — Charles S. Peirce, mathematician and philosopher (Peirce 1902)*

Diagram — A survey of the prolific literature on evaluation in the computer science field of interface usability (Nielsen 1993; Jordan et al. 1996; Diaper & Stanton 2004; Stone et al. 2005; Elmqvist & Yi 2015; Turner 2017), qualitative methods in psychology (Marshall and Rossman 1995; Patton 2001; Willig and Stainton-Rogers 2008; Leavy 2014), and the impact analysis of projects, programs, and policies in econometrics (Rossi et al. 2004; Frölich & Sperlich 2019) shows that quantitative methods dominate the technical disciplines. That most of these methods assume the existence of more or less clear and specific goals and tasks, is, however, contrary to the essence of exploration, as discussed herein.

The author therefore developed a diagrammatic method specifically intended for the analysis of the impact of explorations, such as that enabled by the Document Towers. The use of a diagram couples qualitative and quantitative evaluations, insofar as qualitative results must to be formalized and expressed as physically defined attributes. Thus, the impacts are amenable to both visual analytics and numerical evaluation.

The use of schematic representations of concepts, relationships, and processes to reflect, plan, control, and communicate, i.e. diagrammatics, has a long history, across a great variety of domains. Geometry, for example, is almost unthinkable without diagrams, as is the interweaving of the Feynman diagrams with the early development of quantum physics (Wüthrich 2010), lately also having inspired linguistics (Heunen et al. 2013). Imaging mathematical proofs "without words", using only visual representations, is more than an educational method (Alsina & Nelsen 2006): diagrammatic reasoning has been automated and implemented in mechanical machines and digital computers (Nakatsu 2010), from the medieval theologian Ramon Llull (i.e. directed graphs (Gardner 1958)) to the modern sociologist Pierre Bourdieu (i.e. correspondence analysis diagrams (Bourdieu 1984)). Sketching is also diagrammatics, and interactive charts and maps too; both being excellent examples of how diagrammatics is a kind of thinking-by-doing. For an introduction to diagrammatics read, e.g. (Schneider & Ernst 2016), an anthology of fundamental texts; (Anderson et al. 2002), a multidisciplinary monograph; and (Diagrams 2000–), the Diagrams conference proceedings.

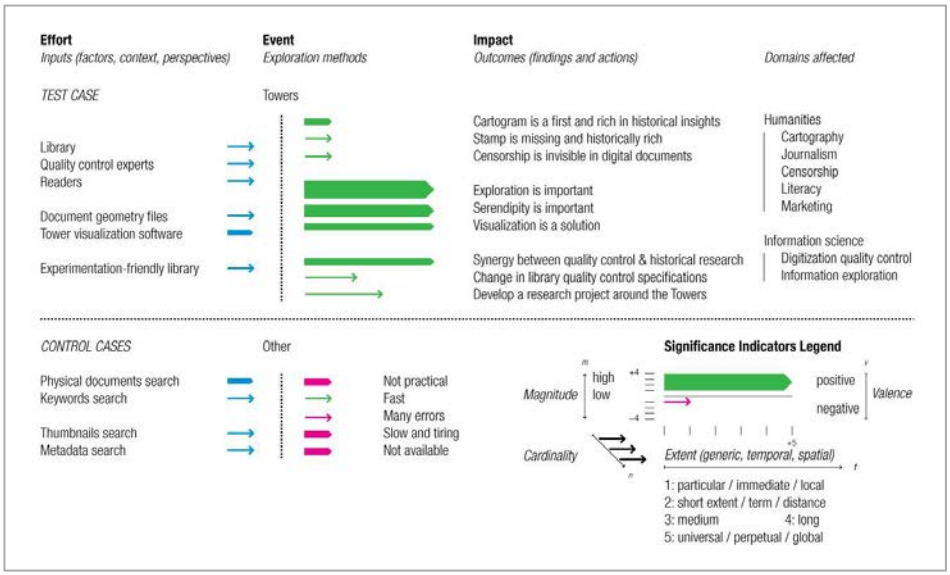


Figure 15. Evaluation of the impact of the Document Towers visualization

Document Towers impact evaluation — The diagram in Fig. 15 serves to evaluate the impact of the Document Towers visualization as a document exploration tool in the context of the experiment described in this article, concerning a dataset of historical newspapers.

- Characteristics — A key diagram feature is that it does not set pre-

defined expectations to evaluate impacts, as these become available only through the process of exploration. From the outset, the diagram defines an analysis framework, a set of significance indicators, and a procedure for graphical representation of variable parameters in support of visual and numerical interpretation. The result is a flexible instrument that combines both qualitative and quantitative characteristics, along with objective, subjective, and contextual aspects. Making subjectivity explicit is valuable, because it clarifies the analyst's standpoint. Fig. 15 presents a concrete example of the proposed diagrammatic analysis applied to the evaluation of the Document Towers. The method's conception was guided by considerations of how to evaluate the outcome of various cases of exploration, from historical examples such as Columbus' discovery of the Americas and space exploration, to the document-related case of a bookshop visit. The choice of impact significance indicators discussed below reflects the intent to provide a fundamental and generic selection, and above all a method that gives users the possibility of describing impacts in their own terms. For instance, the diagrammatic method does not prescribe how to determine the impact magnitude, which can derive from any criteria shared by professionals or decided ad hoc, such as conformance, serviceability, aesthetics, or socio-cultural adaptation, to mention only a few (extended lists in (Schütte 2005, pp. 7–10; Zhang 2008, pp. 252–253)).

- Layout — The analysis diagram is constructed around an event axis, in this case the use of the Document Towers. Left of it are the effort inputs, i.e. factors necessary for the event to take place (such as data, software, and users, a context, such as a library open to experimentation), and perspectives on the nature of the effort (such as from the point of view of the digitization workforce, readers, and library administration). Left of the event are the outcomes, consisting in concrete findings (such as a map), generic findings (such as recognizing the importance of serendipity for quality control), actions (such as changing quality control specifications), and the domains affected by the event (such as cartography within the humanities and digitization within information science). The horizontal symmetrical organization of the diagram facilitates the comparison of effort and impact, in the spirit of a return-on-investment analysis. The upper part of the diagram represents data about the event we want to test, while the lower part contains control data used to compare the impact of the test event with alternative events (here, with different search methods described as a thought experiment in the preceding section).

- Elements — The next elements in the diagram are the significance indicators, represented by graphical arrow markers and textual labels. While the design space of their visual attributes and meaning are provided, the particular values in terms of size and color are the variable parameters to be determined by the analyst. The convention used is to represent the temporal, spatial, societal, or other extent types of the impact as marker length; in this case study, duration is considered along a five-level scale that varies from immediate to perpetual duration of impact. If two extent types (such as time and space) need to be analyzed concomitantly, the markers acquire depth and become three-dimensional. The impact of magnitude is given by marker width, on a four-level scale, and valence by color, with green denoting beneficial outcomes while red outcomes are deemed detrimental from a given perspective; moreover, the cardinality (quantity) of indicator markers reflects the event's impact diversity of outcomes and affected domains. Outcomes and domains can be clustered to provide a sense of hierarchy and sequence, such as for the cause-and-effect diagrams common in process control; these resemble fish skeletons, with each bone a factor that converges into other factors and ends into a final effect (Kume 1985, pp. 26–36). The author however opted for simplicity in order to draw attention to the value of indicators rather than their configuration. The marker units are expressed in words (e.g. "short duration" and "very long duration") and physical units (e.g. 1 or 4 cm). This approach, in the spirit of fuzzy logic, can accommodate both human and machine communication.

- Procedure — The procedure to be followed by analysts for using the diagram is to carry out an exploration, then make a list of outcomes, determine indicator values and translate them to graphical markers, and then fill the analysis frame with markers and labels, before finally evaluating impact significance and trade-off to input effort. The exploration might be supplemented with a number of experiments on control data.

Quantification — The translation of the qualitative analysis expressed in words to a diagram is more than a change of modality: giving visual shape to an abstraction produces the quantification of qualitative information. Thus, we can benefit from the positive properties of both methodological approaches: reduced ambiguity, facilitated interpersonal communication, computer-processable knowledge representation, automation, and large-scale analysis. Let us now see how quantification is realized and define a set of metrics, a terminology, and a notation.

- Vectors — The diagram embodies abstract significance indicators as parameterizable physical markers. These possess quantitative properties — length, width, and color — that make the markers into graphical variables expressible as a three-dimensional vector of scalars:

$$I_i = \{t_i, m_i, v_i\},$$

where I_i is the i-th vector indicator, t is a positive scalar representing the effort or impact extent in time, space, or other units, m is a positive scalar representing its magnitude in (not necessarily physical) energy units, and v is the valence, given as a positive or negative unit value (+1 or −1). It is computationally more efficient to conflate magnitude and valence, cast the valence as a signed scalar (to take positive and negative values), and define a single indicator value as a two-dimensional Euclidean vector with a certain resultant magnitude M:

$$^2M_i = \sqrt{t_i^2 + m_i^2}.$$

The entire set of indicators, ^3S, introduces a third dimension: the cardinality, n, which represents the number of indicators in the left side, right side, or entire diagram. A compact way of describing this vector space is to add values along their respective dimensions and obtain a cumulative resultant, ^3M, in the three-dimensional indicator space:

$$^3M = \sqrt{\left(\sum t_i\right)^2 + \left(\sum m_i\right)^2 + n^2}.$$

The epistemological question arises, however, as to whether it is defensible to integrate dimensions with different units, such as time, energy, and cardinality — or how to interpret constructs that become opaque through dimensionality reduction. It is probably more transparent to express the total effort or impact J_p of an exploration event p as a three-dimensional cumulative vector, where each dimension is the sum of all indicator values along the given dimension:

$$J_p = \left\{\sum t_i, \sum m_i, n\right\}.$$

- Statistics — The indicator space can also be analyzed in terms of ba-

sic descriptive statistics, such as the mean and range of effort or impact duration and magnitude. Given that exploration evaluation diagrams are expected to contain a small number of indicators, the analyst should exercise caution while using statistics. For example, if the diagram has an indicator of one unit duration and another of four units duration, the mean is 2.5 units, which is a value that does not exist in the data and cannot correctly represent the distribution, which is more aptly characterized as bivariate.

- Comparisons — The impact-to-effort trade-off T is given in absolute terms by the difference of their respective vectors:

$$T_{absolute} = J_{impact} - J_{effort}$$

and in terms of relative gain by the impact-to-effort ratio:

$$T_{relative} = J_{impact} \,/\, J_{effort}\,.$$

Positive values and values greater than unity, respectively, indicate that the impact of the exploration event is greater than the effort expended. Division by zero can be ruled out, as the effort will always be positive one way or another. In the same absolute and relative way, the outcome comparison C of two exploration events p and q can be computed as difference and ratio, as follows:

P10 – 216

$$C_{absolute}(p, q) = J_p - J_q\,,$$
$$C_{relative}(p, q) = J_p \,/\, J_q\,.$$

Note that a null absolute difference or a unit relative gain does not necessarily mean that the exploration outcome is neutral and devoid of interest. A change in ontological class between input and output brought about by the exploration might be a desirable outcome; it is indeed useful to feed one unit of grain to a chicken if she consequently lays one tasty egg.

- Perspectives — The linear summation of vectors provides an objective measure, while the summation of logarithms represents the subjective measure experienced by humans participating in, impacted by, or analyzing the exploration:

$$J_{objective} = \left\{ \sum t_i, \sum m_i,\ n \right\},$$
$$J_{subjective} = \left\{ \sum \log_b t_i, \sum log_b m_i, \log_b n \right\}.$$

Typically, an exploration analysis diagram such as that in fig. 15 would use a logarithmic scale for the significance indicator markers. The logarithm base *b* depends on the nature of the data: base ten, for example, is an appropriate choice for contexts where duration is envisioned in decades and centuries, while base sixty, thirty, and seventy-two are suitable for durations experienced in minutes, current average generations, and lifespans, respectively. The difference between the physical and psychological measures is documented by empirical findings about human perception and cognition, many aspects of which can be characterized by a logarithmic or power law, such as acoustical sensitivity and the importance given to events depending on the spatial distance at which they occur. Also, various natural and artifactual phenomena change across logarithmic or power levels of magnitude, from coastline anfractuosity to stock market behavior. The logarithm is furthermore of direct interest for modeling exploration outcomes, insofar as (for reasons similar to those presented here) it defines the fractal dimension of multiscale phenomena such as exploration outcomes (Mandelbrot 1982) and the measure of information entropy related to the uncertainty of exploration outcomes (Shannon 1948, pp. 379–380).

Results — Once the Document Towers' diagram has been produced, what can we learn from Fig. 15 about the impact of using this document visualization? First, we observe the quantity and diversity of findings, actions, and domains affected by the exploration experiment, signifying the broad utility of the Document Towers, which appears especially significant in light of the positive ratio between the large impact and the modest effort involved. Second, the Document Towers are expected to yield outcomes of local importance (such as the discovery of a cartogram of primary interest to Swiss historians), as well as generic and thus long-term outcomes, such as the synergistic benefits between quality control and historical research. Third, the use of the Document Towers as an investigative instrument appears superior (see the amount of green in the test part of the diagram) to many other information search methods (the control part is overwhelmingly red).

From a quantitative perspective, the absolute impact-to-effort trade-off exhibit a clear positive balance:

$$
\begin{aligned}
T_{absolute} &= J_{impact} - J_{effort} \\
&= \{ \sum extent_{impact}, \sum magnitude_{impact}, cardinality_{impact} \} \\
&\quad - \{ \sum extent_{effort}, \sum magnitude_{effort}, cardinality_{effort} \} \\
&= \{ \sum (1,1,1,4,4,4,2,3), \sum (2,1,1,4,3,2,2,1,1),\ 9 \} \\
&\quad - \{ \sum (1,1,1,1,1,1),\ \sum (1,1,1,1,2,1),\ 6 \} \\
&= \{24, 17, 9\} - \{6, 7, 6\} \\
&= \{18, 10, 3\}\,.
\end{aligned}
$$

The comparison in terms of overall vector magnitude between exploration by Document Towers and by all other means is also in favor of the former:

$$
\begin{aligned}
{}^2M_{Document\ Towers} - {}^2M_{Other} &= \sqrt{\left(\sum extent_{Document\ Towers}\right)^2 + \left(\sum magnitude_{Document\ Towers}\right)^2} \\
&\quad - \sqrt{\left(\sum extent_{Other}\right)^2 + \left(\sum magnitude_{Other}\right)^2} \\
&= \sqrt{30^2 + 24^2} - \sqrt{12^2 + 11^2} \\
&= 22.1\,.
\end{aligned}
$$

In conclusion, the diagrammatic and quantitative analysis of the exploration impact formally confirms and expands the results of the previously conducted qualitative evaluations.

Fourth Experiment: Usability Measurements and User Psychometrics

> *"Now, gentlemen, the purpose of this machine, of course, is to free the worker from routine and repetitive tasks and liberate his time for more important work"*
> *— Desk Set (Walter Lang 1957)*

In the preceding sections, we have applied various approaches to demonstrate the effectiveness and efficiency of the Document Towers visualization concept in supporting the quality control of digitization workflows, and provide further insights into digital documents that will be useful for both libraries and their users. We now consider the Document Towers from a user satisfaction perspective. We also investigate psychometric differences between user groups, which may affect usability.

Satisfaction experiment

Participants — Thirteen librarians from two large Swiss libraries (9 females, median *MD* = 47 years, range *RG* = 27–52, standard deviation *SD* = 7.9)

and 8 academics from three Swiss universities (5 females, *MD* = 39.5, *RG* = 22–51, *SD* = 11.5) volunteered for the study (*N* = 21). All librarians were actively involved in digitization projects (e.g. scanning, metadata processing, user requirements, project management), while all academics had a digital humanities background (5 B.A., M.A., or Ph.D. students from an entry-level data visualization class, 2 researchers, 1 professor). Programming competence within both groups ranged from none to basic to professional.

These groups represent two stakeholder communities (librarians and users) in the digital library ecosystem (others include information technology providers and policy-makers). The sampling these specific librarians, who are involved in work on a broad range of technical, content, and management issues, is relevant in the context of evaluating the utility and adoption potential of novel technologies; the digital humanists serve as a digitally savvy comparison reference drawn from among library users. As thanks for their contributions, the participants were offered chocolate at the beginning of the experiments, which took place in the early afternoon.

Constraints — The aspects of the Document Towers that we wished to evaluate were their ability to support document exploration (as opposed to search) and their polyvalence for other tasks, such as document overview and navigation, utility for digitization tasks and low-vision users, aesthetic appeal, etc.

Evaluating an exploration system is challenging, since exploration is by definition a task that is not necessarily reproducible, as well as highly dependent on user subjectivity and the application context. For example, to realize the importance of the lone military map identified in the ZFZ newspaper, an evaluator must have a historian's mindset; a certain technical skillset is required to realize that the ZFZ tax stamp is a digital object missing from the scanned document, while an understanding of project management functions within the library is needed to understand the potential benefits of these findings for the library offerings as a whole, beyond the specifics of quality control. Therefore, typical quantitative evaluation methods (e.g. asking evaluators to identify the first military map in the ZFZ and recording their response time) are not appropriate for evaluating exploration, since specifying the target is a search task, while in exploration mode, the target is unknown.

Furthermore, given the limited time availability of the experimental participants, especially the professional librarians, it would not be possible to evaluate each task for which the Document Towers may be useful and thus to convey the full spectrum of their applications to evaluators. A further experimental constraint is that we want participants to evaluate the utility of the visualization concept, not the idiosyncrasies of a specific software implementation.

Protocol — The solution we developed is a combination of a guided tour, a hands-on workshop, and a satisfaction questionnaire, taking a total of one hour to complete. The method employed may be conceptualized in informal terms as an "assisted drive through an IKEA store".

The "guided tour" consisted in acquainting the participants with the various potential benefits of using the Document Towers, the notion of exploration as opposed to search, and the mechanics of deriving information from the spatial structure of documents. The "tour" was delivered verbally by the evaluator (the author). The instruction material comprised: (1) a cardboard architectural model of a small document used as a material instantiation of the document-as-architecture metaphor, supplemented by pictures of buildings in the shape of books (Atanasiu 2022a); (2) a book and the printed Document Towers representation of its InDesign and PDF-formatted digital files, to demonstrate the differences between digital document formats (Atanasiu & Ingold 2021); (3) a few PDF documents supplied by the participants, which were visualized and interactively manipulated in front of them using the author's Crystal software for Document Towers visualization, as well as opened in Adobe Acrobat and Illustrator to demonstrate the impact of software choice on which parts of document structures are visible (e.g. a document viewer such as Acrobat does not show page areas outside its viewport or "CropBox"); (4) a small wooden tablet laser-engraved with a compact Document Towers representation for the benefit of low-vision document users (fig. 8).

During the hands-on phase, participants were supplied with Document Towers representations of selected documents and invited to interpret the visualization. To encourage participants to focus on the concept, these visualizations were mainly printed on paper rather than being made available via the interactive Crystal human–computer interface. The documents in question were the ZFZ sample (fig. 12), scientific articles (fig. 10), and the user-supplied documents. With respect to the ZFZ documents, the main focus was on evaluating how well participants could identify certain features of interest (i.e., the map, the stamp, the outsized pages, and the illustrations) and what they made of them; other features of interest (i.e., document misclassification, distinction between scanned and native digital documents, non-visible objects) were exemplified using the articles.

The document exploration procedure comprises three steps, proceeding from data gathering to knowledge formation to action-taking (see "Practice: How to explore documents in three steps using the Document Towers"): (1) identification of patterns of interest in the Document Towers visualization (e.g. an outsized blue rectangle); (2) determination of the sources of the observed patterns (e.g. the blue rectangle represents raster images and corresponds to

a map); (3) operationalization of the findings (e.g. further research establishes that the image is the first known appearance of a military map in a Swiss newspaper, and thus represents a historical discovery). A fourth step, specific to the evaluation, is to compare the efficiency of the Document Towers to other methods for eliciting the same findings (in our case, one participant had to use the ZFZ website to find the first map in the newspaper). The evaluator replied to technical questions about the visualization and digital document formats, and guided the participants towards features of interest when necessary.

Table 1. System Usability Scale used for evaluating the Document Towers

1	I think that I would like to use the visualization frequently
2	I found the visualization to be simple
3	I thought the visualization was easy to use
4	I think that I could use the visualization without the support of an expert
5	I found the various functions in the visualization were well integrated
6	I thought there was a lot of consistency in the visualization
7	I would imagine that most people would learn to use the visualization very quickly
8	I found the visualization very intuitive
9	I felt very confident using the visualization
10	I could use the visualization without having to learn anything new

After the training phases were complete, the participants' satisfaction with the Document Towers was evaluated via the System Usability Scale (SUS), a ten-item questionnaire (Table 1) developed for quickly evaluating the quality of interactions between humans and technological systems (Brooke 1996, 2013; Sauro 2011).

SUS is today an industry quasi-standard for usability measurement, as it has been thoroughly statistically validated as an effective and robust evaluation instrument and is often more accurate than alternatives (Sauro & Lewis 2016, pp. 185–248), as well as being appropriate for small sample sizes (as low as eight to 12 users (Tullis & Stetson 2004, pp. 134–136; Sauro 2011)) and a wide range of products (Sauro 2011, pp. 205; Kortum & Bangor 2013). Furthermore, the thousands of experiments that have employed SUS since its inception in the late 1980s provide a basis for benchmarking novel systems against existing ones (Bangor et al. 2008; Sauro 2011, pp. 55, 56, 138, 155).

Throughout the evaluation, participants made unsolicited comments, which later proved to be very valuable in providing context to the interpretation

of the results; moreover, some chose to stay behind after the end of the formal experimental session for further discussions with the evaluator about the Document Towers, in what could be termed "qualitative debriefing" sessions.

Results — The SUS questionnaire was administered to the participants in its all-positive questions version (Sauro & Lewis 2011), with the term "system" replaced by "visualization", and in English (the language for which its validity is best studied, and in which all participants are conversant (Finstad 2006)). Statistics of the SUS scores (i.e., the average of individual item scores (Lewis & Sauro 2018)) are presented in Fig. 16.

The SUS values become meaningful only when assessed against those representing the usability of other reference systems reported in the literature (Kortum & Bangor 2013). The largest such SUS benchmarks, based on $N > 5000$ questionnaires, are approximated by a logarithmically skewed normal distribution with mean $M = 68.1$ (Sauro 2011, pp. 25–26, 31–37, 53–59, 111–115, 155). Its cumulative density function defines the transfer function between SUS scores and benchmark percentile ranks, as depicted by the curve in Fig. 17. The two circle markers indicate the mean scores and percentile ranks for the librarian and academic groups, with the whiskers indicating the confidence intervals.

The adjectives on the y-axis were suggested on an empirical basis based on SUS research into qualifying usability in words rather than numbers (which the author adapted by applying the adjectives suggested in (Bangor et al. 2009; Kortum & Bangor 2013, p. 75) to ranges applicable to the larger-scale benchmark data of (Sauro 2011, pp. 32–34)). Reference technical products are also provided on the y-axis ([Sauro 2011, pp. 205; Kortum & Bangor 2013, pp. 71).

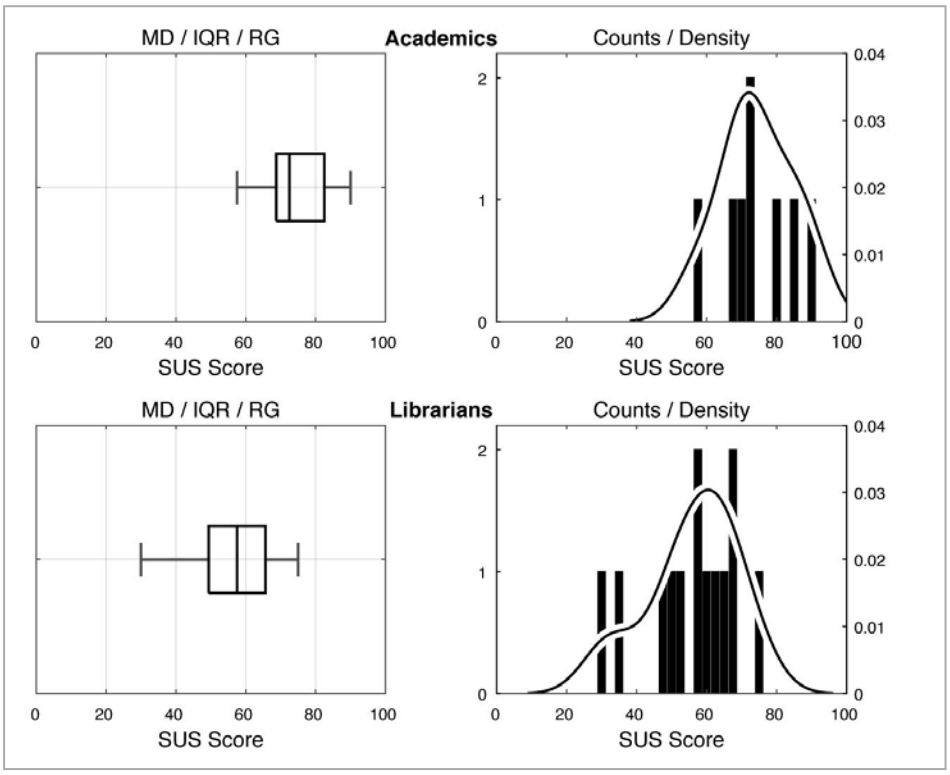


Figure 16. Scores on the System Usability Scale, academics and librarians

P10
–
223

Figure 16 shows the scores on the System Usability Scale for the evaluation of the Document Towers by academics and librarians. Academics: N = 8, Scores M = 74.4, MD = 72.5, SD = 10.3, RG = [57.5, 90.0], IQR = [68.8, 82.5]; Librarians: N = 13, Scores: M = 56.0, MD = 57.5, SD = 12.9, RG = [30.0, 75.0], IQR = [49.4, 65.6].

According to our experimental design, the usability of the Document Towers is "good" in the opinion of academics (74.4 SUS score, above the 68.1 global industry average (Sauro 2011, p. 31), and better than 70.9% of other systems), and "poor" according to librarians (56.0 SUS score, 20.9 percentile rank).

The assessment disparity between the two groups is sizable (Hedges' g = 1.47, with pooled standard deviation (Hedges & Ingram 1985, pp. 78–81; Coe 2002; Kline 2004, pp. 95–142) and cannot be attributed to sampling error (paired t-test, p = 0.0021, at 95% confidence level (Sauro 2011, pp. 122–126)); moreover, it persists even after the two low SUS score outliers in the librarian group are removed. It should be noted here that very few systems are known to have scored in the "best" usability class (Sauro 2011, pp. 33).

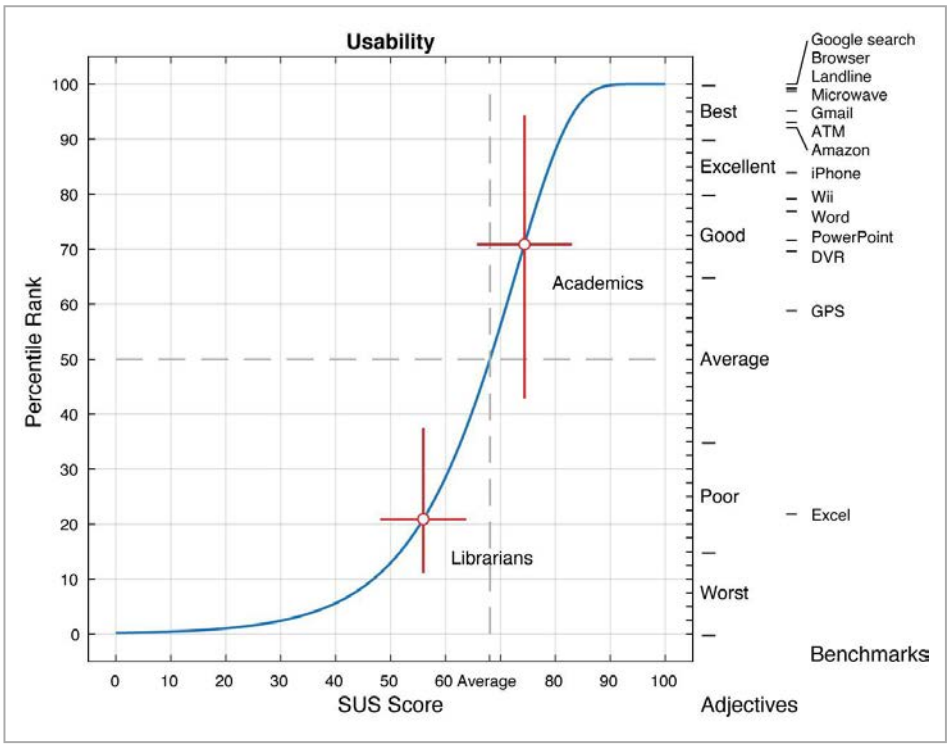


Figure 17. Usability of the Document Towers

Figure 17 illustrates the usability of the Document Towers expressed in adjectives obtained by transforming the participants' SUS scores into percentile ranks of a large benchmark dataset of products. Academics: Good usability; *N* = 8, Scores *M* = 74.4, *CI* = [65.7, 83.0], *MD* = 72.5, *SD* = 10.3, *RG* = [57.5, 90.0], *IQR* = [68.8, 82.5]; Percentiles *M* = 70.9, *CI* = [42.9, 94.3]; Librarians: Poor usability; *N* = 13, Scores: *M* = 56.0, *CI* = [48.1, 63.8], *MD* = 57.5, *SD* = 12.9, *RG* = [30.0, 75.0], *IQR* = [49.4, 65.6]; Percentiles: *M* = 20.9, *CI* = [11.1, 37.4]; Benchmark: Scores: *M* = 68.1; *CL* = 0.95%.

To better understand this disparity and arrive at an informed interpretation of the results, we will need to temporarily set aside the object of evaluation (i.e., the Document Towers) and investigate the evaluators themselves in terms of their psychometric differences.

Psychometric experiments

Predictor selection — To identify which psychological factors might be the best predictors of the experimental participants' usability scores, we drew up a "long list" of salient characteristics of the Document Towers visualization concept: for example, the fact that the Towers are a visual rather than a numerical analysis method, best used for exploration rather than search, require

the comprehension of perhaps unfamiliar concepts (such as deriving information from spatial structure rather than semantic content), and necessitate some familiarity with digital document formats. The selection of these factors was also guided by the unsolicited comments made by participants both during the experiments and in their informal debriefing at the end of the experiments.

Three classes emerged as dominant: skillset, mindset, and experimental setup. We consider that the latter can be excluded as a factor affecting the usability assessment, since the same setup was found to yield similar results in independent evaluation sessions (3 for the librarians and 3 for the academics). The age ranges largely overlapped (27–52 vs 22–51 years), although the median was slightly higher for librarians (47 vs 39.5 years), as was the female-to-male ratio within groups (2.25 vs 1.6).

The skillsets in both groups were mixed (e.g. not all participants were familiar with the digitization standard format ALTO, certain peculiarities of the PDF format (such as the viewport), and even the difference between raster and vector graphics). This being said, the absence of such skills did not seem to preclude the ability to use the Document Towers, as these lacking skills could be compensated for through skill transfer ("The Document Towers could be used for overviewing and navigating in InDesign documents."), generalization ("I am thinking how could I use the Document Towers for my own data."), and collaboration (group members volunteered information to each other; note that collaboration is considered one of the dimensions of creativity (Carroll et al. 2009)).

From the participants' comments, we determined that the major usability difficulties were encountered among librarians: on one hand, with respect to using visualization (as opposed to linguistic and numerical approaches) as an analytic tool, and on the other hand, regarding the need to deal with and extract benefit from the vaguely defined exploration task as opposed to the clearly specified search task.

Representative examples of spontaneous comments made by librarians included: "I am not a visual person", "I am not a creative person, at this time", and "I can get the numbers from the data with a one-line script". This last comment condenses in a single statement the problems encountered by the librarians in the experimental group, namely the problem of visualization not being a necessarily universal cognitive style, the problem of exploration not being a fully understood principle (it would not have been possible to write such a script without the visualization having revealed the issue to be acted upon), and the problem of the resource economy enabled by the Document Towers (numerous such scripts would need to be written to cover all ideas

that might emerge from using the Document Towers, with many of these needing much more than a single line).

By contrast, the academics were more attuned to the Document Towers' visual qualities: "What I found interesting is the global view, to see everything in one glance", and "It is art!"

We thus decided to investigate the mindset-specific factors of mental imagery and curiosity; while both have long been studied in psychology (Berlyne 1960; Pearson & Kosslyn 112), detailed psychometric instruments have only recently been developed.

For example, the two scales we selected (after conducting a survey of the state of the art) for the experiments described below distinguish between two types of visual imagery and five curiosity dimensions. Likewise, there is very little research available on the interaction between related psychological parameters, such as how curiosity may affect creativity (Gross et al. 2020), or how imagery vividness is linked to decision-making (Pham et al. 2001). We were unable to find psychometric scales specifically designed for assessing exploration, considering that exploration is distinct from curiosity (the latter is about acquiring information, while the former is about both acquiring information and transforming it into knowledge that may be acted upon). Interestingly, the curiosity scale we used was developed based on a previous version entitled "Curiosity and Exploration Inventory-II" (Kashdan et al. 2009).

One line of research that comes close to advancing the modeling of the explorative mindset is that constructed around the investment theory of creativity, according to which creativity is related to the amount of investment (of various types) in producing novelties (Sternberg & Lubart 1991).

In addition to the originally proposed investment dimensions (intelligence, knowledge, intellectual style, personality, motivation, and environment), recent work has shown "Openness to Experience" ("reflecting cognitive engagement with perception, fantasy, aesthetics, and emotions"), rather than "Intellectual Curiosity" ("reflecting cognitive engagement with abstract and semantic information, primarily through reasoning"), to be correlated with learning and intelligence in everyday life (von Stumm 2017).

Furthermore, Openness to Experience is more strongly associated with artistic personalities, while Intellectual Curiosity is correlated with scientific ones (Kaufman et al. 2016), meaning that both characteristics are relevant to the type of information representation evaluated here. However, the question of how to measure the effectiveness of the operationalization phase of explo-

ration, after patterns of interest have been identified and their sources understood, remains unanswered.

Figure 18 shows the visual and verbal cognition styles of the evaluation participants profiled via the Object-Spatial Imagery and Visual Questionnaire. (The size of the boxplot whiskers is 1.5 times the upper and lower quantiles, respectively, and values beyond them are considered outliers; the confidence intervals are based on a two-tailed *t*-test at the 95% confidence level.)

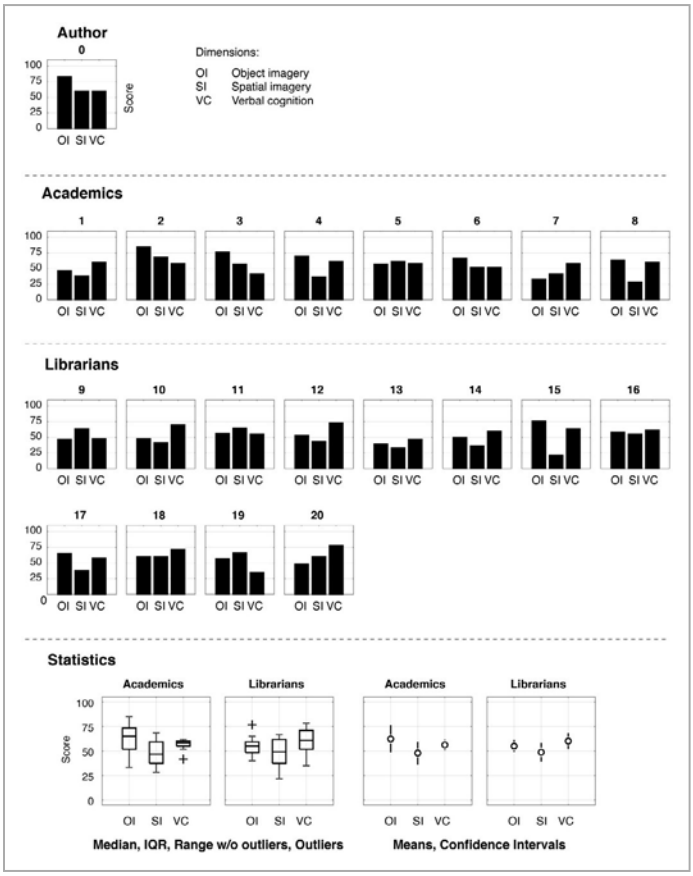


Figure 18. Visual and verbal cognition styles

Mental imagery — The Object-Spatial Imagery and Visual Questionnaire (OSIVQ) is a state-of-the-art, validated, 45-item psychometric questionnaire designed to distinguish between three dimensions of cognitive abilities and preferences for mental representation (Blajenkova et al. 2006; Blajenkova & Kozhevnikov 2009): (1) "Object Imagery" (OI) concerns mental imagery that manifests itself in the form of vivid, colorful, and detailed objects; (2) "Spatial Imagery" (SI) is a type of visual cognition eliciting abstract and schematic mental manifestations of spatial structures and relationships; and (3) "Verbal Cognition" is a dimension that is opposed to the two "visual cog-

nition" dimensions and concerns linguistic preferences and abilities. These three dimensions have been shown to correlate with distinct semiprofessional groups: visual artists score high on "object imagery", scientists on "spatial imagery", and humanists on "verbal cognition".

From among the participants in the usability experiment, $N = 20$ agreed to complete the questionnaire, and the results are shown in Fig. 18. To illustrate the mindset from which this concept originated, the profile of the Document Towers creator (i.e., the author of this article) is included for reference; he identifies with the academics group.

Notably, there are no discernible differences between the profiles of librarians and academics on any of the cognitive dimensions; this is contrary to what we expected given the clear differences in terms of usability satisfaction. The apparent dominance of "Object Imagery" for academics and "Verbal Cognition" for librarians over the other two dimensions is not significant (OI vs SI for academics has a Hedges' $g = 0.89$ and $p = 0.081$; and VC vs SI for librarians $g = 0.48$ and $p = 0.234$; at 95% confidence level).

Curiosity — The Five-Dimensional Curiosity Scale Revised (5CDR) is a state-of-the-art, validated, 24-item psychometric questionnaire designed to measures six aspects of curiosity (Kashdan et al. 2018; Kashdan et al. 2020).

1. The "Joyous Exploration" (JE) dimension refers to a psychological outlook in which the world appears as a fascinating place full of exciting things to explore;

2. "Deprivation Sensitivity" (DS) measures the drive for problem-solving, with high scores being typical among those who "cannot fall asleep if a solution to the problem absorbing them was not found";

3. "Stress Tolerance" (ST) is about coping with the stress induced by novel situations;

4. "Thrill Seeking" (TS) is about deliberately seeking novelties;

5. "Overt Social Curiosity" (OSC) measures the degree to which one seeks to learn more about other people by employing direct questioning and other overt means, and can be contrasted with "Covert Social Curiosity" (CSC), associated with a preference for obtaining this information surreptitiously. The five dimensions may be evaluated individually, as well as together, as part of a personal "profile" (e.g. "Fascinated", "Problem-Solver", "Empathizer", and "Avoider").

As in the case of the mental imagery experiment, the data appears to be heterogeneous, with no distinct trends in dimensions or profiles (fig. 19). Given the small sample size and presence of outliers, we refrain from further statistical analysis beyond visual exploration.

Correlations — In addition to studying the psychometric data per se, it may be instructive to evaluate its correlation with the SUS scores. Again, no clear trends are visible, or, where present, such trends are noticeably weaker than those between the SUS usability scores of the two demographics (fig. 20). Given the combination of low correlation, small sample size, and outliers, it is only possible to hint at some of the factors that might be useful to investigate in future research. Based on the present data, these are "Object Imagery", with low correlation (Kendall's $\tau = 0.335$), and "Deprivation Sensitivity", also with low correlation ($\tau = 0.288$). Kendall's τ is 0 for no correlation, and +1 and −1 for positive and negative correlation; being rank-based, this non-parametric method is robust to outliers. The null hypothesis that "Object Imagery" and SUS scores are correlated cannot be rejected for a two-tailed *t*-test at significance level $\alpha = 0.05$, with $p = 0.042$; however, the correlation hypothesis for "Deprivation Sensitivity" and SUS scores is rejected ($p = 0.090$).

Discussion — How might the apparent lack of differences between the psychological profiles differences of librarians and academics be explained, in view of the clear difference in their attitude towards the Document Towers? Based on the classical nature/nurture dichotomy, we propose two hypotheses.

Our first hypothesis is ecological, and speculates that individuals in the two groups are largely indistinguishable—as the psychometric instruments indeed suggest—as long as they evolve in identical contexts; as soon as these contexts diverge, however, group differences begin to manifest. This seems to be the case for the usability experiment, in which the participants evaluate a visualization concept (the Document Towers) not only from their personal psychological, but also from the perspective of their respective socio-professional contexts. Quite literally, the usability of the Document Towers varies depending on the physical setting in which they are experienced: the library building and the university building.

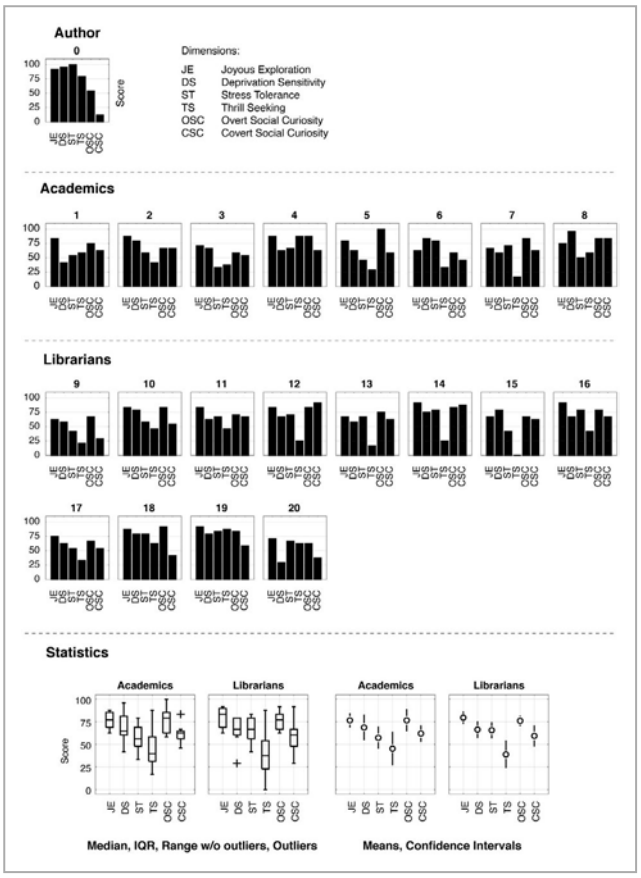


Figure 19. Curiosity styles of the evaluation participants profiled via the Five-Dimensional Curiosity Scale Revised.

The nature of psychological traits is an actively debated subject within the field of psychology, with the current view holding that they may be best described as distributions, whose general statistical makeup is stable for an individual, while comprising variable states; in other words, a psychological profile is modeled as a probabilistic distribution subject to temporal and contextual variation (Fleeson 2001; Bohner & Wänke 2002; Gross et al. 2020). As an example, the SUS scores over a period of one month for the Document Towers as evaluated by the author were 75.0, 77.5, 70.0, and 72.5 (M = 73.8).

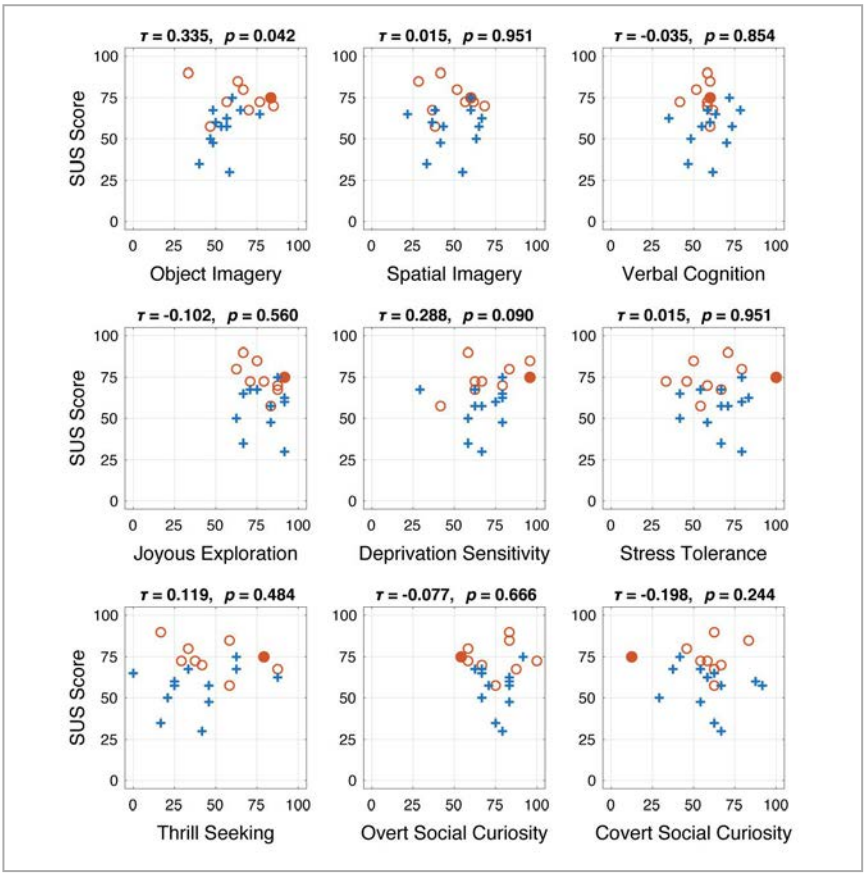


Figure 20. Correlations between psychometric dimensions and subjective usability rating

Figure 20 illustrates the correlations between the psychometric dimensions of the evaluation participants and their subjective usability rating of the Document Towers. The strength of correlation is measured via Kendall's τ, and *p*-values. Red circles represent academics, blue crosses denote librarians, and the red dot symbolizes the author.

Two comments made by participants are representative of the role of socio-professional context in the evaluation of the Document Towers. First, at the beginning of the experiment, a librarian asked if I was aiming to sell them a product ("Are you trying to sell us something?"), indicating that they had construed the evaluation in commercial terms (the misconception that the "guided tour" performance was a salesman's technique was subsequently and duly dispelled). In the context of professional mass digitization, in which "there is a lot of pressure to achieve results within a tight schedule" that obliges employees to work on multiple projects simultaneously [Steffen 2016, p. 2], this comment may also reflect an underlying apprehension that "disruptive technologies" and automation will ultimately increase workloads and stress (Yoose 2015; Rutkowski & Saunders 2019; Panda 2020), in

an already hierarchical and routine-prone environment (Nauratil 1989, pp. 45–46, 53–54, 68–73).

On the other hand, a PhD student mused aloud about possible applications of the Document Towers ("I am thinking how I could use the visualization for my research..."), adopting a "blue sky research" attitude towards the visualization. This participant seemed to enjoy the leisure of sagacity, or the "aptitude for investigation and discovery" (OED Online 2022), which together with chance defines the concept of "serendipity". A recurring answer to the evaluator's question on the roots of the difference between librarians and academics was "They are librarians, not academics!" But what exactly makes librarians "librarians" and academics "academics"?

This brings us to our second hypothesis, which is essentialist in the sense that it considers the differences to be inherent qualities of the participants. An explicit answer provided by an academic located the difference on the librarians' side as "the need to classify, yes or no, with no in-between"—after all, a library cannot adopt serendipity as a classification system (unlike some bookshops (Deutsch 2022, pp. 19–49)). Information classification is indeed a central and time-consuming activity engaged in by librarians, with document misclassification representing a constant source of dread and laborious inventory procedures ("A misplaced book is a lost book" is an often-heard admonition — which obviously ignores the pleasures of finding what has been lost) (Cooper & Wolthausen 1977).

The implied cognitive rigidity of the academic's characterization may be related to the observation that most librarians expect document visualization to be based on the semantics of information rather than its spatial structure, with the latter method deemed "novel", or at least "unexpected". One librarian, after more information on the proposed approach had been supplied during a post-evaluation discussion, commented that "I would have rated the visualization differently, would I have known all this". Moreover, reflecting on the non-predominance of visual representations as information processing tools in libraries, one participant remarked that "after all, libraries are about words, so we use language-based methods". Furthermore, visualization implies a transparency of content that is historically at odds with the material, architectural, technological, and ideological aspects of the library (and even more so of the archive), which encompass the necessity to preserve documents from direct light, the vanishing of documents in virtual clouds, and matters related to censorship.

Since six out of eight academics were either enrolled in a visualization class or had a background in visual art history, it was arguably easier for them to fathom the Document Towers visualization concept—as one participant put

it, the benefit of being able to overview documents is "self-evident" (their education may have contributed towards explaining the small correlation observed between "Object Imagery" and SUS scores, in fig. 20). This is also the spirit in which the author initially conceived of the Document Towers: specifically, contrasting the poor quality of document overview in the typographical layout software InDesign (the author freelances as a graphic designer), as well as digital document readers such as Acrobat, Finder, and DjView, with the availability of the spatial coordinates of document objects in digital file formats such as PDF and InDesign's IDML. Therefore, the difficulties encountered by the librarians—especially when confronted with alternatives that are more time-consuming, resource-intensive, and cumbersome to implement—could seem baffling to other demographics.

While different work contexts may certainly attract different personalities, the explanations collected so far suggest that the observed demographic differences are preponderantly due to educational and socio-professional factors, which encourage a more speculative approach among academics in contrast to more tangible results among librarians. As lightheartedly described in Hazard Adams' classic account of academic life, The Academic Tribes, "Faculties are, after all, composed largely of people who like problems, perhaps even more than solutions, and even to the point of actively seeking them where they have not been recognized" (Adams 1976, p. 11). The data backs up this view to some extent, revealing a moderate correlation between the perceived usability of the Document Towers and a problem-solving-oriented attitude.

Interestingly, the divide revealed by our small usability experiment parallels a much larger and older divide within academia between faculty and librarians—at times acrimonious, despite the common interest in intellectual matters, the shared mission of student education, and suffering from the same time pressures and technology fatigue. Based on this framing, the stereotypical professor would caricature the librarian as conformist, passive, and uninformed (in short, a Shhh-hissing dragon hoarding knowledge), while the stereotypical librarian would lampoon the (shabby) professor as an arrogant, information-technological illiterate jetsetting between far-flung conferences while on tenure track for cozy sabbaticals (Biggs 1981; Jenkins 2005, pp. 1–35). In fact, a dispassionate investigation would expose the source of conflict as a misunderstanding, along with a mutual ignorance of their respective professional work parameters (or, at least, a difficulty in accommodating them): i.e., intellectual hedonism and freedom as the academic's main motivators, and community service within a complex organization as the librarian's job description. Whether related to satisfaction, or lack thereof, with human relations or computer interfaces, these differences are a result (even if not necessarily consciously experienced as such) of the contexts that force

specific perspectives upon the individuals who operate within them. (One additional factor suggested as contributing to the faculty–librarian discord is the pervasive sexism experienced by the (overwhelmingly female) librarians in their interactions with (largely male) academics and IT professionals (Wilson 1979; Brandon et al. 2018)).

More research is needed to clarify these issues; this could involve the study of other groups, such as computer scientists (the developers of the technologies used by the librarians) and graphic designers (a group with strong visual imagery, involved in the creation of human–computer interfaces). In this respect, one aspect to consider is that of avoiding the creation of artificial categories: how much of a "librarian" is the academic who took up the job only recently? or the data scientist who happens to work for a library rather than the history department of a university?

While the results of our psychometric experiments were "negative" (in the sense of lack of clear correlations with usability), we consider them to be valuable research, as they may direct future work in research directions that are hopefully more conclusive. One practical outcome of this experiment was that of creating a concise set of guidelines for readers using the Document Towers (see Section 3.3 "Practice"), in view of their potential and the ability of user training to improve their usability.

A prolific literature on librarianship is available to the interested readers to further their understanding of the sociology (e.g. Nauratil 1989), psychology (e.g. Gullickson Spencer et al. 2015), culture (e.g. Crawford 2015), identity (e.g. Deitering et al. 2017), perception (e.g. Pagowsky & Rigby 2014) and cinematic representation of librarianship (Tevis & Tevis 2005), as well as the faculty–librarian relationship (Jenkins 2005) and the personal experience of information technology work in libraries (Brandon et al. 2018).

Conclusions of user satisfaction and psychometrics experiment

Focusing on two stakeholder demographics of digital libraries, namely academics and librarians, we empirically evaluated the usability of the Document Towers from the point of view of user satisfaction, then interpreted the results in terms of user skillsets and mindsets, experimental setup, and socio-professional contexts, on the basis of psychometric experiments on mental imaging and curiosity, which are two cognitive aspects related to the fundamental features of Document Towers studied in this article: namely, visualization and exploration.

The results suggests that the Document Towers visualization concept has

both "good" and "poor" usability, depending on the target users and application contexts. Academics may benefit more from the Document Towers than librarians; however, training users may improve usability. These findings raise two practical questions of more general import.

First, information systems appear to be more effective if personalized to specific demographics and individuals, according to their various skillsets and, in particular, mindsets and contexts.

Second, it would be useful to consider what changes in the library environment would be conducive to adopting information technologies that seem to be appreciated and effective for information processing outside the library context.

Conclusions

The case study has yielded eight major results, which can be summarized as follows:

1. Confirmation of the utility of Document Towers, along with a list of benefits, conceptual and contextual factors.
2. Drawing attention to the importance of serendipity in document exploration and quality control and the need for tools to support it.
3. Extension of the definition of quality control as a collateral knowledge-producing tool.
4. Demonstrating that even processes, and technologies involved in them, as narrowly defined as document digitization quality control are a dialogue between stakeholders at various stages of the document life cycle.
5. Pointing out the issue of possible disparities between usability as perceived by users and objective effectiveness.
6. Emphasizing the benefits of personalizing information technologies to the users' psychological abilities and preferences, and adapting them to the specifics of socio-professional contexts.
7. Discovery of possibly the first map published in a Swiss newspaper.
8. Some concrete operational changes in quality control at the Swiss National Library, along with the pursuit of future technological collaborations in-

volving the Document Towers between the study participants' respective institutions.

Some concrete operational changes in quality control at the Swiss National Library, along with the pursuit of future technological collaborations involving the Document Towers between the study participants' respective institutions.

Opportunities — In line with the initial exploratory goal, the case study described a surprising trajectory from assessing the Document Towers to a historical discovery and back to its utility for quality control. These results recast quality control as a dialogue between librarians, technologists, and readers that is mutually enriching to each community. The dialogue is dynamic, ever-shifting in terms of individual requirements, and creating significant retroprocessing of the same data, as we saw, making it difficult to predict, and complex to model. There is thus a need for appropriately open and flexible technologies.

The importance of this dialogue has been demonstrated in concrete terms by a number of objectively effective aspects of the Document Towers that are not perceived as such by some user groups (e.g. document overview at a glance vs page scrolling). We have situated the disparity between subjective and objective effectiveness at the confluence of individual psychology and socio-professional contexts, and further proposed to reduce this gap through individual training and adjusting the context.

P10 – 236

Benefits — The Document Towers are specifically appropriate for quality control because they make document structures visible, preserve the serendipity of pattern discovery (as demanded by quality control), and combine the complementary pattern recognition strengths of humans and machines. The use of the Document Towers has the added value of prompting questions that link quality control to the much broader issue of the usability of digital data and computational resources.

The flexibility of the Document Towers contributes to their performance. Their representation is generic, with any kind of spatially defined information being representable. It covers multiple document scales, from individual pages and single documents to collections, can address multiple sensory modalities (most notably vision, audition, and touch), and has variable information density, from three-dimensional Document Towers to two-dimensional Ribbons to one-dimensional Chips (Atanasiu & Ingold 2021). Last but not least, the representation has polyvalent application potential beyond quality control (Atanasiu 2022a).

Paradigm — The observed performance is, firstly, a result of the naturalness and affordance of the documents-as-Document Towers paradigm, and further answers the critical questions "Why 3D? Why Document Towers?" Paginated documents are physical or conceptual three-dimensional objects, meaning their representation as architectural models maintains structural information, while lower-dimensional representations break the holistic integrity and reduce it to specific aspects. It is thus easy to identify, thanks to a three-dimensional representation, the image at the top of the first page of the ZFZ newspaper in Fig. 11, which is replicated across all issues as the publication's logo. It is precisely these three-dimensional spatial attributes that made Napoleon's campaign map stand out (large size and location in the middle of the front page).

Our familiarity with urban Document Towers not only facilitates our interaction with Document Towers (an excellent design feature), but causes us to think about and act differently towards documents in terms of the target domain to which the documents are mapped (much like the reasons for using mathematical transforms). For example, handling XML files becomes far easier simply by referring to them as "tree-like"; similarly, by seeing the structures as "Document Towers", we naturally become interested in measuring information fragmentation. Furthermore, spatial thinking is a fundamental cognitive ability, and its use in information design is well documented, from present times back to antiquity, as mentioned in (Atanasiu 2022a).

In addition to the specific form of the Document Towers metaphor (architecture and urbanism), their superior performance is also due to some of the generic functions of metaphors, such as their propensity to intrigue and induce a change of mental perspective.

Context — While the generic character of the Document Towers support generalizing their utility from this case study, their acceptance and fit is partially due to certain noteworthy local specificities. Switzerland's strong federal political system has heretofore precluded the imposition of a unified national digitization program, unlike, for instance, France, a centralized state, where the French National Library has a greater influence on the development of documentary standards than some of its counterparts. Digitization approaches have proliferated in Swiss cantonal, university, and other types of libraries; only recently, with growing amounts of data, library inter-connectivity projects, and rising costs, have the methods and standardization of digitization quality and interaction ergonomics emerged as real issues (Steffen 2019). In this context, the Swiss National Library has a greater openness for "unconventional solutions" (FS) and hybrid human–machine document management processes such as the Document Towers. The "highly political" nature of mass digitization projects (Thylstrup 2019, p. 4) has thus surrepti-

tiously manifested itself, even in a benign academic research project.

Actions — The success of the Document Towers might also be judged by the fact that, although the setting was that of artificial experiments, they prompted a series of actions in the real-world. As a consequence of this evaluation, an actual digitization quality error was identified with respect to non-editorial objects. Given the interest in the missing data among the library patrons, the head of the Digitization Unit, FS, decided to require this kind of information in future calls for tender from digitization service providers. The Document Towers also prompted further discussions among the participants as to what kind of information would be useful to users (which is a question of system design), how to extract it efficiently (a question of pattern analysis), and how to convey it to users (a question of human–computer interaction). Ultimately, the participants decided to develop the visualization into a commercial product in collaboration with an industrial partner.

On a longer-term horizon, the large difference in the perceived usability of the Document Towers between librarians and academics is of sufficient importance for the general issue of the adoption of novel or non-mainstream information technologies in libraries that it warrants further research. What is at stake was put in the following unambiguous terms by one scholar of librarianship: "An awareness of the structural factors involved in the creation and resolution of [librarianship] problems [...] may well be the sine qua non of our effective survival as a profession" (Nauratil 1989, p. 105).

Acknowledgments

The author wishes to express his gratitude to Andreas Fischer and Florian Steffen, as well as to each and every participant in the evaluation experiments, whose contributions have made this research possible. The reviewers are cordially thanked for their unrelenting efforts to support the development of an empirical usability evaluation. K. F. is thanked for having improved the manuscript's readability. This research was supported by the Hasler Foundation, Switzerland, under grant No. 17039.

References

Adams, H. (1976) The Academic Tribes. Liveright, New York, NY.

Almeida, R. A. de, Alessio, P., Topol, A., and Cubaud, P. (2009) A Visualization Technique for Quality Control of Massive Digitization Programs. Proceedings of the 13th European Conference on Research and Advanced Technologies for Digital

Libraries (ECDL 2009), 27 September – 2 October 2009, Corfu, Greece, pp. 150–155.

Almeida, R. A. de, Cubaud, P., Dupire, J., and Topol, A. (2006) Interactions et métadonnées riches pour les bibliothèques numérisées. Document numérique, 9 (2): 83–109, https://doi.org/10.3166/dn.9.2.83-109.

Alsina, C., and Nelsen, R. B. (2006) Math Made Visual: Creating Images for Understanding Mathematics. Mathematical Association of America, Washington, DC.

Anderson, M., Meyer, B., and Olivier, P. (2002) Diagrammatic Representation and Reasoning. Springer, London.

Apache Software Foundation, The (2022) Apache PDFBox – A Java PDF Library. https://pdfbox.apache.org. Accessed 2022/08/18.

Atanasiu, V. (2014) Expert Bytes: Computer Expertise in Forensic Documents – Players, Needs, Resources and Pitfalls. CRC Press, Boca Raton, FL. https://doi.org/10.1201/b15467.

Atanasiu, V. (2016) Cuckoos among Your Data: A Quality Control Method to Retrieve Mislabeled Writer Identities from Handwriting Datasets. In IS&T International Symposium on Electronic Imaging 2016, Document Recognition and Retrieval XXIII (DRR'16), 15–18 February 2016, San Francisco, CA, article 56, 14 pages.

Atanasiu, V. (2022a) Applications of the Document Towers method of representing document structures. Information Design Journal, in press. https://doi.org/10.1075/idj.20026.ata.

Atanasiu, V. (2022b) The Structural Information Potential and Its Application to Document Triage. IEEE Access, 10: 13103–13138. https://doi.org/10.1109/ACCESS.2021.3133654.

Atanasiu, V., and Ingold, R. (2021) Document Towers: A MATLAB software implementing a three-dimensional architectural paradigm for the visual exploration of digital documents and libraries. SoftwareX, 14: 100684. https://doi.org/10. 1016/j.softx.2021.100684.

Back, M., Jonathan, C., Harrison, S., and Minneman, S. (2006) Speeder Reader: An experiment in the future of reading. Computers & Graphics, 26: 623–627.

Bangor, A., Kortum, P. T., and Miller, J. T. (2008) An Empirical Evaluation of the System Usability Scale. International Journal of Human–Computer Interaction, 24 (6): 574–594. https://doi.org/10.1080/10447310802205776.

Bangor, A., Kortum, P. T., and Miller, J. T. (2009) Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies, 4 (3): 114–123.

Barnett, V., and Lewis, T. (1978) Outliers in Statistical Data (3rd ed). John Wiley & Sons, Chichester.

Bederson, B. B., Clamage, A., Czerwinski, M. P., and Robertson, G. G. (2004) DateLens: A Fisheye Calendar Interface for PDAs. ACM Transactions on Computer–Human Interaction, 11 (1): 90–119.

Berlyne, D. E. (1960) Conflict, arousal, and curiosity. McGraw-Hill, New York, NY.

Biggs, M. (1981) Sources of Tension and Conflict between Librarians and Faculty. The Journal of Higher Education, 52 (2): 182–201. https://www.jstor.org/stable/1981090.

Blajenkova, O., and Kozhevnikov, M. (2009) The New Object-Spatial-Verbal Cognitive Style Model: Theory and Measurement. Applied Cognitive Psychology, 23: 638–663. https://doi.org/10.1002/acp.1473.

Blajenkova, O., Kozhevnikov, M., and Motes, M. A. (2006) Object-Spatial Imagery: A New Self-Report Imagery Questionnaire. Applied Cognitive Psychology, 20: 239–263. https://doi.org/10.1002/acp.1182.

Bogel, E. (1973) Schweizer Zeitungen des 17. Jahrhunderts : Beiträge zur frühen Pressegeschichte von Zürich, Basel, Bern, Schaffhausen, St. Gallen und Solothurn, Schünemann, Bremen.

Bohner, G., and Wänke, M. (2002) Attitudes and Attitudes Change. Psychology Press, Hove.

Bollinger, E. (2018) Freitags-Zeitung. Historisches Lexicon der Schweiz. https://hls-dhs-dss.ch/de/articles/043059/2018-01-11/. Accessed 2022/08/18.

Bonde, T. N. (2019) The Politics of Mass Digitization. MIT Press, Cambridge, MA.

Bourdieu, P. (1984) Distinction: A Social Critique of the Judgement of Taste. Harvard University Press, Cambridge, MA.

Brandon, J., Ladenson, S., and Sattler, K. (2018) We Can Do IT: Women in Library Information Technology. Library Juice Press, Sacramento, CA.

Briet, S. (1951) Qu'est-ce que la documentation ? Éditions documentaires, industrielles et techniques, Paris.

Brooke, J. (1996) SUS: a 'quick and dirty' usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A., and Mcclelland, I. L. (eds), Usability Evaluation in Industry. Taylor & Francis, London, pp. 189–194.

Brooke, J. (2013) SUS: A Retrospective. Journal of Usability Studies, 8 (2): 29–40.

Brysbaert, M. (2019) How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. Journal of Cognition, 2 (1): 1–38. https://doi.org/10.5334/joc.72.

Buckland, M. (1991) Information and Information Systems. Praeger, New York, NY.

Bush, V. (1945) As we may think. Atlantic Monthly, 176 (1): 101–108. http://dougengelbart.org/content/view/ 258/000/.

Card, S. (2008) Information Visualization. In Jacko, J. A., and Andrew, S. (ed), The Human–Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications (2nd ed). Lawrence Erlbaum, New York, 509–543.

Carroll, E. A., Latulipe, C., Fung, R., and Terry, M. (2009) Creativity Factor Evaluation: Towards a Standardized Survey Metric for Creativity Support. Proceedings

of the 2009 ACM SIGCHI Conference on Creativity and Cognition (C&C '09), 26–30 October 2009, Berkeley, CA, USA, pp. 127–136.

Center for Research Libraries (2015) The "State of the Art": A Comparative Analysis of Newspaper Digitization to Date. http://www.crl.edu/sites/default/files/d6/attachments/events/ICON_Report-State_of_Digitization_final.pdf. Accessed 2019/04/02.

Chapman, J., and Leonard, S. (2013) Cost and benefit of quality control visual checks in large-scale digitization of archival manuscripts. Library Hi Tech, 31 (3): 405–418. https://doi.org/10.1108/LHT-01-2013-0002.

Chen, L., Wang, N., Yang, Y., Yang, K., and Yuan, Q. (2022) User Validation of Recommendation Serendipity Metrics. https://doi.org/10.48550/arXiv.1906.11431. Accessed 2022/08/18.

Clausner, C., Papadopoulos, C., Pletschacher, S., and Antonacopoulos, A. (2015) The ENP Image and Ground Truth Dataset of Historical Newspapers. Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR 2015), 23–26 August 2015, Nancy, France, pp. 931–935. https://doi.org/10.1109/ICDAR.2015.7333898.

Clausner, C., Pletschacher, S., and Antonacopoulos, A. (2011) Aletheia — An Advanced Document Layout and Text Ground-Truthing System for Production Environments. Proceedings of the 11th International Conference on Document Analysis and Recognition (ICDAR 2011) (18–21 September 2011) Beijing, China, 48–52. https://doi.org/10.1109/ICDAR.2011.19.

Clausner, C., Pletschacher, S., and Antonacopoulos, A. (2016) Quality Prediction System for Large-Scale Digitisation Workflows. Proceedings of the 12th IAPR International Workshop on Document Analysis Systems (DAS 2016), 11–14 April 2016, Santorini, Greece, pp. 138–143. https://doi.org/10.1109/DAS.2016.82.

Cockburn, A., Gutwin, C., and Alexander, J. (2006) Faster Document Navigation with Space-Filling Thumbnails. Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI '06), 22–28 April 2006, Montréal, Québec, Canada, 10 pages.

Cockburn, A., Karlson, A., and Bederson, B. B. (2008) A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. ACM Computing Surveys, 41 (1): 1–31. https://doi.org/10.1145/1456650.1456652.

Coe, R. (2002) It's the Effect Size, Stupid! What effect size is and why it is important. Proceedings of the Annual conference of the British Educational Research Association, 12–14 Sep. 2002, Exeter, UK. https://cebma.org/wp-content/uploads/Coe-2002.pdf.

Content Conversion Specialists. (n.d.) docWorks. https://content-conversion.com/#docworks-2. Accessed 2019/03/30.

Cooper, M. D., and Wolthausen, J. (1977) Misplacement of Books on Library Shelves: A Mathematical Model. The Library Quarterly: Information, Community, Policy, 47 (1): 43–57. https://www.jstor.org/stable/4306756.

Crawford, A. (2015) The Meaning of the Library: A Cultural History. Princeton University Press, Princeton, NJ.

Deitering, A.-M., Stoddart, R., and Schroeder, R. (2017) The Self as Subject: Autoethnographic Research into Identity, Culture, and Academic Librarianship. Association of College and Research Libraries, Chicago, IL.

Demm, E. (n.d.) Censorship. In International Encyclopedia of the First World War. https://encyclopedia.1914-1918-online.net/article/censorship. Accessed 2019/09/27.

Deutsch, J. (2022) In Praise of Good Bookstores. Princeton University Press, Princeton, NJ.

Diagrams (2000–) Diagrams: International Conference on Theory and Application of Diagrams. https://link.springer. com/conference/diagrams. Accessed 2022/08/18.

Diaper, D., and Stanton, N. A. (2004) The Handbook of Task Analysis for Human–Computer Interaction. Lawrence Erlbaum, Mahwah, NJ.

Drucker, P. F. (2015) Innovation and Entrepreneurship: Practice and Principles. Routledge, London.

Dunning, A., The European Library, and Europeana Foundation (2012) D4.1: European Newspaper Survey Report. http://www.europeana-newspapers.eu/wp-content/uploads/2012/04/D4.1-Europeana-newspapers-survey-report.pdf. Accessed 2019/04/02.

e-newspaperarchives.ch (n.d.) https://www.e-newspaperarchives.ch/. Accessed 2019/03/30.

Elmqvist, N., and Yi, J. S. (2015) Patterns for Visualization Evaluation. Information Visualization, 14 (3): 250–269.

Europeana Newspapers (n.d.) http://www.europeana-newspapers.eu. Accessed 2019/04/02.

Fandom contributors (n.d.) Jedi Archives. https://starwars.fandom.com/wiki/Jedi_Archives/Legends. Accessed 2022/04/27.

Fang, X., Jacquemin, C., Vernier, F., and Luo, B. (2009) A Survey of 3D Document Corpus Visualization. Information Technology Journal, 8 (1): 1–15.

Finstad, K. (2006) The System Usability Scale and Non-Native English Speakers. Journal of Usability Studies, 1 (4): 185–188.

Fleeson, W. (2001) Toward a Structure- and Process-Integrated View of Personality: Traits as Density Distributions of States. Journal of Personality and Social Psychology, 80 (6): 1011–1027. https://doi.org/10.1037//0022-3514.80.6.1011.

Franklin, R. D., Allison, D. B., and Gorman, B. S. (2014) Design and Analysis of Single-Case Research. Psychology Press, New York, NY.

Frölich, M., and Sperlich, S. (2019) Impact Evaluation: Treatment Effects and Causal Analysis. Cambridge: Cambridge University Press.

Gardner, M. (1958) Logic Machines and Diagrams. McGraw-Hill, New York, NY.

Gidalewitsch, N. (1956) Bildbericht und Presse. Ein Beitrag zur Geschichte und Organisation der illustrierten Zeitungen. H. Laupp, Tübingen.

Gross, M. E., Zedelius, C. M., and Schooler, J. W. (2020) Cultivating an under-

standing of curiosity as a seed for creativity. Current Opinion in Behavioral Sciences, 35: 77–82. https://doi.org/10.1016/j.cobeha.2020.07.015.

Guggenbühl, C. (1996) Zensur und Pressefreiheit : Kommunikationskontrolle in Zürich an der Wende zum 19. Jahrhundert. Chronos, Zürich.

Gullickson Spencer, L., VandeCreek, L., and Wright, H. S. (2015) Technology Anxiety in Academic Professional and Paraprofessional Library Staff. Library Juice Press, Sacramento, CA.

Hafner, U. (2015) Subversion im Satz : Die turbulenten Anfänge der "Neuen Zürcher Zeitung" (1780–1798). Neue Zürcher Zeitung, Zürich.

Heath, C., and Luff, P. (2000) Technology in Action. Cambridge University Press, Cambridge.

Hedges, L. V., and Ingram, O. (1985) Statistical Methods for Meta-Analysis. Academic Press, Orlando, FL.

Heunen, C., Sadrzadeh, M., and Grefenstette, E. (2013) Quantum Physics and Linguistics: A Compositional, Diagrammatic Discourse. Oxford University Press, Oxford.

Hoffmann-Feer, E. (1969) Die Typographie im Dienste der Landkarte: Zusammenfassung von bisher erschienenen Veröffentlichungen über typographisch hergestellte Landkarten mit besonderer Berücksichtigung der Haas'schen Blätter. Verlag Helbing & Lichtenhahn, Basel.

Höhener, H.-P. (2018) Alte Karten des Gebiets des heutigen Kantons St. Gallen. Cartographica Helvetica, 56: 13.

i2s (n.d.) LIMB Maestro: Supervising your digitization projects in real time. https://www.i2s.fr/en/product/limb-maestro. Accessed 2019/03/30.

Igarashi, T., and Hinckley, K. (2000) Speed-dependent Automatic Zooming for Browsing Large Documents. Proceedings of the 13th annual ACM symposium on User interface software and technology (UIST '00), 6–8 November 2000, San Diego, CA, United States, pp. 139–148. https://doi.org/10.1145/354401.354435.

Ingold, R., Bloechle, J.-L., and Rigamonti, M. (2014) Reverse-Engineering of PDF Files. In Chaudhuri, B. B. and Parui, S. K. (eds), Advances in Digital Document Processing and Retrieval. World Scientific, Singapore, pp. 249–285.

International Organization for Standardization (2018) ISO 9241-11:2018(E) 2018. Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. Standard. International Organization for Standardization, Geneva.

Jenkins, P. O. (2005) Faculty–Librarian Relationships. Chandos, Oxford.

Jordan, P. W., Thomas, B., Werdmeester, B. A., and McClelland, I. L. (1996) Usability Evaluation in Industry. Taylor & Francis, London.

Kashdan, T. B., Disabato, D. J., Goodman, F. R., and McKnight, P. E. (2020) The Five-Dimensional Curiosity Scale Revised (5DCR): Briefer subscales while separating overt and covert social curiosity. Personality and Individual Differences, 157: 109836. https://doi.org/10.1016/j.paid.2020.109836.

Kashdan, T. B., Gallagher, M. W., Silvia, P. J., Winterstein, B. P., Breen, W. E., Terhar, D., and Steger, M. F. (2009) The Curiosity and Exploration Inventory-II: Development, Factor Structure, and Psychometrics. Journal of Research in Personality, 43 (6): 987–998. https://doi.org/10.1016/j.jrp.2009.04.011.

Kashdan, T. B., Stiksma, M. C., Disabato, D. J., McKnight, P. E., Bekier, J., Kaji, J., and Lazarus, R. (2018) The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. Journal of Research in Personality, 73: 130–149. https://doi.org/10.1016/j.jrp.2017.11.011.

Kaufman, S. B., Quilty, L. C., Grazioplene, R. G., Hirsh, J. B., Gray, J. R., Peterson, J. B., and DeYoung, C. G. (2016) Openness to experience and intellect differentially predict creative achievement in the arts and sciences. Journal of Personality, 84 (2): 248–258. https://doi.org/10.1111/jopy.12156.

Kawaguchi, H., and Suzuki, N. (2018) Recommending Colors and Fonts for Cover Page of EPUB Book. Proceedings of the ACM Symposium on Document Engineering 2018, 28–31 August 2018, Halifax, NS, Canada, article 48, 4 pages. https://doi.org/10.1145/3209280.3229086.

Kitodo. (n.d.) Kitodo.Production: Workflow management for digitisation. http://www.kitodo.org/ en/software/kitodoproduction/. Accessed 2019/03/30.

Kline, R. B. (2004) Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research. American Psychological Association, Washington, DC.

Kortum, P. T., and Bangor, A. (2013) Usability Ratings for Everyday Products Measured With the System Usability Scale. International Journal of Human–Computer Interaction, 29 (2): 67–76. https://doi.org/10.1080/10447318.2012.681221.

Kume, H. (1985) Statistical Methods for Quality Improvement. The Association for Overseas Technical Scholarship, Tokyo.

Lador, P.-Y. (1990) L'abeille et la poule. Arbido-R, 5 (4): 113–115. https://www.e-periodica.ch/digbib/view?pid=arb-012%3A1990%3A0%3A%3A118&referrer=search#134.

Leavy, P. (2014) The Oxford Handbook of Qualitative Research. Oxford University Press, Oxford.

Lee, J., Olwal, A., Ishii, H., and Boulanger, C. (2013) SpaceTop: Integrating 2D and Spatial 3D Interactions in a See-through Desktop Environment. Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI 2013), 27 April – 2 May 2013, Paris, France, 4 pages.

Lehman, S. (2011) Mapping Havana in the Gentleman's Magazine, 1740–1762. In Brückner, M. (ed), Early American Cartographies, University of Carolina Press, Chapel Hill, NC, pp. 339–359.

Lewis, J. R., and Sauro, J. (2018) Item Benchmarks for the System Usability Scale. Journal of Usability Studies, 13 (3): 158–167.

Liebert, W.-A. (1994) Lascaux – A Hypermedia Lexicon of Metaphor Models for Scientific Imagination. Proceedings of the Sixth EURALEX International Congress on Lexicography, 30 August – 3 September 1994, Amsterdam, The Netherlands, pp. 495–500.

Maissen, T. (2005), Die Geschichte der NZZ, 1780–2005, Neue Zürcher Zeitung, Zürich.

Mandelbrot, B. (1982) The Fractal Geometry of Nature. W. H. Freeman, New York, NY.

Marshall, C., and Rossman, G. B. (1995) Designing Qualitative Research (2nd ed). Sage, Thousand Oaks, CA.

Martin, K., and Quan-Haase, A. (2016) The role of agency in historians' experiences of serendipity in physical and digital information environments. Journal of Documentation, 72 (6): 1008–1026. https://doi.org/10.1108/JD-11-2015-0144.

Martin, K., and Quan-Haase, A. (2017) 'A Process of Controlled Serendipity': An Exploratory Study of Historians' and Digital Historians' Experiences of Serendipity in Digital Environments. Proceedings of the 80th Annual Meeting of the Association for Information Science and Technology (ASIS&T), 27 October – 1 November 2017, Washington, DC, USA, pp. 289–297. https://doi.org/10.1002/pra2.2017.14505401032.

McCay-Peet, L., and Toms, E. G. (2018) Researching serendipity in Digital Information Environments. Morgan & Claypool, San Rafael, CA.

Merton, R. K., and Barber, E. (2004) The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science. Princeton University Press, Princeton, NJ.

Monmonier, M. (1989) Maps with the News: The Development of American Journalistic Cartography. University of Chicago Press, Chicago, IL.

Morley, S. (Auth.), Masterson, C., and Chris J. Main, (2017) Single-Case Methods in Clinical Psychology: A Practical Guide. Routledge, Abingdon.

Munzner, T. (2015) Visualization Analysis and Design. CRC Press, Boca Raton, FL.

Mussell, J. (2012) The Nineteenth-Century Press in the Digital Age. Palgrave Macmillan, Houndsmills, UK.

Mühlberger, G., and Hackl, G. (2015) D5.3: Final public release with updated online resource for documentation. http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/D5.3_Final_release_ ENMAP_1.0.pdf. Accessed 2019/04/02.

Nakatsu, R. (2010) Diagrammatic Reasoning in AI. John Wiley & Sons, Hoboken, NJ.

Nauratil, M. J. (1989) The Alienated Librarian. Greenwood Press, New York, NY.

Neue Zürcher Zeitung (1980) 200 Jahre Neue Zürcher Zeitung: 1780–1980. Neue Zürcher Zeitung, Zürich.

Nielsen, J. (1993) Usability Engineering. Morgan Kaufmann, San Diego, CA.

Novaes, A. R. (2019) Maps in Newspapers: Approaches of Study and Practices in Portraying War Since 19th Century. Brill, Leiden.

Oakland, J. (2008) Statistical Process Control (6th ed). Butterworth-Heineman, Amsterdam.

Ong, W. J., and Hartley, J. (2012) Orality and Literacy: The Technologizing of the Word. Routledge, London.

OED Online (2022) sagacity, n. Oxford English Dictionary Online. https://www.oed.com/view/Entry/169718. Accessed 2022/07/16.

Ooi, K. (2019) How adult fiction readers select fiction books in public libraries: a study of information-seeking in context. Master's thesis. School of Information Management, Victoria University of Wellington, Wellington, New Zealand.

Pagowsky, N., and Rigby, M. (2014) The Librarian Stereotype: Deconstructing Perceptions and Presentations of Information Work. Association of College and Research Libraries, Chicago, IL.

Panda, S. (2020) Technostress among Library Professionals: Possible Causes, Symptoms, Coping Strategies, and Future Proposals. International Journal of Information Studies and Libraries, 5 (2): 1–19.

Patton, M. Q. (2001) Qualitative Research and Evaluation Methods (3rd ed). Sage, Thousand Oaks, CA.

Pearson, J., and Kosslyn, S. M. (2015) The Heterogeneity of Mental Representation: Ending the Imagery Debate. Proceedings of the National Academy of Sciences of the United States of America, 112 (33): 10089–10092, https://doi.org/doi.org/10.1073/pnas.1504933112.

Peirce, C. (1902) Logic, Regarded As Semeiotic. Archive document, Memoir 4, Draft C, MS L75.91–92, Carnegie application. Ransdell, J. (ed) https://arisbe.sitehost.iu.edu/menu/library/bycsp/L75/ver1/l75v1-02.htm. Accessed 2023/03/06.

Pham, M. T., Meyvis, T., and Zhou, R. (2001) Beyond the Obvious: Chronic Vividness of Imagery and the Use of Information in Decision Making. Organizational Behavior and Human Decision Processes, 84 (2): 226–253.

Pletschacher, S., and Antonacopoulos, A. (2010) The PAGE (Page Analysis and Ground-Truth Elements) Format Framework. In Proceedings of the 20th International Conference on Pattern Recognition (ICPR 2010), 23–26 August 2010, Istanbul, Turkey, pp. 257–260. https://doi.org/10.1109/ICPR.2010.72.

Reinfurt, D., and Wiesenberger, R. (2014) Messages and Means: Muriel Cooper at MIT. (2014) exhibition.

Rickenbacher, M. (2011) Napoleons Karten der Schweiz: Landesvermessung als Machtfaktor 1798–1815. hier + jetzt, Baden, CH.

Riley, J., and Whitsel, K. (2005) Practical Quality Control Procedures for Digital Imaging Projects. OCLC Systems & Services 21 (1): 40–48.

Rossi, P. H., Lipsey, M. W., and Freeman, H. E. (2004) Evaluation: A systematic Approach (7th ed). Sage, Thousand Oaks, CA.

Rutkowski, A.-F., and Saunders, C. S. (2019) Emotional And Cognitive Overload: The Dark Side of Information Technology. Routledge, Abingdon.

Sadiq, Sh. (2013) Handbook of Data Quality: Research and Practice. Springer, Berlin.

Sauro, J. (2011) A Practical Guide to the System Usability Scale. Measuring Usability LLC, Denver, CO.

Sauro, J., and Lewis, J. R. (2011) When Designing Usability Questionnaires, Does It Hurt to Be Positive? Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2011), 7–12 May 2011, Vancouver, BC, Canada, pp. 2215–2223.

Sauro, J., and Lewis, J. R. (2016) Quantifying the User Experience: Practical Statistics for User Research (2nd ed). Morgan Kaufmann, Cambridge, MA.

Schneider, B., Christoph, E., and Wöpking, J. (2016) Diagrammatik-Reader: Grundlegende Texte aus Theorie und Geschichte. De Gruyter, Berlin. https://doi.org/10.1515/9783050093833.

Schoot, R. van de, and Miočević, M. (2020) Small Sample Size Solutions: A Guide for Applied Researchers and Practitioners. Routledge, Abingdon.

Schütte, S. (2005) Engineering Emotional Values in Product Design: Kansei Engineering in Development. Ph.D. Dissertation. Department of Mechanical Engineering, University of Linköping, Linköping, Sweden. No. 951.

Sellen, A. J., and Harper, R. H. R. (2002) The Myth of the Paperless Office. MIT Press, Cambridge, MA.

Serenson, C. L. (2000) Guides to Quality in Visual Resource Imaging: Planning an Imaging Project. Washington, DC. https://www.oclc.org/research/publications/library/visguides/visguide1.html. Accessed 2020/01/20.

Shannon, E. C. (1948) A Mathematical Theory of Communication. The Bell Systems Technical Journal, 27 (3): 379–423.

Signer, B., and Norrie, M. C. (2010) Interactive Paper: Past, Present and Future. Proceedings of UbiComp 2010: 1st International Workshop on Paper Computing (PaperComp 2010), 26–29 September 2010, Copenhagen, Denmark, 5 pages.

Spence, R., and Witkowski, M. (2013) Rapid Serial Visual Presentation: Design for Cognition. Springer, London.

Staatsbibliothek zu Berlin (2017) Empfehlungen zur Digitalisierung historischer Zeitungen in Deutschland (Masterplan Zeitungsdigitalisierung): Ergebnisse des DFG-Projektes "Digitalisierung historischer Zeitungen" Pilotphase 2013-2015. http://www.zeitschriftendatenbank.de/fileadmin/user_upload/ZDB/z/Masterplan.pdf. Accessed 2019/04/02.

Steffen, F. (2016) Lessons learned in 7 years of newspaper retro-digitization in Switzerland: How to learn lessons from mistakes, how to establish models and guidelines to avoid repeating them, how to share the lessons learned. Proceedings of the 2016 IFLA [International Federation of Library Associations and Institutions] International News Media Conference, 20–22 April 2016, Hamburg, Germany. http://blogs.sub.uni-hamburg.de/ifla-newsmedia/wp-content/uploads/2016/04/Steffen-Lessons-learned-in-7-years-of-newspaper-retro-digitization-in-Switzerland.pdf.

Steffen, F. (2019) Der Computergestützte Digitalisierungsworkflow bei der

Zeitungsdigitalisierung [Computer-Aided Digitization Workflow for Newspaper Digitization]. Master's thesis. Wildau Institute of Technology, Wildau, Germany.

Sternberg, R. J., and Lubart, T. I. (1991) An Investment Theory of Creativity and Its Development. Human Development, 34 (1): 1–31. https://doi.org/10.1159/000277029.

Stockar, C. (2012) Stempelsteuer. Historischer Lexikon der Schweiz. https://hls-dhs-dss.ch/de/articles/013772/2012-02-14/. Accessed 2022/08/18.

Stone, D., Jarrett, C., Woodroffe, M., and Minocha, S. (2005) User Interface Design and Evaluation. Morgan Kaufmann, San Francisco, CA.

Stumm, S. von (2017) Better Open Than Intellectual: The Benefits of Investment Personality Traits for Learning. Personality and Social Psychology Bulletin, 44 (4): 562–573. https://doi.org/10.1177/0146167217744526.

Sugarcube. (n.d.) https://www.sugarcube.ch. Accessed 2020/01/10.

Swiss National Library. (n.d.) https://www.nb.admin.ch/snl/en/home.html. Accessed 2019/03/30.

Tevis, R., and Tevis, B. (2005) The Image of Librarians in Cinema, 1917-1999. McFarland & Co, Jefferson, NC.

Thylstrup, N. B. (2019) The Politics of Mass Digitization. MIT Press, Cambridge, MA.

Tschichold, J. (1995 [1928]) The New Typography: A Handbook for Modern Designers. University of California Press, Berkeley, CA.

Tukey, J. W. (1977) Exploratory Data Analysis. Addison-Wesley, Reading, MA.

Tullis, T. S., and Stetson, J. N. (2004) A Comparison of Questionnaires for Assessing Website Usability. In Proceedings of the 13th Annual Usability Professionals' Association Conference (UPA 2004), 7–11 June 2004, Minneapolis, MN, USA, 12 pages.

Turing, A. M. (1950) Computing Machinery and Intelligence. Mind, 59 (236) Pp. 433–460.

Turner, P. (2017) A Psychology of User Experience: Involvement, Affect and Aesthetics. Springer, Cham.

Weisz, L., and Verein der Schweizer Presse (1933) Zürcher Zeitungen aus den Jahren 1521–1531. Keller, Luzern.

Wikimedia Commons contributors (n.d.) File:Erasmus censored.png. https://commons.wikimedia.org/wiki/File:Erasmus_censored.png. Accessed 2019/06/07.

Wikipedia contributors [n.d., a]. ALTO (XML) Website. https://en.wikipedia.org/wiki/ALTO_(XML). Accessed 2019/04/01.

Wikipedia contributors [n.d., b]. Metadata Encoding and Transmission Standard. https://en.wikipedia.org/wiki/Metadata_Encoding_and_Transmission_Standard. Accessed 2019/04/01.

Wikipedia contributors [n.d., c]. Stamp Act 1765. https://en.wikipedia.org/wiki/Stamp_Act_1765. Accessed 2019/09/27.

Wikipedia contributors [n.d., d]. Think (slogan). https://en.wikipedia.org/wiki/Think_(slogan). Accessed 2022/08/01.

Wikipedia contributors [n.d., e]. Trove. https://en.wikipedia.org/wiki/Trove. Accessed 2022/08/19.

Willig, C., and Stainton-Rogers, W. (2008) The SAGE Handbook of Qualitative Research in Psychology. Sage, Thousand Oaks, CA.

Wilson, P. (1979) Librarians as Teachers: The Study of an Organization Fiction. The Library Quarterly: Information, Community, Policy, 49 (2): 146–162, http://www.jstor.org/stable/4307078.

Wiskeman, E. (1959) A Great Swiss Newspaper: The Story of the Neue Zürcher Zeitung. Oxford University Press, Oxford.

Wüthrich, A. (2010) The Genesis of Feynman Diagrams. Springer, Dordrecht.

Yoose, B. J. (2015) Technology Anxiety in Academic Professional and Paraprofessional Library Staff. In Gullickson Spencer L., VandeCreek, L., and Wright, H. S. (eds), The Psychology Of Librarianship. Library Juice Press, Sacramento, CA, pp. 157–193.

Zhang, J. (2008) Visualization for Information Retrieval. Springer, Berlin.

Cite as

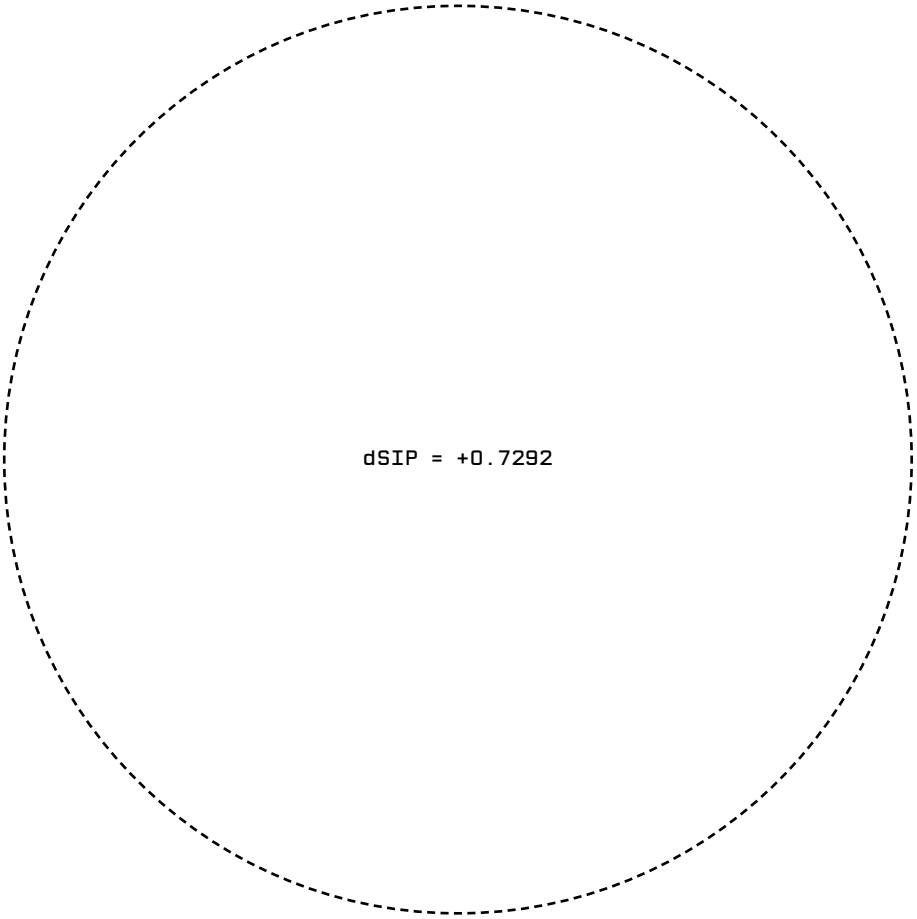
Atanasiu, V. (2022) Improving the Quality Control and Exploration of Digital Documents Using Visualization and Serendipity. Journal of Information Architecture. Vol. 07. Iss. 02. Pp. 7–82. http://journalofia.org/volume7/issue2/02-atanasiu/. doi: 10.55135/1015060901/222.011/2.046.

Vlad Atanasiu

Department of Informatics, University of Fribourg

Vlad Atanasiu, PhD, has a background in the humanities, sciences, and arts. He works in the computer science areas of document analysis, image processing, information visualization, and geographical information systems (PhD candidate, University of Fribourg, Switzerland, 2012–; researcher, University of Berne, 2021–, University of Basel, 2019–2021, Télécom–ParisTech and CNRS, Paris, 2010–11, and the Austrian Academy of Sciences, Vienna, 2006–9). During his postdoctorate, he took classes in cognitive sciences (MIT, 2003–5). He holds a PhD in Arabic Paleography, Art History and History (ÉPHÉ, Paris, 2003), an MA in Lexicology, and a BA in Middle Eastern Studies (Aix-en-Provence University, 1995, 1994). Vlad also does freelance typographic work, graphic design, typeface creation, and calligraphy.

Publication the Eleventh


dSIP = +0.7292

Information Design Journal 27(1), **p. (1–47)**
© John Benjamins Publishing Company
doi.org/10.1075/idj.20026.ata
Publication date: October 03, 2022

Vlad Atanasiu

Applications of the Document Towers method of representing document structures

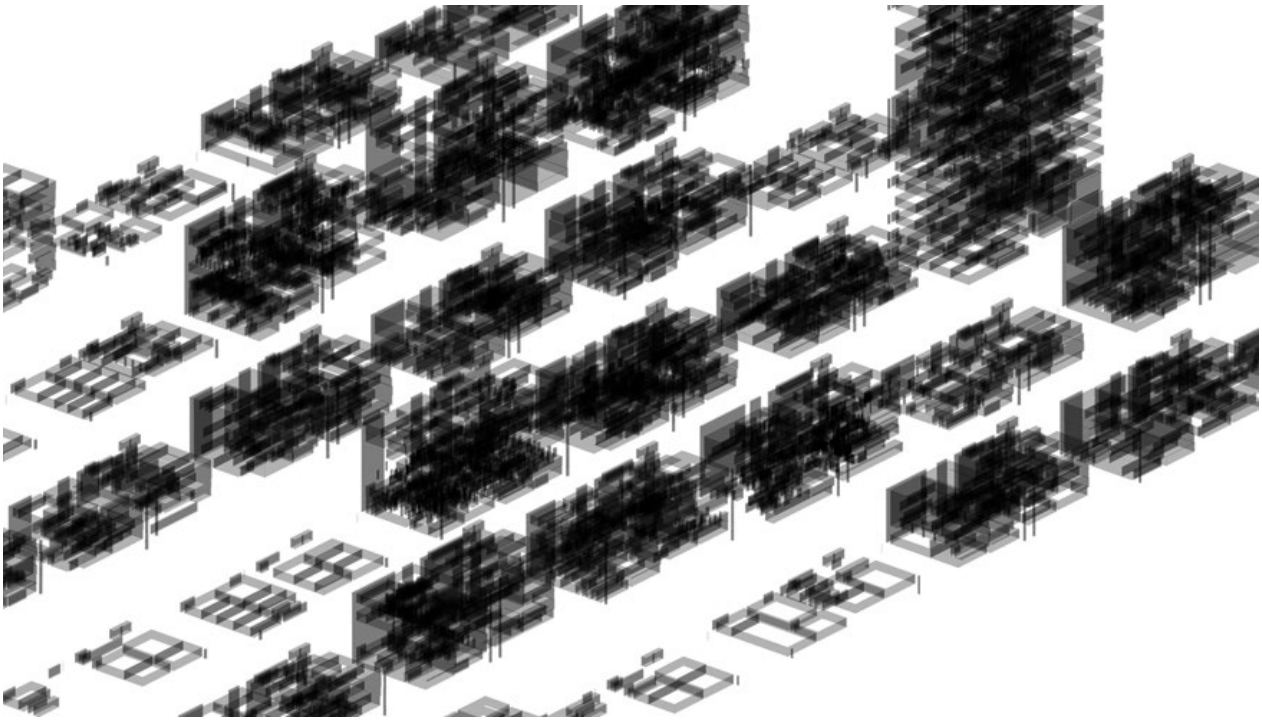


Figure 1. One possible interpretation of the above picture is that of an architectural plan of a cityscape. We see a Cartesian street grid, buildings of various sizes and complexities, and what might be the lower part of a twin-bodied skyscraper. There appear to be rooms and pillars. We might think of dollhouses, of blinking lights and zipping cars, of the electrifying bustle of rush-hour crowds of a metropolis. And, in a sense, we are right, because this visualization paradigm, called Document Towers, represents the bounding boxes of text paragraphs, stacked by successive double pages of selected articles of the Information Design Journal. Rather than the life of city dwellers, the visualization reveals the patterns of written communication and the spatialization of ideas in a scientific journal. Some articles have a more homogeneous, settled organization, while others are more dynamic and fragmented. This type of document structure representation not only fosters insights and prompts questions, but also favors a rich diversity of applications, a central topic of this article.

Keywords: document structure, document representation, document analysis, information systems, visual metaphors

This article argues in favor of representing the spatial distribution of information within and between documents, by surveying a broad variety of potential applications, including the entire document lifecycle, multiple sensory modalities, and a large spectrum of tasks and users. The theoretical explanations of this richness are a further facet of the article, and can be summarized as follows: (1) insights emerge from focusing on information structure, rather than information meaning; (2) spatializing information creates new information; (3) simplification increases the polyvalence of representation models; (4) introducing mystery in communication channels motivates discovery and diversifies insights; (5) approaching information design as a *Gesamtkunstwerk* multiplies the applications; (6) information is a manifestation of a link between structures and the actions these enable, while information design is the art and science of creating such links. The argument is developed around the concrete example of a document structure visualization, the Document Towers, which uses the metaphor of architectural models to represent documents.

> "*Our language can be regarded as an ancient city: a maze of little streets and squares, of old and new houses, of houses with extensions from various periods, and all this surround- ed by a multitude of new suburbs with straight and regular streets and uniform houses.*"
> – Ludwig Wittgenstein, *Philosophical Investigations*, §19

1. Introduction

The reader who begins to think about document structures will soon discover that these lead to a surprisingly large number of applications. This article discusses the utility of representing document structures from three distinct viewpoints relevant to information design: (i) to enable practical acquaintance with a specific document visualization technique, the Document Towers, which also provides a case study around which to organize the material contained herein; (ii) to present an illustrated survey of potential applications of the broad information class represented by document structures, (iii) to rationalize the applications' richness, and champion the inclusion of key factors in what could be termed "the information designer's practical toolbox". Visualization, applications, and theory form a system of interlocking themes that complement and explain each other.

The article's core thesis is that information structure is in itself informative and complements information semantics. From here derives the relevance of document structure representation for information design: after all, what is the essence of information design if not the structure of information, given that the discipline does not concern itself with the production of content. Moreover, information *is* structure. To convince oneself, it will suffice to shuffle the letters in this paragraph, which will change its meaning.

This work is motivated by the apparent lack of a survey assessing the utility of document structure representation. The utility of the Document Towers is also a recurrent discussion topic between their creator and potential users; thus, the matter has been deemed sufficiently consequential to warrant a dedicated study to answer a manifest need.

The findings of our study show that document structures are relevant throughout the lifecycle of documents and related tasks and users. For example, the extraction of physical and logical structures is essential in the fields of document engineering and analysis for digitization and conversion purposes, and accordingly underpins the societal changes driven by digital technologies. The

quality of human interactions with digital document information systems also depends on the user's ability to request, extract, represent, and navigate document structures, from everyday e-reader usage to the analysis of multimedia resources in virtual reality, that electronic avatar of Victorian-era panoramas.

This article focuses on text-based analog and digital documents and libraries, such as books, magazines, forms, posters, and product labels, characterized by a mix of alphanumeric text, lists, tables, pictures, and graphics. Document structure is defined as the spatial distribution of information within a document. The information can be of any kind – commonly semantic (e.g., topics, named entities, connotations) or logical (e.g., titles, paragraphs, notes, tables, images), while the structure is defined by physical attributes (mainly the coordinates of entities within the page and the document, but also colors, typeface names, or descriptors such as perceptual crowding). All sensory modalities of information representation are considered, from visual to audio to tactile. Reviews of document representations can be found in Fekete & Lecolinet (2006) and Keim et al. (2010).

This section has set forth the article's objectives, motivations, relevance, and subject matter definitions. The next section introduces the reader to the Document Towers, as a case study in document structure representation. The following section presents theories generalizing key aspects of the Document Towers, whose operational role is to facilitate better understanding of the proposed applications and the generation of new ones. The subsequent section is the article's *res materia*, in which potential applications of document structure representation are surveyed. An assessment of the work's place in the relevant literature concludes the article.

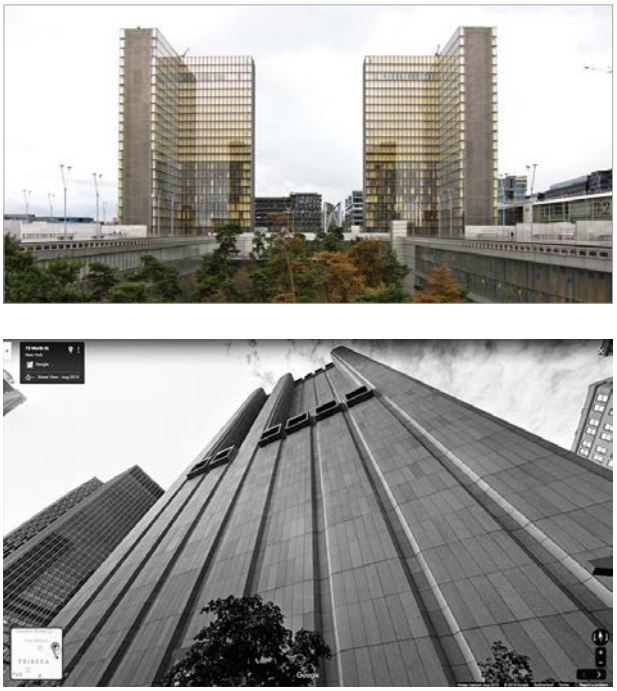


Figure 2. These pictures represent examples of the architecture-as-documents metaphor implemented in reverse in the Document Towers paradigm. The National Library of France in Paris (top picture) stores documents in four glass towers shaped as open books, intended to convey the idea of open culture. The windowless, monolithic AT&T telephone exchange building at 33 Thomas Street in Manhattan, New York (bottom picture), is an architectural exponent of opaque information. Fittingly, it is presumed to house an information surveillance facility of the National Security Agency (Moltke & Poitras 2016). (Credits: V. Atanasiu, Google Street View)

2. Case study

The Document Towers will serve as a case study to anchor the discussion with a concrete example of document structure representation. The Document Towers paradigm, visualization, and open-source software were first introduced in Atanasiu & Ingold (2021), and their usefulness for the quality control of digitization workflows has been evaluated on a real-world historical newspaper dataset at the Swiss National Library (Atanasiu 2022b). This publication in the Document Towers series intends to demonstrate (i) that the number of potential applications is much greater, (ii) that some of its key characteristics are common to other types of document structure representations, and (iii) that it facilitates the investigation of the role of such factors as spatialization, mystery, and *Gesamtkunstwerk* in information design.

We first provide an overview of the techno-cultural background of the Document Towers. We then explain how to create and read the visualization. We use a case study to demonstrate insights obtained from Document Towers. We conclude by situating the paradigm's salient features within the broader perspective of document representation.

The Document Towers visualization implements the architectural metaphor of DOCUMENTS-AS-BUILDINGS, an analogy with a lengthy history. The romantic art critic John Ruskin famously compared Gothic cathedrals to books set in stone, alluding to their shared narrative dimension (Schoenherr 2004). The definition of architecture as "a machine to inform with" provided by the architect Le Corbusier, who was also an accomplished book designer (De Smet 2007), has been extended to documents in a variety of instances, such as the letter-shaped Book Pavilion of the futurist Fortunato Depero at Monza in 1927 (Tavares 2016: 150–153), and Dominique Perrault's French National Library in Paris in 1996, whose high-rise towers resemble open books (Vidler 1993). In the domain of graphic arts, typotecture is a style in which written characters are drawn to mimic individual buildings or whole cities (Heller & Ilić 2013; Jenser 2002); a style popularized by movies such as *Tron* (1982) and *The Matrix* (1999), in which space is made of streams of information. The zeitgeist of the second half of the 20th century abounded in information-spatializing metaphors, both in common and technical language: cyberspace, information superhighway, electronic frontier, goto and return commands, front- and back-end development, file up- and download, network ports and tunnels, Internet websites, firewalls and clouds, the professional field of Information Architecture, the virtual world Second Life, and many more (Wertheim 2000, Munro 1999). Information spatialization appears to be a deeply entrenched technique of human cognition, both cross-culturally and diachronically. One noteworthy example is the mnemotechnique known as the Art of Memory, spanning the period from Greco-Roman antiquity to the Enlightenment, and consisting in placing information in imaginary buildings, which provided an effective system for "squirreling" memories back (Yates 1966). Similarly, the "songlines" of Australian Aborigines on mythical animals and their deeds encode sequences of landmarks to support cross-continental navigation (Chatwin 1988); for the Apache people, places are where wisdom is found in the form of geolocated stories (Basso 1996), a concept not dissimilar to the French historian Pierre Nora's description of how "places of memory" (*lieux de mémoire*), such as statues, monuments, or battlefields, are integral to building modern national identities (Nora & Jordan 2001–2010). The convergence of the spatial, the semantic, and the social in architecture and urbanism is also a continuing source of inspiration for technologists of information systems (Chalmers 1995, 1999). It has led to mainstream human–computer interaction paradigms, via the cognitive mapping of data, software, and

operations to graphical user interfaces, such as windows, desktops, icons, and various interaction patterns, as well as many experiments in three-dimensional representation (Dodge & Kitchin 2001). Representations of text-based documents as three-dimensional architectural models have also been developed in a series of projects at the MIT MediaLab from the 1970s to the 2000s (Cooper & VLW 1994; Sparacino, Davenport, & Pentland 2000; Dragulescu 2009), and more recently, for visualizing the nested structure of HTML entities in the Firefox web-browser (Mozilla 2021) and Java software packages (Alam et al. 2009) as pseudo-urban landscapes.

The Document Towers paradigm is part of the cultural and technological tradition outlined above. Its main contribution is the way in which it provides access to information in documents, by representing the physical structure given by the graphical layout, rather than the logical structure of the linguistic dimension. This difference in access results in a difference in the kind of information made accessible, while the architectural metaphor provides a familiar framework with which to interact with the visualization, i.e., to increase its "affordance" (Gibson 2015: 119–135; Norman 1990).

To visualize documents as Document Towers, the starting point is to obtain the coordinates of the bounding boxes of document objects, such as pages, paragraphs, and images. These are recorded explicitly in electronic document formats such as Adobe's Portable Document Format (PDF) and InDesign Markup Language (IDML) (Adobe 2008, 2006), and are recoverable from document images following conversion to formats such as the Analyzed Layout and Text Object (ALTO) (Library of Congress 2021). For the technical details of the extraction procedure, the reader is referred to Ingold, Bloechle & Rigamonti (2014) and Tang (2012).

Once object coordinates have been extracted, they are employed to create a three-dimensional graphical object; for example, by using the Document Towers open-source MATLAB software (Atanasiu & Ingold 2021). Figure 4 illustrates the process of simplification, extrusion, and projection leading from a page to a wireframe resembling an architectural model, where paragraphs and images correspond to rooms, double pages to floors, documents to buildings, and libraries to cities. This is how the architectural metaphor of documents-as-buildings comes about and how it is read.

To demonstrate how insights can be derived from the visualizations, articles from the Information Design Journal (IDJ) have been visualized as Document Towers (Figure 3). The choice is part autopoetic playful gift from the author to his hosts, part curiosity about the layout evolution of a design journal, and part need to understand its graphic design in order to write this article. To keep the volume of data manageable for the reader, three representative volumes spanning the journal's lifetime were selected.

The low-rise "buildings" are the articles published in 2017, except for the front corner building, which is the inaugural article of the first issue of 1979, while the high-rise "tower" represents the articles of 2002, stacked one on top of the other. The choice between horizontal and vertical presentation depends on the visualization compactness one wishes to achieve, and on how important it is to distinguish the different articles from the same issue from one another. Red slabs represent the boundaries of physical pages, as specified in the PDF files from which the wireframe models were generated; the double pages appear as twin towers, in their logical reading order, top-down and left-to-right. The location of the low-rise Document Towers within the Document City matrix reflects the article sequence within the respective issue. The high-rise Document Tower was placed to avoid obscuring other buildings, a phenomenon typical of three-dimensional representations. To avoid obscuring other buildings, the user can rotate, pan around, and zoom into the visualization.

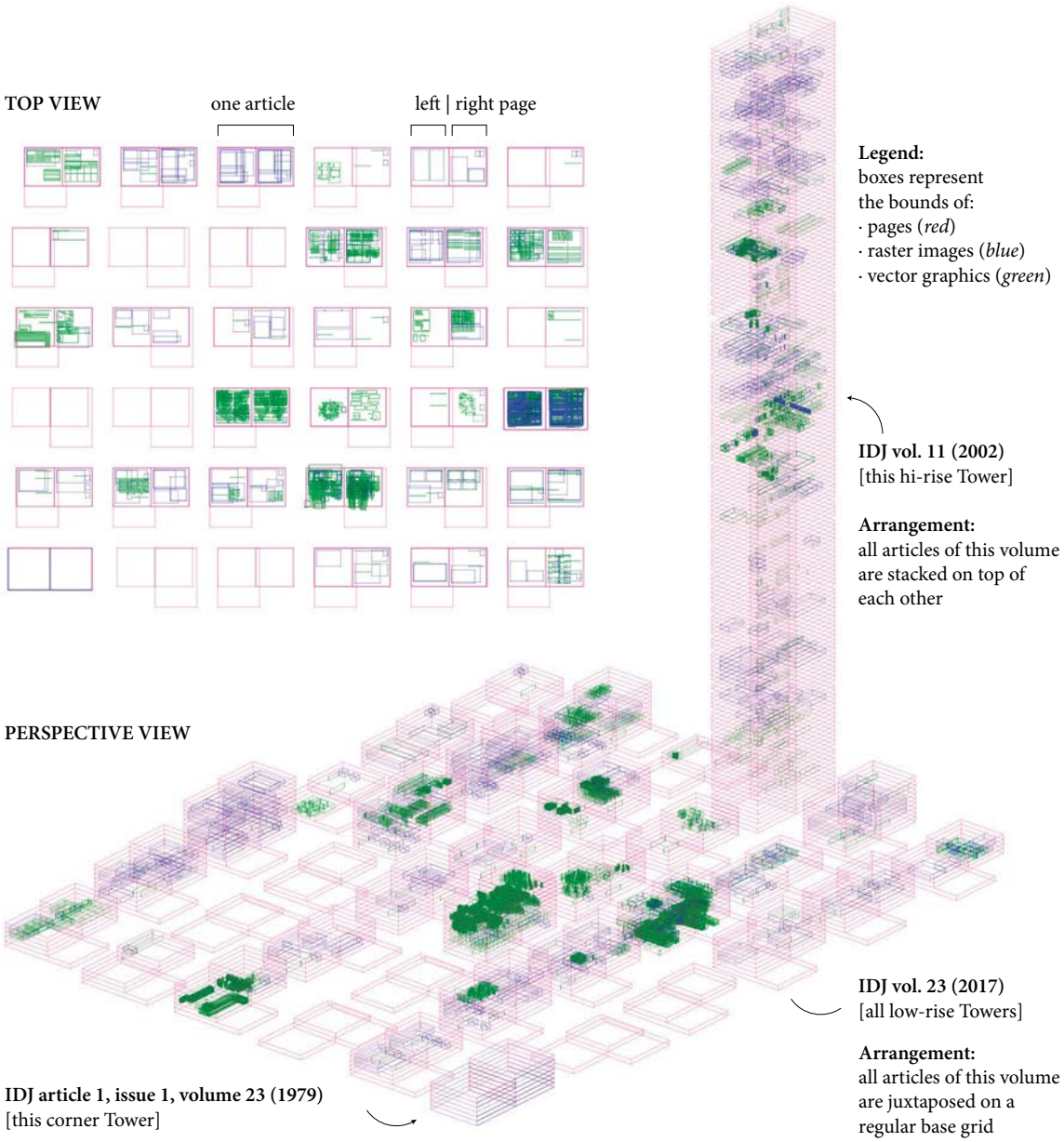


Figure 3. Document structure of selected IDJ articles

The oversize slabs at the bottom of some buildings represent a page containing terms of use appended to the documents at download time by the digital library system. Blue slabs stand for the boundaries of objects identified in the PDF as raster images, such as pictures taken with a camera, while the green slabs are vector graphics, such as those found in diagrams or used as rulers in tables. Figure 1 depicts text paragraph boundaries as gray slabs; these are obtained from the documents in the central area of Figure 3.

Turning now to the insights made possible by the visualization, let us first appraise the visual pattern in Figure 3. This may be best characterized as "uniform heterogeneity", in that it strikes a balance between types of articles. Text-only articles (of predominantly red color) are largely theory articles, reviews, and editorials; image-based articles (mainly blue) are typically practice-oriented; articles using diagrams and tables (with green components) are indicative of qualitative evaluations. In other words, the Document Towers allow for a diagrammatic reading and "visual proof" of the journal's editorial philosophy as a place that offers a diversity of perspectives to readers with diverse backgrounds and interests (Westendorp & van der Waarde 2001). When comparing the visual pattern of articles from 2002 with those from 2017 it seems that the balance of content has remained stable over time, with a slight increase in the number of vector objects; this is possibly due to a proliferation of quantitative evaluations, presented in tabular form. Here the reading made by the author shifts from elementary to speculative, and the visualization from comprehensible signifier to hermetic symbol. This increase in uncertainty—or information entropy—shows that some representations are endowed with the ability to prompt unexpected questions, and this should be welcomed.

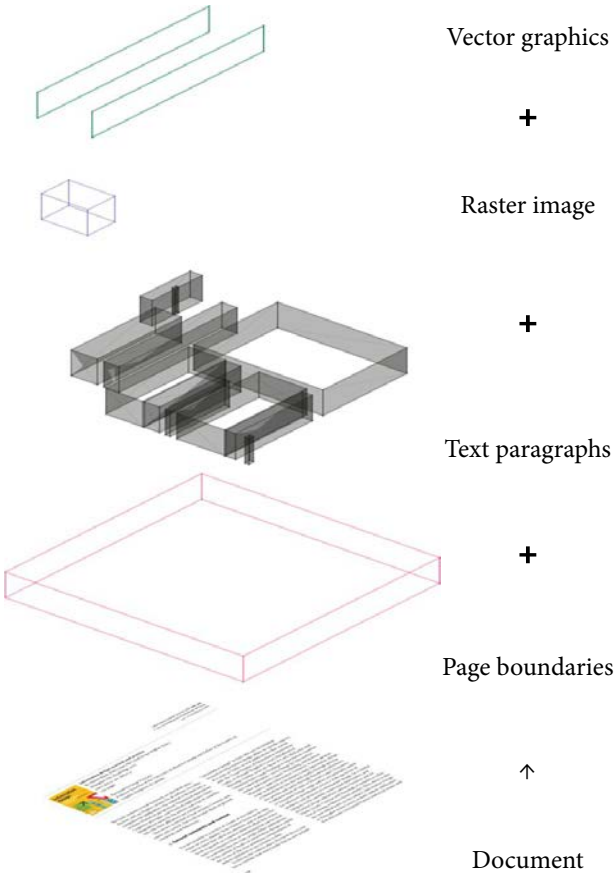


Figure 4. Representation procedure—This figure illustrates how a PDF document page is transformed into elements of a Document Towers visualization. The representation is generic, and any information that can be attributed to a location in space is representable. Some document formats allow to relate object coordinates to additional information (here, for example, the type of object).

Several other features of the visualization catch the eye. For example, the 1979 article is remarkable in that it appears to be composed entirely of raster images covering the whole page surface. In fact, this article is scanned, so that every page is indeed a picture. One application of this observation pertains to the management of digital libraries in terms of costs, resources, and quality. Understanding the percentage of scanned documents in a dataset facilitates the planning of the acquisition and allocation of the resources necessary to perform optical character recognition (necessary for scans, but unnecessary for digitally born documents). It also makes it easier to estimate the quality of answers to textual searches made by humans or machines (different between the two document types). The reason for selecting the 2002 issue for visualization should now be clear. Since articles until that date were scanned, they all exhibit a uniform pattern, therefore displaying this pattern is redundant. However, it is via visualization that the author obtained this valuable information.

The simplicity of producing the Document Towers, the straightforwardness of the metaphor, and its visual attractiveness are also important factors (Turner 2017: ix, 3–8). Figure 3 also reveals a striking fact about document structures, namely, that logical and physical structures are consubstantial, not distinct, that is, they are enforced by conceptual models of document markup languages that separate content from presentation. Because documents designed to communicate with humans have physical materiality, the information they convey must be spatialized to some degree. It is from information spatialization that many applications derive, such as the exploration and discovery via visual search. This contrasts with spoken and written language search interfaces, which demand that users know what to look for in order to formulate queries.

3. Theory

Why does a representation such as the Document Towers "work"? Why is it insightful, and why does it yield so many applications? The reasons advanced here generalize beyond a specific visualization, and thus should contribute more broadly to the theory of information design (Pettersson 2016).

3.1 Agnosticism

The Document Towers paradigm is marginally contingent on document medium (codices, scrolls, hypermedia, and other spatial embeddings of information can be equally represented), on the document encoding (object coordinates can be extracted from many formats), on the type of extracted information (provided that coordinates can be attributed to it), and on the representation modality (Document Towers can be rendered graphically, printed in three dimensions or in relief, and/or sonified). Consequently, the strong agnosticism of the representation makes it highly suitable for a broad range of applications.

3.2 Essentialism

The Document Towers visualization reduces documents to the essential. The wireframe architectural model emphasizes structure by removing the opaque material of paper, leaving behind an object as transparent as a "crystal goblet", to use a famous metaphor defining good typography (Warde 1956). The deliberate choice of simplicity (as opposed to a realistic rendering in the manner of SimCity) intends to diminish visual clutter, avoid complex software requirements, and reduce the cognitive load of interacting with virtual three-dimensional objects. The abstraction process, nevertheless, stops short

of transforming documents into numerical artifacts, however compact and easily manipulable they may be. Rather, the Document Towers provide unmediated and intuitive access to the object's geometry, one benefit of structure-preserving models of information.

3.3 Spatialization

Information spatialization cannot only preserve and reveal information, but also create information (sometimes useful, sometimes spurious). Consider the case of three text paragraphs and the impact of their configuration on their meaning. Arranged vertically and equally spaced, they will be perceived as a single visual and semantic unit. Sliding the middle paragraph laterally will mark it (again in both visual and semantic terms) as an outlier, and perhaps as a marginal gloss. Moving the third paragraph diagonally away from the rest, the hierarchy between them is reinforced; another possible connotation emerges, related to the modernity and sophistication of such a design, in that it is asymmetrical, in contrast to the symmetry observed by classical typography (Tschichold 1995). The principle at work here is that inhomogeneities developing from spatio-temporal relationships have both energetic—i.e., material tension—and informational potential (Atanasiu 2022). The reduction of matter and information to energy is what allowed the computer science pioneer Carl Zuse to formulate the idea of the universe as a computing machine (Zuse 1970; Stonier 1990). The phenomenon is fundamental to cognition, as found by Gestalt psychology in the empirical work on grouping and other perceptual "laws" (Metzger 2006), and to social relations, as epitomized in the work of the sociologist Niklaus Luhmann on "distinction" (Luhmann 2006). Furthermore, the informational properties of spatialization explain the fascination for visualization of diverse thinkers such as the philosopher Charles Pierce, for whom "all mathematical reasoning is diagrammatic" (Peirce 1902); the statistician Edward Tukey, father of the Exploratory Data Analysis (Tukey 1977); the mathematician Benoît Mandelbrot, who declared that "blind analytic manipulation is never enough" and whose fractals have become psychedelic household items (Mandelbrot 2002: 48); and the physicist Richard Feyman, inventor of the eponymous notation for particle physics. Approaching "problem-solving by means of imagistic representations" (Gauker 2011) has often been instrumental in spawning new research fields.

3.4 Mystification

Building the Document Towers paradigm around a metaphor is a well-known design method to facilitate the use of representations (Lakoff & Johnson 1980; Zhang 2008: 215–237). However, combining the metaphor with the sibylline simplicity of the wireframe architectural model engenders mystery. The meaning of a picture such as that in Figure 1 is not instantly comprehensible, but the similarities with a cityscape are sufficient to prompt attempts at explanation. Effective information design does not need to be crystal clear communication (pace Warde); mystery has a role to play, as is well-known in marketing. Mystery first teases and attracts, then motivates the viewer to "crack the nut"; and may finally lead to a great variety of insights, due to the combinatorial variability of subjective and contextual factors (Zhang 2008: 191–213). Information design by deliberate mystification is an appropriate technique to elicit discoveries when the information is unknown or insufficient to the information designer. The Baroque era was a great time for the design of mysteries. Chiaroscuro. Caravaggio. Athanasius Kircher. **Ex tenebris lux!**

3.5 Holism

The art and science of communication have been rightly defined as a *Gesamtkunstwerk* (Flaskamp & Schmidbauer 2003), a term equally applicable to the field of information design, given the role communication plays in it. To take the Document Towers paradigm as an example, we see that it harnesses an extended cross-cultural history and commands an interdisciplinary perspective, that it channels information through visual and haptic modalities, and that it addresses multifarious users and tasks; it is data- and user-centered, situated, and embodied. In brief, the adoption of a holistic, systemic attitude is a significant factor in the emergence of the diversity of applications.

3.6 Enaction

People see little meaning in document structures if they cannot imagine applications; information emerges from structure only if it leads to some form of action, be it physical, symbolic, or delayed. This is known in psychology as "enaction" and it is closely related to embodiment: the idea that motor activity is integral to cognition, perception, and emotion (Coello & Fischer 2016; Stewart, Gapenne, & Di Paolo 2010). It is a notion reverberating through many domains. In linguistics this notion rejoins the stance of pragmatics (from Greek *pragma* "act"), according to which language is about achieving goals, not communicating thoughts (Huang 2017; Gauker 2003). It has been proposed as a theory of visual art as social agency (Gell 1998); and it is also a core tenet of Buddhism, in which non-action (*nirvana*) releases from existential meaninglessness perpetuated by a chain of actions (*karma*) (Collins 1998). Translated to this article's context, enaction creates the conceptual and operational link between document structures and applications, from which information on documents arises. From this perspective, information and information design may be defined as the linking of some structure with some action.

P11
–
262

Figure 5. Document sonification—This is a scene from Wim Wenders' movie *Wings of Desire* (1987) where an angel listens to the flow of the readers' thoughts in the State Library of Berlin. Similarly, Mark Twain said: "In a good bookroom you feel in some mysterious way that you are absorbing the wisdom contained in all the books through your skin, without even opening them." Document structures have both a simplicity and a complexity that make them eminently suitable for sonification, to realize Wenders' and Twain's visions. But the picture also suggests another idea for information designers: Not every piece of information has to be made explicit, and allusions are sometimes more appropriate, as in this metonymical use of an image for a soundtrack that the reader is required to imagine. (Credits: WDR)

4. Applications

It is a truism that document content is more than words and pictures (Ehrmann, Bunout, & Düring 2019): it is narrativity (Pflaeging 2017), materiality (McKenzie 1999), form (Hiippala 2016), affect (Turner 2017), context (Agarwal 2018), and more. Similarly, documents and interfaces are dependent on know-how, resources, management and politics

for their development and use (Thylstrup 2019). In brief, the components of information systems form a multidimensional ecosystem (Dorte 2012). To reflect on this characteristic and its impact on the usefulness of document structure representation, the following sections are organized along the following basic design dimensions: representation modalities, document media, extractable information types, application tasks, and relevant user categories.

Applications are generally illustrated by a use case scenario. Vignettes provide visual demonstrations of relevant topics, selected from the work on Document Towers. This content should be understood as made of pointers to places beyond the horizon, intended to stimulate the imagination of readers. It may be best used by searching for specific items, or by skimming, as one would when using a catalog.

4.1 Modalities

Document structures can be represented in different modalities (i.e., human sensory channels). The advantages and disadvantages of these modalities are well known and widely discussed in textbooks on information design (e.g., Ware 2008). Some of these modalities are discussed here to illustrate their impact on document representation.

Visual
Of all sensory channels, visualization is the one that offers the greatest communication *bandwidth* and *resolution*. Static visualizations facilitate data *comparisons*, as several visualizations can be presented concomitantly in the visual field, in contrast to auditive information, which is deployed over time. However, three-dimensional visualizations, such as the Document Towers, can easily become inscrutable due to *occlusion* and *crowding*. Visualizations also *monopolize attention*.

Audio
Data sonification allows for *simultaneous* attention to information streams delivered through more than one sensory channel. For example, reading two books concomitantly is challenging, whereas reading and listening to music is not. Much like odors, auditive representations can elicit *emotional responses* more easily than visual representations. Audio has, however, *narrower bandwidth* and *spatial resolution*, and is also more *difficult to design* due to psychological, cultural, and subjective variability.

Somatosensory
Haptic feedback (the sensation of weight or proprioception) and *tactile* feedback (texture sensation) are part of document interaction, and as such are natural representation modality choices (e.g., indentations in address books facilitate faster access to alphabetic name groups). Ongoing research looking at tactile sensation stimulated by ultrasound interfaces suggests that future electronic documents could be explored via touch. This would be especially useful for *visually impaired people* and would also relieve the tedium of conducting the basic tasks of browsing and navigating.

4.2 Media

This section discusses the advantages of both media-independent and media-specific representations.

Analog vs digital
In digital publishing, there is no need to physically print a new book if you wish to change the typeface, nor there is any need to wait for a physical copy to arrive in the mail to start reading the book. Although convenient, this can make one forget the advantages of hardcopy visualizations of document structures. Paper-based representations do not need electricity

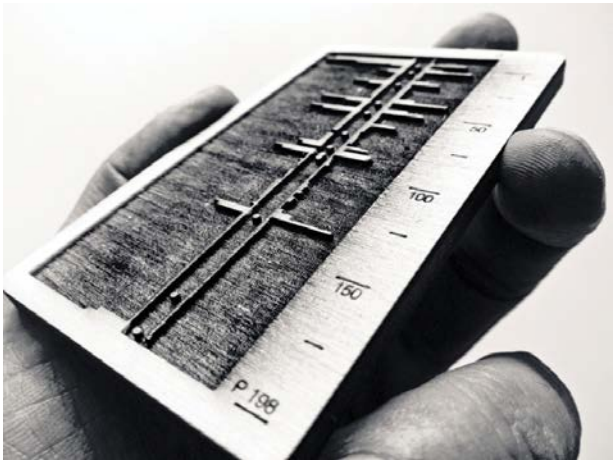


Figure 6. Tactile document structures—This laser-engraved wooden tablet encodes in a histogram-like manner the degree of layout regularity of a books' double-pages. The tactile document interface is a compact alternative structural representation of the Document Towers, especially suitable for document navigation by those who are visually impaired. For such an application, the "historgram" could be embossed on book covers. (Credits: V. Atanasiu)

P11
–
264

to be consulted, can be pinned on a wall as constant reminders, and do not disappear when the computer is shut down. Moreover, paper-based representations do not require regular software updates to remain operational and can be spread across as many tables as there are available to extend the display surface. Paper-based representations do not necessitate huddling around a screen or darkening the room to make a projected page legible, and they allow anybody—not just the person monopolizing the mouse—to point out a passage of interest.

Complexity

Structural complexity and its representations differ depending on whether documents are printed, handwritten, or hyperlinked; and increases based on whether the scale is a character, a page, a document, or an entire collection. Page thumbnails might be sufficient for overview and are certainly a common representation. However, heavily annotated documents are better represented with the transcribed text presented separately on selectable layers, as it is the case with documents found in physical archives. Documents found in online archives are in practice rarely represented in more sophisticated ways than categorized title lists, despite much creative research on the visualization of cyberspace.

Format

The digital format of documents has an impact on both the content being represented and the quality of the representation. For example, the extraction of text from document images is not as simple as the extraction from XML files, reading order in PDFs is often not preserved, and the exact location of paragraphs in InDesign documents cannot be inferred from source files alone (since it depends on the proprietary layout software).

4.3 Information

The information extracted from documents takes the form of *coordinates* and *labels* of *entities*. Entities are any physical, logical or semantic information; or other types of information such as page boundaries, headings, and concepts. Labels are the names by which entities are designated, such as "Heading Level 1". The coordinates define the physical extent of the entities within the pages and the document stack for codices. The physical, logical, and semantic structures are affected by a number of factors, such as psychological, social and cultural factors.

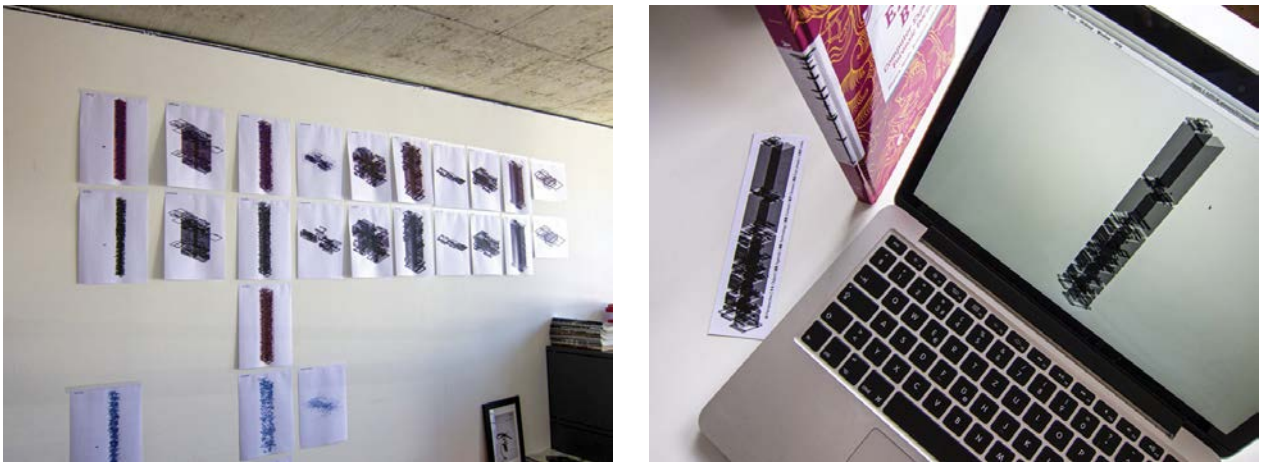


Figure 7. Figure 7. Embodiments—The materiality of document structure representation determines its utility, the information extracted, and the user experience. Hard copies are more persistent visualizations than displays on computer screens, and provide more display space, but computer interfaces allow for real-time interaction with the data. Simply by manufacturing a representation in a specific way may endow it with new usages. The bottom picture presents the structure of one of the author's books, represented as a Document Tower; printed on an oblong, rigid piece of paper, it is simultaneously a handy bookmark and a table of contents. When the Document Towers visualization is reduced to a ribbon pasted on the book spine, its internal structure is made visible even if the book is closed. In this case, we can determine the number and length of chapters based on the number of wider ribbons, which indicate chapter beginnings. (Credits: V. Atanasiu)

Physical structures

Geometry
Three-dimensional *coordinates* of the extent of pages, paragraphs, words, characters, symbols, pictures, and other entities; the shape can be reduced to rectangles, a natural form for many document entities such as pages and paragraphs.

Psychophysical structures

Descriptors
Layout, shape, color, texture, and other document descriptors can be classified, and their spatial distribution represented. Representing font sizes, for example, can reveal the location of titles and footnotes, while italics are indicative of important concepts, citations, or foreign words. These could be correlated with section topics and reveal the style of the author.

Perceptual phenomena
In order to improve readability, the following should be detected in documents: *Saliency* (the degree to which an area in the perceptual field attracts attention), *optical vs geometrical centers, kerning* (character spacing influencing the segmentation of strings in words), *rivers* (spurious curves formed by the random vertical alignment of spaces in contiguous text lines), *crowding* (which impairs

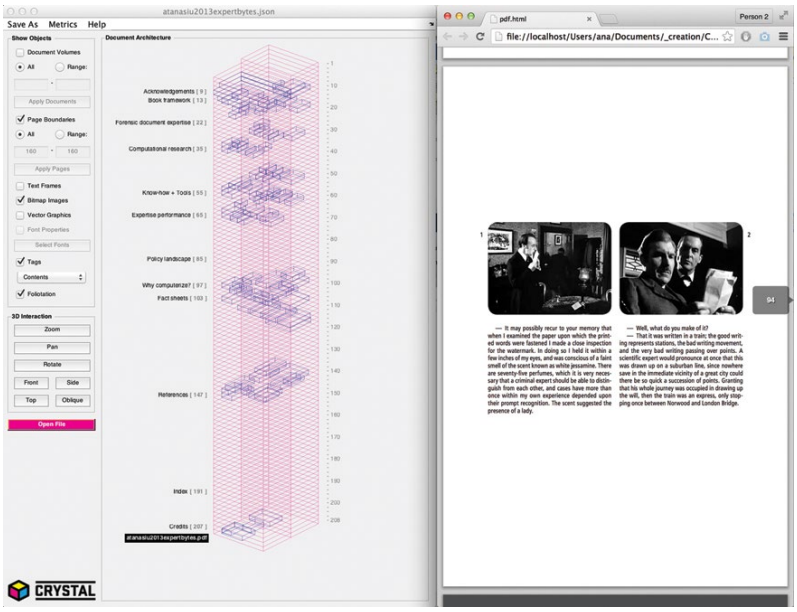


Figure 8. Structural navigation—Through the interactive interface of the Document Towers software various document objects can be selected and the visualizations manipulated. Pages are hyperlinked to the original PDF document, which is shown on the right opened by the user at a specific page in a web browser. This navigation concept would be particularly useful to implement in e-readers. While such devices often have the functionality to present the logical structure of documents (i.e., tables of contents), there is also an interest in displaying the physical structure (the layout), as it provides a different kind of information. The three-dimensional view offers the added benefit of providing an overview of the whole document.

P11 – 266

readability), *occlusion, optical illusions*, and *perceptual effects*. Representing their spatial location helps in selecting the occurrences requiring inspection, which will depend on how important an issue they appear to be and where they occur (the Adobe InDesign layout software has an option to highlight hyphenation and justification violations on each page, but does not allow an overview of their distribution over the entire document to be generated).

Logical structures

Format
An electronic document may contain variously formatted elements: *electronic text, raster images, vector graphics, fonts*, and visible and total *substrate area*. The spatial statistics of these elements are highly informative for many users, as illustrated by the Document Towers case study in Figure 1. To be able to ascertain from a visualization the presence, quantity, and location of images in a document is helpful even to the casual user.

Encoding
Determining the distribution of font code-maps (e.g., ASCII or Unicode) in a document illustrates the usefulness of document structure representation and is of interest to the developers of document-format conversion software.

Functions
An efficient manipulation of documents (e.g., browsing, navigation, bookmarking) relies on the quick and

reliable identification and memorization of functional entities, such as tables of contents, headings, paragraphs, footnotes, page numbers, references, indices, and bar codes. The representation of entity distribution is key for the successful completion of the tasks.

Styles
Functional entities are rarely marked as such for the reader (e.g., there is no label next to a title saying "This is a title"). Instead, they are made visually distinctive (a text line is usually a title if it is set in a larger size than the rest of the text and appears at the top of a page, possibly separated from the subsequent lines by whitespace). Therefore, revealing stylistic structures brings to light the functional structure of a document.

Identities
The distribution of font names, languages, scripts, and other classified objects has a broad spectrum of potential users—e.g., historians studying the diffusion of languages, forensic experts comparing typefaces used in documents under investigation, big data analysts interested in linguistic patterns, researchers engaged in the ground-truthing of documents for the purpose of evaluating novel document analysis algorithms.

Flow
Reading order is a document structure of particular importance in the automated conversion of documents from analog to digital, or between different digital formats. Reading order allows for the successful execution of simple and recurrent tasks such as copying of text from multiple-column PDF documents and reading them in the right order. This is particularly useful for the visually impaired. Visualizing text flow also helps ensure that the document is indeed readable and that it is read in the intended order, even by machines. The visual *scanning path* is distinct from reading order, and is task-, user-, and context-dependent; it represents the sequence of readers' fixation locations as they skim a document to quickly assess its content or find specific information. The physical structure of this process is of interest to psychologists studying reading, as well as designers considering ways to improve readability.

Semantic structures

Semantics
A keyword search within documents is one of the most expected interface features. The results of *a keyword* search can be visually represented, for example, by the use of colored notches next to the scroll bar. A more sophisticated document structure would make it easier to represent the distribution not only of keywords, but also of *topics*.

Named entities
People, organizations, places, and periods all have names, and their distribution within documents is just as useful to learn as that of other keywords.

Sensation structures

One important aspect of literary analysis and screenplays is the ability to represent, mentally or physically, the type and intensity distribution, within a document or movie, of *emotions, atmospheres, colors, sounds*, and plot *pace*.

4.4 Tasks

This section presents a selection of tasks significant to various application domains that could potentially benefit from the representation of document structures at different stages of the *document life cycle*.

General qualities

Polyvalence
Document structures are characterized by a broad set of possible representation modalities and extractable information, as shown above. They are also characterized by having many potential applications and users, as will be seen in this and the next sections.

Ergonomics
The polyvalence of the Document Towers is largely due to their *adequacy* to represent the underlying data. One spatial object (a document) is represented by another spatial object (an architectural model), thus propagating a fundamental property of the source (spatial relationships) in the target domain. The problem resembles that of a map vs a verbal description of a territory in terms of description length.

The ergonomic *advantages* of document representation are those of sensory representation in general, as will be made apparent below. In brief, such representations are appealing, intuitive, holistic, complementary to quantitative analysis, and technically lightweight. The *limitations* are also generic, notably the difficulty of computing with words and images, and the limitations of human perception and cognition.

One important ergonomic aspect of document representation that became apparent while evaluating the Document Towers is their *human and social qualities*. In addition to involving the use of the senses, being aesthetically appealing, and providing a playful experience, the inclusion of interactivity makes the workers feel appreciated for their skills, and, as a result, they become more involved in the project. This would not have been the case under a completely automated system.

P11
–
268

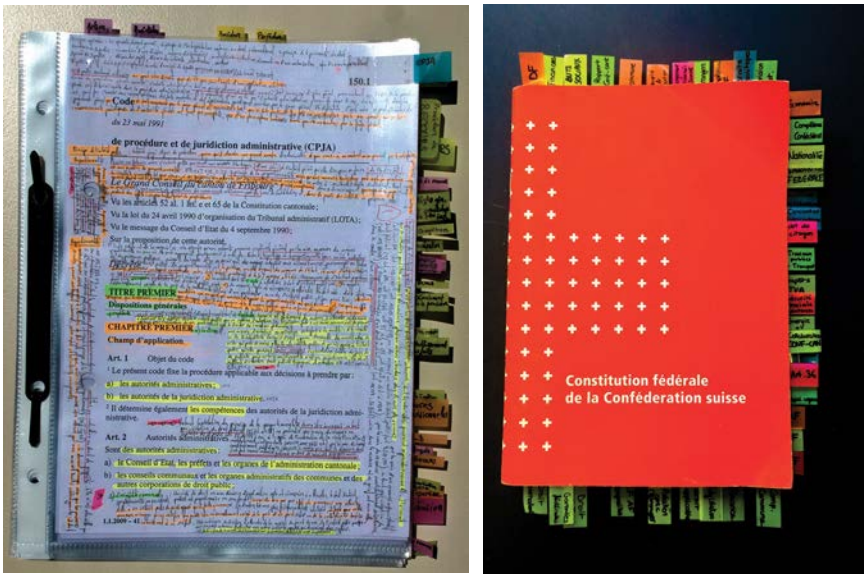


Figure 9. Information from structure—Law texts color-coded and bookmarked by students at the University of Fribourg, Switzerland. The typographical object has been strikingly morphed into a physically very different entity, in which structure is a means of accessing and personalizing information. (Photo: V. Atanasiu)

Efficiency
Another advantage of document representations such as the Document Towers is that their polyvalent and ergonomical aspects come at a low cost. If we take the example of quality control of digitization workflows, it becomes evident that only the development of multiple specialized software could fully encompass the range of issues detectable through the simple Document Towers visualization. Certain representations are therefore less costly and require less parametrization, maintenance, and training; they are characterized by *minimal investment and high yield*.

Create

Design
Representations such as the Document Towers make it possible to view a document in its entirety at a single glance, and this opens up the possibility of *top-down, parametric document design*. Rather than laying out pages individually and retaining only a mental image of its structure, or a verbal description, as is typically the case in graphic design, one could instead imagine a process in which designers work on the overall shape of the document, its visual rhythm and functional structure, leaving their implementation to algorithms.

Modeling
Document structure representation can support document *style transfer* from a source document to a target document at the pre- and postprocessing stages. This would enable understanding of the style characteristics and control of output quality.

Digitization and conversion
The extraction of the logical structure is essential to document digitization and format conversion, principally for *labeling functional entities and styles* and establishing *reading order*. The digitization and conversion of tasks such as project management and quality control, as well as of lower-level tasks, would benefit from document structure visualization. Representations can help to *categorize* data prior to digitization and conversion. Representations can also help to *select* appropriate categories according to various criteria, such as expected quality, processing speed, and costs. These are important commercial factors that are not always explicitly integrated into digitization and conversion workflows. The same visually guided categorization can lead to *more realistic document models* for training neural networks. Measurements of information distribution are additional insight sources for *improving output quality* and stating *confidence levels*; structure visualization is a method for understanding, defining, and testing such measurements.

Management
Being able to visualize document collections can help with *policy making* and *decision support*. Document processing and archiving institutions, in particular, would benefit from this. Of particular interest are the organizational aspects of digitization and conversion projects, specifically for *evaluating* the types, recurrence, and relevance of issues, and for *allocating* financial, human, temporal, and technical resources. The organizational aspects of digitization and conversion projects are also helpful for *planning, monitoring*, and *communicating* between team members, between different levels of management, and between data holders and service providers.

Quality control
Missing pages or images, misclassified documents, and metadata errors are among the anomalies that can be detected through representations without great investment in custom software.

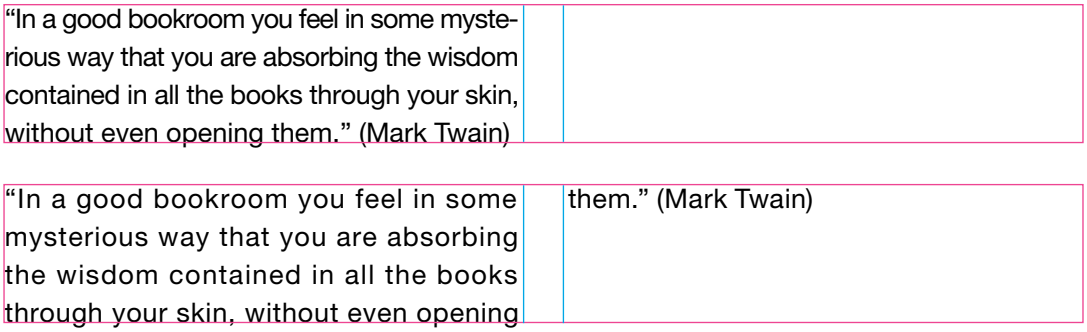


Figure 10. Information uncertainty—This figure exemplifies the impact of document format on the representation of document structures, as well as the limits of visual, non-numerical analysis. The above texts are laid out in InDesign, each placed within a single text frame (red) in two columns (blue), set in Neue Helvetica LT typeface, and kerned according to the font metrics (top) and the proprietary algorithm (bottom). The exact spatial extension of the text cannot be known from the document text file alone, where only the frame location is recorded; thus, the InDesign software is needed to resolve the uncertainty.

P11 – 270

Interact

Attract, fascinate, entice
Representing the structure of books on their spines can be seen as a marketing strategy. It helps customers decide whether to buy or not to buy the product on the shelf.

Intrigue, explore, question
Document structure representations are not always understandable at first sight. *Mysteriousness*, however, can be a desirable cognitive quality, in that it is often a strong incentive to explore the object in question, with the result being visual document analysis and a formulation of questions about the represented documents.

Collaborate, discuss, convince
A document representation is a tangible object, and as such, it is a fertile ground for human *interaction* and *debate*, in a more concrete way compared to an abstract idea about information enclosed in documents.

Images vs numbers
Perceptual representation and quantification are complementary document analysis methods that allow the integration of human and machine information processing capabilities. For example, visualization allows for more *serendipitous* discoveries and *holistic* analysis in comparison to statistical document descriptions. Moreover, the inherent *fuzziness* of visualizations may also contribute to a more *flexible* understanding of data than that permitted by “crisp” numerical analysis. However, statistics *scale up* more easily than visualizations: a number representing the mean of a distribution or a line indicating the distribution in a diagram has practically unbounded summarization power.

Explore vs search
Exploration is, by definition, undefined and open-ended, unlike *search*; the former is *holistic* while the latter is *focused*. Document structure representation supports both these information retrieval modalities, as exemplified by the Document Towers.

Sensory vs language interaction
Interacting *linguistically*, in speech and writing, with a document library typically involves performing *searches*, while *sensory* scanning (e.g., visually) promotes *exploration* (we are unlikely to spend hours chit-chatting with Alexa but can get lost surfing on Pinterest). The fundamental difference between these two tasks is worth supporting in the design of information retrieval systems. Consider, for example, a keyword search in a document collection represented as a document City: the results would light up like apartments illuminated in a *city by night*, an interaction paradigm integrating linguistic search and visual exploration.

Hands-off access
Representations of document structures make documents *transparent*; there is no need to open them if we can peer inside. Visualization can transform opaque gray archive boxes into "crystal goblets" of information.

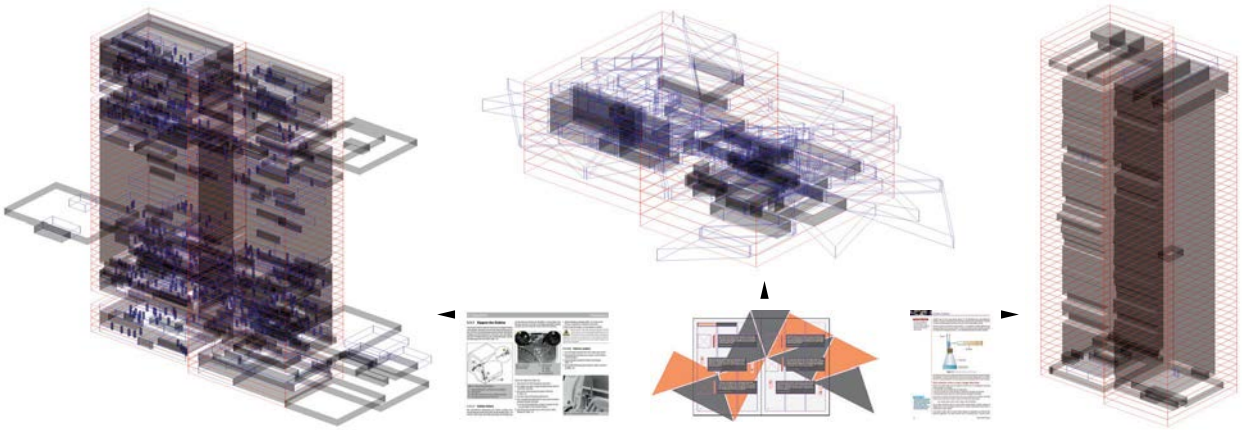


Figure 11. Classification—The 3D representations, based on InDesign files, create differences in documents when visible at a glance.—Left: A truck-driving manual is replete with pictures and diagrams, hence its heterogeneous aspect. Note the frames outside page boundaries, leavings of the book designer; while imperceptible to the reader, they could attract the attention of book historians or forensic investigators.—Center: A complex and dynamic document, representing the portfolio of a graphic artist.—Right: The simplicity of the layout suggests this to be a novel, as there is little visual variation. In reality, it is a biology school manual containing many figures, tables, and sidenotes. The designer has succeeded in hiding them within the flow of a single frame per page. If changes are needed, it is much easier to perform them with this design method than to reposition numerous small frames. The visualization is indicative of the designer's expertise, and it is a tool for managers to evaluate costs based on document complexity.

Process

Index
A search engine that preserves the spatial distribution of information is especially useful for documents, given that these are spatial objects (or, in their digital form, use a spatial mental model imported from the physical structure of analog documents and the physical world). Rather than being a list of document titles, the answer would take the form of a *document City*, with the relevant documents highlighted, while document proximity would provide additional *contextual information*.

Overview
A detail-rich overview of single documents or collections is one of the main qualities of document representations. It avoids viewing and manipulating documents through the "*keyhole*" of the spatially and mechanically limited computer display and hardware interface.

Perspectives
Because representations have a vast design space in both physical and psychological terms, they excel in creating multifarious perspectives on data. The way users would benefit from a Document Towers City that is a static picture is likely to be different from the way they would benefit from an interactive model, or a sonification.

Compare
Structures are the "fingerprints" of documents. Their representation enables visual analytics useful in *stylistic analysis* for literary studies, forensics, and pattern recognition.

Classify
Representations support *human-guided* document classification and clustering. This can be seen by how the Document Towers enabled distinguishing between articles and monographs, and between scanned and native digital documents.

Triage
As shown by the model of medical emergency triage, document triage is a classification type characterized by the *rapid speed* with which it is conducted and the *uncertainty* surrounding the nature of documents (Atanasiu 2022). Document triage for intelligence purposes is another application example; a further example is that of a journalist operating under tight deadlines facing large amounts of documents to sift through. In such cases, visualizing layout complexity helps to quickly identify potentially information-rich items.

Navigate
A document structure representation is similar to a table of contents. A document structure representation, however, is more useful since it allows for arbitrary types of information to be presented concomitantly and at a higher spatial resolution. For example, the Document Towers can use different color codes for text paragraphs, raster images, and vector graphics. Such a *document signage device* complements the classical table of contents and facilitates *navigation* for e-book users and *browsing* for the visually impaired.

Remember
Effective representations often implement familiar *mental models* for data manipulation. For example, the Document Towers use an architectural metaphor to convey the spatial distribution of information in documents. Recovering information becomes a question of recalling where it is and how to get there, a task that is often more straightforward than describing complex

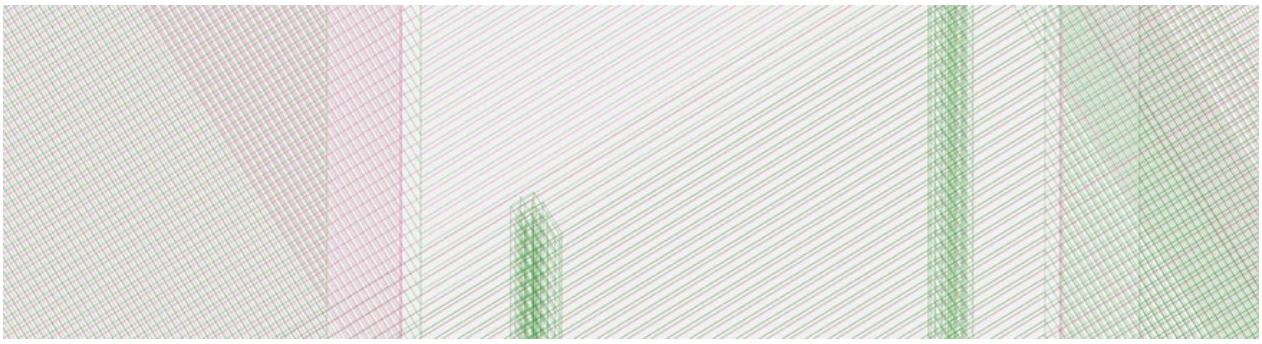


Figure 12. Ars electronica—This tartan-like texture was obtained by zooming into a Document Towers visualization. It could potentially be turned into a tapestry, which would provide an excellent backdrop for a reading room in a library; it is auto-referential scientific art. Another quality which might captivate a psychologist or an art historian: This specific view was identified by the author after some probing into Document Towers. The question therefore arises as to what specific properties made it so aesthetically appealing as to be selected from among many others. The answer should also be of interest to the information designer.

information and scrolling through a long list of results. Information spatialization can help prevent users from getting "*lost in cyberspace*".

Evolution
Document *genesis* is a complex process of confluence in a physical, logical, and semantic structure, often concluded with its publication; this is followed by a phase of *alterations*, some intentional (e.g., annotations) and some accidental (e.g., stains), and ends in *destruction* or *recycling* (e.g., palimpsests). It is not only the humanities that are interested in recovering and monitoring document life cycles, but also publication designers, digitization managers, and forensic experts. Static (images), dynamic (movies), and interactive representations of changes in digital or physical documents are human-accessible means of *keeping records* and visualizing *trends*.

Forensics
Human analysts are also indispensable to forensic practice, and representations can provide intuitive support to the work of these experts. For example, document structures can be easily compared in view of quantitative identification or verification.

E-reading
Until e-readers consisting of a hundred or so individual electronic papers are produced, the browsing of electronic documents will remain a more *tedious* task compared to the browsing of analog paper documents. This is especially the case for professionals consulting large numbers of documents and for the visually impaired. These issues can be mitigated through document panoramas, which facilitate browsing, purposeful navigation, information localization, and memorization.

Access

Access
Document structure representations can act as *substitutes* for the original documents when these are inaccessible—e.g., remotely stored physical documents, fragile documents, documents in exhibition showcases, or confidential documents.

Preservation
Ancient manuscripts and historical publications (such as those printed on brittle 19th century industrial paper) are fragile, valuable, and irreplaceable artifacts. Therefore, librarians want to minimize their physical manipulation. Document structure reproductions help *preserve* cultural heritage while still providing *access* to contents.

Reconstruction
For damaged documents (e.g., by shredding, water, fire, bugs, or fungi) which exist only as fragments, structural representations can function as *virtual models*, similar to reconstructed archaeological vessels or buildings, which are built from pieces of the original and patched with plaster.

Privacy
Representing the document structure of sensitive archives and confidential or private documents requires a compromise between *restricted access* and *anonymization*, and some degree of *disclosure* and manipulability.

Security
Digital texts which use the same color as the background are virtually impossible to read and can pose security problems for unsuspecting readers. This trivial obfuscation technique illustrates how document structure can be an additional investigative tool for *law enforcement* and *intelligence*.

Share

Training
The quality control of digitization workflows comprises manual and automatic steps. Document structure representations can not only be part of the process, but also serve to train workers by *demonstrating* what structures are, what types of issues exist, and how to detect them. This training could help developers of digitization quality control software better *understand* data and tasks to identify, select, and develop appropriate computational tools.

Education
Document representations such as the Document Towers can be used as *teaching material* in areas such as computer science (e.g., visually demonstrating differences in digital document formats and their impact on computational document analysis and usability), the humanities (e.g., book history and paleography), and the arts (e.g., book design and typography). They can also be used in primary education to foster children's enthusiasm for books, for the information contained within them, and for the science behind their production (e.g., presenting documents as doll houses).

Outreach
Visualization and other information representation modalities have promising outreach potential and can reach the broader public by virtue of being founded in *sensory experience*. It is easy to envisage how libraries, archives, museums, organizations, and firms in possession of document collections could use representations to communicate online, in print, or through exhibitions showcasing the value of their holdings, in addition to traditional means such as displaying the documents themselves or by statistics.

Marketing and merchandising
The sensory appeal of document representations can be leveraged for marketing purposes by publishers, bookshops, repositories, and document processing software producers. *3D-printed* Document Towers, or Document Towers *texture* wrapped on various objects (e.g., pens), facilitate the merchandising of information representation.

Art
Making art with information (Wilson 2002; Ars Electronica 2021). The Document Towers patterns can be used for *textile* production or as *murals* in library halls and reading rooms; document Cities can become the electronic versions of 19th-century *panoramas*, and the Document Towers can become the electronic versions of *doll houses; video games* could be played within real book structures, such as Umberto Eco's medieval library-centered novel *The Name of the Rose* or JG Ballard's quasi-Shakespearian social decay drama *High-Rise*, which plays out in a London tower building. Such *narrative* artwork would not be arbitrary, but rather emerge from the data structure. Wim Wenders, whose movie *Wings of Desire* depicts angels listening to the hum of readers' thoughts in the State Library of Berlin, explains that "storytelling is ultimately nothing else than producing meaning" (Stahl 2019).

4.5 Users

Fields
Almost everyone can benefit from document representation: from content creators and document producers, to access providers and developers of computational infrastructure and software, to professional document analysts and casual readers. An overview of the benefits of document representations and who would benefit from them is provided in this section, with references to selected entries in the Tasks section (→ *in italics*).

Benefits
Businesses and organizations can expect the following major benefits from document representations: increased competitiveness, cost reduction, quality improvement, optimization of planning, monitoring, quality control, and communication.

In the field of computer science, three main benefits of using document structure representations have been identified. (1) The routinely large amounts of data, high processing speeds, and unfamiliar heterogeneous data that computer scientists have to deal with objectively hamper their close investigation of said data. *Representations offer tools for data investigation*. Heavily involving the human senses, these tools open up *more perspectives on the data than mathematical models*. (2) Representations *allow humans to intervene in automated processes* at all stages: understanding data, building and parametrizing models, monitoring processes, making decisions, discussing and disseminating results, etc. (3) Representations *complement quantitative and automated approaches*.

Production
Content creators—private writers, organizations → *overview, navigate, evolution*
Publishing—books, journals, magazines, and newspaper publishers; document designers, editors, marketing → *overview, navigate, evolution*
Software—desktop publishing → *overview, navigate, evolution*

Access
Bookshops—physical and online → *attract, intrigue, explore*
Repositories—libraries, archives, museums, organizations, firms, private persons → *digitization, management, overview*

Software—digitization, migration, information retrieval, indexing, search engines → *modeling, digitization, indexing*

Use
Readers—leisure and professional reading; special needs readers: visually impaired → *overview, navigate, e-reading*
Children → *education, art*
Education—book history, visual psychology, document design, information sciences → *education, evolution, reconstruction*
Journalism → *triage*
Forensics, intelligence → *forensics, security, training*
Administrative document flows → *triage, overview, access*
Exhibition visitors → *overview, marketing, art*
Artists → *art*
Software—research in document engineering, document analysis, and information systems → *explore, triage, e-reading*

5. Conclusions

Let us now briefly assess the contribution of this work to the state of the art and consider some possible future directions.

This article has shown that there is a lack of survey data on the potential applications of document structure representations. It has also been shown here that the usefulness of document structure representations is a major concern of computer science researchers, information technology companies, and research funding agencies alike, which indicates that the marketability of ideas is an important dimension of information design. From this viewpoint, the innovation process can be seen as a chain of stages that goes beyond having ideas and prototyping them; it also includes persuading stakeholders and achieving market breakthrough. The survey of applications is a direct product of market constrains intended to assist readers in similar situations. This conceptual meta-level is a supplemental story line that this article begins unwinding: a field of research analyzing intrinsic and extrinsic factors to find out why some information design ideas prosper, while other wane.

The Document Towers paradigm used here, however, is not without its limitations—for example, three-dimensional visualizations are ill-suited to representing very large datasets. The limitations, however, are largely mitigated by the paradigm's ability to create an overview of one or more documents, and by its ability to facilitate unexpected discoveries about these documents, through a relatively simple technological implementation. These strengths compare favorably with document analysis solutions that are specialized, complex, and make use of non-visual, automated approaches, such as optical character recognition and text-based search engines.

As for the theoretical considerations related to document structure representation, some of these are well-known (e.g., the role of simplicity and metaphors in design), while others merit further investigation (e.g., casting information as potential action, using mystery in information design). These considerations exemplify how the purposeful seeking of vantage points outside the mainstream, counterintuitive or even otherworldly, is a design method in itself (Nodder 2013).

We finish by highlighting the features shared by the application survey of document structure representation, the Document Towers paradigm, and information design theories as forms of *design patterns*: standardized, modular, and generic problem-solving methods (Alexander, Ishikawa, & Silverstein 1977; Sedig & Parsons 2013).

Submission: 30 November 2020
Accepted: 30 June 2022

Acknowledgements

The author is grateful to the three anonymous reviewers for having prompted many new ideas. The author is also grateful to K. F. for never tiring of working to improve the readability of the Document Towers manuscripts.

References

Adobe Systems (2006). *IDML file format specification, version 8.0*. San Jose, CA: Adobe Systems. <https://wwwimages.adobe.com/content/dam/acom/en/devnet/indesign/sdk/cs6/idml/idml-specification.pdf>, March 27 2021.

Adobe Systems (2008). *Document management – Portable document format – Part 1: PDF 1.7 (PDF 32000-1:2008)*. San Jose, CA: Adobe Systems. <https://web.archive.org/web/20210319185027/https://www.adobe.com/content/dam/acom/en/devnet/acrobat/pdfs/pdf_reference_1-7.pdf>, March 27 2021.

Agarwal, N. K. (2018). *Exploring Context in Information Behavior: Seeker, Situation, Surroundings, and Shared Identities*. San Rafael, CA: Morgan & Claypool. https://doi.org/10.1007/978-3-031-02313-2

Alam, S., Boccuzzo, S., Wettel, R., Dugerdil, Ph., Gall, H., & Lanza, M. (2009). EvoSpaces: Multi-dimensional Navigation Spaces for Software Evolution. In D. Lalanne, J. Kohlas (eds.), *Human Machine Interaction: Research Results of the MMI Program*. Berlin: Spinger, 167–192. https://doi.org/10.1007/978-3-642-00437-7_7

Alexander, Ch., Ishikawa, S., & Silverstein, M. (1977). *A Pattern Language: Towns, Buildings, Construction*. New York, NY: Oxford University Press.

Anderson, M., Meyer, B., Olivier, P. (eds.) (2002). *Diagrammatic Representation and Reasoning*, London: Springer. https://doi.org/10.1007/978-1-4471-0109-3

Ars Electronica (2021). *Ars Electronica*. <https://ars.electronica.art/about/en/>, April 4 2021.

Atanasiu, V. (2022). The Structural Information Potential and Its Application to Document Triage. *IEEE Access*, 10: 13103–13138. https://doi.org/10.1109/ACCESS.2021.3133654

Atanasiu, V. (2022b). Improving the Quality Control and Exploration of Digital Document Collections Using Visualization and Serendipity, *Journal of Information Architecture*, 6 (2). https://doi.org/10.55135/1015060901/222.011/2.046.

Atanasiu, V., & Ingold, R. (2021). Document Towers: A MATLAB software implementing a three-dimensional architectural paradigm for the visual exploration of digital documents and libraries. *SoftwareX* 14: 100684. https://doi.org/10.1016/j.softx.2021.100684

Basso, K. H. (1996). *Wisdom Sits in Places: Landscape and Language Among the Western Apache*. Albuquerque, NM: University of New Mexico Press.

Chalmers, M. (1995). Design perspectives in visualizing complex information. In S. Spaccapietra & R. Jain (eds.), *Visual Database Systems 3, Visual information management: Proceedings of the third IFIP 2.6 working conference on visual database systems, 1995* (pp. 103–111). Dordrecht: Springer-Science+Business Media.

Chalmers, M. (1999). Informatics, Architecture and Language. In A. J. Munro, K. Hook, & D. Benyon (eds.), *Social Navigation of Information Space* (pp. 55–79). London: Springer. https://doi.org/10.1007/978-1-4471-0837-5_4

Chatwin, B. (1988). *The Songlines*. New York, NY: Penguin.

Coello, Y., & Fischer, M. H. (2016). *Perceptual and Emotional Embodiment: Foundations of Embodied Cognition*, Abingdon: Routledge.

Collins, S. (1998). *Nirvana And Other Buddhist Felicities: Utopias of the Pali imaginaire*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511520655

Cooper, M., & Visible Language Workshop (1994). *Information Landscapes*, video. Cambridge, MA: MIT Media Lab. <https://www.youtube.com/watch?v=Qn9zCrIJzLs>, March 27 2021.

De Smet, C. (2005). *Le Corbusier, Architect of Books*. Baden: Lars Müller.

Dodge, M., & Kitchin, R. (2001). *Atlas of Cyberspace*. Harlow: Addison-Wesley.

Dorte, M. (2012). Interdisciplinarity in the Information Field. *Proceedings of the 75th Annual Meeting of the American Society for Information Science and Technology (ASIS&T), 28–31 October 2012, Baltimore, MD, USA*, 49, 4 p.

Dragulescu, A. C. (2009). Data Portraits: Aesthetics and Algorithms, M.A. thesis. Massachusetts Institute of Technology, Cambridge, MA.

Ehrmann, M., Bunout, E., & Düring, M. (2019). Historical Newspaper User Interfaces: A Review. *World Library and Information Congress, 85th IFLA General Confrence and Assembly, 24–30 August 2019, Athens, Greece*, 10 p.

Fekete, J.-D., & Lecolinet, É. (eds.) (2006). *Document Numérique: "Visualisation pour les bibliothèques numériques"* 9(2).

Flaskamp, J., & Schmidbauer, K. (2003). *Kommunikation als Gesamtkunstwerk: Praxisleitfaden für die Umsetzung von integrierter Kommunikation*. Berlin: Vistas.

Gauker, Ch. (2003). *Words without Meaning*. Cambridge, MA: MIT Press.

Gauker, Ch. (2011). *Words and Images: An Essay on the Origin of Ideas*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199599462.001.0001

Gell, A. (1998). *Art and Agency: An Anthropological Theory*. Oxford: Oxford University Press.

Gibson, J. J. ([1979] 2015). *The Ecological Approach to Visual Perception*. New York, NY: Psychology Press.

Heller, S., & Ilić, M. (2013). *Lettering Large: The Art and Design of Monumental Typography*. New York, NY: The Monacelli Press.

Hiippala, T. (2016). *The Structure of Multimodal Documents: An Empirical Approach*. New York, NY: Routledge.

Huang, Y., (ed.) (2017). *The Oxford Handbook of Pragmatics*. Oxford: Oxford University Press.

Ingold, R., Bloechle, J.-L., & Rigamonti, M. (2014). Reverse-Engineering of PDF Files. In B. B. Chaudhuri & S. K. Parui (eds.), *Advances in Digital Document Processing and Retrieval* (pp. 249–285). Singapore: World Scientific. https://doi.org/10.1142/9789814368711_0011

Jenser, A. (2002). *Typotecture: Typography as Architectural Imagery*. Zürich/Baden: Museum für Gestaltung Zürich & Lars Müller.

Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (eds.) (2010). *Mastering The Information Age: Solving Problems with Visual Analytics*. Goslar: Eurographics Association.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Library of Congress (2021). *ALTO: Technical Metadata for Layout and Text Objects*. <https://www.loc.gov/standards/alto/>, March 27 2021.

Luhmann, N. (2006). System as Difference. *Organization* 13 (1): 37–57. https://doi.org/10.1177/1350508406059638

Mandelbrot, B. (2002). *Gaussian Self-Affinity and Fractals: Globality, The Earth, 1/f Noise, and R/S*. New York, NY: Springer.

McKenzie, D. F. (1999). *Bibliography and the Sociology of Texts*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511483226

Metzger, W. (2006). *Laws of Seeing*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/4148.001.0001

Moltke, H., & Poitras, L. (2016). *Project X, documentary short film*, Praxis Films. <https://www.youtube.com/watch?v=SOJugaGKoeU>, March 27 2021.

Mozilla developers (2021.03.03). Firefox developer tools: 3D view. <https://developer.mozilla.org/en-US/docs/Tools/3D_View>, March 31, 2021.

Munro, A. J., Hook, K., & Benyon, D. (eds.) (1999). *Social Navigation of Information Space*. London: Springer. https://doi.org/10.1007/978-1-4471-0837-5

Nodder, Ch. (2013). *Evil by Design: Interaction design to lead us into temptation*. Indianapolis, IN: John Wiley & Sons.

Nora, P., & Jordan, D. P. (eds.) (2001–2010). *Rethinking France: Les lieux de mémoire*. Chicago, IL: University of Chicago Press.

Norman, D. A. (1990). *The Design of Everyday Things*. New York, NY: Doubleday.

Peirce, Ch. (1902). *Logic, Regarded as Semeiotic*. <http://www.iupui.edu/~arisbe/menu/library/bycsp/L75/ver1/l75v1-02.htm>, April 2, 2021.

Petterson, R. (2016). *Information Design Theories*. Tullinge: Institute for Infology.

Pflaeging, J. (2017). Tracing the Narrativity of National Geographic Feature Articles in the Light of Evolving Media Landscapes. *Discourse, Context & Media* 20: 248–261. https://doi.org/10.1016/j.dcm.2017.07.003

Schneider, B., Ernst, Ch., & Wöpking, J. (2016). *Diagrammatik-Reader: Grundlegende Texte aus Theorie und Geschichte*. Berlin: De Gruyter. https://doi.org/10.1515/9783050093833

Schoenherr, D. E. (2004). A Note on Burne-Jones's 'Pocket Cathedral' and Ruskin. *Journal of William Morris Studies* 15 (4): 91–93.

Sedig, K., & Parsons, P. (2013). Interaction Design for Complex Cognitive Activities with Visual Representations: A Pattern-Based Approach. *AIS Transactions on Human-Computer Interaction* (5)2, 84–133. https://doi.org/10.17705/1thci.00055

Sparacino, F., Davenport, G., & Pentland, A. (2000). City of News: cataloguing the World Wide Web through Virtual Architecture. *KOS* 9: 179–180.

Stahl, A. (2019). Filmen ist ein aggressiver Akt. *Neue Zürcher Zeitung* 2019.05.15: 41a.

Stewart, J., Gapenne, O., & Di Paolo, E. A. (2010). *Enaction: Toward a New Paradigm for Cognitive Science*, Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262014601.001.0001

Stonier, T. (1990). *Information and the Internal Structure of the Universe: An Exploration into Information Physics*. London: Springer Verlag. https://doi.org/10.1007/978-1-4471-3265-3

Tang, Y. Y. (2012). *Document Analysis and Recognition with Wavelet and Fractal Theories*. Singapore: World Scientific. https://doi.org/10.1142/8434

Tavares, A. (2016). *The Anatomy of the Architectural Book*. Montréal & Zürich: Canadian Centre for Architecture & Lars Müller.

Thylstrup, N. B. (2019). *The Politics of Mass Digitization*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/11404.001.0001

Tschichold, J. ([1928] 1995). *The New Typography*. Berkeley, CA: University of California Press.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Turner, Ph. (2017). *A Psychology of User Experience: Involvement, Affect and Aesthetics*. Cham: Springer. https://doi.org/10.1007/978-3-319-70653-5

Vidler, A. (1993). Books in Space: Tradition and Transparency in the Bibliothèque de France. *Representations* 42: 115–134. https://doi.org/10.2307/2928626

Warde, B. (1956). The Crystal Goblet or Printing Should be Invisible. In H. Jacob (ed.), *The Crystal Goblet: Sixteen Essays on Typography by Beatrice Warde* (pp. 11–17). Cleveland, OH: The World Publishing Company.

Ware, C. (2008). *Visual Thinking for Design*. Burlington, MA: Morgan Kaufmann.

Wertheim, M. (2000). *The Pearly Gates of Cyberspace: A History of Space from Dante to the Internet*. New York, NY: W. W. Norton.

Westendorp, P., & van der Waarde, K. (2001). Editorial introduction. *Information Design Journal* 10(1): 1–3. https://doi.org/10.1075/idj.10.1.01wes

Wilson, S. (2002). *Information Arts: Intersections of Art, Science, and Technology*. Cambridge, MA: MIT Press.

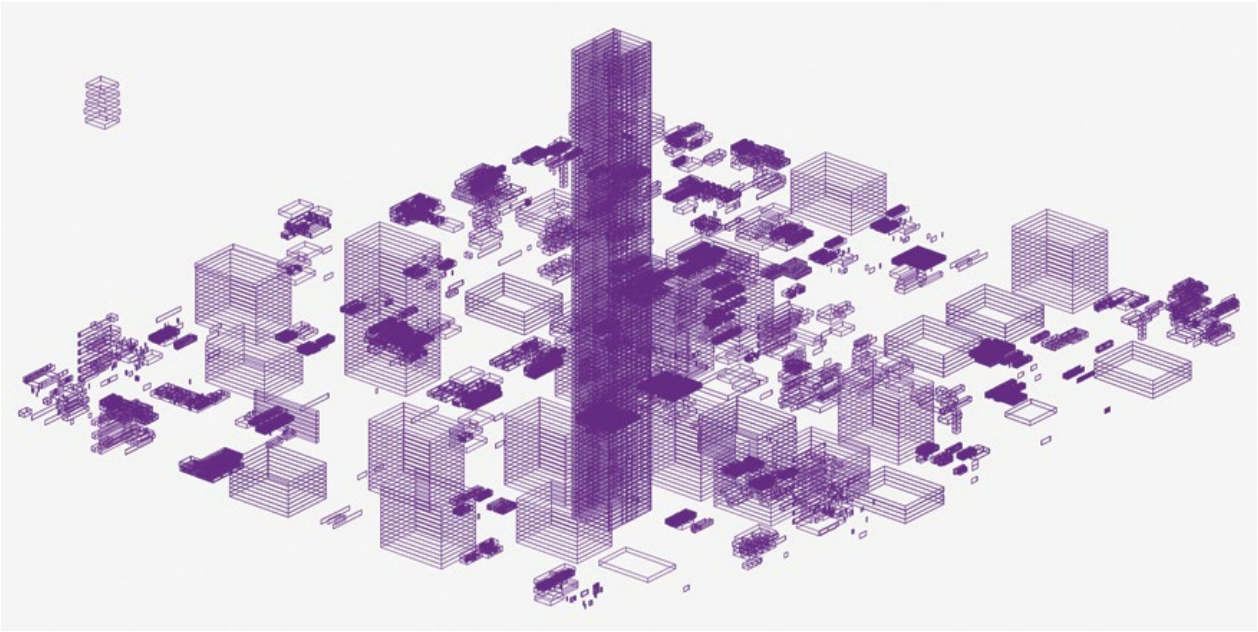
Yates, F. A. (1966). *The Art of Memory*. London: Routledge.

Zuse, K. (1970). *Calculating Space*. Cambridge, MA: Massachusetts Institute of Technology.

Zhang, J. (2008). *Visualization for Information Retrieval*. Berlin: Springer. https://doi.org/10.1007/978-3-540-75148-9

Annex

Dreams, Architecture, and Information Systems: A brief review of the utility of document exploration, visualization, and metaphors



P11
–
280

Figure A1. Document City—The Towers shown here stand for eighty-nine PDF documents on the topic of machine learning in a directory ordered by file name on the author's computer; the blue slabs represent the location and extent of raster images in the digital documents. The picture exemplifies three benefits of visualizing document structures.—Misclassification: The collection was supposed to contain only articles, but the presence of a high-rising Tower reveals a monograph among them. This misclassification would most likely have gone unnoticed if not for its serendipitous discovery due to visualization.—Quality: The regular Towers are scanned documents, where each page is a single raster image, while the fragmented Towers represent native electronic documents, in which images cover only a portion of pages if present at all. The implication for a library wishing to offer its readers searchable digital documents is that scanned documents must be identified and the text and logical structure extracted, which means time, costs, and an imperfect search quality. Visualization offers a quick and lightweight solution to anticipate document quality.—Steganography: The small Tower in the upper left corner, attracting attention by its outlying location, is not an individual document, but rather represents images outside the visible frame of the PDF documents. What secrets might they hide?

Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing.
Vannevar Bush (Bush 1945: 106)

It is a misunderstanding that you go to the library to look for a book whose title you know.
Umberto Eco (Eco 1995: 21)

Prologue

How useful would you expect the oracle to be that exclusively answers your questions and keeps silent about everything else of import—a chasm, maybe, to your right, a slumbering beast to your left, and a wall of fire straight ahead.

The currently dominant digital search engine paradigm behaves in precisely this way. It compels people to view the world through the narrow slit of an input field for keywords selected from the lexicographical stock of specific languages and users, offering little understanding of nuance, ambiguity, humor, mystery, seduction, playfulness, aesthetics, imperfection, inspiration, imagination, creativity, serendipitous perspectives, unplanned exploration, provocation and evocation, allusions and delusions, or the emotional and intellectual hedonism of surfing the Web, of being a *flâneur* on the information boulevards—in short, the oneiric functions of information systems that are diametrically opposed to the conscious goal-directed information-seeking strategies of Cyberman zipping over information highways towards his next waypoint over the horizon of his todo-checklist.

Meanwhile, a panoramic view of the data, such as that enabled by the urban metaphor utilized in the Document Towers, invites the curious eye to make discoveries, the ear to listen as though the data speak for themselves, and the tongue to utter "Eureka!" These are unconscious cognitive processes whose power resides in their complex pattern detection capabilities, the creation of optimal linkages within vast conceptual spaces, and the secretion of the motivational substance oxytocin. They are thus essential to human information processing.

Introduction

This section provides context for the article's general problem of understanding digital documents more generally and the solution provided by Document Towers specifically. It reviews arguments as to why the contemporary emphasis on search tools in computer science research and commercial software should be complemented with support for other tasks, such as exploration, and have other characteristics, such as eliciting serendipity or playfulness. It also reviews arguments in favor of complementing numerical analyses of documents with sensorial interaction modalities, primarily visualization, and demonstrates their suitability for supporting the exploration task, and having such characteristics where imagination and emotion dominate over rational approaches. Moreover, a survey of the use of metaphors, architectural and urban in particular, for understanding and interacting with digital documents and libraries highlights an important ingredient in the creation of effective visualizations and explains why such a metaphor was adopted for the Document Towers technique.

Exploration

Put simply, the difference between searching and exploration is that in the former case, you know what you are looking for, while in the latter you do not (for fine-grained definitions, see references in Bates (2007) and McKay et al. (2019); the linguistic evolution of the terms is retraced in the OED (2019: search, *v.*; explore, *v.*); on theories of information see Buckland (1991) and Marchionini (1991: 5–9); on information-seeking behavior: Fisher et al. (2006) [encyclopedical], Kuhlthau (2004)

[monographical], Spink (2010) [evolutionary], Spink & Cole (2006) [new trends], Kalbach (2006) [uncertainty measurement]). Information searching is one of the oldest problems in information science as reflected by the variety of knowledge classification schemes and document indexing methods invented since the first archives of Sumer and even before (Piggott 2012; Lesk 1997: 99–123). Exploration is studied in various fields beyond information science, e.g., psychology, computer interface design, consumer, audience, and organizational research, and environmental planning and architectural design (Rice, McCreadie, & Chang 2001: 173–215; Bates 2007). Browsing, in the case of documents, is less understood, despite being a primary information strategy in physical documentation environments such as bookshops, newspaper kiosks, and library stacks (McKay et al. 2019: 1384). It is also a cognitive strategy more closely associated with work in archives than libraries, given the former's greater potential for surprise, which underlines the importance of translation into the digital realm of such physical clues as "cobwebs in dark corners", the unmistakable signs of forgotten information.

P11 – 282

The growing interest in browsing among information science professionals is nurtured by the limited support for browsing in digital information systems (Bates 2007; Race & Makri 2015: 2; McKay et al.: 1384), and the resultant loss of a critical information strategy; a particularly worrisome evolution, considering that physical "library shelves are almost perfectly designed as serendipity engines" (McKay et al. 2019: 1390). The irony of these belated laments is that engaging in exploration by accessing the stacks has been the subject of passionate debate throughout the history of libraries, with consensus shifting between open shelves (until the end of the 18th century, and since in the United States) and closed shelves (until recently in Europe and communist countries), depending on the number of books, cultural progressive attitude, and/or financial resources available (Rovelstad 1976; Atkins 2012: 29–30). Library exploration, and even more so archive exploration, is a privilege in the physical world, and for the time being also remains one in digital environments.

Serendipity is closely related to exploration and has been aptly described as the contrary of "problem-solving information", for which "individuals must experience a 'problem situation'" (Ross 1999: 784). Already possessed of a distinguished socio-historical pedigree (Merton & Barber 2004; Eco 1998), the advent of the World Wide Web (a name itself conveying the image of a spider waiting to see what chance will deliver to its encyclopedic net) has seen considerable attention paid to the concept of serendipity, for search engines in particular, and as a scientific method in general, as reflected in the hundreds of articles on the subject (Erdelez et al. 1997) and its status as a compelling topic for popular science books (Roberts 1989; Hofmann 2013). In the realm of household technologies, it is among the endearing idiosyncrasies of Apple's voice-operated digital agent Siri and her text-based ancestor Eliza, the make-believe psychoanalytical chatbot of the 1960s (Wikipedia: "Siri"; Weizenbaum 1966; Wikipedia: "ELIZA"). Circumscribed to "discoveries, by accident and sagacity" in Horace Walpole's original definition from 1754 (McCay-Peet & Toms 2018: 2–3), serendipity is the animating principle of exploration, but also of search (even the search of extraterrestrial intelligence (Tarter 1984)) and other information-related activities, and has a variety of definitions reflecting the complexity of factors at play (McCay-Peet & Toms 2018: 1; Witten, Gori, & Numerico 2006: 24; Race & Makri 2015: 15–26; Taramigkou et al. 2013: 3). The Document City in Figure A1 demonstrates the salient characteristics of serendipity: detection of abnormality (tall or eccentric Document Towers), sensitivity of the observer to the recognition of valuable objects (awareness of quality and costs of OCR-ed documents), reactivity to and acting upon opportunities (use of the

visualization to illustrate this article), an incubation process leading to value recognition (the interpretation of the visualization), utility (correcting misclassification), an unplannable process, unexpected findings, and results that are not guaranteed. Current research is concerned with better understanding serendipity through empirical and theoretical studies (Erdelez 1997), developing methods to induce serendipity (Erdelez 2004; Campos & de Figueiredo 2002), measure serendipity (de Figueiredo & Campos 2001; McCay-Peets & Toms 2012), and design serendipitous learning, research, and information environments (Race & Makri 2015; O'Connor 1988). Document-centered technologies generally use the selection and ranking of text content to create serendipity, rather than letting it emerge spontaneously from the data, as occurs under the Document Towers paradigm (Cooper & Prager 2000). The "Bohemian Bookshelf" is an example of an interface designed for serendipitous book discoveries that explicitly takes advantage of physical aspects (cover color and number of pages), in addition to content information (authors, keywords, and timelines) (Thud, Hinrichs & Carpendale 2012). The Document Towers further reveal other aspects of serendipity that are latent in the literature: opportunistic scanning of the visual field (an information strategy equated with "berrypicking" (Bates 2007: 410)), playfulness that invites the use of the technology; intriguing mystery arising from extreme visual simplification, stimulating the "thoughtful activity" that is visual perception (Noë 2006); bonding with the technology, which is useful to sustaining long-term usage (Ross 1999); and physical embodiment through laser engraving and three-dimensional printing to induce recurrent interaction and persistent memories.

Let us briefly speculate on the broader psycho-societal implications of an imbalance between the ability to perform information search and exploration, as these have a direct impact on appreciating the utility of the Document Towers as document visualization. Search and exploration are generic categories and reflect different goals, as well as specific individual and group behaviors and mind frames. Search is an activity in which you "ask questions from the data", while in exploration, you "listen to the data speak"; one is about points and strict demands, the other is about relationships and unexpected receiving. In information theory terms, the former yields low amounts of information and is rather akin to recovery, while the latter is full of surprises. From a psycho-social perspective, these approaches represent the dichotomy of ego-centric interests and environmental awareness; politically, that of dictatorship and democracy, as the American chemist and Nobel laureate Irving Langmuir remarked in the early years of the Cold War (while defining freedom as "the opportunity to profit from the unexpected", with the underlying message being a plea to government and industry to finance blue-sky academic research, as opposed to mission-oriented projects) (Langmuir 1956: 416).

That the former system usually keeps this activity shrouded in secrecy (e.g., restricted access to high-quality maps (Monmonier 1991: 113–123)) is not without parallels to how digital search engines give users a false sense of power over their online lives, transforming them into small tyrants who speak in "code" and computer "commands", and destructuring their linguistic abilities to the level of "keywords"; the answers, moreover, are manipulated by incognoscible algorithms obscured behind the wizardly interfaces of Oz. Their addictive allure comes from construing the clicking of hyperlinks as "surfing", the serendipitous experience of "exploration" and "interconnectedness" suggested by the names of major web browsers (e.g., Mosaic, Netscape Navigator, Internet Explorer, Safari) (Wikipedia: "List of web browsers"). Here, lodged at the heart of online information-seeking, we may fathom the oxymoron of

(a) targeted searching brought about by search engines and (b) serendipitous exploration, which is the ostensible function of web browsers.

Visualization

"Blind analytic manipulation *is never enough*." (Mandelbrot 2002: 48). The italics are by Benoît Mandelbrot, whose mathematical objects, the "fractals", became popular as home-grown psychedelic images with the spread of personal computers in the 1980s. For the statistician John W. Tukey, whose book from 1977, *Exploratory Data Analysis*, popularized the eponymous concept in reaction to the hegemony of hypothesis testing or confirmatory analysis in statistics, "the greatest value of a picture is to force us to notice what we never expected to see" (Tukey 1977: vi). In the intervening decades since these statements were made, visualization—alongside representations addressing other sensory modalities, such as sonification and haptic interfaces—has become a common technique of human–data interaction (Fekete et al. 2008; Keim et al. 2010; Ware 2008), including in the domain of documents (Fekete & Lecolinet 2006), as attested by the appearance of the term "visual document analytics" (Heyer & Scheuermann 2009). Visual analytics covers a broad spectrum of activities, from data preprocessing and representation, through quantitative and visual exploration, to interpretation, decision-making, and communication (Keim et al. 2008: 77). Its umbrella research field is diagrammatics, and it deals with the theoretical and practical aspects of the design and reasoning through physical diagrams and abstract schema. Diagrammatics has roots in antiquity and across cultures. It stands at the intersection of the broadest range of disciplines, notably philosophy, logic, mathematics, artificial intelligence, physics, psychology, design, information visualization, and media studies (Schneider, Ernst, & Wöpking 2016; Anderson, Meyer, & Olivier 2002; Varia 2000).

The ongoing work of applying visualization to understanding data processing in one of the most impactful computer science fields, namely neural networks (Rauber 2017; Yosinski et al. 2015), underscores another tenet of this article, namely the actuality of technological transparency with respect to the degree and manner in which the inner workings and the encapsulated and linked data are revealed (Figure 9). Inscrutable but easy-to-use "black boxes" (Latour 1987) oppose insight-rich but complex "crystal goblets"—the ideal of typography that, as Beatrice Warde of the Monotype typesetting corporation remarked in 1930, should convey the crimson shine of its content unhindered to the avid reader (Warde 1956)—and their computer counterparts, the graphical user interface "windows" that create the illusion of data and software covered in transparent skin.

The materiality of documents, the spatial organization of collections, the remoteness of off-site storage, the access rights, the limited screen size, and the intangibility of digital documents are among the factors that make information opaque, even invisible. With light and unauthorized eyes being their enemies, one struggles to produce examples of transparent documents, libraries, and archives (apart from, maybe, the glass slabs constituting the soon-to-be-disconnected memory of the psychotic computer Hal in *2001: A Space Odyssey* (1968), and an artist fashioning the worried astronaut's words "Hal? Hal!" as a glass object (Klante et al. 2004: 126–127), or the giant showcase holding King George III's book collection in the entrance hall of the British Library and the rare books at Yale (The British Library 2019; Campbell & Pryce 2013: 248, 268–269)). It is not even clear whether readers prefer the transparency of sunny, open-plan reading rooms (perhaps, as seen at the National Library of France, surrounded by bucolic vistas of gardens populated by licentious rabbits and

squabbling parrots, living modern versions of medieval tapestries (Anonymous n.d.)), or the tiny alcove at the Shakespeare and Co. bookshop vis à vis of Notre-Dame in Paris (Halverson 2016: 289)—cozy cone of electric light surrounded by darkness, "bony O's" (to borrow the Shakespearian image of skull as theater in the Prologue to *Henry V* (Richard 2019: 200–201)) in the seclusion of which they can work on their "imaginary forces" without the distraction of visual stimuli (theater, incidentally, has been shown to bear striking similarities with information processing by computers (Laurel 2014)). Although windows have replaced stone as the building' outer layer in the time between ziggurats and Manhattan, Philip Johnson, who created the icon of American architecture that is the transparent Glass House (1949), still built for himself a library with just one window, explaining that "you're distracted by wind, the sun, the waving of trees. It's very hard to work where you have nature…. And it's very important to get isolation… for contemplation." (Devens et al. 1991). It is indeed a tomb-like isolation that buildings housing intelligence agencies are sometimes intended to provide through their architectural style (Figure 2). If information visualization is born out of the opaqueness and immateriality of documents and their contents, the question of how much clarity is desirable remains a critical parameter of digital information systems.

The success of visualization and other representations relates to how human sensory modalities produce different perspectives on data and have different degrees of processing power, as well as the filtering and transformation necessary to make data acquirable by these sensory systems (Zhang 2008). Transformations also play a role in mathematics and computer science, since finding solutions becomes more affordable by moving problems from one domain to another. One example is the Fourier transform and its fast algorithmic implementation by James Cooley & John Tukey, through which data can be analyzed in either the spatio-temporal or frequency domains, with regularities being easier to identify in the latter (Cooley & Tukey 1956). Natural languages proceed similarly, with metaphors facilitating comprehension by presenting the unfamiliar in familiar guises (Lakoff & Johnson 1980; Indurkhya 1992; Hofstadter & FARG 1995). This principle is also at the root of good design, where it is known as "affordance": the more the form and behavior of objects are familiar to users, more they are likely to use them efficiently (Gibson 2015: 119–135; Norman 1990; Zhang 2008: 215–237; Overhill 2012). It should therefore come as no surprise that the drive to grasp abstract informational entities by projecting them into the tangible space of architecture is deeply ingrained. However, the importance of transformation is far more fundamental than the diverse domains discussed above, in fact, it is the principle of information, energy, and life itself.

Metaphors

Regarding the Document Towers, in addition to the method of visual document exploration that defines the conceptual framework, their substance is three metaphors reviewed hereafter: information-as-place, information-as-document, and information-as-architecture. Their role is to provide efficient mappings between physical or digital documents and cognitive models of documents (Norman 1990: 23–27, 75–79; Dillon 204: 63–64). Metaphors are also compact substitutes for narratives, which play their own roles in information behavior and human—computer interaction (Doty & Broussard 2017; Tufte 1997). For a discussion of the various types of information in the context of information science, see Buckland (1991) and Resnikoff (1989); for metaphors in general, see Gibbs (2008); for their role in visualizations for information retrieval, see Zhang (2008: 215–237); for a human factors perspective on the spatial structure of information, refer to Dillon (2004); on the role of

Figure A2. Mixed-reality library—In a project from the early 2000s of the Conservatoire national des arts et métiers (CNAM), Paris, using the VisonStation hardware by Elumens, a collection of physical books was digitized, rendered as facsimiles, and projected back into the physical space on a half-dome screen (Almeida et al. 2006). Note the axial arrangement of the pages on the left, reminiscent of old-fashioned museal equipment used for the display of individual folios of illuminations or engravings. (Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer, Technologies for E-Learning and Digital Entertainment by Z. Pan et al., CC-BY-NC 2006)

P11
–
286

metaphors in computer science, see Colburn & Shute (2008); Gillan & Bias (1994); Johnson (1994); Wozny (1989), and Weber (1986); on the theory of the design and evaluation of metaphorical interfaces, see Chatwin (1988); Mehrpooya & Nowroozzadeh (2013); Marshall, Nelson, & Gardiner (1987); L'Abbate & Hemmje (1998), and Liebert (1994).

Place

The representation of INFORMATION-AS-PLACE is a deeply ingrained cognitive model in computer science, as evidenced in the technical and popular terminology "goto" and "return" commands, "front" and "back end", "up-" and "download", "physical" and "logical layers", "file paths" and "Universal Resource Locators", computing "platforms" and "pipelines", network "ports", "gateways", "tunnels", "nodes", and "routers", "shells", "firewalls", and "clouds", "chat rooms", "portals", "home pages", "web sites", "under construction", "navigation breadcrumbs", "information superhighway", "Internet café", "the global village", and "cyberspace" (a "Heavenly City" and abode of a "Second Life" that "will feel like Paradise", were only some of the ecstatic visions from before the turn of the millennium (Benedikt 1991: 16, 52), although others were less enthusiastic: "Gomorrah… Hell" (Wertheim 2000: 290–299), "the new domain that [the West] has colonized" (Sardar 1995: 778)) (Bittarello 2009; Stefik 1997: xv–xxiv; Targowski 2016: 171–180; Meyer 2000). Entire scientific fields and professions—consider Information Architecture (Resmini & Rosati 2011) and Document Engineering—bear names that frame information and documents as spatial structures and

machines. As graphical operating systems became mainstream in the 1980s, the concept of spatial data management was increasingly applied to problems of digital libraries (Lesk 1997: 163), resulting in a broad spectrum of three-dimensional document representations (Fang et al. 2009; Card 2008). During the 1990s the need to imagine the abstract structure and flow of information on the Internet blossomed into a cornucopia of visualizations employing spatial metaphors (Dodge 2001). Research into hardware solutions designed to master the snowballing deluge of digital data led to immersive physical information exploration environments, room- and building-sized, with names such as "Pocket Cathedral" and "CAVE", cultural references to the Gothic Revival and Plato (Peddie 2013: 419–423) (Figure A2). One aspect for the creators and users of these apparently benign similes to reflect on is their potential impact, which has unexpected legal, policy, and social consequences (Hunter 2003; Osenga 2013; Graham 2013).

It is tempting to see this terminology as a product of the "young, anti-authoritarian" culture of computing and the Internet (Meyer et al. 1998: 525), with a whiff of California coolness, entirely oblivious to the Latin and Greek spoken in the club of elder sciences. Instead, information spatialization should be considered an unsurprising propensity, given the importance of spatial cognition in human life (Landau 2002). Information spatialization has been developed by many *cultures* (Wertheim 2000), especially as an instrument of *memorization*. Such spatialization may provide valuable inspiration for technologists, insofar as our present information systems remove, via hyperlinks or cloud storage, the necessity for the user to remember where information was found (spatial memory) in favor of knowing how to find it (procedural memory). For example, in the Western world, from Greco-Roman antiquity until the Enlightenment, there existed a set of mnemotechniques known as the Art of Memory. These consisted in attaching the information to be remembered, from simple dates to lengthy conference proceedings, to physical or imaginary objects, then placing these in a specific order with reference to real or imagined buildings for later retrieval (Yates 1966). Similary, the "songlines" of Australian Aborigines pertaining to mythical animals and their deeds encode strings of landmarks to support cross-continental navigation (Chatwin 1988). For the Apache people, places are where wisdom is found in the form of geolocated stories (Basso 1996), a concept not dissimilar to the French historian Pierre Nora's description of how "places of memory" (lieux de mémoire), such as statues, monuments, or battlefields, are integral to building national identities (Nora & Jordan 2001–2010) (Figure 2).

Documents

INFORMATION-AS-DOCUMENT representation is also common in computer science, a bibliomorphism that has as many proselytes as detractors (Dillon 2004: 116–117; PopularLibros.com 2010). It ranges from the now-standard, operating system interfaces, where application windows are laid out on screens like papers on a desk (Clarkson 1991), to three-dimensional desktops (Wikipedia: "Project Looking Glass"), to versions with the form and feeling of naturalistic libraries and bookshelves (Card, Robertson, & York 1996; Rauber & Bina 2000; Cubaud, Stokowski, & Topol 2002; Cubaud 2008). Using architecture, urban planning, and landscape as metaphors for organizing digital information continues to be fruitful. Desktops have been represented as floor plans (Henderson & Card 1986), and the structure of software code libraries rendered as stylized cities (Alam et al. 2009) (Figure A3). Moreover, "landscape" is the driving metaphor for the visualization of Self-Organizing Maps (Wise et al. 1995), and multimedia objects can be navigated using maps built with Geographical Information Systems software (Laufer, Halacsy, & Somlai-Fischer 2011).

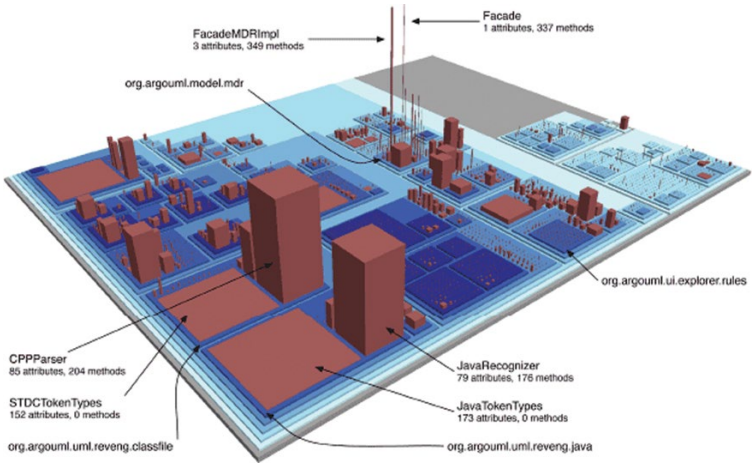


Figure A3. Software architecture—EvoSpaces uses the city metaphor to visualize software properties, here of the ArgoUML Java libraries: Building heights represent the number of methods, base sizes the number of attributes, and their spatial location the nesting level of packages (Alam et al. 2009: 174). The striking similarity between EvoSpaces and Document City reveal that natural language texts and computer code share fundamental structural characteristics. (Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer, Human Machine Interaction by Denis Lalanne and Jürg Kohlas, CC-BY-NC 2009)

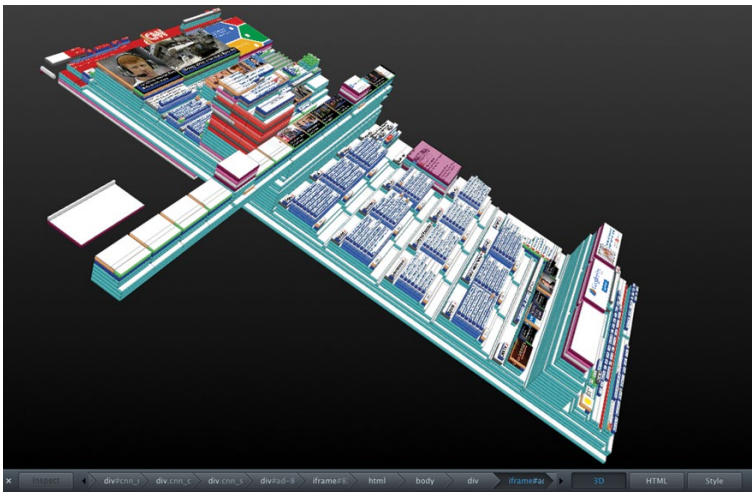


Figure A4. Webpage architecture—The "3D view" extension to the FIrefox webbrowser (until version 46) provided a visualization of the structure of HTML objects within a webpage's document object model, similar to what the Document Towers does for PDFs and other paginated document formats

P11
–
288

Spatial metaphors have been used to great effect in visualizing the structure of the World Wide Web, reflecting the awareness of its pioneers of living on an "electronic frontier" (yet another spatial analogy) (Munro, Höök, & Benyon 1999; Markham & Tindberg 2020; Maglio & Matlock 1998; Meyer, Zaluski, & Mackintosh 2008; Markham 2003; cross-cultural perspectives: Galichkina 2016; Meyer et al. 1998). Examples include the representation of webpage elements as a perspective of extruded structures in the Firefox webbrowser (Mozilla n.d.) (Figure A4), the visualization of websites as technical drawings (Dodge & Kitchin 2001: 81–91;

Durand & Kahn 1998), the conceptualization of navigation between Internet domains as a walk through tunnels between various rooms using a game engine Selfridge & Kirk 1999), and the fanciful conception of information as a landscape of swamps, oceans, forests, roads, barns, villages, and liquid architectural structures (Dieberger 1994: 62–65).

Architecture

Body, text, architecture, and urbanism have been interchangeable metaphors in many cultures. A book has a "head", a "spine", and a "tail", while its content has "headings", text "body", and "footnotes" (Milon & Perelman 2012). This bibliographical terminology was reimagined by Archimboldo in "The Librarian" as a human made entirely of books (Wikipedia: "The Librarian (painting)"). Four centuries later, the same cross-mirroring produced the metaphors of computers-as-brains and brains-as-computers (Denett 1984; Searle 1990; Gigerenzer & Goldstein 1996; Massaro 1986; Berman 1989; Casey & Moran 1989; Coleman & Fraser 2011). The architectural metaphor can be traced from the very definition of language as "an ancient city: a maze of little streets and squares, of old and new houses" presented by Wittgenstein (*PI*: § 19; Wittgenstein 2009: 11), and as a tower reaching in vain for divine knowledge in the myth of the Tower of Babel (Atanasiu 2001), to the peculiar eroticism of the Biblical *Song of Songs* ("Your neck is the tower of David") (Anonymous 1925; § 3.4) and 15th century Burgundian poetry ("il n'y a plus nobles logis que sont les cons") (Cowling 1998: 23), down to the topos of the book as a "pocket cathedral" popularized by the Romanticist art critic John Ruskin, painter Edward Burne-Jones, and novelist Victor Hugo (Schoenherr 2004; Le Men 1998: 14, 20, 21), and to "typotecture", the fusion of typography and architecture in bricks-and-mortar and graphic arts buildings and cityscapes of the 20th and 21st centuries (Heller & Ilić 2013; Jenser 2002). Cult movies also use the same metaphors: in *Tron* (1982) humans are envisioned as data packets battling inside computer architectures, while the virtual reality emerging from streams of Japanese characters in *The Matrix* (1999) defines the fabric of space-time itself as information (Dodge & Kichin 2001: 234–237).

At the most general level of INFORMATION-AS-ARCHITECTURE, it suffices to paraphrase the architect Le Corbusier by pointing out that documents and libraries are "machines to inform with" (Hillier 2007: 292). Not unlike other architects who designed their books themselves (Rem Koolhaas (O. M. A., Koolhaas, & Mau 1995)) or had an impact on book history (Renaissance architects (Carpo 2001), the Bauhaus movement (Hollis 2006)), Le Corbusier vouched for the isomorphism between information architecture and living architecture (Le Corbusier 1950: 9; De Smet 2007), taken literally in such famous monuments as the letter-shaped futurist Book Pavilion by Fortunato Depero (Monza, 1927) (CIMAD 2019; Tavares 2016: 150–153) and the book-shaped National Library of France by Dominique Perrault (Paris, 1996) (Figure 2). Computer scientists have examined the commonalities between architectural facades and computer interfaces—their pragmatic aspects, designer vision vs user wishes, aesthetics and politics—to learn from the work of architects (Hooper 1986).

Three specific computational implementations, selected for illustrative purposes, originate with researchers from the MIT MediaLab working at the crossroads of technology and the arts (Figure A5). "Books without Pages" was a project spanning from the 1970s to the early 2000s that attempted to replicate key qualities of the physical document and bodily interactions in its digital avatar and computational infrastructure (Reinfurt & Weisenberger 2014: 19–20). One particularly visionary concept, and technical feat for the times, was to render text columns as surfaces in a navigable three-dimensional space (Small 1999; Cooper & VLW 1994). "City of News"

P11
–
290

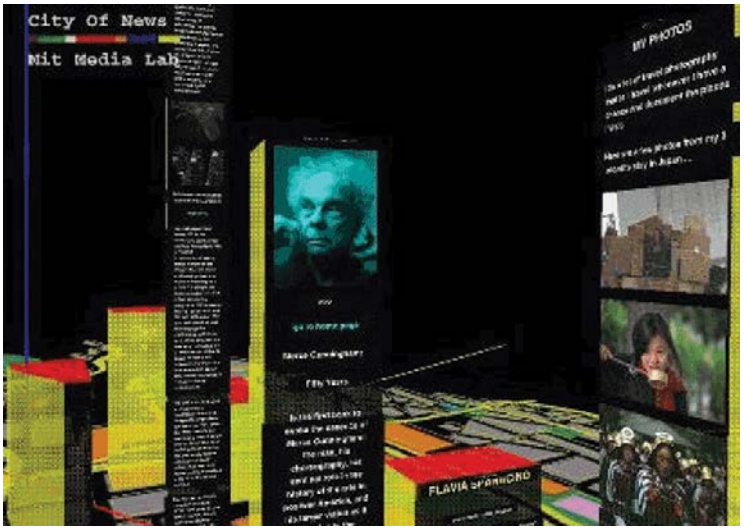


Figure A5. Information Cities—Three projects on information visualization from MIT MediLab using the architecture and city metaphors: Information Landscapes, City of News, and Cityscape
(Cooper & VLW 1994; Sparacino, Davenport, & Pentland 2000: 179; Dragulescu 2009: 74).
(Credits: Muriel Cooper and the Visible Language Workshop / Flavia Sparacino / Alexandru C. Dragulescu)

uses the metaphor of the urban space, INFORMATION-AS-CITY, as an intuitive spatial model for information display. Remarkably, its design was explicitly grounded in concepts from Classical antiquity and the Renaissance, science fiction, and Cognitive Science (Sparacino, Davenport, & Pentland 2000); we can observe that one of these is the metaphor of INFORMATION-AS-BODY, as addressed by the "Cityscape" project (Dragulescu 2009: 66–76). Here, the representation of feeds on social media as towers resembling futuristic cairns serves to individualize users by "abstract architectural bod[ies]". The city-of-skyscrapers information visualization metaphor results from "the idea that the data we generate online is the result of us living in various habitats, using various tools".

At work in these projects is the construction of a reality in which the physical reflects the virtual and vice-versa, as in a hall of mirrors. The magic exuded by the visualizations is a kind of modern technological exorcism that makes architecture, documents, and interfaces more "user-friendly" through familiar behavior models (Meyer et al. 1998: 524–525).

Parsimony

While document structures have more degrees of freedom when represented as three-dimensional objects, lower-dimensional representation spaces have their own advantages, notably more spatially compact representation, an absence of occlusion, and they allow to visualize the model from all angles without manipulation. The Ribbons technique (Figure 6) belongs to this rich class of tiny diagrams, some of which are ubiquitous—the visualization of location and keyword frequency in document search results, located next to the sliding bar of application windows such as a web pages (Hearst 1995)—while others are experimental: document thumbnails with semantic highlighting, variable text size, and selection of significant graphics (Stoffel et al. 2010; Stoffel et al. 2012; Stoffel et al. 2009), a dashboard of topic distribution in documents (Humphreys et al. 2018) depicted as plots of acoustic signals (Favata et al. n.d.), chat logs timelines (Donath & Viégas 2002), line modification frequency in software code (Eick, Steffen, & Summer 1992), or dominant colors in movie frames (Brodbeck 2011).

Phenomenology

Playfulness is important in the creation of the Document Towers, as shown in its phenomenology. The Document Towers are experienced by the author as a mix of aesthetic pleasure derived from their form, epistemological thrill provided by their function as a source of information, and the satisfaction derived from cognitive action and physical interaction. A similar concept is advocated by the Japanese Kansei approach to industrial design, which emphasizes bonding between users and technology (Schütte 2005) (as illustrated by the care-exigent cyberpet Tamagotchi (Wikipedia: "Tamagotchi")), with the aim of stimulating social knowledge transfers, thus contributing to product adoption and fitness and (hopefully) more harmonious societies. Current research in user experience is operationalizing, for its own ends, the work of Western philosophers, such as Heidegger, Dewey, and Merleau-Ponty, on phenomenology. The aim is to challenge the very definition of what interfaces ought to be: rather than hardware and software, "our thesis is that user experience consists of three core, inter-related elements: involvement, affect and aesthetics" (Turner 2017: ix, 3–8).

The Document Towers' shape is defined by the physical structure of a particular document format, namely the codex, raising the question of what mood and connotations it imparts to users. People confronted with this visualization for the first time often report that the modernist architectural style, the Manhattanesque look and feel of the Document City exhibits a fascinating symbiosis of crystal rationality and lively chaos, and as a result, they draw closer to investigate the visualization at close range (personal observation). Here, we witness

how the properties of the cognitive model domain propagate to the source domain of document interpretation. Understanding the response to the styles of built environment and comparing reactions in the physical and virtual worlds is an active research topic in computer games, cultural heritage, architecture, and urban planning (Houtkam 2005; Botton 2014).

Conclusions

This overview of document-related information representations has revealed how the Document Towers paradigm is rooted in an ancient and sprawling cultural and technological ecology. While fascinating, it remains (intriguingly) populated primarily by prototypes.

Perspectives

The opening argument of the article was that information behavior extends far beyond the question-asking offered by today's digital search engines. Based on the case studies of exploration, visualization, and metaphor, the word "dreaming" may be proposed as an overall concept describing what future technologies should support. Could we, in these closing lines, speculate on what that might mean in practice?

P11
–
292

"Dreaming" in the context of information systems is defined here as the state between waking and sleep and is characterized by (a) the awareness of both conscious control of thought and unconscious activity and (b) an unusual, "Alice in Wonderland" type of reality. This state is conductive to exploration, since the reality it offers is new, and to solving problems, as it encourages adopting new perspectives—this propensity of dreams being well-known, including in scientific discovery (Samuels & Samuels 1975). Exploration and problem solving imply human—machine integration that goes beyond targeted querying. The information system becomes a system that you think with, rather than simply ask from.

This article has presented some of the methods by which oneiric information systems could be designed, identifying key elements such as mystery (to elicit curiosity), pleasure (to create conditions for playfulness), metaphors (to create new perspectives), and serendipity (to allow for stumbling across useful things one did not think to search for). The role of the physical interface is equally important as regards aspects such as available display space and the haptic exploration and manipulation of data and methods.

Thus, to conclude, future digital information systems would benefit from preserving aspects of organization that have proven useful in physical libraries—which, let us remember, have always been great places for dreaming (Figure 5).

References

Alam, S., Boccuzzo, S., Wettel, R., Dugerdil, Ph., Gall, H., & Lanza, M. (2009). EvoSpaces: Multi-dimensional Navigation Spaces for Software Evolution. In Denis Lalanne, Jürg Kohlas (eds.), *Human Machine Interaction: Research Results of the MMI Program*. Berlin: Springer, pp. 167–192. https://doi.org/10.1007/978-3-642-00437-7_7

Almeida, R. A. de, Cubaud, P., Dupire, J., & Topol, A. (2006). Interactions et métadonnées riches pour les bibliothèques numérisées. *Document numérique* 9 (2): 83–109. https://doi.org/10.3166/dn.9.2.83-109

Anderson, M., Meyer, B., & Olivier, P. (eds.) (2002). *Diagrammatic Representation and Reasoning*. London: Springer. https://doi.org/10.1007/978-1-4471-0109-3

Anonymous (1925). *The Song of Songs*. Illustrated by Eric Gill. Waltham St. Lawrence: Golden Cockerel Press.

Anonymous (n.d.). Document exceptionnel : le lapin sauvage de la Bibliothèque François Mitterrand. *YouTube*. https://www.youtube.com/watch?v=7PQZaiFzvBU (2020.01.25)

Atanasiu, V. (2001). 5170 marches jusqu'à Dieu – Traditions musulmanes relatives à la Tour de Babel et à la confusion des langues. *Annales Islamologiques* 35 (2001): 1–16.

Atkins, A. N. (2012). Discretion in Russian Librarianship: Pre-Soviet, Soviet, Post-Soviet. M.A. thesis, Slavic and East European Studies, Ohio State University, Columbus, OH. http://rave.ohiolink.edu/etdc/view?acc_num=osu1343769040 (2020.01.26)

Basso, K. H. (1996). *Wisdom Sits in Places: Landscape and Language Among the Western Apache*. Albuquerque, NM: University of New Mexico Press.

Bates, M. J. (2007). What is browsing – really? A model drawing from behavioral science research. *Information Research* 12 (4), paper 330. http://www.informationr.net/ir

Benedikt, M. (ed.) (1991). *Cyberspace: First Steps*. Cambridge, MA: The MIT Press.

Berman, B. (1989). The computer metaphor: Bureaucratizing the mind. *Science as Culture* 1 (7): 7–42. https://doi.org/10.1080/09505438909526258

Bittarello, M. B. (2009). Spatial Metaphors describing the Internet and religious Websites: sacred Space and sacred Place. *Observatorio* 11 (4): 1–12.

Botton, A. de. (2014). *The Architecture of Happiness*. London: Penguin.

Brodbeck, F. (2011). *cinemetrics*. B.A. project, Royal Academy of Arts, Den Haag, The Netherlands. http://cinemetrics.fredericbrodbeck.de (2019.03.29)

Buckland, M. (1991). *Information and Information Systems*. New York, NY: Praeger.

Bush, V. (1945). As we may think. *Atlantic Monthly* 176 (1): 101–108. http://dougengelbart.org/content/view/258/000/ (2020.01.26)

Campbell, J. W. P., & Pryce, W. (2013). *The Library: A World History*. Chicago, IL: University of Chicago Press.

Campos, J., & de Figueiredo, A. D. (2002). Programming for Serendipity. *Proceedings of the 2002 AAAI Fall Symposium on Chance Discovery: The Discovery and Management of Chance Events*, AAAI Technical Report FS-02-01, pp. 48–60. https://doi.org/10.2139/ssrn.1385402

Card, S. (2008). Information Visualization. In S. Andrew, J. A. Jacko (ed.), *The Human–Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*. New York: Lawrence Erlbaum, 2nd ed., pp. 509–543.

Card, S. K., Robertson, G. G., & York, W. (1996). The WebBook and the Web Forager: An Information Workspace for the World-Wide Web. *Procedings of the ACM Confernce on Human Factors in Computing Systems (CHI'96), 13–18 April 1996, New York, NY, USA*, pp. 111–117. https://doi.org/10.1145/238386.238446

Carpo, M. (2001). *Architecture in the Age of Printing: Orality, Writing, Typography, and Printed Images in the History of Architectural Theory*. Cambridge, MA: The MIT Press.

Casey, G., & Moran, A. P. (1989). The Computational Metaphor and Cognitive Psychology. *Irish Journal of Psychology* 10 (2): 143–161. https://doi.org/10.1080/03033910.1989.10557739

Chatwin, B. (1988). *The Songlines*. New York, NY: Penguin.

Chisholm, R. M. (1986). Selecting Metaphoric Terminology for the Computer Industry. *Journal of Technical Writing and Communication* 16:3: 195–220. https://doi.org/10.2190/XAJ3-8THN-7QGM-93U5

[CIMAD] Center for Italian Modern Art and Designers & Books (2019). *Depero Futurista: The Bolted Book, A New Facsimile*. https://www.boltedbook.com (2019.07.02)

Clarkson, M. A. (1991). An Easier Interface. *Byte* 16 (2): 277–282.

Colburn, T. R., & Shute, G. M. (2008). Metaphor in computer science. *Journal of Applied Logic* 6: 526–533. https://doi.org/10.1016/j.jal.2008.09.005

Coleman, D., & Fraser, H. (eds.) (2011). *Minds, Bodies, Machines, 1770–1930*. Houndmills: Palgrave Macmillan. https://doi.org/10.1057/9780230307537

Cooley, J. W., & Tukey, J. W. (1956). An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation* 19 (90): 297–301. https://doi.org/10.1090/S0025-5718-1965-0178586-1

Cooper, J. W., & Prager, J. M. (2000). Anti-serendipity finding useless documents and similar documents. *Proceeding of the 33rd Hawaii Interantional Conference on System Sciences (HICSS), 4–7 January 2000, Maui, Hawaii, USA*, pp. 1–9. https://doi.org/10.1109/HICSS.2000.926691

Cooper, M., Visible Language Workshop (1994). *Information Landscapes*. Video, MIT Media Laboratory, Cambridge, MA: MIT. Digitized from VHS tape by David Young, facsimile from original code by Michael McKenna. http://infosthetics.com/archives/2010/02/information_landscapes_1994_muriel_cooper_mit_visible_language_workshop.html; https://www.youtube.com/watch?v=Qn9zCrIJzLs; http://smalldesignfirm.com/InformationLandscapes/ (2020.01.26)

Le Corbusier (1950). *Le Modulor*. Basel: Birkhäuser.
Cowling, D. (1998). *Building the Text: Architecture as Metaphor in Late Medieval and Early Modern France*. Oxford: Oxford University Press.
Cubaud, P. (2008). 3D Interaction for Digital Libraries. *Digital Libraires*. London & Hoboken, NY: ISTE & John Wiley & Sons, pp. 123–143. https://doi.org/10.1002/9780470611302.ch8
Cubaud, P., Stokowski, P., & Topol, A. (2002). Binding Browsing and Reading Activities in a 3D Digital Library. *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'02), 13–17 July 2002, Portland, OR, USA*, pp. 281–282. https://doi.org/10.1145/544220.544282
De Smet, Catherine. (2007). *Vers une architecture du livre. Le Corbusier: édition et mise en pages 1912–1965*. Baden: Lars Müller.
Dennett, D. C. (1984). The Role of the Computer Metaphor in Understanding the Mind. In H. R. Pagels (ed.), *Proceedings of a symposium on Computer Culture: the Scientific, Intellectual, and Social Impact of the Computer*. New York, NY: Annals New York Academy of Sciences, pp. 266–275. https://doi.org/10.1111/j.1749-6632.1984.tb16524.x
Devens, E., Schultz, F., Shaw, J., & Sanchis, F. (1991). Walking Tour with Philip Johnson, 1991. In *National Trust for Historic Preservation and The Glass House, The Glass House: Studio Overview*. http://theglasshouse.org/explore/studio/ (2019.10.06)
Dieberger, A. (1994). Navigation in Textual Virtual Environments using a City Metaphor. Ph.D. Dissertation, Vienna University of Technology, Vienna, Austria.
Dillon, A. (2004). *Designing Usable Electronic Text: Ergonomic Aspects of Human Information Usage*. Boca Raton: CRC Press, 2nd ed.
Dodge, M., & Kitchin, R. (2001). *Atlas of Cyberspace*. Harlow: Addison-Wesley.
Donath, J., & Viégas, F. B. (2002). The Chat Circles Series: Explorations in designing abstract graphical communication interfaces. *Proceedings of the 4th Conference on Designing Interactive Systems, 25–28 June 2002, London, UK*, pp. 359–369. https://doi.org/10.1145/778712.778764
Doty, Ph., & Broussard, R. (2017). Fiction as Informative and its Implications for Information Science Theory. *Proceedings of the 80th Annual Meeting of the Association for Information Science and Technology (ASIS&T), 27 October – 1 November 2017, Washington, DC, USA*, 57: 61–70. https://doi.org/10.1002/pra2.2017.14505401008
Dragulescu, A. C. (2009). Data Portraits: Aesthetics and Algorithms. M.A. thesis, Massachusetts Institute of Technology, Cambridge, MA.
Durand, D., & Kahn, P. (1998). MAPATM: A System for Inducing and Visualizing Hierarchy in Websites. *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia, 20–24 June 1998, Pittsburgh, PA, USA*, pp. 66–76.
Eco, U. (1995). *De bibliotheca*, Milano: Biblioteca comunale.
Eco, U. (1998). *Serendipities: Language and Lunacy*. New York, NY: Columbia University Press. https://doi.org/10.7312/ecou11134
Eick, S. G., Steffen, J. L., & Summer, Jr., E. E. (1992). SeeSoft – A Tool for Visualizing Software. *IEEE Trans. on Software Engineering* 18 (11): 957–968. https://doi.org/10.1109/32.177365
Erdelez, S. (1997). Information Encountering: A Conceptual Framework For Accidental Information Discovery". In P. Vakkari, R. Savolainen, B. Dervin (eds.), *Information-seeking in context: Proceedings of an international conference on research in information needs, seeking and use in different contexts (ISIC '96), 14–16 August 1997, Tampere, Finland*. London: Taylor Graham, pp. 412–421.
Erdelez, S. (2004). Investigation of information encountering in the controlled research environment. *Information Processing & Management* 40 (6): 1013–1025. https://doi.org/10.1016/j.ipm.2004.02.002
Erdelez, S., Heinström, J., Makri, S., Björneborn, L., Beheshti, J., Toms, E., & Agarwal, N. K. (2016). Research Perspectives on Serendipity and Information Encountering. *Proceedings of the 79th Annual Meeting of the Association for Information Science and Technology (ASIS&T), 14–18 October 2016, Copenhagen, Denmark* 53 (1), 5 p. https://doi.org/10.1002/pra2.2016.14505301011
Fang, X., Jacquemin, C., Vernier, F., & Luo, B. (2009). A Survey of 3D Document Corpus Visualization. *Information Technology Journal* 8 (1): 1–15. https://doi.org/10.3923/itj.2009.1.15
Favata, Marc, Dojcinoski, G., Drywa, N., & Moore, B. [n. d.]. *Detector Sounds: Sounds from the initial LIGO detectors, Sounds of Spacetime*. https://www.soundsofspacetime.org/detector-noise.html (2019.03.31)
Fekete, J.-D., & Lecolinet, É. (2006). *Document Numérique: "Visualisation pour les bibliothèques numériques"* 9 (2).
Fekete, J.-D., van Wijk, J. J., Stasko, J. T., & North, Ch. (2008). The Value of Information Visualization. In K. Andreas, J. T. Stasko, J.-D. Fekete, Ch. North (eds.), *Information Visualization:*

Human-Centered Issues and Perspectives. Berlin: Springer, pp. 1–18. https://doi.org/10.1007/978-3-540-70956-5_1

Figueiredo, A. D. de, & Campos, J. (2001). The Serendipity Equations. In R. Weber, C. Gresse (eds.), *Proceedings of the Workshop Program at the Fourth International Conference on Case-Based Reasoning (ICCBR 2001)*. Technical Note AIC-01-003, Washington, DC: Naval Research Laboratory, Navy Center of Applied Research in Artificial Intelligence, pp. 121–124.

Fisher, K. E., Erdelez, S., & McKechnie, L. (E. F.) (2006). *Theories of Information Behavior*. Medford, NJ: Information Today.

Galichkina, E. N. (2016). Internet Conceptual Metaphors. *Journal of Siberian Federal University: Humanities & Social Sciences* 9 (2): 470–480. https://doi.org/10.17516/1997-1370-2016-9-2-470-480

Gibson, J. J. (2015 [1979]). *The Ecological Approach to Visual Perception*. New York, NY: Psychology Press.

Gibbs, R. W., Jr. (2008). *The Cambridge Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511816802

Gigerenzer, G., & Goldstein, D. G. (1996). Mind as Computer: Birth of a Metaphor. *Creativity Research Journal* 9 (2–3): 131–144. https://doi.org/10.1207/s15326934crj0902&3_3

Gillan, D. J., & Bias, R. G. (1994). Use and abuse of metaphor in human-computer interaction. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 2–5 October, San Antonio, TX, USA*, pp. 1434–1439. https://doi.org/10.1109/ICSMC.1994.400047

Graham, M. (2013). Geography/internet: ethereal alternate dimensions of cyberspace or grounded augmented realities? *The Geographical Journal* 179 (2): 177–182. https://doi.org/10.1111/geoj.12009

Halverson, K., ed. (2016). *Shakespeare and Company, Paris: A History of the Rag & Bone Shop of the Heart*. Paris: Shakespeare and Company.

Hearst, M. A. (1995). TileBars: Visualization of Term Distribution Information in Full Text Information Access. *Proceedings of the ACM 1995 Conference on Human Factors in Computing Systems (CHI'95), 7–11 May 1995, Denver, CO, USA*, pp. 59–66. https://doi.org/10.1145/223904.223912

Heller, S., & Ilić, M. (2013). *Lettering Large: The Art and Design of Monumental Typography*. New York, NY: The Monacelli Press.

Henderson, D. A., Jr., & Card, S. K. (1986). Rooms: The Use of Multiple Virtual Workspaces to Reduce Space Contention in a Window-Based Graphical User Interface. *ACM Transactions on Graphics* 5 (3): 211–241. https://doi.org/10.1145/24054.24056

Heyer, G., & Scheuermann, G. (2009). *Visual Document Analytics Workshop (VDA 2009), 8 June 2009, Leipzig, Germany*. http://vda.informatik.uni-leipzig.de (2019.07.08)

Hillier, B. (2007). *Space is the machine: A configurational theory of architecture*. London: Space Syntax.

Hofmann, A. (2013). *LSD: My problem child*. Oxford: Oxford University Press.

Hofstadter, D., & Fluid Analogies Research Group (1995). *Fluid Concepts & Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York, NY: Basic Books.

Hollis, R. (2006). *Swiss Graphic Design: The Origins and Growth of an International Style, 1920–1965*. New Haven, CT: Yale University Press.

Hooper, K. (1986). Achitectural Design: An Analogy. In D. A. Norman, S. W. Draper (eds.), *User Centered System Design: New Perspectives on Human–Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum, pp. 9–26. https://doi.org/10.1201/b15703-1

Houtkam, J. (2005). Engineering the Affective Appraisal of 3-D Models of Buildings. In H. Van Oostendorp, L. Breure, A. Dillon (eds.), *Creation, Use, and Deployment of Digital Information*. Mahwah, NJ: Lawrence Erlbaum, pp. 93–199.

Humphreys, A., Spencer, Ch., Brown, L., Loy, M., & Snyder, R. (2018). Reimagining the Digital Monograph: Design Thinking to Build New Tools for Researchers. *Journal of Electronic Publishing* 21 (1). https://doi.org/10.3998/3336451.0021.102

Hunter, D. (2003). Cyberspace as Place and the Tragedy of the Digital Anticommons. *California Law Review* 91 (2): 439–519. https://doi.org/10.2307/3481336

Indurkhya, B. (1992). *Metaphor and Cognition: An Interactionist Approach*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-2252-0

Jenser, A. (2002). *Typotecture: Typography as Architectural Imagery*. Zürich/Baden: Museum für Gestaltung Zürich & Lars Müller Publishers.

Johnson, G. J. (1994). Of metaphor and the difficulty of computer discourse. *Commuincations of the ACM* 37 (12): 97–102. https://doi.org/10.1145/198366.198388

Kalbach, J. (2006). On Uncertainty in Information Architecture. *Journal of Information Architecture* 1 (1): 48–56.

Keim, D. A., Kohlhammer, J., Ellis, G., & Mansmann, F. (eds.) (2010). *Mastering The Information Age: Solving Problems with Visual Analytics*. Goslar: Eurographics Association.

Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., & Ziegler, H. (2008). Visual Analytics: Scope and Challenges. In S. J. Simoff, M. H. Böhlen, A. Mazeika (eds.), *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*. Berlin: Springer, pp. 79–90. https://doi.org/10.1007/978-3-540-71080-6_6

Klanten, R., Mischler, M., Bilz, S., & Thoenen, N. (2004). *Type One: discipline and progress in typography*. Berlin: Die Gestalten Verlag.

Kuhlthau, C. C. (2004). *Seeking Meaning: A Process Approach to Library and Information Services*. Westport, CT: Libraries Unlimited, 2nd ed.

L'Abbate, M., & Hemmje, M. (1998). *VIRGILIO: The Metaphor Definition Tool*. Sankt Augustin: GMD – Forschungszentrum Informationstechik GmbH, GMD Report 15, 47 p.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Landau, B. (2002). Spatial Cognition. In V. S. Ramachandran (ed.), *Encyclopedia of the Human Brain*. San Diego, CA: Academic Press, vol. 4., pp. 395–418. https://doi.org/10.1016/B0-12-227210-2/00326-5

P11 – 296

Langmuir, I. (1956). Freedom: the opportunity to profit from the unexpected. In C. G. Suits, H. E. Way (eds.), *The Collected Works of Irving Langmuir, Vol. 12: Langmuir, the Man and the Scientist*. Oxford: Pergamon Press, pp. 411–416.

Latour, B. (1987). *Science in Action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Laufer, L., Halacsy, P., & Somlai-Fischer, A. (2011). Prezi Meeting: Collaboration in a Zoomable Canvas Based Environment. *Proceedings of the ACM 1995 Conference on Human Factors in Computing Systems (CHI'11), 7–12 May 2011, Vancouver, BC, Canada*, pp. 749–752. https://doi.org/10.1145/1979742.1979673

Laurel, B. (2014). *Computers as Theatre*. Upper Saddle River, NJ: Addison-Wesley, 2nd ed.

Le Men, S. (1998). *La Cathédrale illustrée de Hugo à Monet: Regard romantique et modernité*. Paris: CNRS Éditions.

Lesk, M. (1997). *Practical Digital Libraries: Books, Bytes, Bucks*. San Francisco, CA: Morgan Kaufmann.

Liebert, W.-A. (1994). Lascaux – A Hypermedia Lexicon of Metaphor Models for Scientific Imagination. *Proceedings of the Sixth EURALEX International Congress on Lexicography, 30 August – 3 September 1994, Amsterdam, The Netherlands*, pp. 495–500.

Maglio, P. P., & Matlock, T. (1998). The Conceptual Structure of Information Space. In A. J. Munro, K. Höök, D. Benyon (eds.), *Social Navigation of Information Space*. London: Springer, pp. 155–173.

Mandelbrot, B. (2002). *Gaussian Self-Affinity and Fractals: Globality, The Earth, 1/f Noise, and R/S*. New York: Springer.

Marchionini, G. (1995). *Information-seeking in Electronic Environments*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511626388

Markham, A. (2003). Metaphors reflecting and shaping the reality of the Internet: Tool, place, way of being. *Paper presented at the conference of the International Association of Internet Researchers*, October 2003, Toronto, Canada. http://markham.internetinquiry.org/writing/MarkhamTPW.pdf (2019.10.15)

Markham, A., & Tiidenberg, K. (eds.) (2020). *Metaphors of Internet: Ways of Being in the Age of Ubiquity*. Bern: Peter Lang. https://doi.org/10.3726/b16196

Marshall, Ch., Nelson, C., & Gardiner, M. M. (1987). Design guidelines. In M. M. Gardiner, B. Christie (eds.), *Applying Cognitive Psychology to User-interface Design*. Chichester: Wiley & Sons, pp. 221–278.

Massaro, D. W. (1986). The computer as a metaphor for psychological inquiry: Considerations and recommendations. *Behavior Research Methods, Instruments, & Computers* 18 (2): 73–92. https://doi.org/10.3758/BF03201006

McCay-Peet, L., & Toms, E. G. (2012). The Serendipity Quotient. *Proceedings of the 75th Annual Meeting of the Association for Information Science and Technology (ASIS&T), 9–13 October 2011, New Orleans, LA, USA*, 49, 4 p. https://doi.org/10.1002/meet.2011.14504801236

McCay-Peet, L., & Toms, E. G. (2018). *Researching serendipity in Digital Information Environments*. San Rafael, CA: Morgan & Claypool. https://doi.org/10.1007/978-3-031-02312-5

McKay, D., Chang, Sh., Smith, W., & Buchanan, G. (2019). The things we talk about when we talk about browsing: An empirical typology of library browsing behavior. *Journal of the Association for Information Science and Technology* 70 (12): 1383–1394. https://doi.org/10.1002/asi.24200

Mehrpooya, A., & Nowroozzadeh, N. (2013). Metaphor-Laced Language of Computer Science and Receptor Community Users. *Journal of Technical Writing and Communication* 43 (4): 403–423. https://doi.org/10.2190/TW.43.4.d

Merton, R. K., & Barber, E. (2004). *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science*. Princeton, NJ: Princeton University Press.

Meyer, I. (2000). Computer Words in Our Everyday Lives. How are they interesting for terminography and lexicography? *Proceedings of the Ninth EURALEX International Congress on Lexicography, 8–12 August 2000, Stuttgart, Germany*, pp. 39–58.

Meyer, I., Zaluski, V., & Mackintosh, K. (1997). A Conceptual and Structural Analysis of Metaphorical Internet Terms. *Terminology* 4 (1): 1–33. https://doi.org/10.1075/term.4.1.03mey

Meyer, I., Zaluski, V., Mackintosh, K., & Foz, C. (1998). Metaphorical Internet Terminology in English and French. *Proceedings of the Eight EURALEX International Congress on Lexicography, 4–8 August 1998, Liège, Belgium*, pp. 523–531.

Milon, A., & Perelman, M. (eds.) (2012). *Le livre au corps*. Nanterre: Presses universitaires de Paris Nanterre. https://doi.org/10.4000/books.pupo.4271

Monmonier, M. (1991). *How to Lie with Maps*, Chicago, IL: Chicago University Press.

Mozilla developers (n. d.). "Firefox Developer Tools: 3D view". *Mozilla Developer Network*. https://developer.mozilla.org/en-US/docs/Tools/3D_View (2019.03.29)

Munro, A. J., Höök, K., & Benyon, D. (eds.) (1999). *Social Navigation of Information Space*. London: Springer. https://doi.org/10.1007/978-1-4471-0837-5

Noë, Alva. (2006). *Action in Perception*. Cammbridge, MA: The MIT Press.

Norman, D. A. (1990). *The Design of Everyday Things*. New York: Doubleday.

Nora, P., & D. P. Jordan (eds.) (2001–2010). *Rethinking France: Les lieux de mémoire*. Chicago, IL: University of Chicago Press.

O'Connor, B. (1988). Fostering creativity: Enhancing the browsing environment. *Interantional Journal of Information Management* 8 (3): 203–210. https://doi.org/10.1016/0268-4012(88)90063-1

OED (2019). *OED Online [Oxford English Dictionary]*. Oxford University Press. https://www.oed.com (2019.11.12)

O.M.A., Koolhaas, R., & Mau, B. (1995). *S,M,L,XL*. New York, NY: The Monacelli Press. https://oma.eu/lectures/s-m-l-xl (2020.01.27)

Ooi, K. (2008). How adult fiction readers select fiction books in public libraries: A study of information-seeking in context. M.A. thesis, Victoria University of Wellington, Wellington, New Zealand. http://hdl.handle.net/10063/1268 (2020.01.21)

Osenga, K. (2013). The Internet is Not a Super Highway. Using Metaphors to Communicate Information and Communications Policy. *Journal of Information Policy* 3: 30–54. https://doi.org/10.5325/jinfopoli.3.2013.0030

Overhill, H. (2012). J.J. Gibson and Marshall McLuhan: A survey of terminology and a proposed extension of the theory of affordances. *Proceedings of the 75th Annual Meeting of the American Society for Information Science and Technology (ASIS&T), 28–31 October 2012, Baltimore, MD, USA*, 49, 4 p. https://doi.org/10.1002/meet.14504901340

Peddie, J. (2013). *The History of Visual Magic in Computers: How Beautiful Images are Made in CAD, 3D, VR and AR*. London: Springer. https://doi.org/10.1007/978-1-4471-4932-3

Piggott, M. (2012). *Archives and Societal Provenance: Australian Essays*. Oxford: Chandos Publishing. https://doi.org/10.1533/9781780633787

PopularLibros.com (2010). *Did you know the BOOK? YouTube*. https://www.youtube.com/watch?v=iwPjoqgvfIs (2020.01.27)

Race, T. M., & Makri, S. (eds.) (2015). *Accidental Information Discovery: Cultivating Serendipity in the Digital Age*. Cambridge, MA: Chandos.

Rauber, A., & Bina, H. (2000). 'Andreas, Rauber'? Conference pages are over there, German documents on the lower left…": an "old-fashioned" approach to Web search results visualization. *Proceedings of the 11th International Workshop on Database and Expert Systems Applications, 4–8 September 2000, Greenwich, UK*, pp. 615–619. https://doi.org/10.1109/DEXA.2000.875088

Rauber, P. E., Fadel, S. G., Falcão, A. X., & Telea, A. C. (2017). Visualizing the Hidden Activity of Artificial Neural Networks. *IEEE Transactions on Visualization and Computer Graphics*, 23 (1): 101–110. https://doi.org/10.1109/TVCG.2016.2598838

Reinfurt, D., & Wiesenberger, R. (2014). *Messages and Means: Muriel Cooper at MIT*. Exhibition, 25 February – 28 March 2014, Arthur Ross Architecture Gallery, Columbia University, New York, NY.

Resmini, A., & Rosati, L. (2011). *Pervasive Information Architecture: Designing Cross-Channel User Experiences*. Burlington, MA: Morgan Kaufmann.

Resnikoff, H. L. (1989). *The Illusion of Reality*. New York, NY. https://doi.org/10.1007/978-1-4612-3474-6

Rice, E. R., McCreadie, M., & Chang, Sh.-J. L. (2001). *Accessing and Browsing Information and Communication*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/1066.001.0001

Richard, J.-P. (2019). *Shakespeare pornographe: un théâtre à double fond*. Paris: Éditions Rue de l'Ulm / Presses de l'École normale supérieure.

Roberts, R. M. (1989). *Serendipity: Accidental Discoveries in Science*. Hoboken, NJ: John Wiley & Sons.

Ross, C. S. (1999). Finding without seeking: the information encounter in the context of reading for pleasure. *Information Processing and Management* 35 (6): 783–799. https://doi.org/10.1016/S0306-4573(99)00026-6

Rovelstad, M. V. (1976). Open Shelves/Closed Shelves in Research Libraries. *College & Research Libraries* 37 (5): 457–467. https://doi.org/10.5860/crl_37_05_457

Samuels, M., & Samuels, N. (1975). *Seeing With The mind's eye: The History, Techniques and Uses of Visualization*. New York: Random House.

Sardar, Z. (1995). alt.civilization.faq: Cyberspace as the darker side of the West. *Futures* 27 (7): 777–794.

Schneider, B., Ernst, Ch., & Wöpking, J. (2016). *Diagrammatik-Reader: Grundlegende Texte aus Theorie und Geschichte*. Berlin: De Gruyter. https://doi.org/10.1515/9783050093833

Schoenherr, D. E. (2004). A Note on Burne-Jones's 'Pocket Cathedral' and Ruskin. *Journal of William Morris Studies* 15 (4): 91–93.

Schütte, S. (2005). Engineering Emotional Values in Product Design: Kansei Engineering in Development. PhD dissertation, #951, Department of Mechanical Engineering, University of Linköping, Sweden.

Searle, J. R. (1990). Cognitive Science and the Computer Metaphor. In B. Goranzon, M. Florin (eds.), *Artificial IntelligenceI Culture and Language: On Education and Work*. Berlin: Springer, pp. 23–34. https://doi.org/10.1007/978-1-4471-1729-2_4

Selfridge, P., & Kirk, Th. (1999). COSPACE: Combining Web Browsing and Dynamically Generated, 3D, Multiuser Environments. *Intelligence* 10 (1): 24–32. https://doi.org/10.1145/298475.298488

Small, D. L. (1999). Rethinking the Book. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.

Sparacino, F., Davenport, G., & Pentland, A. (2000). City of News: cataloguing the World Wide Web through Virtual Architecture. *KOS*, 9: 179–180.

Spink, A. (2010). *Information Behavior: An Evolutionary Instinct*. Heidelberg: Springer. https://doi.org/10.1007/978-3-642-11497-7

Spink, A., & Cole, Ch. (eds.) (2006). *New Directions in Human Information Behavior*. Dordrecht: Springer. https://doi.org/10.1007/1-4020-3670-1

Stefik, M. (1997). *Internet Dreams: Archetypes, Myths, and Metaphors*. Cambridge, MA: The MIT Press.

Stoffel, A., Kinnemann, H., Spretke, D., & Keim, D. A. (2010). Enhancing Document Structure Analysis using Visual Analytics. *Proceedings of the 2010 ACM Symposium on Applied Computing (SAC), 22–26 March 2010, Sierre, Switzerland*, pp. 8–12. https://doi.org/10.1145/1774088.1774091

Stoffel, A., Strobelt, H., Deussen, O., & Keim, D. A. (2012). Document Thumbnails with Variable Text Scaling. *Proceedings of the 14th Eurographics Conference on Visualization (EuroVis 2012)*, 5–8 June 2012, Vienna, Austria, 31 (3): 1165–1173. https://doi.org/10.1111/j.1467-8659.2012.03109.x

Strobelt, H., Oelke, D., Rohrdantz, Ch., Stoffel, A., Keim, D. A., & Deussen, O. (2009). Document Cards: A Top Trumps Visualization for Documents. *IEEE Transactions on Visualization and Computer Graphics* 15 (6): 1145–1152. https://doi.org/10.1109/TVCG.2009.139

Taramigkou, M., Bothos, E., Apostolou, D., & Mentzas, G. (2013). Fostering Serendipity in Online Information Systems. *Proceedings of the 2013 IEEE International Technology Management Conference (ITMC) and 19th International Conference on Engineering, Technology and Innovation (ICE), 24–26 June 2013, The Hague, The Netherlands*, pp. 1–10. https://doi.org/10.1109/ITMC.2013.7352707

Targowski, A. (2016). *The History, Present State, and Future of Information Technology*. Santa Rosa, CA: Informaing Science Press.

Tarter, J. C. (1984). SETI and Serendipity. *Acta Astronomica* 11 (7–8): 387–391. https://doi.org/10.1016/0094-5765(84)90079-1

Tavares, A. (2016). *The Anatomy of the Architectural Book*. Montréal & Zürich: Canadian Centre for Architecture & Lars Müller Publishers.

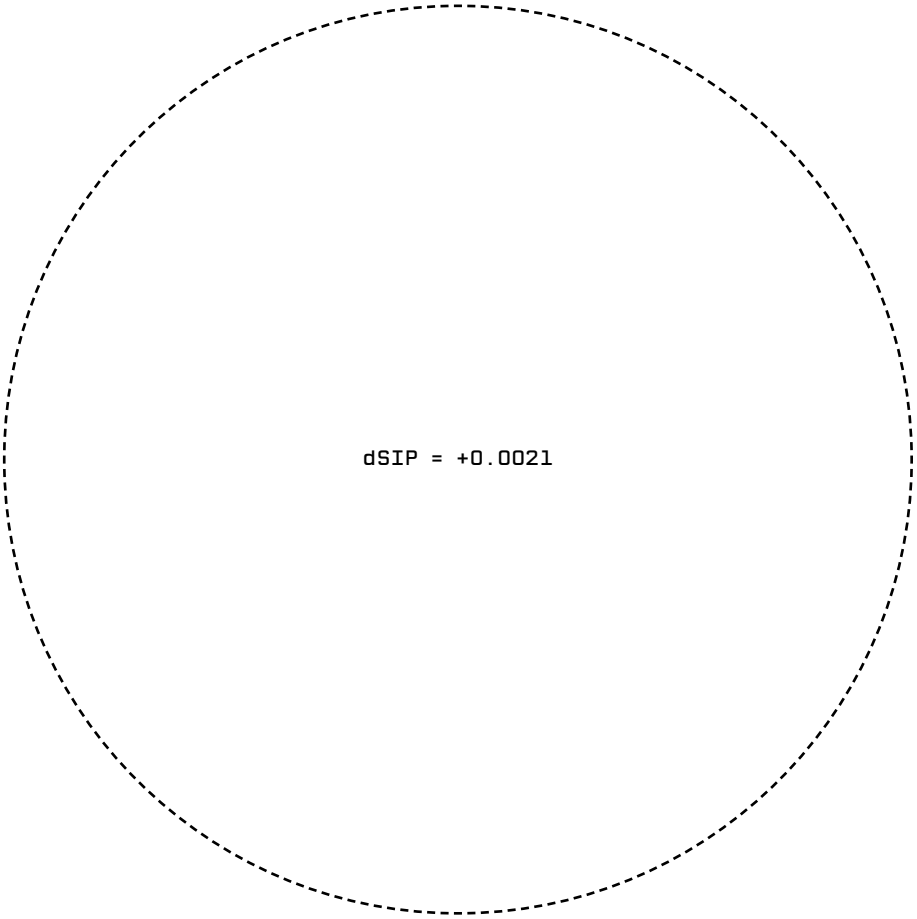
The British Library (2019). *The King's Library*. https://www.bl.uk/collection-guides/the-kings-library (2019.10.06)
Thudt, A., Hinrichs, Uta, Carpendale, Sh. (2012). The Bohemian Bookshelf: Supporting Serendipitous BookDiscoveries through Information Visualization. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12), 5–10 May 2012, Austin, Texas, USA*, pp. 1461–1470.
Tufte, E. R. (1997). *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphis Press.
Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
Turner, Ph. (2017). *A Psychology of User Experience: Involvement, Affect and Aesthetics*. Cham: Springer. https://doi.org/10.1007/978-3-319-70653-5
Varia (2000–). *Diagrams: International Conference on Theory and Application of Diagrams*. Cham: Springer. https://link.springer.com/conference/diagrams
Warde, B. (1956). The Crystal Goblet or Printing Should be Invisible. In H. Jacob (ed.), *The Crystal Goblet: Sixteen Essays on Typography by Beatrice Warde*. Cleveland, OH: The World Publishing Company, pp. 11–17.
Ware, C. (2008). *Visual Thinking for Design*. Burlington, MA: Morgan Kaufmann.
Weber, R. S. (1986). Information and Experience in Metaphor: A Perspective From Computer Analysis. *Metaphor and Symbolic Activity* 1 (4): 227–270. https://doi.org/10.1207/s15327868ms0104_1
Weizenbaum, J. (1966). ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine. *Communications of the ACM* 9 (1): 36–45. https://doi.org/10.1145/365153.365168
Wertheim, M. (2000). *The Pearly Gates of Cyberspace: A History of Space from Dante to the Internet*. New York: W. W. Norton.
Wikipedia contributors. ELIZA. *Wikipedia*. http://en.wikipedia.org/wiki/ELIZA (2019. 04.01)
Wikipedia contributors. List of web browsers. *Wikipedia*. https://en.wikipedia.org/wiki/List_of_web_browsers (2019.07.06)
Wikipedia contributors. Project Looking Glass. *Wikipedia*. https://en.wikipedia.org/wiki/Project_Looking_Glass (2019.10.01)
Wikipedia contributors. Siri. *Wikipedia*. https://en.wikipedia.org/wiki/Siri (2019.10.01)
Wikipedia contributors. Tamagotchi. *Wikipedia*. https://en.wikipedia.org/wiki/Tamagotchi (2019.10.06)
Wikipedia contributors. *The Librarian* (painting). *Wikipedia*. https://en.wikipedia.org/wiki/The_Librarian_(painting) (2019.10.06)
Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., & Crow, V. (1995). Visualizing the Non-Visual: Spatial Analysis and Interaction with Information from Text Documents. *Procedings of the First Information Visualization Symposium (InfoVis'95), 30–31 October 1995, Atlanta, GA, USA*, pp. 51–58. https://doi.org/10.1109/INFVIS.1995.528686
Witten, I. I., Gori, M., & Numerico, T. (2006). *Web Dragons: Inside the Myths of Search Engine Technology*. San Francisco, CA: Morgan Kaufmann.
Wittgenstein, L. (2009). *Philosophical Investigations*. Blackwell: Malden, MA.
Wozny, L. A. (1989). The application of metaphor, analogy, and conceptual models in computer systems. *Interacting with Computers* 1 (3): 273–283. https://doi.org/10.1016/0953-5438(89)90015-5
Yates, F. A. (1966). *The Art of Memory*. London: Routledge.
Yosinski, J., Clune, J., Nguyen, A., Fuchs, Th., & Lipson, H. (2015). Understanding Neural Networks through Deep Visualization. *Deep Learning Workshop, 32nd International Conference on Machine Learning (ICML), 6–11 July 2015, Lille, France*, poster presentation. https://arxiv.org/abs/1506.06579 (2020.01.27)
Zhang, J. (2008). *Visualization for Information Retrieval*. Berlin: Springer. https://doi.org/10.1007/978-3-540-75148-9

About the author

Vlad Atanasiu works in the computer science areas of document analysis, image processing, information visualization, and geographical information systems. He holds a PhD in Arabic Paleography, Art History and History (ÉPHÉ, Paris, 2003), an MA in Lexicology, and a BA in Middle Eastern Studies (Aix-en-Provence University, 1995, 1994). Vlad does freelance typographic work—graphic design, typeface creation, and calligraphy.

Email: atanasiu@alum.mit.edu

Publication the Twelfth


dSIP = +0.0021

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 6, 2021, accepted November 22, 2021, date of publication December 7, 2021, date of current version February 4, 2022.

Digital Object Identifier 10.1109/ACCESS.2021.3133654

The Structural Information Potential and Its Application to Document Triage

VLAD ATANASIU
Department of Informatics, University of Fribourg, 1700 Fribourg, Switzerland
e-mail: atanasiu@alum.mit.edu

This work was supported by the Document, Image and Video Analysis (DIVA) Research Group, Department of Informatics, University of Fribourg, Switzerland.

ABSTRACT This article introduces structural information potential (SIP), a measure of information based on the potential of structures to be informative about their content. An example of this concept is the clustered appearance that typically characterizes the first page of scientific articles, which summarizes the article's contents and provides additional data, yielding potentially the largest and most diverse amount of information from a single page in the shortest time with the least effort. This characteristic makes SIP particularly well-adapted to triage tasks (i.e., rapid decision-making under conditions of uncertainty and limited resources), an application illustrated by means of a case study on classifying document images. The SIP method consists in unifying the Shannon entropy, the Fourier transform, the fractal dimension, and the golden ratio into a single equation and several algorithmic components. While the application domain is document images, the concept has generic character. The method results in a mathematically and perceptually coherent pattern space, characterized by continuous transition between uniform, clustered, and regular configurations, and corresponding to a structural information potential with a well-defined maximum. The maximum SIP leads to the identification of shapes and patterns with minimal structural redundancy, termed ''fluorescent objects'' as a complement to regular graphs and the Platonic solids.

INDEX TERMS Information theory, structural entropy, spectral entropy, Fourier transform, fractals, graph theory, golden ratio, pattern analysis, image classification, document analysis, layout analysis, document triage, digital libraries.

Form is the visible shape of content.
— Ben Shahn

I. INTRODUCTION

This article introduces a new measure of information, the structural information potential (SIP). SIP defines information as the potential of structures to be informative about their nature and utility. A readily available example is the clustered appearance typical of the first page of scientific articles, such as this one; these pages summarize the article's content and provide additional metadata, thus yielding potentially the largest and most diverse amount of information from a single page in the shortest time with the least effort. The optimality of the data transmission rate, as influenced by the structure of the communication channel represented by documents, is also a critical factor in document triage, the main application studied in this article to evaluate the SIP measurement method.

The associate editor coordinating the review of this manuscript and approving it for publication was Khmaies Ouahada.

Code 1

```
function loopyloops(n); message = {'Hello world!',
'I like apple pi.'}; for loop1 = 1:length( message );
for loop2 = 1:n; disp( message{loop1} ); end; end;
```

Code 2

```
function loopyloops(n);
message = {'Hello world!', 'I like apple pi.'};
for loop1 = 1:length( message );
    for loop2 = 1:n;
        disp( message{loop1} );
    end;
end;
```

FIGURE 1. Demonstration of the structural information potential principle: While the above codes are identical both content-wise and functionally, the structured layout of code 2 facilitates faster analysis by programmers, enabling them to notice at a glance that there are potentially seven instructions and two nested loops.

In the document domain, the above observation is supported by compelling empirical evidence from the practice [1]–[4], theory [5]–[7], psychology [8]–[10], and history [11], [12] of document design (these references are

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

a small representative sample from a substantially larger bibliography). For example, much of the effort invested in microtypography [13], [14] (already an aesthetic quality and commercial factor in the early days of typography) concerns the spacing of characters (kerning) and words (justification), character ligatures, hyphenation rules (ladders), trailing paragraph lines (widows and orphans), and so on, for the purpose of producing visually homogeneous pages of text (a characteristic denoted as ''page gray'') [15], [16]. The principal motivations for this painstaking effort are the desire to keep semantic units visually grouped, as well as to prevent vertical streaks (rivers) from emerging due to chance alignments of spaces or similar letters, drawing the reader's attention towards spurious shapes devoid of content-related information (Fig. 2) [17]. Conversely, an intentionally clustered page pattern is the result of graphic designers arranging distinct informational units so as to augment hierarchical and typological distinctiveness; the goal is to improve legibility, speed up access to information, and guide the reader's gaze with minimal interference. This strategy for written communication is a product of evolution, a centuries-long shift away from homogeneous layouts driven by the increasing availability of written information. Clustered patterns can also emerge as a natural part of the document life-cycle; these may be introduced in the post-production stage either intentionally (e.g., by layers of annotations) or accidentally (e.g., due to physical degradation). In the absence of specific search goals or prior knowledge about the content, the most informative documents are those with clustered patterns.

P12 – 304

It is possible to generalize beyond documents, given that uniform and regular signals, images, objects, and events are in general less informative than structured entities. The paradigm resulting from these insights postulates a correspondence between informativeness, structure, a uniform–clustered–regular pattern continuum, scale-space filling, and structural redundancy. This article accordingly aims to devise a quantitative pattern description method for ordering patterns along said continuum, and further develops a conceptual framework to aid in identifying structures with minimal redundancy.

What this approach cannot provide is an estimation of information potential in the absence of structure, or a semantic content analysis. SIP is no substitute for text recognition and visual scene interpretation. Instead, its purpose is to characterize information at the level of structural organization.

Great strides in characterizing structural informativeness have been made in various scientific fields. However, the application of the proposed solutions to the task of classifying images, particularly text-based document images, has been found to be insufficient, as no prior approach has been able to satisfy both mathematical and perceptual desiderata. The present article substantiates this claim and presents a solution.

In essence, the proposed method characterizes a distribution in the scale-space domain with respect to the degree of redundancy. This characterization is achieved by unifying a number of classical concepts from the fields of information theory (Shannon entropy), pattern analysis (Mandelbrot's fractals), signal processing (the Fourier transform), combinatorics (the golden ratio), and graph theory (the chromatic concept) into a single analytic formula and several algorithmic components.

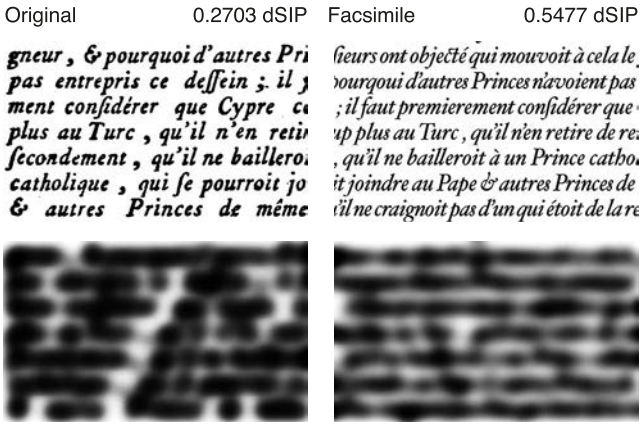


FIGURE 2. Example of spurious information emerging from a salient visual structure, and its detrimental effect. — Left: The smooth flow of reading this text is perturbed due to the reader's attention being reflexively diverted towards three asemantical structures resulting from chance diagonal alignments of white interword spaces (a typographical "river"), commas (an "island"), and the character q (a "ridge"). — Bottom: The perceptual effect is increased when squinting (here, simulated via Gaussian blur), which acts as a low-pass filter that makes the river into a salient structure in the visual field (a structure that is also large and has a different orientation than the text lines). — Right: The artifacts in the original document are removed by creating a facsimile where the spacing has been carefully adjusted to make the distribution of ink more homogeneous overall (typeset in Adobe InDesign). — Top: The SIP measure correctly reflects in a quantitative manner the observed difference in perceptual homogeneity between the "noisy" original (0.2703 dSIP) and the artifact-free facsimile (0.5477 dSIP). However, the lower dSIP value also suggests a higher structural information potential, which from the point of view of the text content is an illusion created by the typographical artifacts. Since these artifacts are structures absent from the facsimile, the original document is indeed richer in information, but of a kind other than linguistic. This information is, for example, useful for evaluating the typographical quality of the document and the evolution of this quality across time and space (accordingly, it may be valuable to e.g. historians and antique dealers). In this respect, the high prices of "incunabula" books published by famous 15th century printers, such as Aldus Manutius of Venice, are in part due to the high typographical quality of their products, including a remarkably even page "gray" that remains a model of exquisite quality even today. The role of this historical note is to explain through concrete examples the relevance of SIP for a broad range of applications. — Credits: Baron de Zur-Lauben, *Mémoires et letters de Henri Duc de Rohan, Sur la Guerre de la Valteline*, Geneva & Paris: Vincent, 1758, vol. 1, p. lxxxvii.

The application domain of the method is restricted in this article to images, more specifically to text-based document images. Given the generic nature of some of the core concepts, the concluding section considers its application to other media, such as video and three-dimensional data.

The relevance of the present article stems from the status of information measurement as a fundamental theoretical and practical issue across a broad range of scientific and technical fields. Its foremost contribution consists of a practical method to measure informativeness from structure, for application to images and possibly other data types. The elaboration of the pattern phase space and the identification of fundamental

shapes with minimal structural redundancy are further theoretical contributions. Finally, a survey of the quantification of irregularity links various fields, and places this research in a wider scientific perspective.

This article demonstrates a practical application of SIP to document triage via a real-life case study. The task in question is a type of classification similar to triage in emergency medicine, defined as *rapid decision-making for critical matters under conditions of uncertainty and with limited resources*. While triage cannot yield optimal solutions due to the constraints under which it operates, it is a useful and sometimes necessary step before more sophisticated procedures (such as semantic document analysis) are implemented. The proposed method fulfills the task in an explainable way, with reduced algorithmic complexity and assumptions.

Section I, ''Introduction'', has acquainted the reader with the concept of SIP and some of its empirical basis. Section II, ''Related work'', presents the advantageous and limiting factors of major existing approaches to information measurement, going on to explain the need for a new method while introducing elements used in the proposed method. Section III, ''Method'', is the theoretical core of the article, in which the analytical formula and algorithmic components of the SIP measurement method are described and justified; the design of structures with minimal redundancy is also discussed. Section IV, ''Experiments'', is devoted to the quantitative and qualitative evaluation of the proposed method through a case study in document triage, as well as to providing the exemplary discussion of a few other applications. Section V, ''Discussion'', concerns technical matters and future work. Finally, Section VI, ''Conclusions'', summarizes the theoretical and practical significance of the proposed method and its applications.

II. RELATED WORK

A. PATTERN IRREGULARITY

1) INFORMATION THEORY

Information theory has a central place among the domains relating pattern structures to informativeness. It is rooted in the *Shannon entropy*, H [18], which defines the amount of information, uncertainty, and choice, and quantifies it via the well-known equation $H = -\sum_{k=1}^{n} p_k \log_2 p_k$, where p_k is the occurrence probability of data class k from among n classes [19, pp. 393]. The normalized variant is given by the relative entropy, $H_r = H/H_{max}$, $H_{max} = \log_2(n)$ [19, p. 398]. This formulation produces a data distribution in which the extrema are on one hand the uniform distribution of values over all possible classes ($H_r = 1$), and on the other hand the concentration of values into a single class ($H_r = 0$). In the context of grayscale images, these correspond respectively to a uniformly monochromatic image and an image in which the number of present gray-level values equals the number of pixels. The application domain of this equation is nominal data, i.e., independent categories, such as the values of a fair dice. Therefore, this equation (and many of its variants) is not directly applicable to ordinal and structural data; given that the measure is independent of the configuration of sampling points (such as pixels in an image), it is not suitable for the goal pursued in this article.

Nevertheless, it is worth mentioning some notable entropy definitions as a way to exemplify the extent and vitality of research in this field, to point to research directions, to underscore the interdisciplinary character, and to identify ideas related to this article's topic. The Rényi and Tsallis entropies are generalizations of the Shannon entropy and define a parametrized family of entropies [20], [21]. Aiming at quantifying biodiversity, the field of ecology has contributed to the research on entropy with rich theories of diversity, as well as with formalisms, such as allowing for weighted probability classes, which express the similarity between species [22], [23]. From the field of nuclear physics comes the *strength of structure* [24, pp. 137–144] defined in 1939 by Satosi Watanabe (then a student of Werner Heisenberg) as J, the difference between the sum of entropies of u parts of a system, each containing v_u entities, and the entropy of the whole, containing $n = \sum_{i=1}^{u} v_{u_i}$:

$$J = -\sum_{i=1}^{u}\sum_{j=1}^{v} p_{ij} \log_2 p_{ij} + \sum_{k=1}^{n} p_k \log_2 p_k. \qquad (1)$$

This method has been utilized in document layout segmentation [25]. The inclusion of both parts and the whole in the characterization of structures is also important to the concepts addressed this article.

Approximate entropy (ApEn) is a measure of ''irregularity'' created to extend information entropy to structural data [26]–[28]. ApEn has been refined through many parametric [29], [30] and algorithmic variants [31]–[33] and compares well with other methods [34], [35]. From its inception onwards, it has been successfully applied to various biosignals [34], [35], and later to a broad range of other applications such as online signature verification [36], speaker recognition [37], radar jamming [38], earthquake prediction [39], and cryptography [28], [40]. In brief, ApEn is obtained by sliding a signal over itself, measuring the distance between the two within a window of given length according to the difference between the maxima, and computing a logarithmic average; this process is carried out for two different window lengths and the irregularity index is obtained as the difference between these partial results. For a finite discrete signal, there exists both a lower and upper bound; this is zero for perfectly periodic patterns, with higher values indicating greater irregularity. One very interesting aspect of this measure is that it is possible to compute number sequences with maximal irregularity (this upper bound is neither white noise nor a deterministic fractal) [41], [42]. The method may be extended to the analysis of images using vectorized bidimensional windows [43], [44]. This approach, in addition to improved parametrization [29], was used in this article to process document images. The results are discussed in Section IV, ''Experiments'', and illustrated in Fig. 15. While clustered documents are classified at one end of the spectrum,

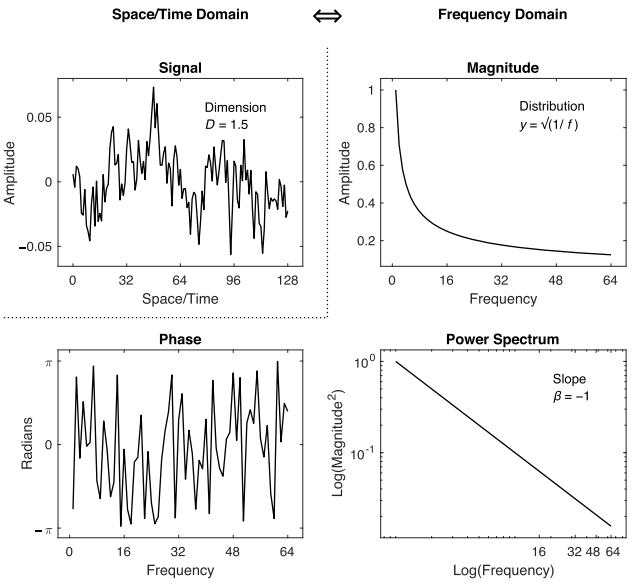


FIGURE 3. Sample fractal with dimension $D = 1.5$ represented in the space/time and frequency domains. The signal was synthesized from the magnitude and phase via inverse Fourier transform. The signal has the characteristics of $1/f$ noise, a power law distribution visible in the shape of the magnitude and confirmed by the linearity of the power spectrum on log–log scales; the phase is random. Only half of the symmetrical spectrum is displayed. The DC is set to zero, hence the zero mean amplitude of the signal.

P12 – 306

as desired, a mix of uniform (predominately empty) and homogeneous (predominately text) patterns appear at the other end, which is undesirable, since these patterns differ both in terms of appearance and information potential. This counter-intuitive behavior has been explained in the past by the observation that ApEn is a measure of irregularity rather than complexity [45].

2) FRACTALS

Fractals [46] are deterministic or stochastic self-similar or self-affine mathematical objects with an interesting property from the point of view of information theory: their clustered structure, which results from their infinite filling of the scale-space, maximizes their information potential. This is one of the reasons why fractal-like structures abound in nature; for example, the energetic intake of plants is optimized via the organization of branches and leaves around stems according to a power law [47], [48].

The dimension of a fractal is a fractional value, D, bounded by the fractal's topological and Euclidean dimensions, $D_T < D \leq D_E$. E.g., $1 < D \leq 2$ for signals, and $2 < D \leq 3$ for images. One of the most reliable methods for determining the fractal dimension uses spectral analysis and defines it as the slope, $\beta < 0$, of the linear fit of the log-power spectrum vs log-frequency: $D = (c + \beta)/2$, where $c = 6$ for an image, and $c = 4$ for a signal [49], [50, pp. 54, 97–114]. Two requirements must be satisfied if the measured entity is to be deemed a fractal: namely, the phase must be uniformly random [49, pp. 99] and the power spectrum must follow a power law, f^{α}, assuming $\alpha = \beta$ [51]–[53] (Fig. 3). By definition, this is not the case for non-fractal structures that are nevertheless clustered, and even less so for other pattern types along the uniform–clustered–regular continuum. The use of the fractal dimensions is therefore not appropriate for use in characterizing patterns of such a broad spectrum. However, the concept of fractality does provide a useful framework for thinking about SIP, especially as concerns its maximal value.

3) SPECTRAL ANALYSIS

Spectral analysis is useful for the characterization of clusteredness. This is because, unlike spatial entropy, the frequency domain captures spatial organization in a compact manner that is amenable to mathematical manipulation, and also captures scale variation (in a similar way to the fractals). Two popular spectral analysis methods are spectral entropy and spectral flatness.

Spectral entropy, H_s, is [54] defined as the Shannon entropy of the probabilities, P, associated with the frequency components of the power spectrum, S, given by the discrete Fourier transform, $\mathcal{F}$, of the data, X, of length n:

$$S = |\mathcal{F}(X)|^2, \quad P = S / \sum_{i=1}^{n} S_i, \; H_s = -\sum_{i=1}^{n} P_i \log_2 P_i. \qquad (2)$$

Variations of the definition include obtaining the power spectrum from the discrete cosine transform or by way of autocorrelation, along with using entropies other than Shannon's. Understanding the properties of these analytical expressions so as to be able to thoroughly explain their effects on empirical data remains an active research field [55]. This is especially critical for the analysis of biomedical data, a major application domain of spectral entropy [56]–[58], where it is used in particular for the clinical interpretation of EEG signals (e.g., for monitoring depth of patient sedation during anesthesia [59]). Audio signals analysis (e.g., urban soundscape classification [60], dolphin whistle segmentation [61], abnormal milling sounds detection [62]) and speech analysis (e.g., speaker identification [63], noise quality assessment [64]) are other common application domains. Spectral entropy has been less frequently applied in image processing, but is used for image quality assessment [65], scene saliency analysis [66], and camera focus estimation [67]. The common goal is the need to distinguish between regular and irregular patterns, where the latter are usually of interest, a task for which spectral entropy has been found useful. The problem remains that ''irregularity'' has many possible formal definitions, and means different things for different data types, tasks, and contexts; moreover, the spectral entropy equation has its own peculiar effects on the data, with the interaction between the two not always being well understood. When applied to image classification, for example, it results in a mix of homogeneous and uniform patterns (see Section IV-A, ''Comparison Of Methods'').

Spectral flatness, or *SF* [68]–[70, pp. 112–115], is a widely used measure of signal structuredness [71]–[73] (e.g., as an

audio descriptor in the MPEG-4 file format [74]). It is defined as the ratio of the geometric and arithmetic mean of a signal's power spectrum, S, taken over its n frequency components:

$$SF = \frac{(\prod_{i=1}^{n} S_i)^{\frac{1}{n}}}{\frac{1}{n}\sum_{i=1}^{n} S_i} = \frac{\exp\left(\frac{1}{n}\sum_{i=1}^{n} \log_2 S_i\right)}{\frac{1}{n}\sum_{i=1}^{n} S_i}. \quad (3)$$

Its principal utility comes from the bounds of the expression being zero for an impulse and one for a uniform distribution in the frequency domain, which correspond in the spatial domain to a regular signal and a impulse, respectively. *SF* relates to SIP in that both have the same bounds. However, just like spectral entropy, the spectral flatness also results in a perceptually unsatisfactory classification of images (Fig. 15).

Point set analysis and *spatial statistics* are concerned with the characterization of the distribution of point-like objects and events in space [75]–[77], while the related field of *discrepancy theory* deals specifically with characterizing the irregularities of distribution [78], [79]. Geographical information systems, ecology, astrophysics, and material science are among the typical application domains. We will herein briefly focus on the latter, as it bears a direct relation to both the topic and methods of the present article. Based on the empirical observation of the structure of physical matter, the continuous pattern space extending from uniform to clustered to homogeneous has been identified as a useful classification concept in material science to characterize such properties as surface roughness and particle dispersion. Much effort has therefore been invested in quantifying these patterns, with some of the classical methods being based on the nearest-neighbor distribution, morphological operations (dilation followed by counting), and Dirichlet tessellation [80]. To date, some of the best-performing approaches rely on spectral fractal analysis, using variations of the methods described in the preceding paragraphs [81] [82, pp. 81–98] [49, p. 108]. The utility of the fractal paradigm has however been called into question by practitioners on practical grounds, due to the difficulty of measuring fractality, as well as on theoretical grounds pertaining to its appropriateness as a model of the observed data [83] [49, p. 109]. In summary, we note that spectral analysis is a powerful method of characterizing structures, that better methods are required, and that methods and data must be compatible.

4) GRAPH THEORY

Graph theory can be applied to model discrete patterns, such as the individual pixels of digital document images, as well as the visual and logical entities of document layouts [84]. A dynamic sub-field studies irregular or color graphs, whose properties derive from the value of their edges; for example, a rainbow graph is one with distinct edge values [85], [86]. The topic relates to this article not only because of its focus on pattern irregularity (as opposed to regular structures, such as the Platonic regular bodies), but also because it deals with determining maximal irregularity, which the approaches reviewed above have yet to achieve. After an extensive literature survey (see below), however, it was not possible to find previous work on maximal irregularity applicable to patterns such as document images, even for basic shapes (such as the triangle). Furthermore, another well-known practical element makes the graph theory approach to pattern classification problematic: the degree of computational complexity for data with millions of sampling points, such as document images, creates high computational costs, especially in comparison with other methods such as spectral analysis.

Survey — We commence by stating here that our goal is the study of *the irregularity of continuous chromatic graphs*, and then proceed by commenting on the aesthetic dimension of this mathematical research, which also serves to introduce the notions vehiculated by the terminology. We next expose the historical origins of chromatic graph theory [87] and the interdisciplinary strands from which the modern research on irregular graphs emerged; finally, we conclude by discussing some particularities of our research in respect to graph theory.

There exists a little theorem that all it says is: ''No graph is irregular'' [88, p. 25] [89, p. 24] [85, pp. 36–37]. What it means is that there exists no graph with two or more vertices for which the vertices have distinct numbers of adjoining edges (i.e. distinct degrees). The simplicity of this statement is exquisite, moreover, and perhaps unsurprisingly, its corollary—that every graph with two or more vertices has at least two vertices with the same degree—has been voted one of the twenty most beautiful theorems of all time [88], [90], [91, p. 25]. Even the field of chromatic graph theory [92] to which these two theorems belong is not bereft of metamathematical charms, which exude from the colorful terminology employed to designate various types of graphs. For illustrative purposes, the author composed out of graphical terms the following pangram (a kind of linguistic color graph, in which all letters of the alphabet must appear): ''Graph Euler cycles over the rainbow to join the flower snarks in a zero-vertex quasi-bramble.''

The question has been asked as to whether it is possible to color a cartographic map with no less than four colors, such that no two adjacent countries have the same color. In the form of this Four-Color Problem, graph coloring famously made its public debut on the 23th of October 1852 as an intriguing mapmaking problem among British mathematicians, albeit in the absence of any specific request from Her Majesty's Ordnance Survey [93, pp. 1–26]. Its impact would be profound, with ''many of the concepts, theorems, and problems of Graph Theory'' being said to ''lie in the shadows of the Four-Color Problem'' [93, pp. 2]. The study of graph irregularity adds new layers to the already consequential applications of coloring (e.g., timetabling, sequencing, and scheduling [94, p. xv]), such as drug design [95, p. 600] and secure intelligence and military communication networks [93, p. 76] [94, p. v–vi]. The topic remains, however, outside of the mainstream: in a hefty 1633-pages handbook on graph theory published in 2014 the word ''irregular'' occurs only

once [96, p. 1032], and not at all in many other books of its kind.

A series of fundamental problems in *combinatorial geometry* and *geometric graph theory* were published in the 1930s and 1940s, dealing with the number of distinct colors that color the plane in such a way that no two blocks of the same color are at unit distance [94], [97]–[100]. The research grew around what became known as the Hadwiger–Nelson problem of the chromatic number of the plane, providing nourishing theoretical ground for the later development of chromatic graph theory and graph irregularity. The earliest trace of *irregular graphs* identified by this author dates back to Germany during the Second World War [101, p. 76–77]. The concept later resurfaced independently in the late 1980s in the work of the American-Iranian mathematician Yousef Alavi, and was elaborated by his colleagues at Western Michigan University, including Paul Erdős himself [102, p. 235]. Their approach to the problem is both astute and unintentionally revealing about the nature of mathematics. Rather than asking ''what *is* an irregular graph?'', their question is ''what *could* an irregular graph be?'', or in the slightly more normative verbatim: ''How *should* one define an irregular graph?'' (emphasis added). They followed up by stating their expectations: ''In research, the goal is not only to come up with a definition that seems natural but to arrive at a class of graphs with interesting, and perhaps even some surprising, properties.'' [85, p. 39]. These propositions seem to advance the mutually exclusive views of mathematics as an activity of *invention* (the ludic and aesthetic motivations are explicit), and as one of *discovery*, in line with Erdős' concept of an immanent Book of mathematical proofs (''Mathematics is there.'' [103], [104, p. 27]). In the end, their answer is a multiplicity of definitions of graph irregularity, with various degrees of ''interestingness'' and ''surprise''. One of the researchers, Gary Chartrand, would subsequently coin the term ''rainbow graph'' [93], [105], [106], labor extensively in this research field [89], [105], [107], [108], popularize graph irregularity [88], and point to its practical usefulness [93].

P12 – 308

In Yugoslavia, meanwhile, members of the Zagreb Mathematical Chemistry Group developed measures of molecular complexity based on their own indices of graph irregularity, the *Zagreb indices*, which remain a mathematical staple among chemists [109]–[113]. Only in the late 1990s did the three research strands converge [114]–[119, p. 45]; after this point, the focus begins to shift towards special topics, such as determining the maximal and minimal irregularity bounds [114], [120], [121], local versus global irregularity [122], irregular graph assignments [123], and many more [124].

The measure explored by this author differs in an essential way from previous graph irregularity concepts. The Zagreb indices are variously the sum of the squared vertices degrees, the sum of products of adjacent vertices degrees, or the sum of absolute differences of adjacent vertices degrees. In other words, they represent the valence of entities, and are therefore suited to applications in chemistry, especially given the prevalence of irregular molecular graphs of interest to chemists [118, p. 222]. The rainbow measure is given by the number of distinct colors in a graph or along specific paths, which is appropriate for routing problems [107], [125], [126]. In both cases the actual distances between vertices are disregarded and the graph's topology is one of distanceless connections. There are instances, however, where edges represent not discrete label classes, but *continuous distance magnitudes*, notably in pattern and shape analysis. This is the case and particularity of the research presented in this article.

Note: Graphs embedded in the Euclidean space are *geometric graphs*, those with other topologies are *topological graphs* [99, p. 465], and both are *distance graphs* [100, pp. 429–430]. Graphs optionally embedded in a space and with values attached to the edges are *chromatic graphs* [87]. Colors may represent both *continuous* values and *categories*. We will say therefore that we study the *ir|regularity of continuous chromatic graphs*.

5) OTHER FIELDS

For the sake of completeness, it is worth mentioning other important fields that deal with irregularity (in the graph theoretical sense) and redundancy (in the information theoretical sense): *combinatorics* (particularly *combinatorial geometry*) [46], [127, pp. 78–81], *packing* [128], *tiling* [129], *tessellation* [130], *coding* [131], [132], and *randomness* [133]–[136].

Details — Irregularity can be conceptualized from many different perspectives. In its simplest form, line segmentation, it can be approached from the points of view of graph theory, point spread, and combinatorics. It is indeed an old problem of *combinatorial geometry* [137], counting among its earliest investigators the Polish mathematician Wacław Sierpiński (1882–1969; cryptanalyst during the First World War and later father of the eponymous ''gasket'' and ''carpet'' fractals [138, pp. 78–81]) and his Swiss-Russian mentee Sophie Piccard (1904–1990; author of a monograph on collinear point sets [127], [139], [140]). The combinatorial aspects of point sets were also a long-standing interest of Paul Erddős [141]–[145]. The absence of mentions of maximal dissimilarity in his publications [146], [147] leaves open the possibility that ''the prince of problem solvers and the absolute monarch of problem posers'' [148, p. 64] may have considered the topic a pleasant surprise, worthy of *The Book*, the imaginary object where all mathematical proofs are kept.

In the fields of *tessellation* and *tiling*—of interest to computer graphics, surface modeling, and material sciences, for example—there is a tangential interest in the irregularity of these planar graphs, albeit typically with the aim of suppressing it so as to achieve maximal regularity [129], [130].

Discrepancy theory is another field that appeared an auspicious match with the search for a model of pattern irregularity, considering its generic framing as ''a measure of the deviation of a point set from a uniform distribution'' [149], [150] [78, pp. 1–3]. With roots in the 19th and early 20th centuries, the mathematical theory of discrepancy has experienced an increase in research interest over

the last decades (it has its own conferences and journal [151]) thanks its many links to other areas of mathematics (e.g. number theory, graph theory, and Ramsey theory), as well as the development of instruments fundamental to computer science (such as for numerical integration, Monte Carlo simulations, or randomization), which find applications in computational physics, computer graphics, mathematical finance, and cryptography, among many others [152, pp. xi–xiv] [78, pp. vii–viii, 22–35] [153, p. 1]. There exist a host of methods for computing discrepancy, including utilizing spectral analysis, be it the Fourier or the Haar transform [79, pp. 621–678] [78, pp. 213–240]. For practical reasons related to ease of computability, the most widely used method is the corner method or L_2-discrepancy, which is based on the number of points falling within rectangles anchored at each of these points and the bounding box of all points (for details see [78, pp. 2–3, 10, 12–16] [79, pp. 623–624]). Unfortunately, the pattern spectrum defined by the discrepancy measure differs from the perceptual spectrum uniform–clustered–regular targeted in this article and defined by the structural information potential. For example, maximum regularity in the sense of discrepancy is not a pattern of equally spaced points (a triangular lattice), as for SIP, but rather one with inhomogeneous density [79, p. 3].

Given the importance of the golden ratio and the most irregular triangle for the structural information potential, it was anticipated that some references would also be found in the plethoric literature on the *golden ratio* [154]–[157] and *triangle geometry* [158]–[161], both of which also bear some relevance to graph theory. One tantalizing possible avenue that might lead to the discovery of new mathematical visions is that of looking into cultures with limited global contact, such as pre-modern Japan, when the distinctive *sangaku* geometry was developed in temples [162]. This search, however, was less than fruitful.

The one domain that is almost pathologically fascinated by structural irregularity is that of the *arts* and *architecture* (Fig. 4). East Asian calligraphy, for example, is theoretically grounded in a spectrum extending from almost mechanical regularity to highly irregular brush patterns, while contemporary Western typography is built on the visual tension between homogeneous and hierarchical layouts. Even the movement of people within movie frames may be analyzed in terms of dynamic changes in graph irregularity, as a means of expressing meaning and emotion (the sleek geometrical compositions of Michelangelo Antonioni's films are good examples). Architecture makes irregularity concrete, such as in the fractal geometry of Islamic *muqarnas* (stalactite-like wall decorations) and the irregular volumes of the villas designed a century ago by Ludwig Wittgenstein and Le Corbusier (the creator of the golden ratio-based Modulor building module), and nowadays with the aid of computers by Zaha Hadid and Frank Gehry.

FIGURE 4. A photographic study of natural irregularity, in which a visual framing is found such that the pairwise length between the three rocks, along with their size differences, is maximized. The theme is reminiscent of similar experiments related to the aesthetics of perspective conducted by the American photographer Ansel Adams, as well as the spatial organization of traditional Japanese gardens. — Credits: Vlad Atanasiu, 2016, Golden Gate Bridge, San Francisco.

6) MACHINE LEARNING

This section addresses the possibility of using machine learning for classifying documents on the uniform–clustered–regular pattern spectrum and determining graphs with minimal redundancy.

Classification — If the goal is to solve the classification problem at hand, rather than to use a specific method to solve it, then machine learning is not necessary since we have already identified a solution that does not require learning. Machine learning may however be useful if it could demonstrate certain advantages over the classical approach. The author searched, but could not find a suitable solution of this kind; on the contrary, a number of issues were instead identified that have the potential to diminish the classification quality and complicate the solving process.

The first step of learning—constituting a representative dataset—is already problematic. Since the concept of structural information potential (SIP) is applicable to any kind of pattern (e.g., time series, images, movies, three-dimensional objects, and social networks), in any dimension, the typological diversity and size of the training and testing dataset is at the limit of practicality. Even when restricting ourselves to the domain of documents, there is a significant diversity of document types, including types of noise, which represent real possible impediments to knowledge generalization.

The second learning step—producing the groundtruth—is expensive and possibly suboptimal. It is expensive because the classification examples to be emulated by the machine must be generated by humans; this will result in small dataset sizes, which may reduce the learning quality. Additionally, the machine would be trained on human behavior, but physical and perceptual pattern measurement are not identical.

Moreover, the influence of perceptual variability across a wide range of human factors (e.g., age, gender, culture, motivation, human–machine interfaces) would need to be assessed and accounted for.

Minimal redundancy — The contribution of this article is not only a practical method of pattern classification, but also a theoretical work on defining minimal structural redundancy. Of course, emulation and trial-and-error are fundamental learning methods; however, there were no preexisting examples to emulate in the minimal redundancy definition process, nor a set of reward-and-penalty criteria to use. Instead, the author developed a rational argument about what ought to be a solution that is both mathematically provable and useful in practice. It was through a cognitive argumentation process, rather than learning, that it was possible to define the SIP value with maximal information potential, which is an important component of the SIP theory, the SIP value with maximal information potential, which is an important component of the SIP theory, the SIP measurement method, and the uniform–clustered–regular pattern spectrum.

7) CONCLUSIONS

As a concluding remark, it can be stated that a common goal of the reviewed perspectives is the characterization of structuredness, as distinct from uniformity, regularity, and randomness, and under the moniker ''information,'' as a proxy for the utility that can be derived from pattern analysis. The goal is thus doubly defined in terms of methods and applications. The review has highlighted how a thorough understanding of the empirical application domain supports the development of successful theoretical methods. This insight is reflected in the work reported herein, particularly in the description of the design, psychology, history, and life cycle of documents that has shaped the theory of structural information potential.

P12 – 310

B. LAYOUT-BASED DOCUMENT TRIAGE

The classification of documents with respect to their informativeness is a task common to a number of applications, notably document overview, retrieval, summarization, and presentation. The problem becomes more critical as the number of documents to be processed increases (e.g., within organizations, libraries, and archives), but is equally relevant for a single document, such as when browsing a digital book. Informativeness is a relational quality, in that it depends on both the stimulus and the observer. In cases where prior knowledge about the user is lacking, or where there are many users with heterogeneous interests, a user-independent estimation of document informativeness is necessary. An additional constraint is the speed with which users may acquire the information presented to them. From a statistical perspective, the most commonly applied solutions rely on some form of data summarization and sampling.

Summarization consists in beginning with a given document set or item and synthesizing a new, more compact one that retains the characteristic features of the original. For text documents, this typically involves removing text chunks so as to reduce the overall redundancy and fit the text in a smaller spatial frame, producing something like a more or less extended abstract or even a title [163], [164]. Image summarization is conceptually similar but more difficult to realize; for example, representing a person by her or his face and a wood by a single tree [165], [166], removing empty areas from document pages [167], replacing a color image with a black-and-white line art sketch, or reducing an image to its dominant colors [168]. Documents representing three-dimensional data (e.g., buildings or landscapes) and multimodal documents (e.g., videos) are far more complex to represent compactly (a high-quality movie trailer goes beyond simple cut-and-paste; it is an artistic project in itself) [169], [170].

This article's contention that spatial organization is informative has also been applied to document summarization. For example, this author has introduced the Document Towers visualization paradigm, which represents the three-dimensional structure of bounding boxes of paragraphs, images, and other entities in paginated documents as architectural wire-mesh models that resemble buildings and cities [171]. The quantified structural information potential is encoded as a color-coded ''ribbon'', which allows users to take stock of features such as document fragmentation, regularity, and outliers without opening the document itself. In addition to facilitating overview and navigation, this information enables document type, quality, and other insights to be inferred, often serendipitously.

Sampling differs from summarization in that it does not create new entities but instead aims to identify a limited amount of existing document parts that are representative of the whole. A further difference can be compared to the distinction between statistical expectation and the probability density function: while summarization often ends up presenting the average content or the most informative (e.g., the table of contents), sampling may provide the full range of content types (e.g., cover, text, figure, index). Semantic layout analysis is therefore a dominant method used to sample document images, as it benefits from not requiring character recognition (a substantial argument in support of its performance and quality, particularly given that noisy, historical, and/or handwritten documents still present challenges for this process) [172], [173]. Another sampling approach is pattern-driven; for example, a handwriting dataset can be compactly represented by a few ''vantage point'' samples [174]. This is a classical pattern classification and clustering problem, resulting in an ordering specific to an application domain and dataset. A good example of the complexity of defining interestingness is given in [9], where the interaction of factors as diverse as color, layout, content, reading ergonomy, and readership is analyzed. The contrast along multiple dimensions in terms of types and number of entities between neighboring document pages has also been used as a criterion for informativeness, in the framework of Shannon's information

theoretical definition of information as the amount of ''surprise'' [175]. The work presented in [176] is the closest to a pure pattern-based measure of document informativeness such as that described in this article. Although it aims at being a fast, simple, and approximate method (in the spirit of the triage task), it does not address spatial organization. Instead, page informativeness is quantified as the degree of (chromatic) saturation and (achromatic) lightness computed over the connected components of binarized pages and weighted by their size.

This succinct survey reveals how the quantification of informativeness has been approached at various points on the spectrum, ranging from pattern to semantic to contextual analysis. However, the quantification of spatial organization and the derivation of insights therefrom remains a fruitful research direction for computational document analysis.

III. METHOD

The empirical insights expounded in the introduction suggest that pattern informativeness varies along a uniform–clustered–regular continuum. The goal of this section is to introduce a quantitative description method, namely the structural information potential, which facilitates the ordering of patterns along this continuum. A discussion of the correspondence between maximal SIP and minimal structural redundancy is also presented, as the latter provides a theoretical foundation for the former.

A. STRUCTURAL INFORMATION POTENTIAL

We begin by introducing the SIP equations, then explain the general rationales behind them, after which we discuss each element of the procedure in detail.

1) CORE PROCEDURE

The core mathematical machinery of the structural information potential measurement consists of the Shannon entropy of the logarithm of the power spectrum of a binary image's two-dimensional Fourier transform. Hence, it is a mixture of information-theoretical concepts and signal processing, with a fractals perspective and properties of the golden ratio playing also a role, as shall be seen. Formally, the value $\mathrm{SIP} \in [0, 1]$ is obtained through a number of equations and algorithmic steps:

$$\mathrm{SIP} = 1 - |\mathrm{dSIP}|, \tag{4}$$

$$\mathrm{dSIP} = T\left(H_r\left(\log_2\left(1 + V\left(|\mathcal{F}(B(I))|^2/n\right)\right)\right)\right), \tag{5}$$

where I is an intensity image, B a binarization process, $\mathcal{F}$ the image's Fourier transform [177], $|\cdot|^2$ the power spectrum (S), n the number of image pixels, V a vectorization algorithm for the input matrix (described below), H_r the relative Shannon entropy, T a transfer function used for convenience purposes (also described below), and $\mathrm{dSIP} \in [-1, +1]$ the divergence from maximum SIP. The utility of dSIP is to concomitantly provide information about the intensity of the information potential as given by the absolute value, |dSIP|, and locate the pattern type on the uniform–clustered–regular continuum, made explicit by the signum, sgn dSIP. For the sake of simplicity, we will use the same notation (''SIP'') to denote both the concept of structural information potential and its instantiation in a particular mathematical expression; when necessary, a qualifier will provide clarification.

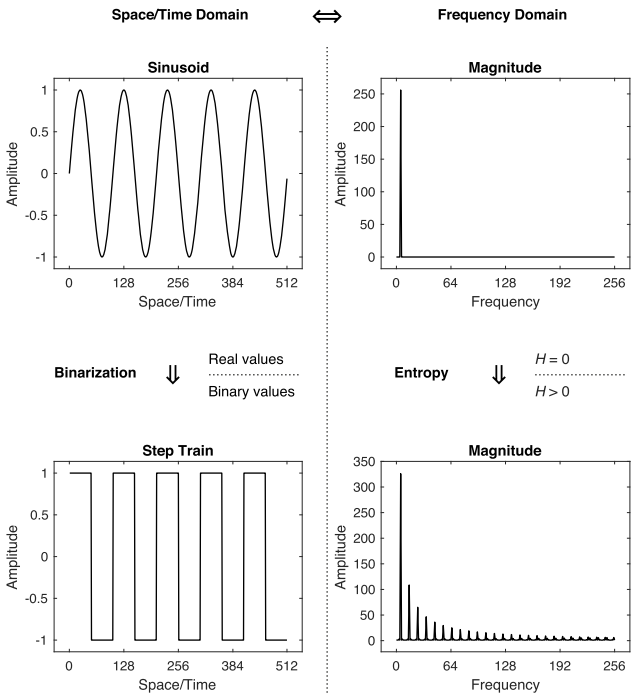


FIGURE 5. Illustration of the effect of binarization of a pure sine signal on the Shannon entropy, H, of its frequency domain magnitude.

2) FREQUENCY DOMAIN AND ENTROPY

The choice of the frequency domain for pattern analysis has a number of benefits [178]: it facilitates the integration of sample points across space, and thus the description of spatial structures in terms of regularity; it enables characterization of the degree of clustering via the scale-space spectrum of frequencies; finally, it allows for a pattern phase space to be obtained for which the extreme values correspond to regular and uniform patterns in the spatial domain. (Recall that the magnitude of the Fourier transform does not encode the spatial location of structures, but rather contains information about their form. For example, the magnitude represents the frequency and amplitude of the sine wave, but provides no information regarding the location of the local maxima; this information is carried by the phase component of the Fourier transform. From a SIP perspective, however, it is the form and not the location of a pattern that is of primary concern.)

The use of the logarithm of the power spectrum stems from the method (outlined above) for determining the fractal dimension and is intended to allow the characterization of clustered patterns. Normalizing the power in equation 5 makes it independent of image size; adding one to the squared magnitude avoids the logarithm of zero and negative output

values. To put SIP in a fractal perspective, SIP is a measure of how far a pattern is from potentially being a fractal.

In addition to having a dimensionality reduction effect, the role of the Shannon entropy is to help devise a linear space in which patterns are ordered from uniform to clustered to regular. This can be achieved if a transform is found such that the extrema of the pattern space correspond to a vector of zeros except for one value (i.e., an impulse) and a vector with equal values (i.e., uniform), respectively; in that case, the entropy will range from zero to one.

3) BINARIZATION

A uniform signal in the time domain has as its pair in the frequency domain an impulse with frequency of zero [179, p. 33]. As this frequency is ignored in the SIP measurement method, the Shannon entropy value of the remaining power spectrum will also be zero, as desired. Considering that the Fourier transform uses sines and cosines as base functions, the sinusoid is the regular pattern with the fewest spectral artifacts (e.g., aliasing, harmonics, Gibbs effect [180, pp. 194–200, 218–222]). It becomes an impulse in the frequency domain [179, p. 29], and thus has zero entropy and represents an undesirable outcome for our goal. However, by binarizing the sinusoid, a step train is obtained, which corresponds in the frequency domain to a set of harmonics of the form $1/(f\,\pi), f \in \mathbb{N}_{\text{odd}}$ [181, pp. 102–113] [180, p. 257], the effect of which is to increase the signal's spectral entropy (Fig. 5). A slightly different, but relevant, regular signal is the pulse train, where the step length differs from the distance between consecutive steps. As the spatial length of the pulses expands and the pattern becomes more uniform within the finite signal bounds, the power distribution is increasingly compressed towards the lower frequencies [180, p. 201] and the spectral entropy decreases, as desired.

In summary: (*a*) we translate the problem of devising an analytic formulation of the pattern–informativeness space in the frequency domain in order to be able to characterize spatial structures; (*b*) we employ the entropy because its extrema are the impulse and the uniform distributions; (*c*) we apply data binarization to transform the spectral representation of regular patterns from the impulse to the uniform distribution, so that the spectral entropy yields the desired uniform–clustered–regular continuum. Taking advantage of the spectral artifacts introduced by binarization is key to obtaining measurements of the spectral entropy that order patterns in a perceptually consistent manner.

The global Otsu binarization algorithm [182] has been used to produce the SIP measurements presented in the figures of this article. This general-purpose method was appropriate for our intention to preserve document noise, a particularly important aspect of the card images presented in the case study, as noise was found to have a direct impact on the quality of the optical character recognition. In the case of the head pictures of Fig. 21, however, the background shadow was impinging on the preservation of facial details that were the focus of interest during binarization; for this reason, the locally adaptive algorithm of Raleigh was chosen [183]. Conversion from color images to grayscale is realized by converting the images to the perceptual CIELAB color space and extracting the lightness channel, L^* [184, pp. 30, 95, 200–212]. The binarization step of the SIP measurement is not necessary for data that is already binary.

4) DIMENSIONALITY REDUCTION

The dimensionality reduction of the two-dimensional power spectrum to a vector is performed to ensure the rotation-independent measurement of patterns. The step consists in averaging data points of identical frequency. Due to quantization, however, digital images have sparse spectral representations; for example, the lowest frequency is expressed only at two orientations, 0 and π radians, in the usual Cartesian representation of the Fourier transform. To guarantee sufficient data and avoid flattening the spectrum (which artificially increases entropy), the frequencies are therefore rounded to the nearest integer prior to averaging the corresponding power spectrum values. We also disregard the direct component (DC; it has a frequency of 0) from the computation, which has no impact on the image pattern, since it represents the mean signal power. The vectorization procedure can be formally expressed as

$$w = \text{int}(\omega + 0.5), \tag{6}$$

$$n = \text{card}(\text{unique}(w)), \tag{7}$$

$$V = \sum_{i=1}^{n} P_{w_i}/\text{card}(w_i), \tag{8}$$

where the vector ω contains the frequencies corresponding to the power spectrum values S, the matrix into which the vector is indexed, with $\omega \in \mathbb{R}_{\geq 0}$; w is defined over the integer frequencies, $w \in \mathbb{N}_{\geq 0}$; n is the index of the rounded Nyquist frequency into the unique values of w; and $\text{card}(w_i)$ is the number (cardinality) of samples for a given integer frequency. (The Nyquist frequency for an image is half the number of pixels of the image diagonal.)

5) MAXIMAL SIP

Given that the SIP method orders patterns along the uniform–clustered–regular continuum, the following two fundamental questions arise: ''What is the pattern corresponding to maximal clustering?'' and ''What dSIP value does this pattern have before it was calibrated to 0?'' The second question will be answered here, while a tentative answer to the first question is provided in section III-C4.

Referring to the dSIP equation (5), we are interested in finding the value $\mathfrak{f}$, such that $T(\mathfrak{f}) = 0$. Whatever this value might be, it should correspond to a maximally clustered pattern for any length n of the power spectrum (S) of equation (5), including for $n = 2$. Anticipating the explanations to follow, maximal clustering corresponds to minimal redundancy; for the division of a whole into two parts (hence, $n = 2$) this is the reciprocal, Φ, of the golden ratio, ϕ, $\Phi = 1/\phi = (\sqrt{5}-1)/2 \approx 0.6180$ (see section III-C1). We further

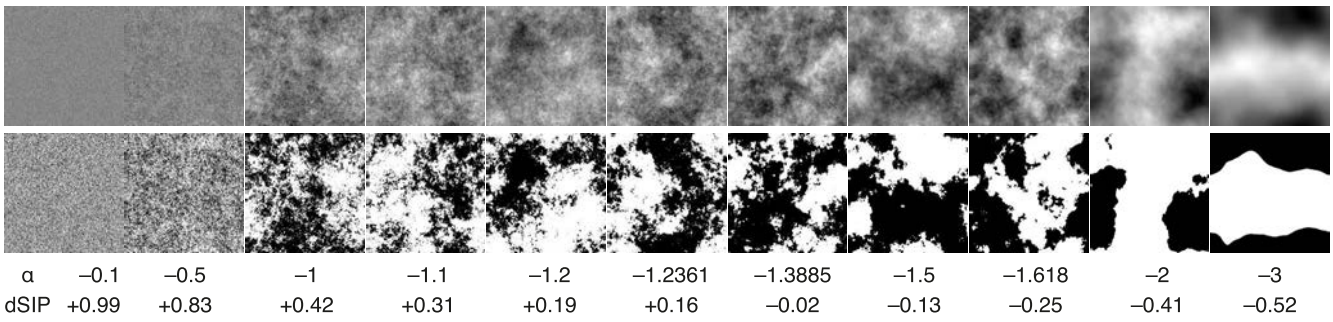


FIGURE 6. Intensity (top row) and binarized images (bottom row) of random fractals, with various power spectrum exponents, α, and their corresponding dSIP values. The third value is the negative of the golden ratio ($-\phi \approx -1.618$), the fifth value is the negative of twice the reciprocal of the golden ratio ($-2\Phi = \sqrt{5} - 1 \approx -1.2361$), and the sixth value is twice the negative of the value $\mathfrak{f}$ of equation 9 ($-2\mathfrak{f} = 2\log_2 \Phi \approx -1.3885$). The fractal images are obtained by synthesis in the frequency domain, through inverse Fourier transform from $\sqrt{f^{\alpha}}$ magnitude, random phase, and DC = 0; a subsequent zero-level cut produces the binary images [51, pp. 49–50] [49, p. 122]. The fractal dimension, D, of the image corresponding to $\alpha = -2\mathfrak{f} \approx -1.3885$ is $D = (6 + \alpha)/2 = 2.3058$. For the "paradoxon" of images with fractional dimensions $D \notin [2, 3]$, i.e. outside the expected bounds defined by the topological and Euclidean dimensions, see [53].

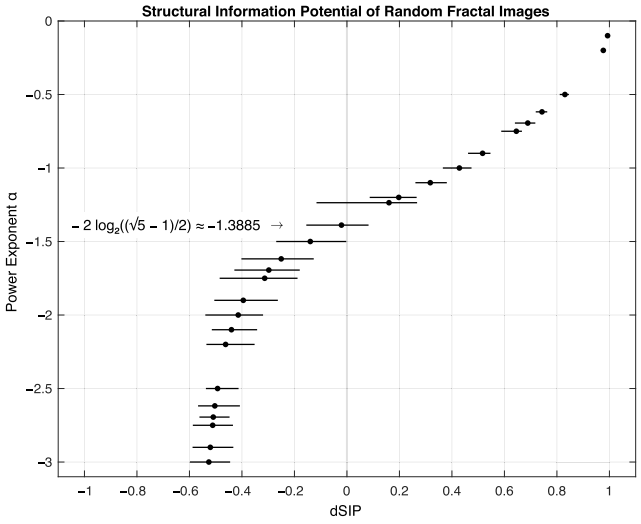


FIGURE 7. Median (dot markers) and range (bar markers) of dSIP values of one hundred binary random fractal images, a sample of which are shown in Fig. 6, vs the exponent α of the fractals' power spectrum power law distribution, f^{α}. The median dSIP of the images with $\alpha \approx -1.3885$ is -0.0208.

note that the input to T is modified logarithmically by the relative entropy H_r. Thus, we define $\mathfrak{f}$ (let it be called the "*fluorescent number*", as related to the most conspicuously clustered patterns, and following the graph theory terminology applied to the rainbow family of graphs) as the logarithm of the reciprocal of the golden ratio, or, equivalently, as the logarithm of the golden ratio:

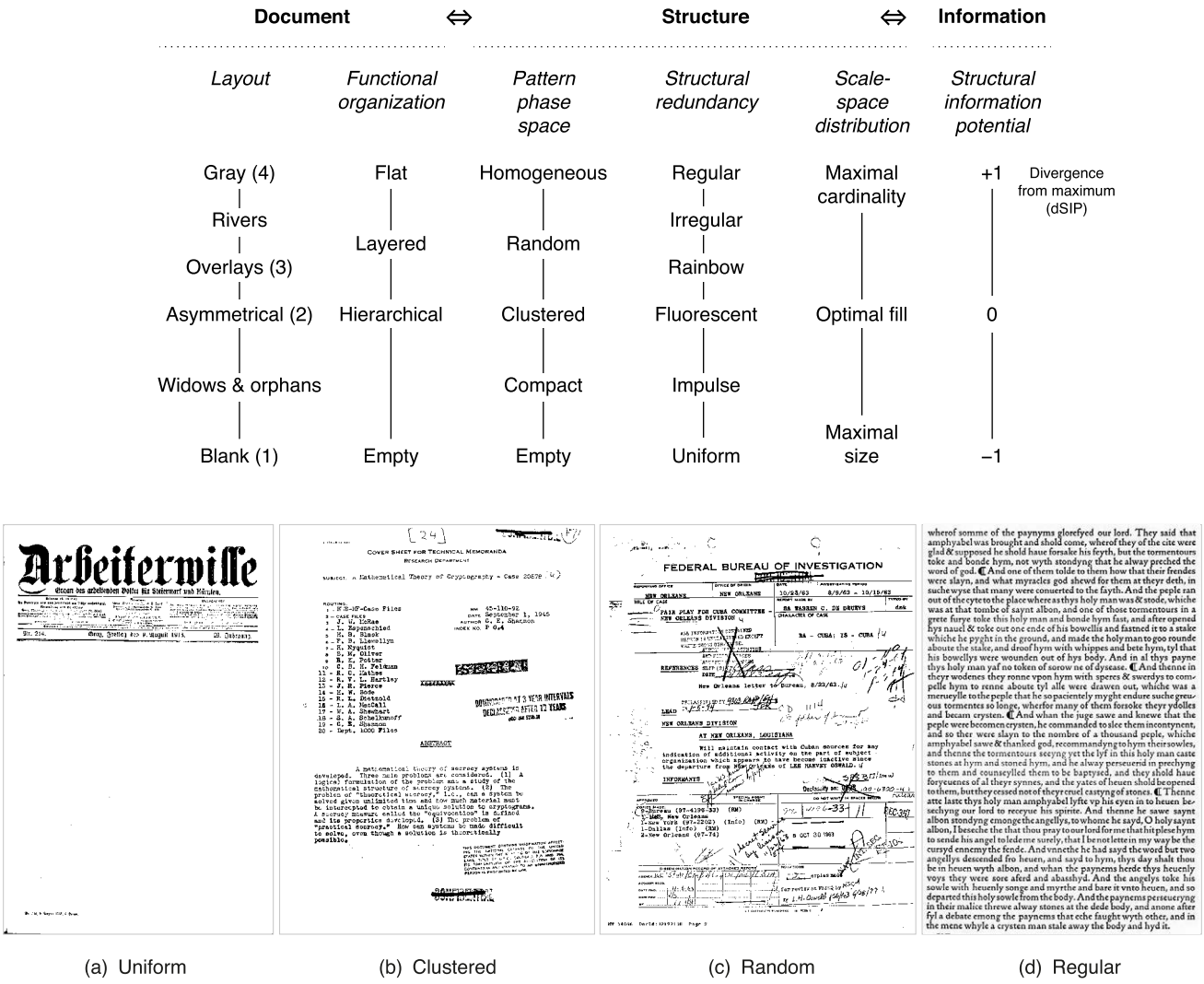
$$\begin{aligned} \mathfrak{f} &= -\log_2 \Phi = -\log_2((\sqrt{5} - 1)/2) \\ &= \log_2 \phi = \log_2((\sqrt{5} + 1)/2) \\ &\quad 0.6942. \end{aligned} \tag{9}$$

We have empirically validated this logic multiple times with the numerous datasets we have processed. The reader may ascertain for her- or himself that the patterns with a dSIP close to 0 do indeed present the highest degree of clustering. A good example reference is the cover drawing of the magazine shown in Fig. 15 (framed in red), which is highly clustered due to the variety of shape sizes and the perspective view. It has a dSIP value of $+0.0339$, which is close to minimal, as expected. Another example drawn from the case study is the card with minimal dSIP and high clustering in Fig. 18e and 19. Furthermore, the argument is also supported from an analytical point of view by the following observation. If we consider the epitome of clustered patterns — the fractals — then we can observe that by using the value $-2\mathfrak{f}$ as the exponent of their defining power law distribution of the power spectrum, $f^{-2\mathfrak{f}}$, we can obtain fractal images with a median dSIP close to zero, dSIP $= -0.0208$ (the doubling of $\mathfrak{f}$ is due to images having an exponent twice the value it has for signals [51, pp. 49–50] [49, p. 99, 105, 122]; Fig. 6, Fig. 7).

P12 – 313

6) TRANSFER FUNCTION

The transfer function T calibrates the structural information potential value to facilitate easier mathematical manipulation and cognitive interpretation. However, the following issue arises: due to the use of the relative entropy H_r in equation 5, the values within the transfer function T are bound to the [0, 1] range, the extrema of which correspond to uniform and regular patterns (see section III-B). As explained above, the pattern with maximal clustering (and, therefore, maximal structural information potential) has the value $\mathfrak{f} \approx 0.6942$. However, $\mathfrak{f}$ is not the center of the [0, 1] range, and the numerical pattern informativeness levels computed with the SIP method before calibration (say 0.4936 and 0.8107) are neither immediately comprehensible nor directly comparable. The solution we propose is to use a transfer function, T, to remap the values such that maximal clustering is attained for 0, maximal uniformity for -1, and maximal regularity for $+1$. We will then say that dSIP is a measure of *divergence* from the maximum structural information potential, with its sign indicating the pattern towards which it tends: positive for a more regular pattern and negative for a more uniform pattern.



P12
–
314

FIGURE 8. Conceptual schema of the generic relationship between information and structure, along with the correspondence of the visual and functional organization of documents. — (a) Uniform pattern, −0.106 dSIP; censored newspaper front page; Austria, August 9, 1918 (Austrian National Library). — (b) Clustered pattern, +0.173 dSIP; memorandum frontispiece, including institutional header, summary, circulation list, and secrecy stamps (Claude Shannon, "A Mathematical Theory of Cryptography"', 1945, AT&T Bell Laboratories) [185], [186]. — (c) Random pattern, +0.276 dSIP; official FBI printed form, heavily annotated by handwriting and stamps (Kennedy Assassination Records Collection, National Archives and Records Administration). — (d) Regular pattern, +0.424 dSIP; typography inspired by medieval manuscripts, by the Arts and Crafts movement exponent William Morris. ("The Golden Legend", 1892, London, Kelmscott Press/Quaritch Books; author's personal collection).

Furthermore, to honor the linguistic expression ''maximal (or minimal) structural information potential''—which is awkwardly 0 (respectively both [*sic*] −1 and +1) when computed via dSIP—we will define SIP as given by equation 4, which is maximal at 1 and minimal at 0, representing a more intuitive situation.

To map the values from the range [0, ƒ, 1] to [−1, 0, +1], we fit a second-order polynomial to the three value pairs and obtain the following coefficients for an input value x (where x represents the input into T in equation 5) and output y:

$$y = 1.83\,x^2 + 0.1699\,x - 1. \tag{10}$$

As a further refinement, we counterbalance the non-linearity resulting from equation 10 by taking the logarithm of the output y for the final definition for the transfer function T:

$$T = \begin{cases} \log_2(y+1), & \text{for } y \geq 0. \\ -\log_2(|y|+1), & \text{for } y < 0. \end{cases} \tag{11}$$

B. PATTERN PHASE SPACE

Consider a raster image in which the pixels take the values of zero and one. The most *homogeneous* pixel distribution is obtained for a checkerboard pattern where every pair of adjacent pixels have distinct values; in other words, when the pattern is *regular*. Any disturbance of this regularity will decrease the homogeneity, including *random* disturbances. When the disturbance is not deterministically or stochastically uniformly distributed across the image, then pixel *clusters* will emerge, with maximal clustering attained when

the clusters have sizes at all scale-space levels (under these circumstances, the entropy of the scale-space will also be maximal). Aside from the regular pattern that exhibits pixel objects of only a single size, and is thus located at a single scale-space level (and has zero entropy), another pattern with this characteristic is the *uniform* pattern. It appears that when pixels of identical value ''coagulate'' into a *compact* mass, the surrounding of this mass becomes uniform. These distributions define a unidimensional space of pattern ''phases'' that vary from regular to random to clustered to compact to uniform.

Fig. 8 provides a practical example of the pattern phase space described above in theoretical terms. It shows the SIP classification of sample document images along a unidimensional pattern space. The progression from regular pattern to clustered to uniform is readily observable and conforms to the problem requirements. Note in the upper part of the image the parallel between the terminologies from various fields (preeminently typography, signal processing, and material sciences). The conceptual system introduced here, and further discussed below, may be used as a model to describe the correspondence between the visual and functional organization of documents. At the paragraph level, the penmanship and typographical ideals are the production of a homogeneous pattern (appearing as gray when observed from a distance), while the functional hierarchy is reflected in a clustered layout (the technical term is ''asymmetrical''). Post-production annotations and degradations disturb the intended (ir)regularity, introducing randomness. Blank pages (etymologically French for ''white'', another chromatic term in the field of documents) mark the end of a functional unit, as well as containing a reduced amount of semantic information.

The uniformly *random* pattern is located between the clustered and the regular, since it does not comprise large uniform areas. As such, their dSIP value may be very high (Fig. 6). Random patterns can occur due to signal noise or annotations (in the case of text documents). Note that other pattern spaces are possible: the uniform–random spectrum [187] is appropriate for patterns with statistically constant homogeneity, such as homogeneously distributed line segments of quasi-equal length and variable orientation.

Patterns with *semantic* value, such as images of faces, are obviously very important for human beings, and can be located anywhere in the proposed pattern space. Because these patterns require semantic analysis and contextual information, they cannot be considered from a low-level computer vision perspective such as that described here, just as the Shannon entropy casts information in purely mathematical terms.

C. GRAPH STRUCTURES WITH MINIMAL REDUNDANCY

While previous sections dealt with the measurement of SIP, this section concerns the definition and design of mathematical objects with minimal structural redundancy; that is, those that maximize the structural information potential.

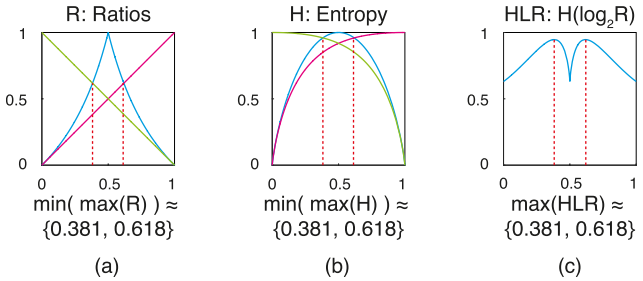


FIGURE 9. **(a) Given the unit segment *c* and its division in two segments *a* and *b*, the figure shows the values of the ratios *a*/*c* (red), *b*/*c* (green), and *a*/*b* for *a* ∈ [0, 0.5] and *b*/*a* for *a* ∈ [0.5, 1]. The minima of the maximum of the ratios correspond to the reciprocal and the reciprocal conjugate of the golden ratio (dotted lines). — (b) Relative Shannon entropy of the segments taken pairwise. — (c) Relative Shannon entropy of the logarithm of the pairwise ratios.**

While measurement may be employed independently of design, a discussion of the latter supports a firmer understanding of the former. In particular, such discussion provides a formal rationale for the calibration of SIP values, as well as visual evidence and quantitative characterization of objects with maximal SIP value. Furthermore, it reveals links between SIP and some interesting mathematical concepts, thereby creating an opening for generalizing SIP to structures other than images.

We will refer to objects with minimal structural redundancy as ''*fluorescent*'' objects, in reference to the ''rainbow'' graphs (which have edge values that are distinct, but not necessarily maximally distinct), to the name of the symbol, *phi*, denoting the golden ratio (which minimal redundancy properties as discussed below), and to the fact that fluorescence maximizes perceptual color discrimination. The proposed notation is $\mathfrak{F}^{v\mathrm{V}}_{d\mathrm{D}}\{\cdot\}$, where d is the object's embedding dimension, v the number of vertices, and the placeholder $\{\cdot\}$ may be used to specify edges. For example, $\mathfrak{F}^{3\mathrm{V}}_{2\mathrm{D}}\{\mathrm{e}_{12}, \mathrm{e}_{13}, \mathrm{e}_{23}\} : \{x_1, y_1, \ldots, x_3, y_3\}$ designates a fluorescent triangle and the vertices' coordinates.

In the following, the argument will proceed from one-, to two-, to three-dimensional objects, from shapes to patterns, and will conclude with some remarks of a more general order.

1) ONE-DIMENSIONAL FLUORESCENCE FOR THREE POINTS

How might a whole be divided into two parts so as to maximize the difference between the parts, while concomitantly maximizing their respective sizes? The first condition of this problem is satisfied when either of the parts vanishes in the limit, while the second condition corresponds to the two parts being equal. The overall solution lies in between these extrema and is found by determining the value at the intersection of the functions representing the conditions, i.e., $f(x) = 1 - x$ and $f(x) = x/(1 - x)$, for $x \in [0, \frac{1}{2}]$ and $f(x) = x$ and $f(x) = (1 - x)/x$, for $x \in [\frac{1}{2}, 1]$, or solving the equation $x^2 - 3x + 1 = 0$ and $x^2 + x - 1 = 0$. For the definition domains, these have the solutions $x_1 = (1-\sqrt{5})/2+1 \approx 0.3819$ and $x_2 = (1+\sqrt{5})/2-1 \approx 0.6180$.

P12 – 315

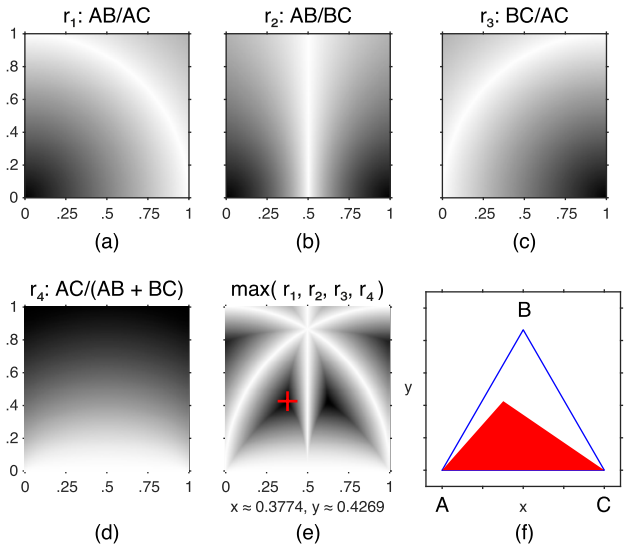


FIGURE 10. Design process of the fluorescent triangle. — (a)–(c) Pairwise edge ratios of a triangle ABC, with unit base AC, for all locations of vertex B within a unit square. Grayscale tones encode values between 0 (black) and 1 (white). — (d) Ratio of base and sum of opposing edges. — (e) Maximum of the surfaces defined by the edge ratios. The minimum of this surface for $x \in [0, \frac{1}{2}]$ is marked by a + symbol, and its coordinates are specified next to the ordinate. — (f) The triangles with the most and least dissimilar edge lengths, i.e., the isosceles triangle (blue outlines) and the fluorescent triangle (solid red). Basic mensurations: apex: $x \approx 0.3774$, $y \approx 0.4269$; edge lengths: $AB \approx 0.5698$, $BC \approx 0.7548$, $AC = 1$; angles: $\alpha \approx 45.5199°$, $\beta \approx 97.0399°$, $\gamma \approx 34.4400°$; area: $T \approx 0.2134$.

P12 – 316

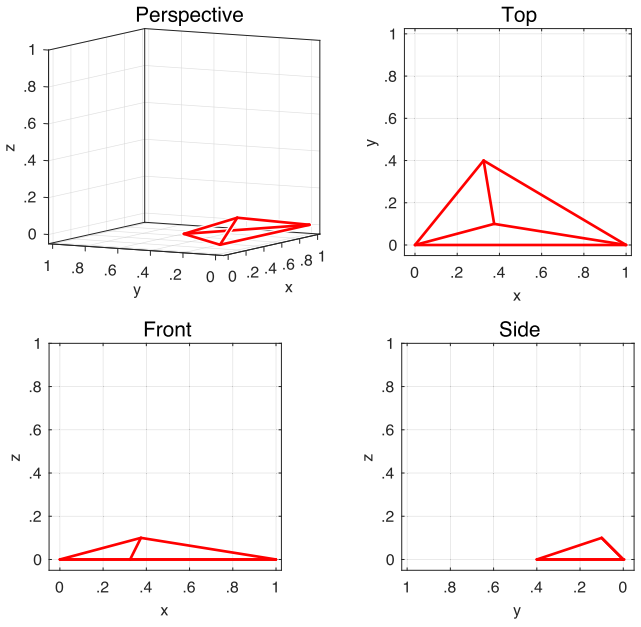


FIGURE 11. Fluorescent tetrahedron based on minimization of pairwise edge ratios and ratios of one edge and two adjacent edges. Coordinates of base triangle apex on a grid with resolution of 0.025 units: $x = 0.325$, $y = 0.4$, $z = 0$; tetrahedron apex: $x = 0.375$, $y = 0.1$, $z = 0.1$.

These values are also known, respectively, as the conjugate of the reciprocal, Φ', and the reciprocal, Φ, of the golden ratio, $\phi = (1 + \sqrt{5})/2 = 1/\Phi = 1/(1 - \Phi') \approx 1.6180$. A diagrammatic representation of these solutions (Fig. 9a) allows

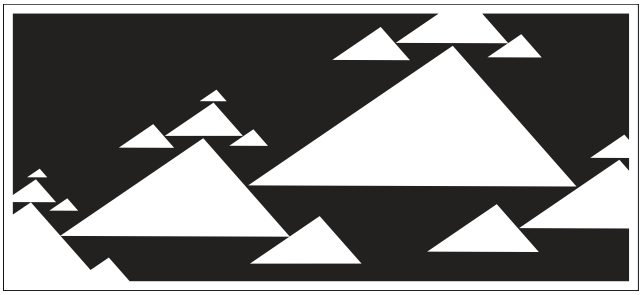


FIGURE 12. Fluorescent pattern generated from a fluorescent triangle. Note the similarity with the naturally occurring patterns of seashells [188].

us to state the criterion for determining the minimal structural redundancy (or maximal SIP) as the minimum of the maximum of the ratios of the part and the whole: $R = \{a/(a + b), b/(a+b), \min(a, b)/\max(a, b)\}$, $\text{SIP}_{max} = \min(\max(R))$, where $a+b = 1$, and $(a, b) \in [0, 1]$. In other words, maximal SIP corresponds to the minimum of the range of the ratios of a system's components. The use of these ratios is equivalent to using the relative Shannon entropy, H, for each edge pair (Fig. 9b). The max formulation can be avoided by combining ratios and entropy: $\text{SIP}_{max} = \max(H(-\log_2(R)))$ (Fig. 9c). The similarity between this equation and equation 5 indicates that maximal SIP is expected for $-\log_2(\Phi)$, which is the reason why this value was used for calibration in the SIP measurement algorithm.

Remark — Note the inclusion of the whole as a third element in the measure of the sizing of the two segments. This is explained by the whole representing the highest scale of the scale-space domain defined by the segments, thus making it part of the system. This is the case in many application domains, particularly those that are subject to human factors, such as document design and perception. For instance, the sizing according to the golden ratio of the width of a text column and a figure placed next to each other ensures maximal legibility for each while giving prominence to one of them.

2) TWO-DIMENSIONAL FLUORESCENCE FOR THREE POINTS

Let us now extend the problem to two dimensions and ask the following question: *What is the triangle with the most dissimilar edges?* Above, we identified the solution for three collinear points that define a degenerate triangle. While this triangle has minimal redundancy between its parts when taken pairwise, it also has zero area, unlike common triangles. More importantly, the degenerate triangle has maximal redundancy between a part and a subset: the triangle base AC equals the sum of the other edges, AB and BC. Therefore, we introduce the ratio of the base and the sum of the other edges as part of the problem formalism, with the effect of increasing the size of the triangle to a certain equilibrium point below that of an equilateral triangle, for which redundancy is maximal. We now solve the equations $AB/BC = BC/AC = AC/(AB + BC)$, given the coordinates x and y of the triangle apex, with $AB = \sqrt{x^2 + y^2}$, $BC = \sqrt{(1 - x)^2 + y^2}$, and $AC = 1$. The solutions in the abscissa

interval $[0, \frac{1}{2}]$ yield $x \approx 0.3774$ and $y \approx 0.4269$. This is the elemental shape with maximal SIP, as per our definition (Fig. 10). A remarkable trait of this triangle is that it extends an essential property of the golden ratio—that of minimal redundancy between parts and whole—to two dimensions.

3) THREE-DIMENSIONAL FLUORESCENCE FOR FOUR POINTS

Fig. 11 illustrates the extension of the fluorescence principle to three dimensions, here for a tetrahedron with minimal edge redundancy. This tetrahedron represents a pendant of the equilateral tetrahedron, one of the five regular Platonic solids. The locations of the vertices of this graph have been computationally determined by placing vertices v_1 and v_2 at two adjacent nodes of a unit cube, then placing vertices v_3 and v_4 in turn at all locations of an orthogonal grid with resolution of 0.025 units. To avoid triangles with identical shape but different sizes and orientations (metamers), locations for which the length of edges ending in vertices v_1 or v_2 is more than unity were excluded. The measure of redundancy was based on the ratio of pairwise edges, as well as the ratio of one edge taken at a time and its adjacent two edges. The latter criterion is adopted from the triangle case discussed above, and its effect is to avoid facets with zero area. This benefits the fluorescence measurement, as it neatly facilitates the quantification of redundancy across dimensions and structure types, such as mesh surfaces and solids. (This observation highlights the possibility of using criteria other than the ratio of edge subsets, such as the perimeter and the volume, to avoid shape degeneration; these would lead to different solutions, but would be more loosely consistent with the problem definition in terms of edge ratios alone. Instead, they may serve special applications, such as maximization of the volume generally resulting in convex solids.)

The inquiry into minimal structural redundancy can be pursued for graphs with an arbitrary number of edges, such as polygons and mesh surfaces, or of arbitrary structure, such as circular, tree-like, or network-shaped graphs. In this article, the goal is limited to open a window into these possibilities.

4) TWO-DIMENSIONAL PATTERN FLUORESCENCE

We investigated minimal structural redundancy for only the most basic graphs (line bisection, triangle, and tetrahedron), stopping short at identifying the fluorescent pattern(s) with arbitrary size. A family of fractals is the likely answer—but is there a fractal dimension more apt to produce fluorescence than another? Furthermore, considering deterministic fractals (such as the Koch curve [46, pp. 87–91]), it could be argued that there may exist a single fractal that minimizes structural redundancy. I propose that the design of such a pattern may consist in taking the simplest structure of a given dimension (e.g., the fluorescent triangle for two dimensions) and replicating it infinitely with a certain size change rate (Fig. 12). The result consists, in fact, in an affine transform of the module of well-known fractals (Cantor dust, Sierpiński gasket and tetrahedron, and Julia set [46, pp. 65–79, 120–123]), so as to morph them into a fluorescent shape.

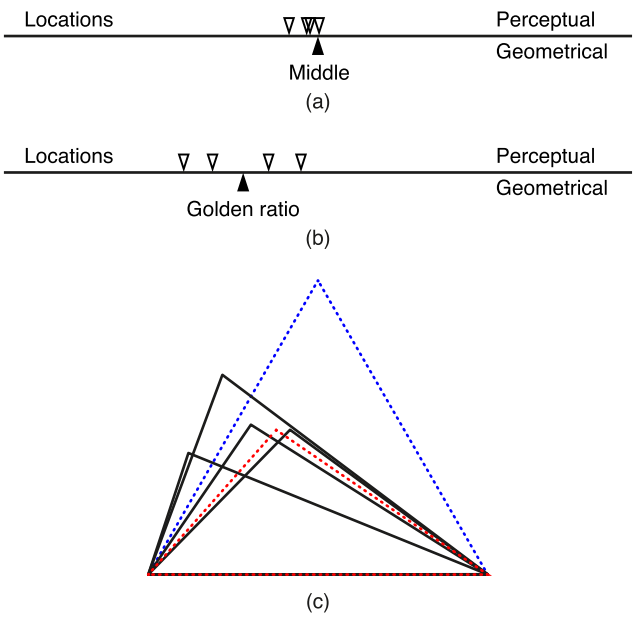


FIGURE 13. Comparison of mathematical and perceptual fluorescence. — (a) Human experimental participants tendentially place the middle of a line left of its geometrical location. Here, an example from four male computer scientists, aged late 20s to late 40s, in which the outlined markers above the line represent their choices. — (b) The same group was tasked to segment a line into the two most dissimilar segments with respect to each one, as well as to the whole line, i.e. according to the golden ratio. — (c) In this task the participants are instructed to draw a triangle with the most dissimilar edges, i.e. a fluorescent triangle, shown in dotted red, as opposed to the equilateral dotted blue triangle.

P12 – 317

5) FLUORESCENCE AND SIP

Let us examine the link between SIP and minimal graph redundancy, two major themes of this section.

The concept of structural information potential, on one hand, posits a correlation between structure and information, based on the scale-space distribution of the structure, which defines a pattern spectrum ranging from compact to homogeneous via clustered and random. This perspective on patterns is quantified by the SIP method in the frequency domain using the Fourier transform and the Shannon entropy. The concept of fluorescence, on the other hand, attempts to characterize structures in terms of component redundancy, and to this end adopts a graph theoretical formalism of the ratio of combinations of edge subsets.

The integration of these two perspectives is realized through the concept of scale-space redundancy, which subtends the pattern spectrum and has a direct information theoretical meaning (Fig. 8). Specifically, a pattern that maximizes both the *number* and the *size* of constituent entities will fill the scale-space in an optimal manner while having maximal structural information potential and minimal structural redundancy. This is the fluorescent clustered pattern, which stands in opposition to the uniform pattern of a maximally

sized single entity, as well as the homogeneous pattern of minimally sized and maximally numerous entities.

In conclusion, the study of graph redundancy offers a powerful and versatile instrument to understand the elemental levels from which clustered and structurally informative patterns emerge. One goal of this section is to provide precisely such a low-level explanation of the SIP method.

6) FLUORESCENCE AND HUMAN PERCEPTION

Human estimation of pattern redundancy differs from mathematically derived values [189], [190], which is an important factor to consider when measuring man-made patterns (such as documents) or evaluating psychophysically formal methods of redundancy measurement. A well-known example is the systematic bias in line bisection (Fig. 13a) [191]. The location of the fluorescent line partition and the drawing of the fluorescent triangle exhibit similar variability and divergence from the mathematically defined shapes (Fig. 13b, c). Psychological aspects, furthermore, interact with cultural and social ones, forming a much more complex and dynamic ecosystem of constraints upon informativeness than the basic, low-level uniform–clustered–regular pattern-informativeness spectrum. The subjective and contextual dimensions of human pattern perception may nevertheless be directly relevant to the evaluation of the SIP measurement method; for instance, to explain why observers might disagree about the redundancy of a given pattern. A converse issue is that of human-designed fluorescent configurations, such as the text/figure sizing in document layout mentioned above, whose divergence from the mathematical model may even be intentional. In particular, this may often be the case with regard to the uncritical application of rules often shunned by artists, keen observers of form, as illustrated by the words of the illustrious French photographer Henri Cartier-Bresson: ''I hope that we will never see the day when the merchants will sell [golden ratio] diagrams engraved on camera displays.'' [192, pp. 26–27]. The quote exemplifies the need to develop models of layout irregularity that integrate human factors. When applied to documents, the desideratum has been carried out in the design of the SIP formalism, which emerged from empirical observation and reflects low-level perception. At such levels of complexity, cognitive automatisms are more amenable to modelization by the pattern–informativeness spectrum. The task of document triage, with its emphasis on fast temporal information processing, is a good case study for testing the model and shall be discussed next.

P12 – 318

7) REPRESENTATION

The opposing states of minimal and maximal redundancy may be represented in a single view. Fig. 14 presents the panchromatic prism, whose top is an equilateral triangle (i.e., ''gray'' in chromatic graph theory terminology), while the base is a triangle with most dissimilar sides (i.e., ''fluorescent''). This mathematical object is visualized in the manner of Leonardo da Vinci's drawings of the five Platonic regular bodies [193].

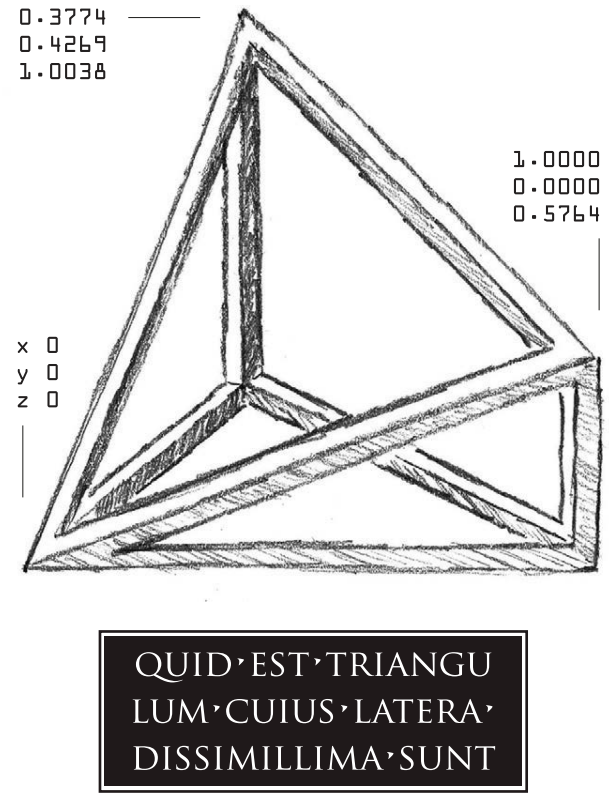


FIGURE 14. The panchromatic prism.

IV. EXPERIMENTS

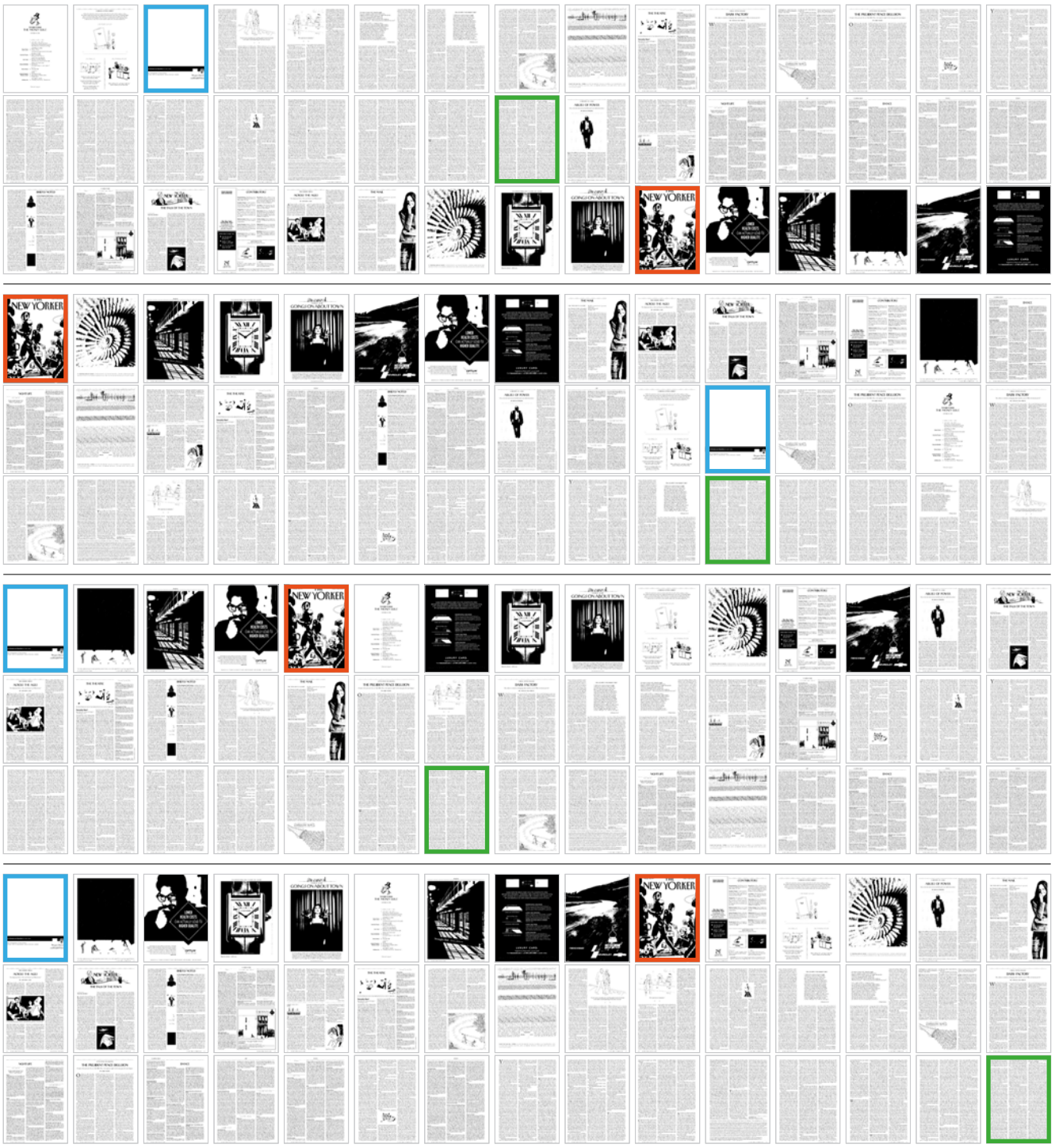
In this section, we apply the method of document ordering developed above to a small set of representative real data, then compare the results to those of other methods. Next, we use a case study to test the robustness of our approach on a large dataset with respect to the triage task. Finally, we demonstrate that the proposed concept may be generalized to objects beyond text-based document images and applications other than triage.

A. COMPARISON OF METHODS

1) VISUAL EVIDENCE

Fig. 15 juxtaposes document pages ordered using the major methods discussed in this article: structural information potential, approximate entropy, spectral flatness, and the ratio of ink pixels to page area. The data are sampled from the 104 binarized pages of an issue of the *New Yorker* magazine (please refer to Fig. 26 to 31 to view all pages). This particular dataset was chosen for display among the hundred analyzed owing to its diversity of text, drawing, and image patterns, which provides exemplary illustration of the entire uniform–clustered–regular spectrum against which we want to test our methods.

We first observe that the ink/page ratio method leads to a mix of structurally unrelated patterns, and to a split of



P12
–
319

FIGURE 15. Ordering (left to right and top to bottom) using the ink-to-page-area ratio (first top block), approximate entropy (second block), spectral flatness (third block), and structural information potential (last block; $dSIP \in [-0.3109, +0.6523]$, $\min(dSIP) = -0.0073$, corresponding to the eighth image in the first row). Colors mark the most uniform (blue) and homogeneous (green) pages, and the highly clustered cover page (red). — Credit: *The New Yorker*, vol. 93, no. 33, October 23, 2017; Condé Nast.

homogeneous clusters, such as the mostly empty pages (the third image and the third from last image of the top-most block in the figure). Approximate entropy is very good at clustering multiscale patterns, such as the first three pages of the sequence in the second block, which depict either entities with different absolute sizes (e.g., people and robots in the street, the rays of the trilobite), or a diversity of sizes resulting from a perspective view (rows of bars in a prison

hallway). However, the method fails to group together homogeneous patterns (text-only pages are interspersed with text and illustration pages) and empty pages. Spectral flatness (third block) groups empty pages and follows them with multiscale patterns, but fails to consistently group homogeneous patterns (the result of spectral entropy is similar, but less good). Structural information potential (last block) achieves a perceptually gradual ordering of the pages from empty to multiscale to homogeneous.

2) QUANTITATIVE EVALUATION

The previous section has presented visual evidence that the SIP method provides a better ordering of document pages on the uniform–clustered–regular pattern spectrum than the other methods considered. We now quantify the difference in suitability between the methods.

Method — Let the monotonously increasing or decreasing ranking of n items $P = \{p_1 \prec p_2 \prec \ldots \prec p_n\}$ and $P' = \{p_n \succ p_{n-1} \succ \ldots \succ p_1\}$, respectively, be the baseline orders of comparison. Any permutation Q of the items will be a less desirable ordering to a degree in accordance with a certain criterion of desirability. In the framework of the pattern spectrum, this is the case when two patterns from the opposite sides of the spectrum appear next to each other following permutation; that is, when the difference between their original rankings is maximal. For example, given an ordering of achromatic chips from white to gray to black, placing the black chip next to the white will disturb the monotonicity (Fig. 16). In terms of document pages, this corresponds to placing a homogeneous text page next to a mostly empty page. Our goodness criterion is, therefore, the difference between two adjacent ranks of the permutation Q of the order P (or its reflection P').

P12 – 320

We repeat the process of flanking items from opposing extremities of the monotonous ranking and obtain the following permutation sequences, which represent the least desirable orderings: $R = \{p_{k+2}, p_k, \ldots, p_5, p_{n-4}, p_3, p_{n-1}, p_1, p_n, p_2, p_{n-3}, p_4, \ldots, p_{k-1}, p_{k+1}\}$, and its reflection R', where $k = n/2$ (Fig. 16d). Having established the best and worst orderings, we next set about developing a quantitative measure of ordering goodness.

Given that the objective is to maximize the difference between adjacent items of a permutation Q, we generalize and compute the sum of the differences between the ranks r_i of consecutive items, $D^1 = \sum_{i=1}^{n-1} r_{i+1} - r_i$; this allows us to distinguish between monotonous and other orderings, since the former have a computed value of n, $D^1(P) = n$, and zero for a second-order differencing, $D^2(P) = D_{\min} = 0$. To distinguish between the remaining patterns we apply the differencing process $n - 1$ times, yielding the scalar D^{n-1}. In other words, we compute at increasingly higher scales the length of a shape with coordinates given by the order indices in Q and the ranking values in P. The permutations with the highest absolute value will correspond to the least desirable permutation R, and its isomorphisms, defined above: $D_{\max(n)} = D^{n-1}(R)$.

It is clear that this is the case considering that the n-minus-one-th order differencing of even n and $k = n/2$ has the form $D^{n-1} = -c_1p_1 + c_2p_2 + \ldots - c_{n-1}p_{n-1} + c_np_n = (c_{k+1}p_{k+1} + c_{k-1}p_{k-1} + c_{k+3}p_{k+3} + \ldots + c_2p_2 + c_np_n) - (c_1p_1 + c_{n-1}p_{n-1} + c_3p_3 + c_{n-3}p_{n-3} + \ldots + c_{k+2}p_{k+2} + c_kp_k)$; here c denotes the binomial coefficients for order n, with $c_{k+1} = c_k > c_{k-1} = c_{k+2} > \ldots > c_n = c_1 = 1$; odd indices are located in the first half of the series, and even indices in the second half. The series is maximized for the ordering R and its reflection $R' = \{p_{k+1}, p_{k-1}, p_{k+3}, \ldots, p_2, p_n, p_1, p_{n-1}, p_3, p_{n-3}, \ldots, p_{k+2}, p_k\}$. For example, for $n = 10$, we have $D^9 = (126p_6 + 84\,p_4 + 36\,p_8 + 9\,p_2 + 1\,p_10) - (1p_1 + 9\,p_9 + 36\,p_3 + 84\,p_7 + 126\,p_5)$. Since $p_6 > p_4 > p_8 > p_2 > p_10 > p_1 > p_9 > p_3 > p_7 > p_5$ and $p_np_i \in \mathbb{N}$, $i \in [1, 6]$, then $p_6 = 10, p_4 = 9, p_8 = 8, p_2 = 7, p_10 = 6, p_1 = 5, p_9 = 4, p_3 = 3, p_7 = 2, p_5 = 1$, yielding the sequence $R = \{5, 7, 3, 9, 1, 10, 2, 8, 4, 6\}$, whose $D = 1930$ is maximal for all permutations. To account for isomorph permutations with the same value but opposing signs, we take the absolute value, $|D|$; while to compensate for the exponential growth induced by the binomial coefficients that affects the distribution homogeneity, we take the logarithm, $\log(|D| + 1)$.

We are now able to characterize in the interval [0, 1] the goodness D of any permutation Q of length n, following normalization by $D_{\max(n)}$ computed from the R of length n:

$$D(Q_n) = \log(|D^{n-1}(Q_n)/D^{n-1}(R_n)| + 1). \tag{12}$$

One limitation of this method is that it is only valid for an even number of items. For odd numbers of items, the permutations with maximal D will include a permutation for which there is a difference of one between adjacent items, which is not the maximal departure from monotonicity. For large n, this is a benign limitation, since one item can be safely removed if selected so that it has the least change in rank across the permutations to be compared.

The author created this ordering distance measure after being unable to find a suitable solution in the literature on non-parametric statistics. However, in the field of combinatorics, topological entropy exists as a measure of the complexity of dynamical systems in the one-dimensional interval, for which the extreme values correspond to the same permutations for which our measure D is minimal and maximal, respectively [47; 3; 72; 1]. The need to develop a new method for apparently the same result is based on the different justifications made by the two methods as to why permutation R is the ''worst'' case: there is no a priori rationale for considering that the permutation maximizing adjacent differences is the same as that which maximizes complexity, as defined by topological entropy. A further reason is that our method is more straightforward than computing the topological entropy, which involves deriving a peculiar ''induced'' matrix from the permutation and finding its maximum eigenvalue. For large datasets, such as all permutations for a given n, the differencing method is also substantially faster, since it is easily vectorizable. Nevertheless, the identical results in respect to the extreme permutations suggest an equivalence between the

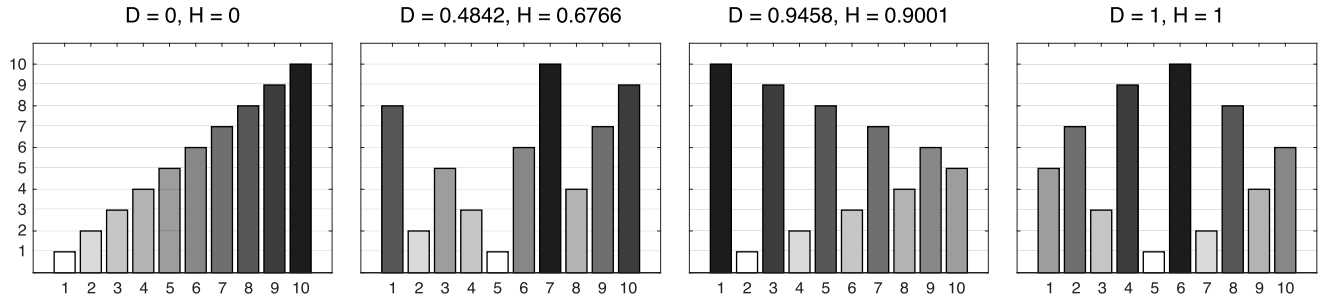


FIGURE 16. Various permutations and their departure from monotonicity, computed with the differencing (*D*) and topological entropy (*H*) methods.

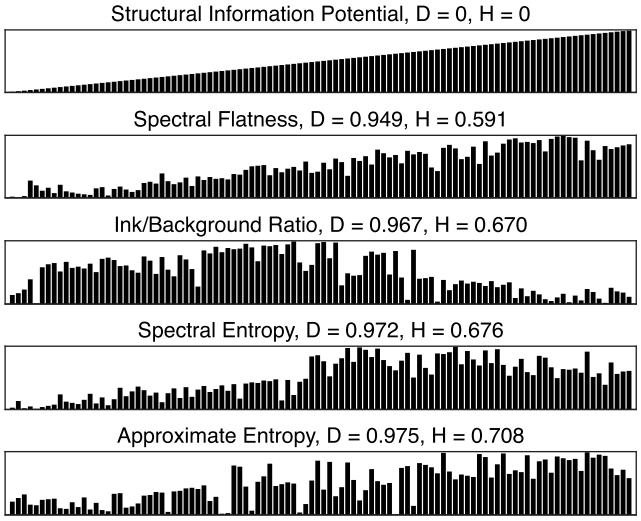


FIGURE 17. Departure from the ordering induced by the SIP pattern classification method of four comparison methods based on the document pages of Fig. 27–31.

n-minus-one-th-order differencing and eigen-analysis, perhaps in the scale-space characterization of patterns that both undertake. It may also be noted that topological entropy is not equivalent to SIP or minimal redundancy as discussed in this article. This can be determined by considering that topological entropy (and adjacency difference) increases when two items a and b are maximally different (that is, when $a \xrightarrow{\lim} 0$ and $b \xrightarrow{\lim} \infty$), while SIP is maximal when both items and their sum are maximally different ($a/b = b/(a+b)$).

I have made both the differencing and entropy methods available as open-source Matlab functions [6].

Results — Fig. 17 provides a graphical representation of the shuffling of the document pages ordered with the SIP method, as realized by the four comparison methods. The numerical values indicate the departure from monotonicity of the respective orderings. Approximate Entropy induces the ordering most different from SIP, while that of Spectral Flatness is the most similar. We can observe the clear juxtaposition of pages from the opposite sides of the uniform–clustered–regular spectrum in the Ink/Background and Approximate Entropy orderings, along with the shifting of page blocks in the case of Ink/Background and Spectral Entropy, which disturb monotonicity at small and large scales, respectively. Both distance measuring methods indicate that, overall, there is a substantial ordering difference between SIP and the other methods.

B. CASE STUDY

This section presents a real-world application of the structural information potential measurement method. The beneficiary is the Swiss National Library, and the goal is to merge the information of corresponding bibliographical records in electronic and analog formats [194]. The case study covers the pilot phase of an ongoing project, in which computational methods are assessed in view of deciding on the next steps.

P12 – 321

The analog records are palm-sized paper cards, typewritten or printed, with handwritten annotations, crossings-out, stamps, bar codes, rulings, and other graphical elements (Fig. 18c–h). A total of 1.2 million cards were scanned from high-contrast black-and-white films (accessible at http://siibns.ch/french/cat1_frame.htm), at a resolution of circa 500 by 300 pixels; this introduces various artifacts, such as background noise and an irregular border around the cards. Some cards contain no records, but only captions for a sequence of cards in the wooden trays that were accessed by library patrons searching the catalog. The texts often mix together two or more of the four national languages of Switzerland (French, German, Italian, and Romanche), as well as English, Latin, and other languages. The majority describe monographs and serial publications, but also maps; the content is inconsistently semantically structured, is generally not made up of full sentences, abounds in entity names, alphanumeric shelf marks, ISBNs, price tags in various currencies, and other codified data, and is rich in typographical formatting of logical entities.

Such a wide typological variety of information, brevity, and visual complexity presents a challenge for automatic recognition. Attempts to perform optical character recognition (OCR) on the whole dataset, using the open-source Tesseract software [195] and the commercial Google Vision [196], revealed that the obtained text is not directly exploitable in the library's public catalog (Fig. 18i). It was therefore envisaged to provide users with digital images of cards along the electronic record in the same graphical interface window. However, which analog

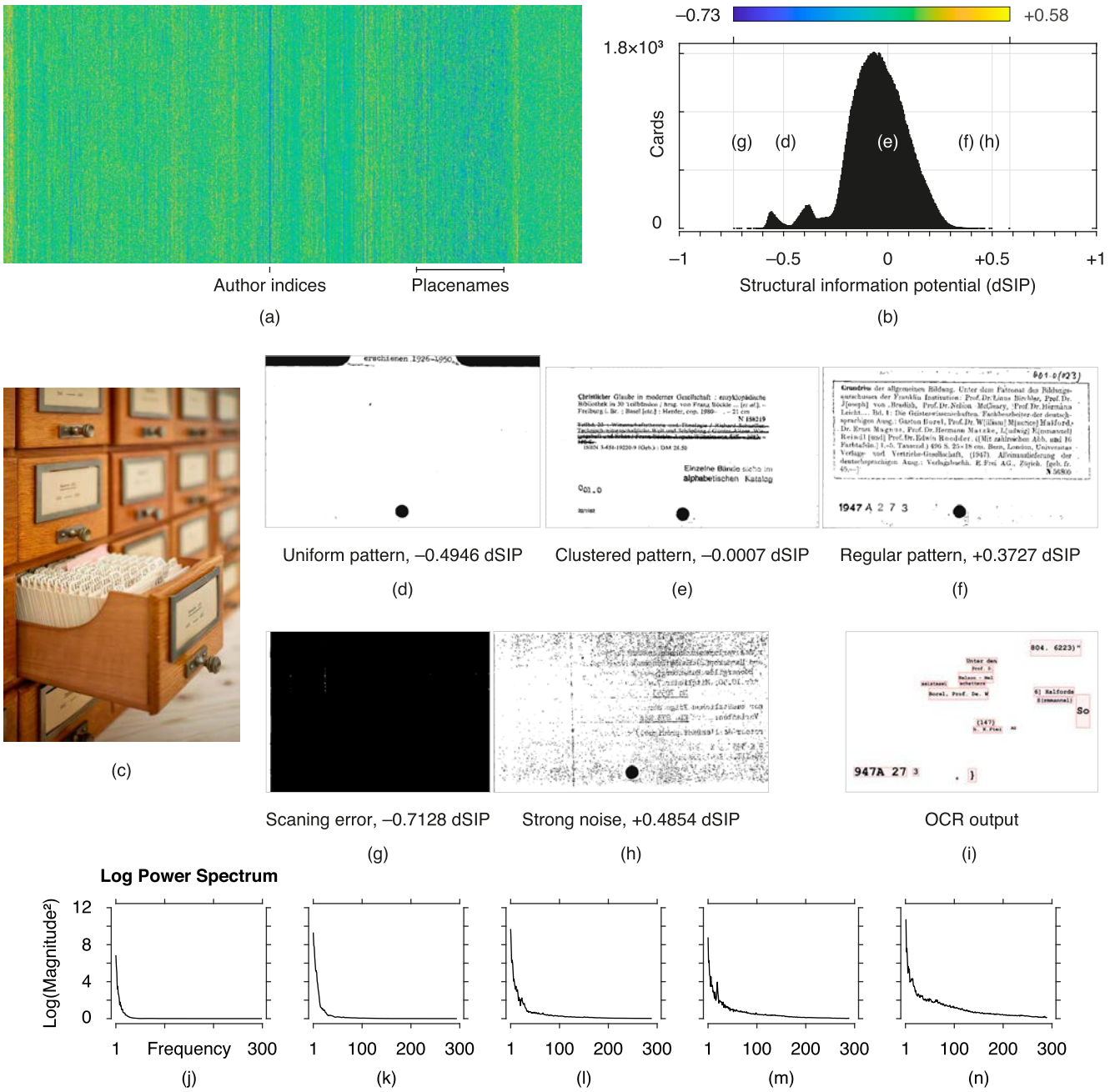
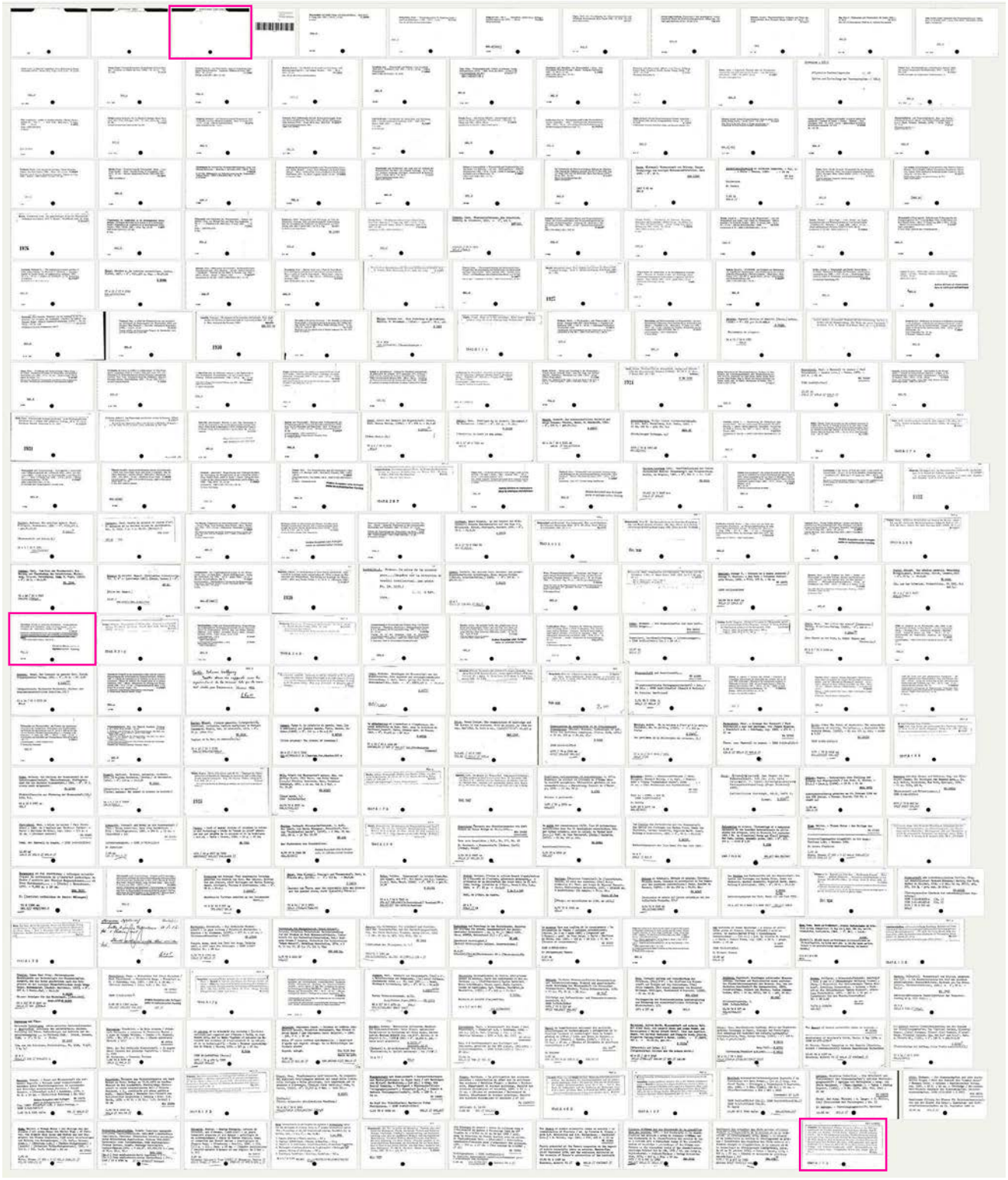


FIGURE 18. (a) 1.2 million library index card images, at one image per pixel, color-coded by their structural information potential, in physical sequential order, rearranged for space reasons top-down and left to right, and shown at reduced size. The markers indicate thematic ranges. — (b) Histogram of dSIP values and colormap for the data range. The letters indicate the approximate locations of the five cards shown above. — (c) Paper cards in the library catalog trays. (Credit: Swiss National Library) — (d) – (f) Sample cards of the main visual pattern types, with their dSIP values. — (g), (h) Two outlier types, found at the extrema of the dSIP distribution. (i) Tesseract OCR output for the card (f). — (j) – (n) Log power spectra of the card images (g), (d), (e), (f), and (h) (to obtain one-dimensional spectra, the two-dimensional spectra are averaged and rounded to the nearest integer frequency).

P12 – 322

record corresponds to which electronic record is unknown, and the two sets do not overlap. Consequently, the technical project objective of matching document images and electronic texts is preceded by a feasibility analysis. This requires a fast and appropriate classification of the cards. Here, "fast" concerns readying the technical resources, the human interaction with the data, and the lax requirements for classification quality and sophistication—hence, a *triage* task.

The dSIP values of the dataset images were measured and the cards located in the pattern–informativeness space. The ordering of the 1.2 million items was checked visually for perceptual consistency and found to be satisfactory; a sample of the results are presented in Fig. 19. The patterns



P12
–
323

FIGURE 19. Cards of one of the library catalog thematic sections, classified according to their structural information potential. The three highlighted cards are shown in higher resolution in Fig. 18d–f. The second card is the closest to 0 dSIP, i.e. considered as the most clustered.

in this sample are representative (*a*) in semantic terms, as the sample represents an entire bibliographical index section, containing both various header cards and reference cards (section CDU 00000001), (*b*) in document typological terms, as it contains printed and handwritten text, noisy images, and other significant features, and (*c*) in statistical terms,

as the sample is roughly uniformly distributed over the pattern spectrum (see the location of the three highlighted cards in the overall distribution of Fig. 18b). The ordering is meaningful as a typological categorization of bibliographical cards and suggests several potential matching strategies between physical and electronic records. The cards with clustered visual patterns are so because, usually, they exhibit greater typological information variety; in such cases, the matching algorithm could be tuned to privilege the logical structure as the matching criterion over linguistic matching. Cards with homogeneous patterns are more likely to contain longer coherent linguistic sequences, such as sentences, for the matching of which syntactic analysis would be preferable. Cards with low SIP value are either empty, are non-record section headers, or contain very little text; as a result, they may yield very poor matching quality, and therefore may best be visually inspected and excluded from matching.

The classification of the cards using the measure of structural information potential was appreciated by both software engineers and library managers for multiple reasons. It is independent of the OCR output; it produces useful results under conditions of uncertainty about the cards' content; moreover, it allows for rapid overview of a large digital image collection that would otherwise remain largely invisible, and further involves the human in the matching process. Thus, card triage by SIP became a complement to OCR in decision-making on tasks such as go/no-go for automatic record matching, selection of a card subset with an expected matching quality, estimation of expected matching quality, and evaluation of project resources (e.g., costs and duration of groundtruthing and quality control).

P12 – 324

Fig. 18a shows the color-coded dSIP values for the 1.2 million bibliographical cards in their physical sequence in the original library card trays. Runs of similar dSIP values are apparent, such as the conspicuous blue (i.e., low dSIP) central column corresponding to author indices and mostly empty cards. The heterogeneously colored area on the right-hand side corresponds to cards indexing placenames, which have varying degrees of visual density and clustering due to the variety of information sources and the continuous updating of the card information, involving many writing technologies and annotation layers. The visualization's pixels are interactively linked to the card images, so that a visual investigation may reveal that the blue run represents author index cards and can thus be removed from the matching process. This visualization is useful for a context-oriented analysis of the card dataset. If, for example, an equivalence has been established between dSIP values and expected matching quality, then it may be used to predict the matching quality of the various thematic classes of the cards.

Fig. 18b represents the dSIP values of the card dataset in histogram form. This aggregated data yields several insights. By extracting sample cards along the distribution to determine the patterns to which the values correspond, a visual and quantitative estimation can be made regarding the amount of various card types and the expected matching quality.

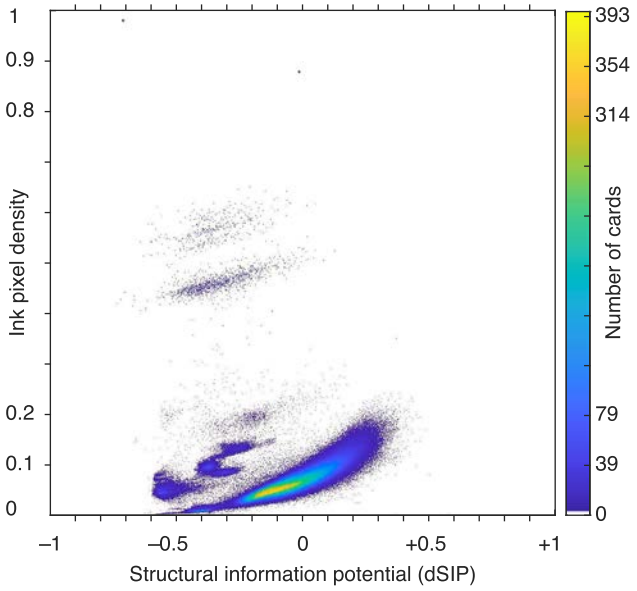


FIGURE 20. This graphic shows the density distribution of the structural information potential plotted against the ink pixel density in the digital images of the library card dataset. An "ink pixel" represents an inked area of the card surface, as opposed to the non-inked writing substrate; in the cards illustrating this article, ink pixels appear in black; noise modifies the groundtruth value of pixels. For legibility purposes, pixels above the density level of 0.4 ink pixels have been slightly dilated. Note the two pixels above level 0.8. The 1.2 million data points were aggregated for visualization in a 1000-by-1000-pixel raster.

It can be observed that the bulk of the cards have a clustered appearance, while cards with homogeneous information distribution are not predominant. A further operationalizable insight enabled by the histogram pertains to outlier detection (Fig. 18g, h). In this case, a first observation concerns the extreme dSIP values: scanned images with the objects of interest partially out of frame are found in the lower values, while the higher values contain images with strong noise. Second, the left-most cluster of the distribution predominantly comprises section headers. While these do not contain bibliographical information, and should thus be excluded from the matching process to avoid a detrimental impact, these cards may be useful for matching, since they identify the topic to which the subsequent cards belong. This topic may then be related to those extracted from the electronic records, thereby increasing the probability of correct matches. The bivariate plot in Fig. 20 supports a finer analysis of the histogram, specifically that it is the result of a mixture of pattern classes characterized by different ink densities, as well as non-linear and loose covariation between the amount and distribution of ink on the cards.

The question might be asked as to whether simply classifying the binarized card images by the amount of black pixels (i.e. ink) would not provide insights similar to the structural information potential. Fig. 20 demonstrates that this is not the case: documents with identical ink density can have very different ink distributions, and vice-versa, as a result of distinct clusters with irregular shape and spread. May, then,

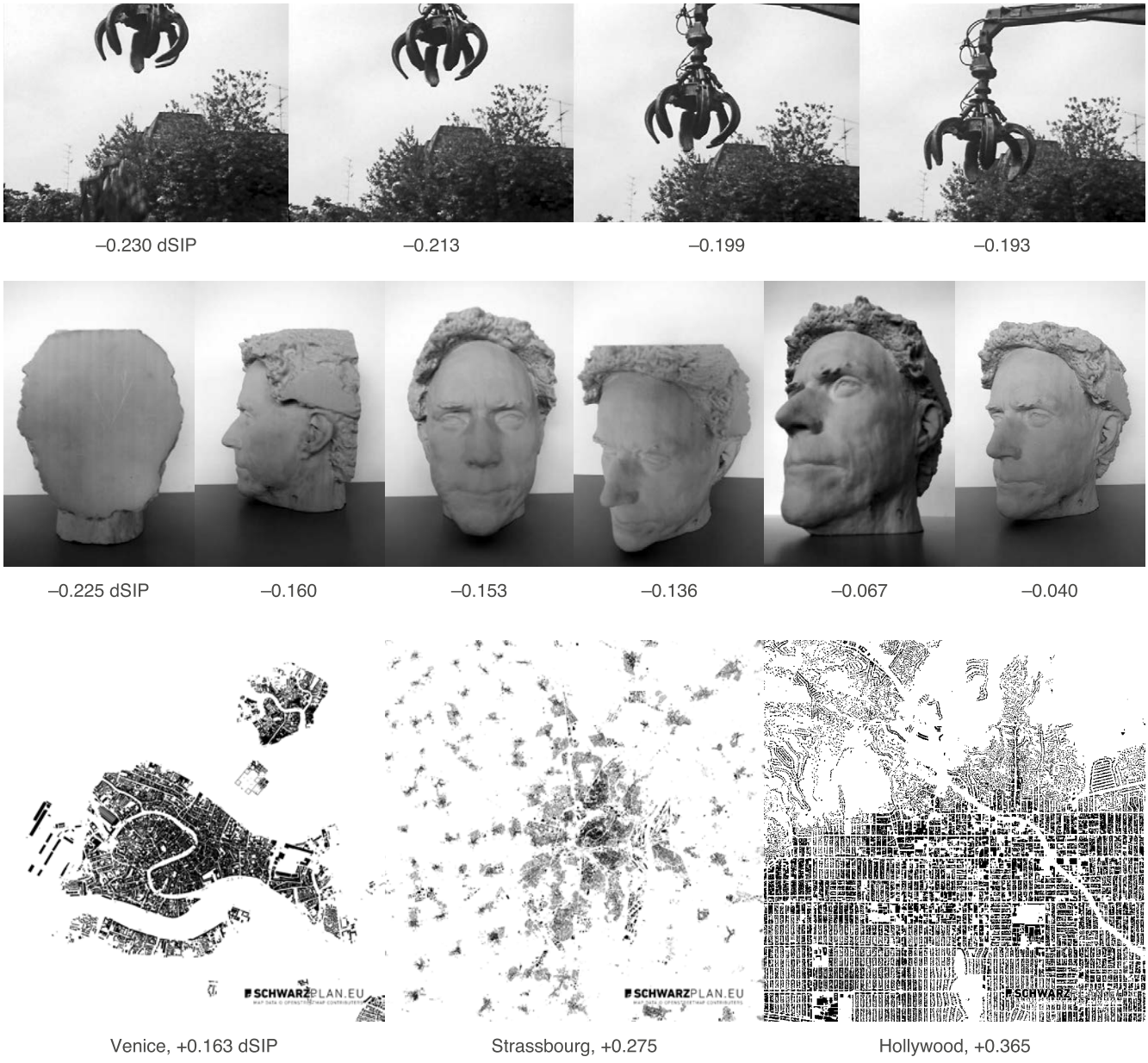


FIGURE 21. Application of the structural information potential to video frames (top), three-dimensional objects (middle), and urban layouts (bottom). — Credits: video: Trigon Film; sculpture: unknown artist; plans: SCHWARZPLAN.eu.

P12 – 325

a classification by the number of recognized characters be a sufficient estimate of matching quality? As illustrated by Fig. 18i, this would not be effective either; even cards with rich linguistic contextual information can result in very few characters being identified.

From a machine learning perspective, the classification of document images according to SIP may be utilized to optimize the sampling of the training and test images. For example, a uniform random sampling of the card dataset (i.e. the typical sampling procedure) may result in an underrepresentation of less numerous but semantically important classes, such as scanning errors and cards with strong noise or that are largely empty, which are visible at the distribution extrema of the dataset SIP visualization in Fig. 18b. However, by using the SIP distribution in conjunction with visual exploratory data analysis, one can apply different sampling rates to different parts of the dataset, thereby acquiring sufficient samples for training the algorithms with the best possible accuracy given the data.

In conclusion, the measurement of structural information potential, coupled with interactive data analysis, provided computer scientists with an inexpensive tool to support record matching, particularly for sample selection and expected quality estimation. It also aided library managers in making faster and more informed decisions regarding an information technology project on issues such as go/no go, expected

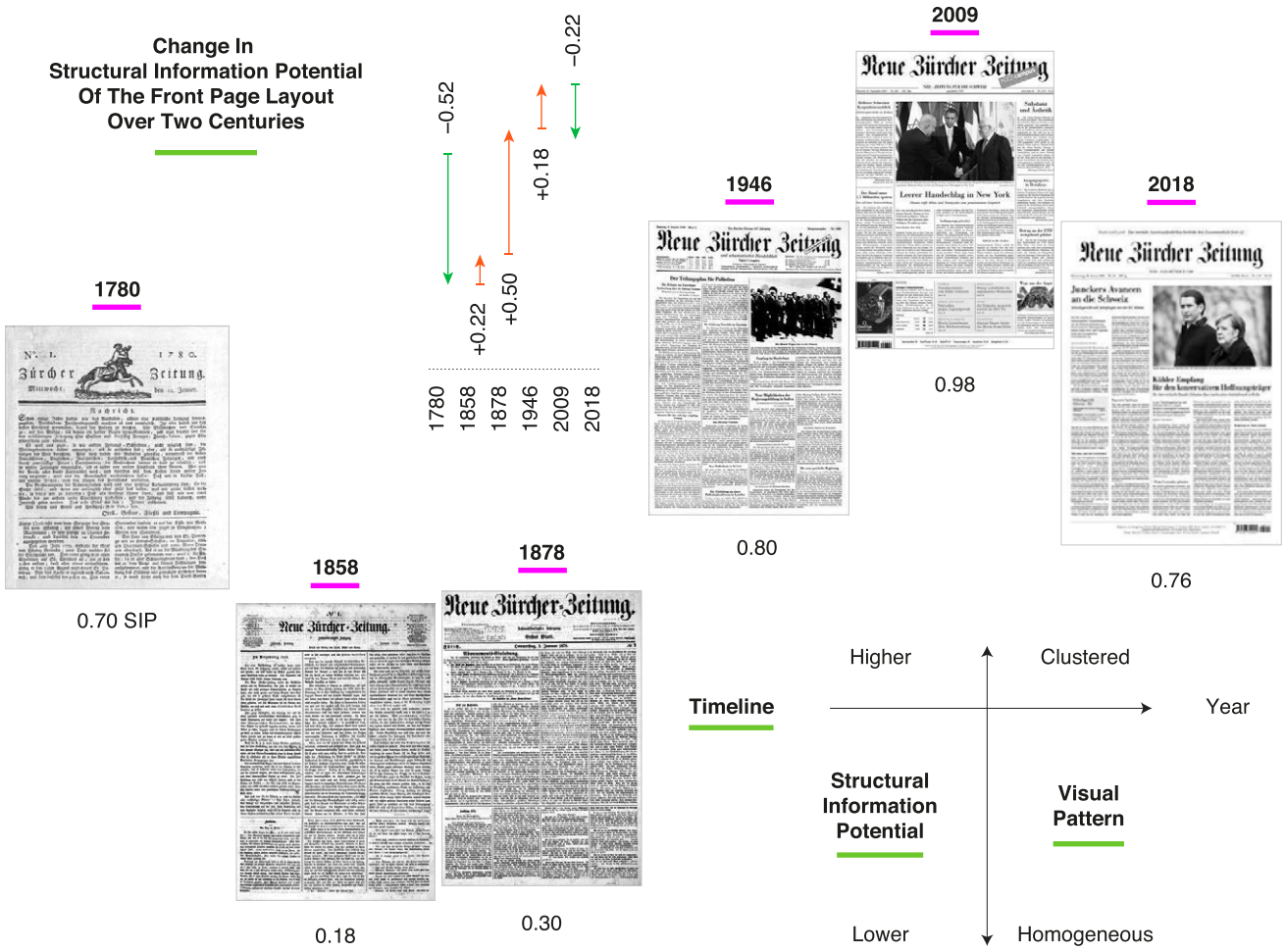


FIGURE 22. Timeline of the layout evolution of the Swiss *Neue Zürcher Zeitung* newspaper, with the SIP value of the front page. — Credits: Neue Zürcher Zeitung.

P12 – 326

quality, and resource estimation. This case study is an example of the usefulness of SIP for both quickly extracting information from patterns and acting upon this information.

C. GENERALIZATION

The generic nature of the structural information potential concept as defining a pattern distribution space allows it to be applied to a variety of tasks and data types. Its main utility advocated in this article is as a quantitative measure of informativeness, which makes it appropriate for triage tasks, as discussed above. Notably, however, the above case study found that SIP was also useful as part of a decision-making process, involving the estimation of OCR quality and needed resources, and as a measure of image quality, given its ability to distinguish images with uniform noise and erroneously framed scans. Furthermore, SIP offers a simple computational solution for classifying document pages according to the predominance of text, the presence of illustrations, the number of post-production visual artifacts (e.g., annotations, stamps, signatures) and the amount of noise (i.e., traces of document degradation); these abilities may be translated into document navigation functionalities and implemented in document readers. The characterization of document layouts is also valuable to historians, as it supports a quantitative analysis of the evolution of written communication.

Fig. 21 illustrates the application of SIP to data types other than documents and tasks other than triage. According to the SIP concept, the patterns with dSIP values closest to 0 are the most informative. The top row illustrates the case of image retrieval from a large set of very similar samples. The SIP-based automatic extraction of keyframes from videos facilitates the identification of frames with potentially high information content. The frame with minimal dSIP value indeed shows more visual details than other frames, such as text, cables, and pistons. In the middle row, SIP is used to determine the most informative point of view of a three-dimensional object, perhaps by an examining robotic camera. For a head sculpture this is the three-quarter view, which integrates anatomical elements of both face and profile in a single shot. The bottom row shows how SIP may

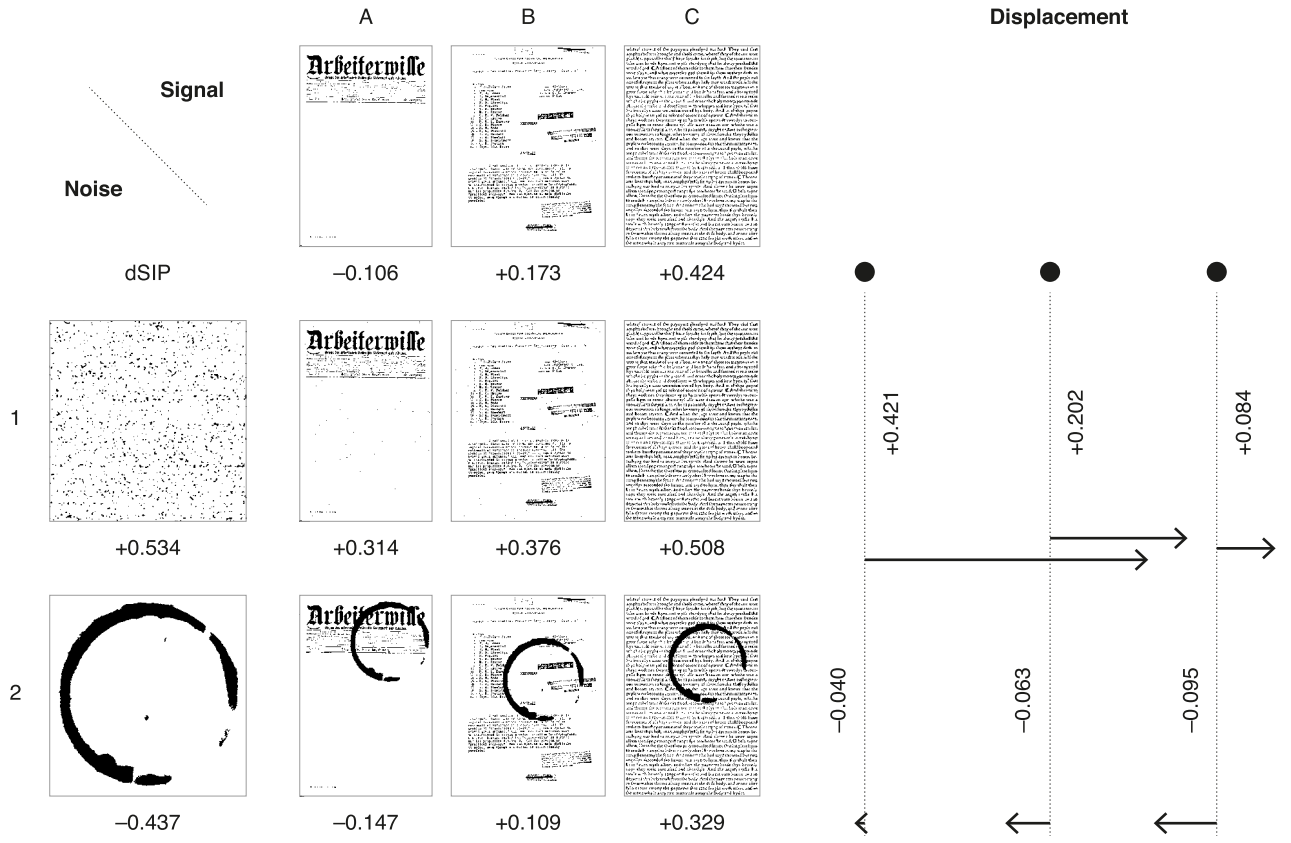


FIGURE 23. The figure illustrates the impact of two common document noise types on the dSIP values of basic document patterns. — Top row: Three noise-free documents, with uniform (A), clustered (B), and regular (C) patterns. — Left column: (1) Digital camera noise (obtained by taking a picture with the camera lens covered, due to which the pattern registered in the digital file is not that of incident light, but rather the results of electronic, thermal, software, and other similar artifacts; postprocessing consisted in median filtering and binarization). (2) Coffee stain. Both left-most images are enlarged for clarity. — Center: Pair-wise mixing of signals and noise, with dSIP values indicated. — Right: The arrows show the direction and magnitude of the change in dSIP due to the addition of noise to the signals.

P12 – 327

be used for classifying aerial images of urban settlements on the uniform–clustered–regular continuum. The impact of the natural topography and cultural–historical factors on the hierarchical urban clustering come readily to mind when contemplating this classification; for example, the meandering water channels of Venice versus the commercial hub of medieval Strasbourg reflected in a spiderweb road network versus the plain of Los Angeles making a flat urban grid affordable.

Fig. 22 showcases the usefulness of the SIP measure to research in humanities on the history of newspaper layout. The figure shows selected front pages of the leading Swiss newspaper *Neue Zürcher Zeitung*, first published in 1780. The principal transformation of the layout concerns the increase in visual and semantic hierarchical structuring through textual elements, the addition of pictures, and greater use of empty space. Some of the evolutionary factors are clearly technological, such as the invention of photography, while others are societal. For example, the accelerating rhythm of modern living encourages the development of visual communication methods that facilitate faster navigation of information, prompting the shift from homogeneous to hierarchically clustered layouts. SIP provides a means to quantify and compare

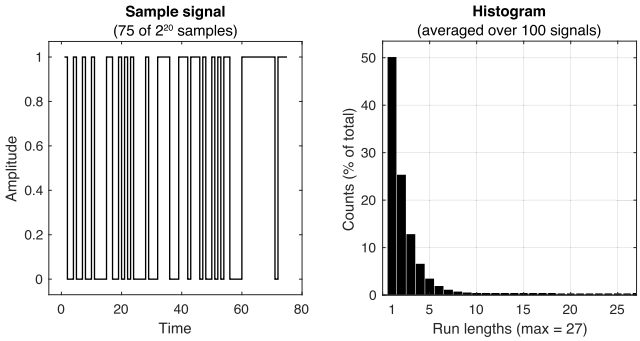


FIGURE 24. Detail of binary white noise signal (left), and the averaged histogram of the lengths of consecutive samples with identical values (runs) obtained from 100 such signals (right).

changes both within a single newspaper and between different publications.

Moving beyond documents, the classification into compact, clustered, and homogeneous distributions is of direct interest to materials science. In telecommunications, the spatial or temporal distance between three senders/receivers—a basic configuration of networks that recalls our investigation

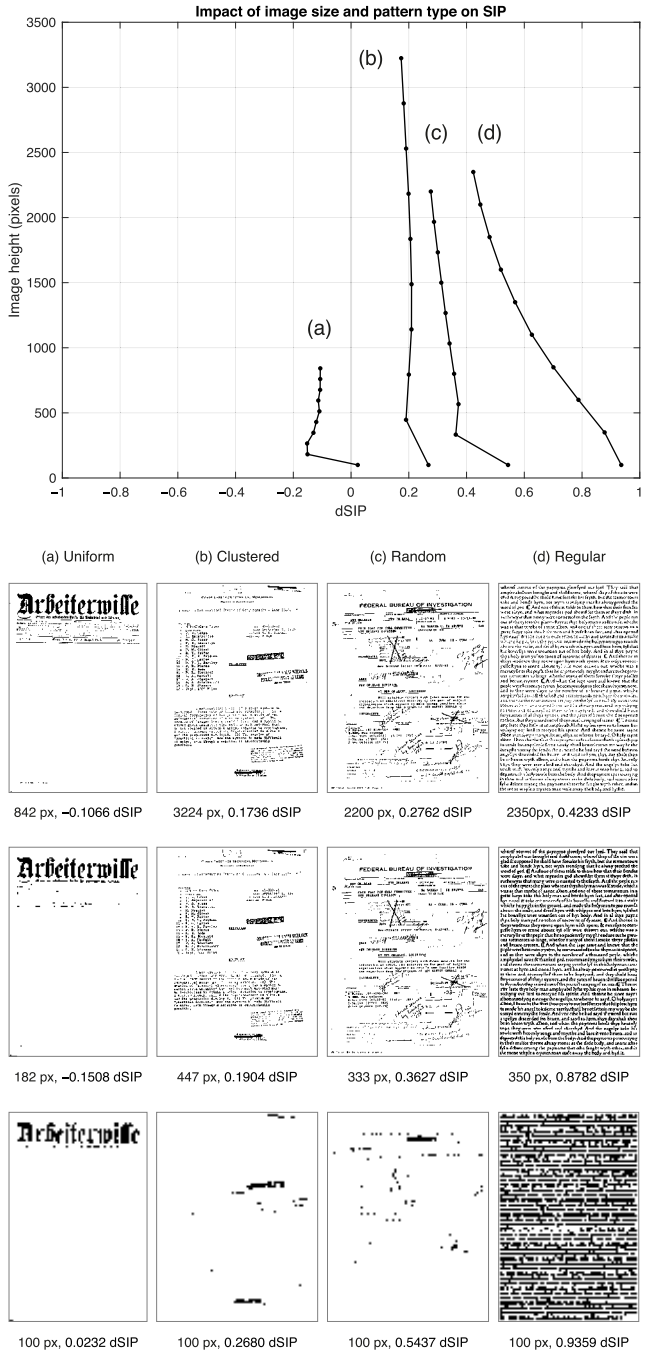


FIGURE 25. The above diagram illustrates how downscaling raster images affects their dSIP according to their pattern type. The images displayed are those with the highest and two lowest heights.

of triangles in the preceding section—affects data transmission quality and the propagation patterns of messages. This ''layout'' phenomenon is fundamentally similar in other types of networks, such as transportation, social, or epidemic. Symmetry plays a great role in physics (for example, for the growth of crystals); however, irregular configurations (minimally redundant) may be equally interesting to study. As for a table with minimally redundant edge lengths, to take an example from psychology, this is a recipe for emotional stress for a party of three sitting at its respective corners.

D. DECISION-MAKING

After having encountered a great variety of applications of structural information potential throughout the article, let us here briefly systematize how this concept might be the basis of ''rapid decision-making for critical matters under conditions of uncertainty and with limited resources'', as stated above in the introduction.

The structural information potential enables decisions to be made through the classification of patterns in the uniform–clustered–regular space, which relates to their degree of informativeness. What the users may do with this information is, however, largely task-specific, as suggested by the diversity of applications and contexts encountered in this article.

SIP-based decision-making may be rapid, first of all, from the semantic point of view, since the patterns are ordered by level of informativeness, which is desirable for the triage task. The SIP approach is also rapid from a technical perspective; this is because its implementation and use require limited resources, and further avoids expensive data preprocessing and other frequently unreliable types of processing, such as document recognition. Finally, SIP is rapid because it may be used to automatically select data with specific levels of informativeness or pattern types.

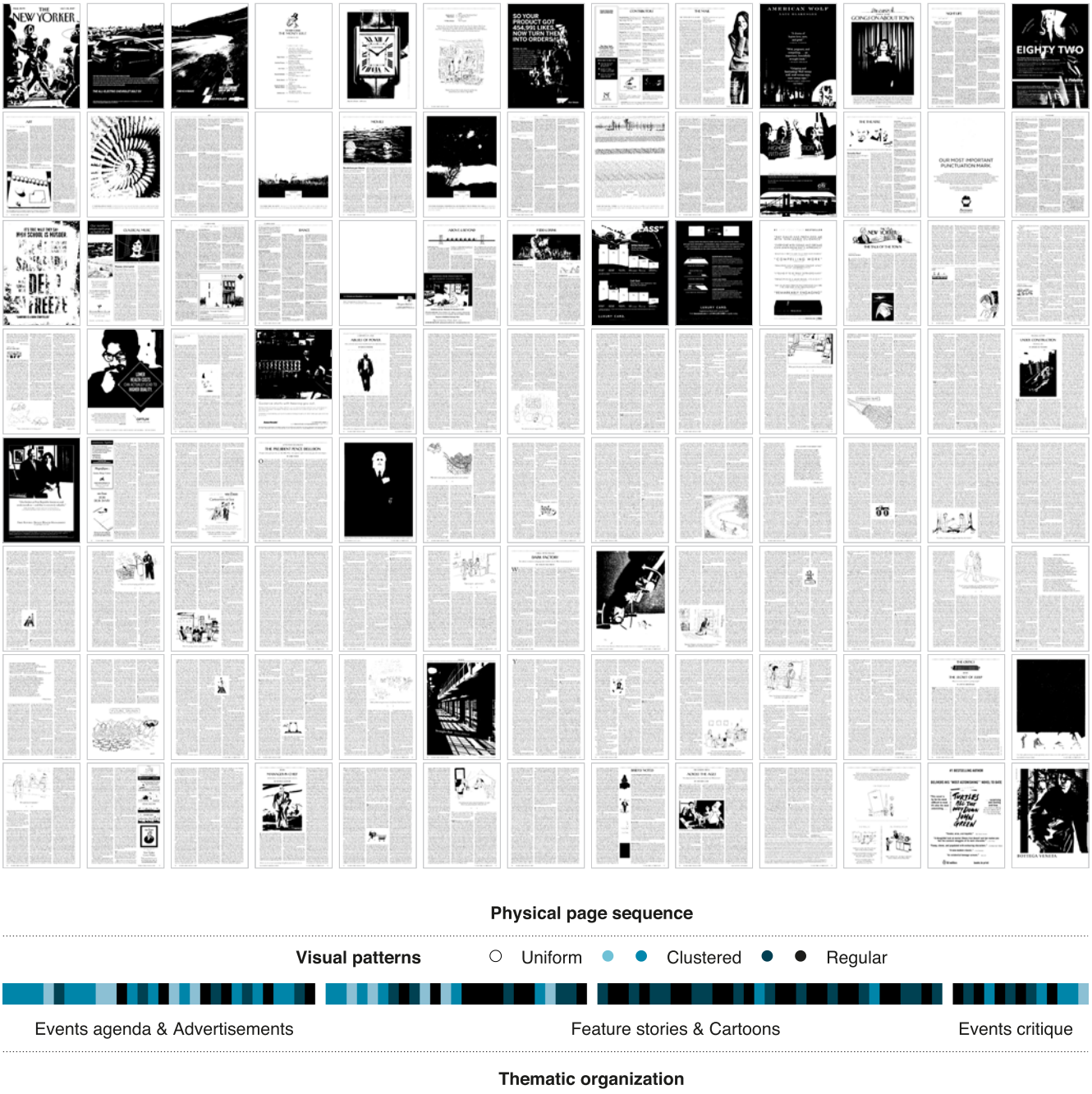
Structural information potential is a quantitative and visual tool for exploratory data analysis, i.e. situations in which the data content and task specification are uncertain. For such cases, there is no single preexisting quantitative decision-making formula that can be applied; serendipity needs to be acknowledged and actively sought, and insights emerge from the subjective interaction between humans and data. Within its limits, SIP enables users to act efficiently in an uncertain environment.

V. DISCUSSION

This section provides an in-depth discussion of some important rationales and implications of the SIP measurement. It thus helps better understand the behavior of this instrument, and opens directions for future research.

A. NOISE

How the addition of noise to a pattern affects its structural information potential depends on a number of factors, primarily the type, extent, and location of the noise, as well as the pattern type of the signal. Fig. 23 illustrates the interaction between two types of noise common in documents and the three basic patterns: uniform, clustered, and regular. Imaging artifacts, such as camera noise (depicted here), may introduce quasi-uniform, quasi-random noise that extends over the entire document and has high spatial frequency. This noise type invariably increases the regularity of the pattern and its dSIP value. Stains and blotches have limited extent and lower spatial frequency, such that their impact is largely determined by their location: they may make a clustered



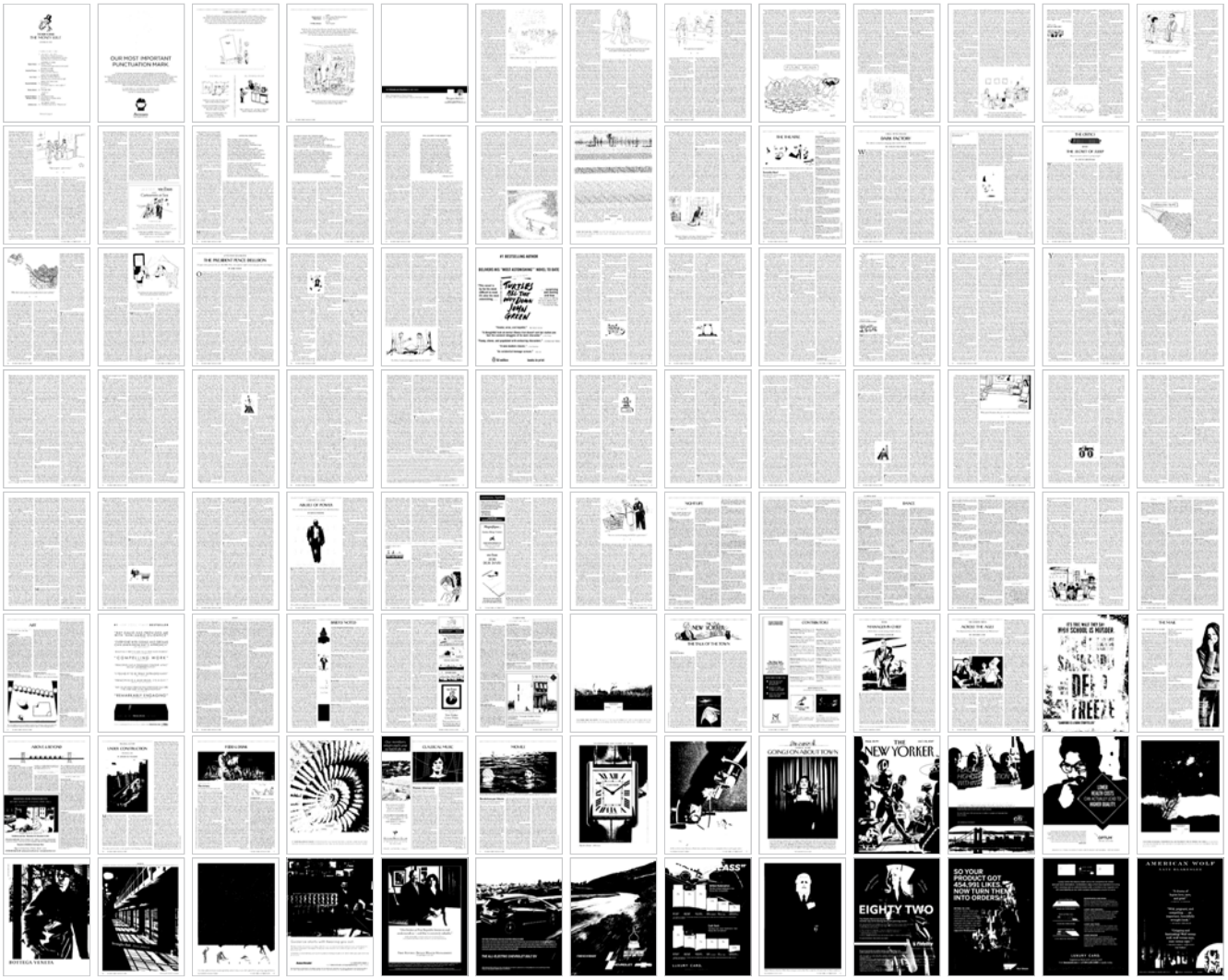
P12 – 329

FIGURE 26. Original page order and document structure visualization — Reading order: left to right, top to bottom. The color-coded ribbon indicates the pattern type of each page according to its structural information potential. Observable is a stretch extending to a third of the document made from mixed uniform, clustered, and homogeneous page patterns, followed by a sequence where homogeneous pages predominate, and ending with several clustered pages. These visual patterns correspond to semantic patterns with different granularity levels: many short agenda items and full-page advertisements, text-based stories interspread with cartoons, and a critique section on multiple topics. The cover is a typical example of a clustered pattern, due to objects of different sizes, and a perspectival view.

pattern appear even more clustered and a uniform pattern more compact. When appearing in conjunction with a regular pattern, however, they invariably increase the clusteredness of the pattern irrespective of location. As a practical take-away message, understanding the interaction between noise and signals aids in interpreting SIP values and selecting data.

B. RANDOMNESS

The attentive reader may have noticed that random patterns have remarkably high dSIP values. Fig. 6, for instance, shows an example where dSIP is practically maximal (0.99). Given the expectation that a regular pattern has maximal value (1), how is this phenomenon possible, and what does it mean for the theory and practice of SIP?



P12 – 330

FIGURE 27. Page order by ratio of ink pixels and page area.

An infinite signal with uniform random binary amplitude values and uncorrelated samples (white noise) has an equal amount of each possible ''run length'', i.e. consecutive samples with identical value. However, a finite signal exhibits a different number of runs of equal length. Specifically, the distribution is defined as:

$$R_i \approx N/2^i, \ \sum_{i=1}^{k} R_i = N, \tag{13}$$

where R is the number of runs of length i, N is the signal length, and k is the maximum run length (Fig. 24). This means that, theoretically, half of all two consecutive samples — a substantial share — have alternating values, just as in the case of a regular pattern: 0, 1, 0, 1, 0, 1, ... Moreover, a finite signal has a maximal run length, which is quite small; e.g., 18 for 1000 samples, and 27 for 2^{20} samples. Consequently, the extent of uniform sequences is very limited in respect to the signal length, further increasing the homogeneity and redundancy of the pattern. This phenomenon is the underlying reason for random patterns having dSIP values very close to those of regular patterns.

C. SIZE

It is an inherent aspect of quantization that modifying the resolution of raster images affects the shape of the pixel structures in the images; in other words, unlike vector graphics, shapes in bitmap graphics are not scale-invariant. Therefore, their SIP will also change with scale. The nature of this change depends on the pattern type: downscaling increases the homogeneity of homogeneous patterns (dSIP $>$ 0), has a relatively low impact on clustered patterns (dSIP $\approx$ 0), and increases the uniformity of uniform patterns (dSIP $<$ 0; Fig. 25). These changes are the most pronounced at small image resolutions, i.e. those below about 200 pixels. When image downsampling is performed for performance reasons, it is preferable to use images with height or width beyond the said threshold.

P12 – 331

FIGURE 28. Page order by approximate entropy.

D. LIMITATIONS

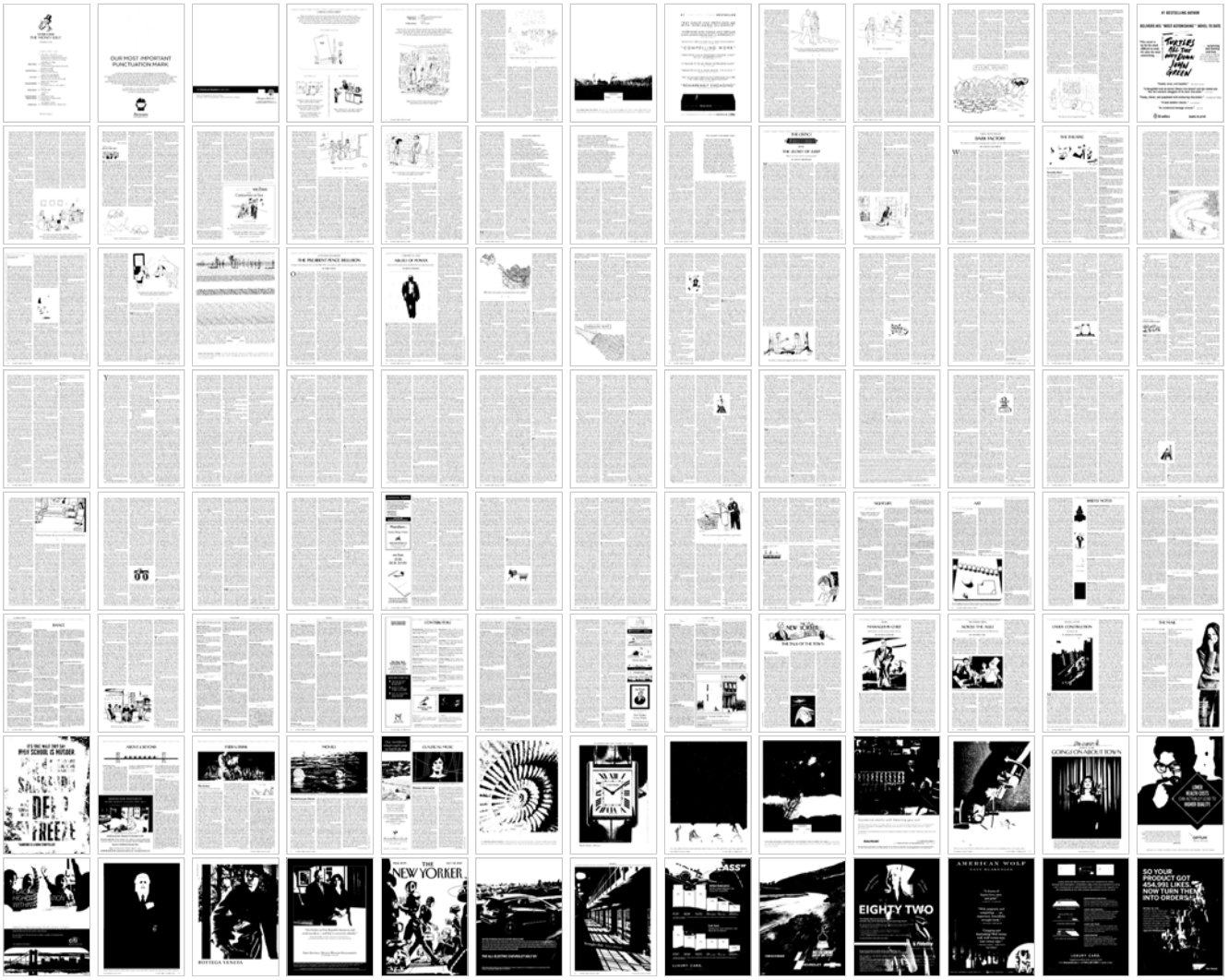
Two patterns generate aberrant SIP values. These are, however, peculiar enough to not be of concern in most practical cases. The first is the perfectly regular structure of the checkerboard and stripes when aligned with the orthogonal image raster. At half-cycles of one and two pixels, their frequency spectra are impulses, hence the entropy has value zero rather than the expected value one. The second problematic pattern is a single square aligned with the image raster; due to the step function of the function having the sinc function $\mathrm{sinc}(a) = sin(a\pi)/(a\pi)$ as frequency domain pair [177, pp. 212–215], its harmonics increase entropy to nearly one, contrary to what would be expected from a pattern with a large uniform surface.

E. BINARIZATION

It may be useful to develop a SIP measurement without binarization. One reason is that it would remove a certain degree of uncontrollable data distortion inherent in the strong reduction of its dynamic range. Another reason is that it would facilitate the application of the SIP method to signals, for which binarization constitutes a drastic distortion. However, developing a method for data with large dynamic ranges creates the need for a second method for binary data (a common data type in many application domains, including document processing). Here we can observe an advantage of the binarization step, which enables both intensity and binary data to be handled with a single method.

F. METAMERISM

Because of dimensionality reduction (from two dimensions to a scalar in the case of images), many different patters will have an identical measured value (the metamerism effect). This is one reason why triage is a good application for the SIP method presented here, namely because it is tolerant to imprecision. One can imagine incorporating parameters into the SIP method to allow the description of various pattern features (for example, pixel density).



P12 – 332

FIGURE 29. Page order by spectral entropy.

G. PSYCHOPYSICS

As the pattern ordering resulting from the SIP is primarily intended to have humans as end users, it is desirable to investigate the relevance of modifying the method to account for human pattern perception. Configurations are, however, difficult psychophysical research stimuli; this is more so the case for real data, such as documents, where semantics, aesthetics, and user experience (among many other factors) play an important role in their evaluation.

H. PHASE

The reader may have noticed that the SIP measurement method disregards the phase information arising from the image transform from the spatial to the frequency domain. This may be inconsequential for speech, where information is largely carried by the frequency spectrum [180], [197], [198, pp. 355–358]; however, phase is critical for image processing, in that radically different patterns result from identical frequency spectra with different phases, while edge location and strength are moreover strongly phase-dependent [180], [199]–[202, pp. 355–358]. This difference may explain to some extent why the spectral methods discussed in the related work section are successful in audio processing but less commonly used for image classification. From a practical point of view, however, the examples presented in this article demonstrate the robustness of the method when phase is discarded. This surprising fact demands a theoretical explanation.

As a first remark, it has been experimentally demonstrated that some binary image types may be reconstructed from frequency magnitude only and zero or random phase, and that this may further depend on the presence or absence of a single pixel [203]. While this phenomenon has yet to be elucidated, it constitutes evidence in favor of magnitude-only image analysis. Second, the extrema of the uniform–clustered–regular pattern space are phase-independent

FIGURE 30. Page order by spectral flatness.

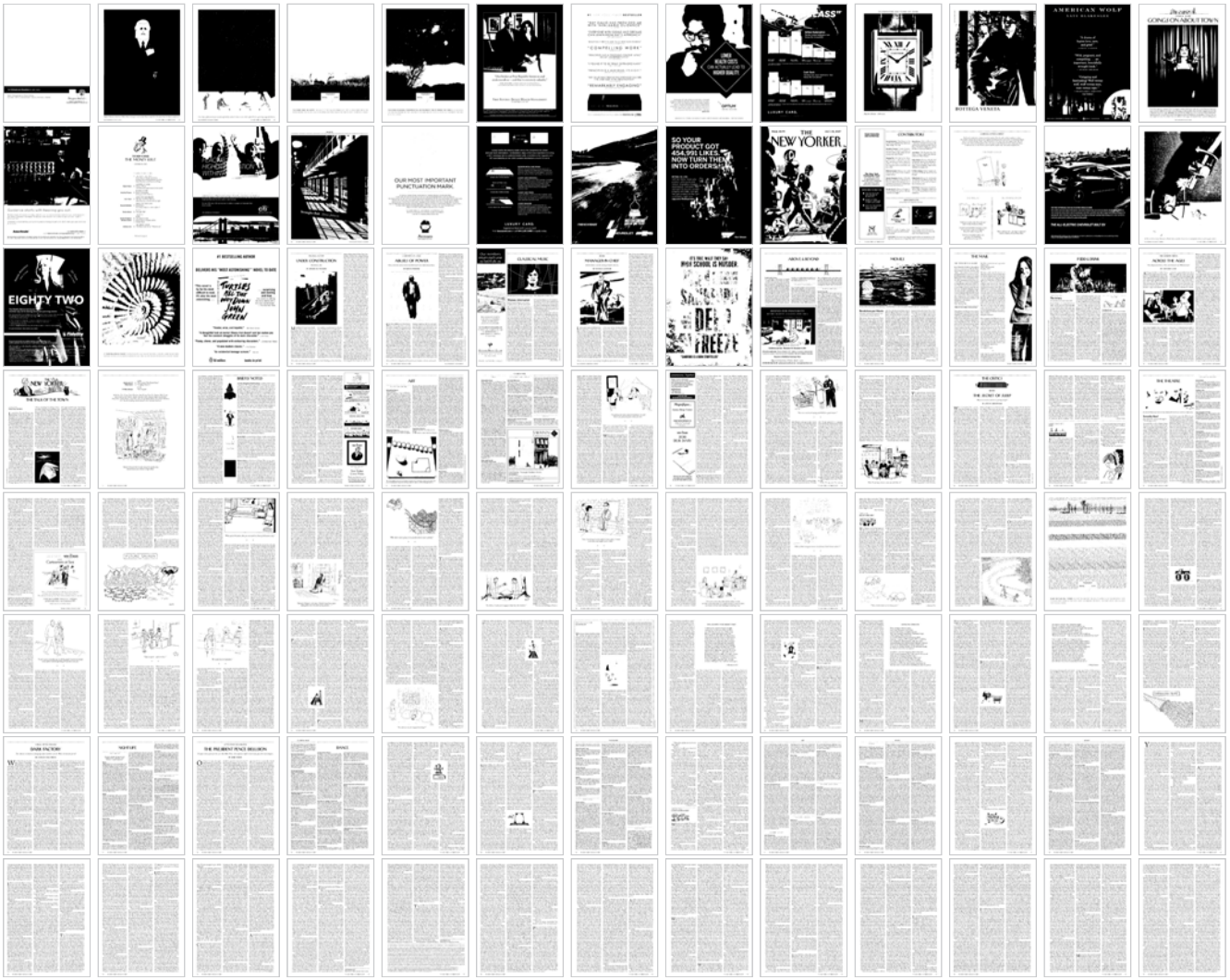
[177, pp. 74–81, 106–107], which reduces the problem to one of explaining why clustered patterns occur around the center of the pattern space. We may recall that fractals (the patterns that maximize clustering) depend in their overall shape on the exponent of the power distribution of the frequency spectrum and a random phase. In fact, any clustered image pattern will exhibit some degree of power magnitude distribution. Moreover, the multitude of harmonics introduced by binarization create a phase that is increasingly better modelled by a random distribution. Taken together, these aspects suggest that within the limits of finite and discrete images, patterns approach fractality at and in the vicinity of maximal SIP or minimal spatial redundancy. This may be an important reason why it is practically possible to robustly order images along the uniform–clustered–regular pattern space on the basis of magnitude alone.

I. EPISTEMOLOGY

One fascinating mathematical aspect of minimal structural redundancy is how one might go about thinking about it; in other words, its epistemology. Specifically, it is richer to conceive the problem as the *design* of fluorescent structures than their *discovery*. Asking "What might the definition of minimal structural redundancy be?", rather than "Which is the structure with the least redundancy?", explicitly embeds the possibility of multiple answers into the inquiry process, as well as contextualizing the problem in respect to the questioner, the data, and the application. For example, a fluorescent graph becomes regular in terms of edge length when the number of vertices tends to infinity! If this behavior is not desired, a new definition of minimal redundancy may be created (for example, stipulating only local fluorescence). This same experimental conceptualization of mathematical definitions was followed by Gary Chartrand, Paul Erddős and Ortrud Oellermann in their article "How to Define an Irregular Graph": "In research, the goal is not only to come up with a definition that seems natural but to arrive at a class of graphs with interesting, and perhaps even some surprising, properties." [85, pp. 39].

VI. CONCLUSION

This article has introduced structural information potential (SIP), a measure of information based on pattern



P12 – 334

FIGURE 31. Page order by structural information potential.

configuration. Its utility was illustrated through a real-life case study for the task of document image triage. On this task, SIP performs better than other methods in both mathematical and perceptual terms.

The main theoretical significance of the work consists in (*a*) the development of a formalism that defines the uniform–clustered–regular pattern-informativeness space, which organizes fundamental pattern types in a mathematically and perceptually coherent fashion and relates them to an information potential, and (*b*) the development of a conceptual basis and analytical methods for the identification of shapes and patterns with minimal structural redundancy.

In practical terms, structural information potential is a useful classification method for triage-like conditions, characterized by decision-making under conditions of uncertainty and time pressure, when it becomes efficient to generate information about content through the analysis of structures. The generic nature of SIP makes it appropriate for many other applications, such as image quality assessment (to detect noisy and erroneously imaged data), predicting OCR output quality before applying OCR, identifying informative keyframes in video streams, and as a document navigation functionality. Given the generic nature of the structural information potential, the author believes that applications to fields as diverse as mathematics, physics, telecommunication, and psychology may be discovered in the future.

ACKNOWLEDGMENT

The author is grateful to Rolf Ingold, of the Informatics Department, University of Fribourg, Switzerland, for having facilitated the funding of the Open Access publication of this article. The author would like to thank Andreas Fischer of the University of Applied Sciences and Arts Western Switzerland (HES-SO), Fribourg, for sharing many ''reflective reflections'', and for having offered to work on the library card dataset. Irvin Schick, formerly with Bolt Beranek and Newman, Cambridge, MA, USA, now at the École des Hautes Études en Sciences Sociales, Paris, France, has been, over the years and continents, a reliable devil's advocate for testing the author's hypotheses. The anonymous reviewers are equally

thanked for having contributed through their comments to improve the manuscript.

REFERENCES

[1] T. Samara, *Making and Breaking the Grid: A Graphic Design Layout Workshop*. Gloucester, MA, USA: Rockport, 2002.
[2] K. Elam, *Grid Systems: Principles of Organizing Type*. New York, NY, USA: Princeton Architectural Press, 2004.
[3] F. Franchi, *Designing News*. Berlin, Germany: Gestalten, 2013.
[4] B. Schwesinger, *Formulare Gestalten*. Mainz, Germany: Hermann Schmidt, 2007.
[5] K. A. Schriver, *Dynamics in Document Design: Creating Text for Readers*. Chichester, U.K.: Wiley, 1997.
[6] A. Dillon, *Designing Usable Electronic Text: Ergonomic Aspects of Human Information Usage*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2004.
[7] R. Pettersson, *Information Design Theories*. Tullinge: Institute for infology, 2016.
[8] F. Richaudeau, *La lisibilité: Langage, Typographie, Signes Lecture*. Paris, France: Gauthier, 1969.
[9] P. Cleveland, "How much visual power can a magazine take?" *Des. Stud.*, vol. 26, no. 3, pp. 271–317, May 2005.
[10] J. R. Pomerantz and M. Kubovy, "Theoretical approaches to perceptual organization: Simplicity and likelihood principles," in *Handbook of Perception and Human Performance: Cognitive Processes and Performance*, vol. 2, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds. New York, NY, USA: Wiley, 1986, ch. 36, pp. 1–46.
[11] A.-M. Christin, *A History of Writing: From Hieroglyph to Multimedia*. Paris, France: Flammarion, 2002.
[12] R. Kinross, *Modern Typography: An Essay in Critical History*, 2nd ed. London, U.K.: Hyphen Press, 2004.
[13] G. Dowding, *Finer Points in the Spacing & Arrangement of Type*, 3rd ed. Vancouver, BC, Canada: Hartley & Marks, 2008.
[14] J. Hochuli, *Detail in Typography*. Paris, France: Éditions B42, 2008.
[15] *Hz-Program: Micro-Typography for Advanced Typesetting*. Hamburg, Germany: URW Software & Type GmbH, 1993.
[16] H. T. Thành, "Micro-typographic extensions to the TEX typesetting system," Ph.D. dissertation, Masaryk Univ., Brno, Czech Republic, Oct. 2000.
[17] B. Shaw, *Bernard Shaw on Modern Typography*. Cleveland, OH, USA: The Printing Press, 1915.
[18] J. Aczél and Daróczy, *On Measures of Information and Their Characterizations*. New York, NY, USA: Academic, 1975.
[19] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
[20] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, Berkeley, CA, USA, vol. 4.1, Jun./Jul. 1960, pp. 547–561.
[21] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *J. Stat. Phys.*, vol. 52, nos. 1–2, pp. 479–487, 1988.
[22] A. E. Magurran, *Measuring Biological Diversity*. Oxford, U.K.: Blackwell, 2003.
[23] T. Leinster and M. Meckes, "Maximizing diversity in biology and beyond," *Entropy*, vol. 18, no. 3, p. 88, Mar. 2016.
[24] S. Watanabe, *Pattern Recognition: Human and Mechanical*. New York, NY, USA: Wiley, 1985.
[25] Y. Y. Tang, *Document Analysis and Recognition With Wavelet and Fractal Theories*. Singapore: World Scientific, 2012.
[26] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Nat. Acad. Sci. USA*, vol. 88, pp. 2297–2301, Mar. 1991.
[27] S. Pincus, "Approximate entropy (ApEn) as a complexity measure," *Chaos*, vol. 5, no. 1, pp. 110–117, Mar. 1995.
[28] S. Pincus and R. E. Kalman, "Not all (possibly) 'random' sequences are created equal," *Proc. Nat. Acad. Sci. USA*, vol. 94, no. 8, pp. 3513–3518, Apr. 1997.
[29] K. H. Chon, C. G. Scully, and S. Lu, "Approximate entropy for all signals," *IEEE Eng. Med. Biol. Mag.*, vol. 28, no. 6, pp. 18–23, Nov./Dec. 2009.
[30] F. Kaffashi, R. Foglyano, C. G. Wilson, and K. A. Loparo, "The effect of time delay on approximate & sample entropy calculations," *Phys. D, Nonlinear Phenomena*, vol. 237, no. 23, pp. 3069–3074, Dec. 2008.
[31] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000.
[32] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 2, pp. 266–272, Jun. 2007.
[33] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, Jul. 2002, Art. no. 068102.
[34] Z. Liang, Y. Wang, X. Sun, D. Li, L. J. Voss, J. W. Sleigh, S. Hagihira, and X. Li, "EEG entropy measures in anesthesia," *Frontiers Comput. Neurosci.*, vol. 9, no. 16, pp. 1–17, 2015.
[35] M. Borowska, "Entropy-based algorithms in the analysis of biomedical signals," *Stud. Log., Grammar Rhetoric*, vol. 43, no. 1, pp. 21–32, 2015.
[36] O. A. Rosso, R. Ospina, and A. C. Frery, "Classification and verification of handwritten signatures with time causal information theory quantifiers," *PLoS ONE*, vol. 11, no. 12, Dec. 2016, Art. no. e0166868.
[37] H. Aronowitz and D. Burshtein, "Efficient speaker recognition using approximated cross entropy (ACE)," *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, no. 7, pp. 2033–2043, Sep. 2007.
[38] Y. Xingwei, Z. Jun, L. Dawei, and W. Jianwei, "Radar jamming detection based on approximate entropy and moving-cut approximate entropy," in *Proc. IET Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, Shenzhen, China, Dec. 2012, pp. 1–6.
[39] R. Hernández-Pérez, L. Guzmán-Vargas, A. Ramírez-Rojas, and F. Angulo-Brown, "Pattern synchrony in electrical signals related to earthquake activity," *Phys. A, Stat. Mech. Appl.*, vol. 389, no. 6, pp. 1239–1252, Mar. 2010.
[40] G. S. Raman and R. T. Subhalakshmi, "Active steganalysis based on adapted lempel-ziv complexity and approximate entropy estimation," in *Proc. IEEE Conf. Inf. Commun. Technol. (ICT)*, Thuckalay, India, Apr. 2013, pp. 917–922.
[41] M. Spindler, "On concepts of randomness in finite sequences," Ph.D. dissertation, Swiss Federal Inst. Technol. (ETHZ), Zürich, Switzerland, 2009.
[42] A. Carpi and A. de Luca, "Uniform words," *Adv. Appl. Math.*, vol. 32, no. 3, pp. 485–522, Apr. 2004.
[43] B. H. Singer and S. Pincus, "Irregular arrays and randomization," *Proc. Nat. Acad. Sci. USA*, vol. 95, no. 4, pp. 1363–1368, Feb. 1998.
[44] C. J. Moore, "A threshold structure metric for medical image interrogation: The 2D extension of approximate entropy," in *Proc. 20th Int. Conf. Inf. Vis. (IV)*, Lisbon, Portugal, Jul. 2016, pp. 336–341.
[45] A. L. Goldberger, C.-K. Penga, and L. A. Lipsitz, "What is physiologic complexity and how does it change with aging and disease?" *Neurobiol. Aging*, vol. 23, no. 1, pp. 23–26, 2002.
[46] H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and Fractals: New Frontiers of Science*, 2nd ed. New York, NY, USA: Springer, 1992.
[47] T. Okabe, "Biophysical optimality of the golden angle in phyllotaxis," *Sci. Rep.*, vol. 5, no. 1, p. 15358, Dec. 2015.
[48] T. Okabe, "Vascular phyllotaxis transition and an evolutionary mechanism of phyllotaxis," 2012, *arXiv:1207.2838*.
[49] J. C. Russ, *Fractal Surfaces*. Boca Raton, FL, USA: Plenum Press, 1994.
[50] H. E. Schepers, J. H. G. M. van Beek, and J. B. Bassingthwaighte, "Four methods to estimate the fractal dimension from self-affine signals (medical application)," *IEEE Eng. Med. Biol. Mag.*, vol. 11, no. 2, pp. 57–64, Jun. 1992.
[51] H.-O. Peitgen and D. Saupe, *The Science of Fractal Images*. New York, NY, USA: Springer, 1988.
[52] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, 2009.
[53] S. E. Hough, "On the use of spectral methods for the determination of fractal dimension," *Geophys. Res. Lett.*, vol. 16, no. 7, pp. 673–676, Jul. 1989.
[54] G. E. Powell and I. C. Percival, "A spectral entropy method for distinguishing regular and irregular motion of Hamiltonian systems," *J. Phys. A, Math. Gen.*, vol. 12, no. 11, pp. 2053–2071, Nov. 1979.
[55] V. Jäntti and S. Alahuhta, "Spectral entropy—What has it to do with anaesthesia, and the EEG?" *Brit. J. Anaesthesia*, vol. 93, no. 1, pp. 150–152, Jul. 2004.
[56] C. Nicolini, G. Forcellini, L. Minati, and A. Bifone, "Scale-resolved analysis of brain functional connectivity networks with spectral entropy," *NeuroImage*, vol. 211, May 2020, Art. no. 116603.

[57] V. Zakeri and A. J. Hodgson, "Automatic identification of hard and soft bone tissues by analyzing drilling sounds," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 2, pp. 404–414, Feb. 2019.
[58] Z. Longfu, S. Yi, H. Sun, L. Zheng, H. Dapeng, and H. Yonghe, "Identification of bowel sound signal with spectral entropy method," in *Proc. 12th IEEE Int. Conf. Electron. Meas. Instrum. (ICEMI)*, Qingdao, China, Jul. 2015, pp. 798–802.
[59] A. Anier, T. Lipping, R. Ferenets, P. Puumala, E. Sonkajärvi, I. Rätsep, and V. Jäntti, "Relationship between approximate entropy and visual inspection of irregularity in the EEG signal, a comparison with spectral entropy," *Brit. J. Anaesthesia*, vol. 109, no. 6, pp. 928–934, Dec. 2012.
[60] T. Giannakopoulos, G. Siantikos, S. Perantonis, N.-E. Votsi, and J. Pantis, "Automatic soundscape quality estimation using audio analysis," in *Proc. 8th ACM Int. Conf. Pervasive Technol. Rel. Assistive Environ. (PETRA)*, Corfu, Greece, Jul. 2015, pp. 1–9.
[61] G. Qiao, T. Ma, S. Liu, N. Zheng, Z. Babar, and Y. Yin, "Spectral entropy based dolphin whistle detection algorithm and its possible application for biologically inspired communication," in *Proc. OCEANS*, Marseille, France, Jun. 2019, pp. 1–6.
[62] Y. Ji, X. Wang, Z. Liu, Z. Yan, L. Jiao, D. Wang, and J. Wang, "EEMD-based online milling chatter detection by fractal dimension and power spectral entropy," *Int. J. Adv. Manuf. Technol.*, vol. 92, nos. 1–4, pp. 1185–1200, Sep. 2017.
[63] R. Jahangir, Y. W. Teh, H. F. Nweke, G. Mujtaba, M. A. Al-Garadi, and I. Ali, "Speaker identification through artificial intelligence techniques: A comprehensive review and research challenges," *Expert Syst. Appl.*, vol. 171, Jun. 2021, Art. no. 114591.
[64] G. Mittag and S. Möller, "Quality estimation of noisy speech using spectral entropy distance," in *Proc. 26th Int. Conf. Telecommun. (ICT)*, Hanoi, Vietnam, Apr. 2019, pp. 197–201.
[65] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process., Image Commun.*, vol. 29, no. 8, pp. 856–863, 2014.
[66] K. T. Park, M. S. Park, J. H. Lee, and Y. S. Moon, "Detection of visual saliency in discrete cosine transform domain," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Las Vegas, NV, USA, Jan. 2012, pp. 128–129.
[67] M. Kristan, J. Perš, M. Perše, and S. Kovačič, "A Bayes-spectral-entropy-based measure of camera focus using a discrete cosine transform," *Pattern Recognit. Lett.*, vol. 27, no. 13, pp. 1431–1439, Oct. 2006.
[68] J. I. Makhoul and J. J. Wolf, "Linear prediction and the spectral analysis of speech," Bolt, Beranek Newman, Cambridge, MA, USA, Tech. Rep. BBN-TR-2304, Aug. 1972.
[69] T. Ziyaee, "Unsupervised denoising via Wiener entropy masking in the STFT domain," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Baltimore, MD, USA, Oct. 2014, pp. 467–472.
[70] N. Madhu, "Note on measures for spectral flatness," *Electron. Lett.*, vol. 45, no. 23, pp. 1195–1196, Nov. 2009.
[71] N. Obin and M. Liuni, "On the generalization of Shannon entropy for speech recognition," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Miami, FL, USA, Dec. 2012, pp. 97–102.
[72] J. J. Burred and A. Lerch, "Hierarchical automatic audio signal classification," *J. Audio Eng. Soc.*, vol. 52, nos. 7–8, pp. 724–739, 2004.
[73] O. Tchernichovski, F. Nottebohm, C. E. Ho, B. Pesaran, and P. P. Mitra, "A procedure for an automated measurement of song similarity," *Animal Behav.*, vol. 59, no. 6, pp. 1167–1176, Jun. 2000.
[74] H.-G. Kim, N. Moreau, and T. Sikora, *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. Chichester, U.K.: Wiley, 2005.
[75] A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp, Eds., *Handbook of Spatial Statistics*. Boca Raton, FL, USA: CRC Press, 2014.
[76] P. J. Diggle, *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, 3rd ed. Boca Raton, FL, USA: CRC Press, 2014.
[77] T. Wiegand and K. A. Moloney, *Handbook of Spatial Point-Pattern Analysis in Ecology*. Boca Raton, FL, USA: CRC Press, 2014.
[78] J. Matoušek, *Geometric Discrepancy: An Illustrated Guide*. Berlin, Germany: Springer, 2010.
[79] W. Chen, A. Srivastav, and G. Travaglini, Eds., *A Panorama of Discrepancy Theory*. Cham, Switzerland: Springer, 2014.
[80] M. T. Shehata, "Characterization of particle dispersion," in *Practical Guide to Image Anal.*, J. J. Friel, Ed. Materials Park, OH, USA: ASM International, 2000, ch. 6, pp. 129–144.
[81] F. Nekka and J. Li, "Various mathematical approaches to extract information from textures of increasing complexities," in *Fractals in Engineering: New Trends in Theory and Applications*, L.-V. Jacques and L.-V. Evelyne, Eds. London, U.K.: Springer, 2005, pp. 255–270.
[82] D. J. Whitehouse, *Handbook of Surface and Nanometrology*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2011.
[83] D. J. Whitehouse, "Fractal or fiction," *Wear*, vol. 249, nos. 5–6, pp. 345–353, Jun. 2001.
[84] A. Fischer, K. Riesen, and H. Bunke, "Graph similarity features for HMM-based handwriting recognition in historical documents," in *Proc. 12th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Kolkata, India, Nov. 2010, pp. 253–258.
[85] G. Chartrand, P. Erdös, and O. R. Oellermann, "How to define an irregular graph," *College Math. J.*, vol. 19, no. 1, pp. 36–42, Jan. 1988.
[86] A. Ali, G. Chartrand, and P. Zhang, *Irregularity in Graphs*. Cham, Switzerland: Springer, 2021.
[87] G. Chartrand and P. Zhang, *Chromatic Graph Theory*. Boca Raton, FL, USA: CRC Press, 2009.
[88] A. Benjamin, G. Chartrand, and P. Zhang, *The Fascinating World of Graph Theory*. Princeton, NJ, USA: Princeton Univ. Press, 2015.
[89] G. Chartrand, L. Lesniak, and P. Zhang, *Graphs & Digraphs*, 6th ed. Boca Raton, FL, USA: CRC Press, 2015.
[90] D. Wells, "Which is the most beautiful?" *Math. Intelligencer*, vol. 10, no. 4, pp. 30–31, Sep. 1988.
[91] D. Wells, "Are these the most beautiful?" *Math. Intell.*, vol. 12, no. 3, pp. 37–41, 1990.
[92] Wikipedia. (Jul. 24, 2017). *Glossary of Graph Theory Terms*. [Online]. Available: https://en.wikipedia.org/wiki/Glossary_of_graph_theory_term
[93] G. Chartrand, G. L. Johns, K. A. McKeon, and P. Zhang, "The rainbow connectivity of a graph," *Networks*, vol. 54, no. 2, pp. 75–81, Sep. 2009.
[94] T. R. Jensen and B. Toft, *Graph Coloring Problems*. New York, NY, USA: Wiley, 1995.
[95] E. Estrada, "Randić index, irregularity and complex biomolecular networks," *Acta Chim. Slovenica*, vol. 57, pp. 597–603, Jan. 2010.
[96] J. L. Gross, J. Yellen, and P. Zhang, *Handbook of Graph Theory*. Boca Raton, FL, USA: CRC Press, 2014.
[97] A. Soifer, *The Mathematical Coloring Book: Mathematics of Coloring and the Colorful Life of Its Creators*. New York, NY, USA: Springer, 2008.
[98] J. Pach and P. K. Agarwal, *Combinatorial Geometry*. New York, NY, USA: Wiley, 1995.
[99] J. Pach, "The beginnings of geometric graph theory," in *Erdős Centennial*, L. Lovász, I. Z. Ruzsa, and V. T. Sós, Eds. New York, NY, USA: Springer, 2013, pp. 465–484.
[100] A. M. Raigorodskii, "Coloring distance graphs and graphs of diameters," in *Thirty Essays on Geometric Graph Theory*, J. Pach, Ed. New York, NY, USA: Springer, 2013, pp. 429–460.
[101] L. Collatz and U. Sinogowitz, "Spektren endlicher Grafen," *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, vol. 21, no. 1, pp. 63–77, 1957.
[102] Y. Alavi, G. Chartrand, F. R. K. Chung, P. Erdös, R. L. Graham, and O. R. Oellermann, "Highly irregular graphs," *J. Graph Theory*, vol. 11, no. 2, pp. 235–249, 1987.
[103] P. Hoffman, *The Man Who Loved Only Numbers: The Story of Paul Erdős and the Search for Mathematical Truth*. London, U.K.: Fourth Estate, 1999.
[104] M. Aigner and G. M. Ziegler, *Proofs From the Book*, 5th ed. Berlin, Germany: Springer, 2014.
[105] G. Chartrand, G. L. Johns, K. A. McKeon, and P. Zhang, "Rainbow connectivity of cages," in *Proc. 38th Southeastern Int. Conf. Combinatorics, Graph Theory Comput.*, Boca Raton, FL, USA, no. 184, Mar. 2007, pp. 209–222.
[106] G. Chartrand, G. L. Johns, K. A. McKeon, and P. Zhang, "Rainbow connection in graphs," *Math. Bohemica*, vol. 133, no. 1, pp. 85–98, 2008.
[107] G. Chartrand, M. S. Jacobson, J. Lehel, O. R. Oellermann, S. Ruiz, and F. Saba, "Irregular networks," *Congressus Numerantium*, vol. 64, pp. 355–374, 1988. [Online]. Available: https://www.researchgate.net/publication/265701559_Irregular_networks/references
[108] G. Chartrand, L. Lesniak, D. W. VanderJagt, and P. Zhang, "Recognizable colorings of graphs," *Discussiones Mathematicae Graph Theory*, vol. 28, no. 1, pp. 35–57, 2008.
[109] I. Gutman, B. Ruščić, N. Trinajstić, and J. C. F. Wilkox, "Graph theory and molecular orbitals. XII. Acyclic polyenes," *J. Chem. Phys.*, vol. 62, no. 9, pp. 3399–3405, 1975.
[110] I. Gutman and N. Trinajstić, "Graph theory and molecular orbitals. Total ϕ-electron energy of alternant hydrocarbons," *Chem. Phys. Lett.*, vol. 17, no. 4, pp. 535–538, 1972.

[111] S. Nikolić, G. Kovačević, A. Miličević, and N. Trinajstić, "The zagreb indices 30 years after," *Croatica Chem. Acta*, vol. 76, no. 2, pp. 113–124, 2010.
[112] G. H. Fath-Tabar, "Old and new Zagreb indices of graphs," *MATCH Commun. Math. Comput. Chem.*, vol. 65, no. 1, pp. 79–84, 2011.
[113] D. Janežič, A. Miličević, S. Nikolić, and N. Trinajstić, *Graph-Theoretical Matrices in Chemistry*. Boca Raton, FL, USA: CRC Press, 2015.
[114] F. K. Bell, "A note on the irregularity of graphs," *Linear Algebra Appl.*, vol. 161, pp. 45–54, Jan. 1992.
[115] M. O. Albertson, "The irregularity of a graph," *Ars Combinatoria*, vol. 46, pp. 219–225, 1997. [Online]. Available: http://www.combinatoire.ca/ArsCombinatoria/TOC.html and http://www.combinatoire.ca/ArsCombinatoria/Vol46.html
[116] V. Nikiforov, "Eigenvalues and degree deviation in graphs," *Linear Algebra Appl.*, vol. 414, no. 1, pp. 347–360, Apr. 2006.
[117] M. Krivelevich and R. Yuster, "The rainbow connection of a graph is (at most) reciprocal to its minimum degree," *J. Graph Theory*, vol. 63, no. 3, pp. 185–191, 2010.
[118] I. Gutman, P. Hansen, and H. Mélot, "Variable neighborhood search for extremal graphs. 10. Comparison of irregularity indices for chemical trees," *J. Chem. Inf. Model.*, vol. 45, no. 2, pp. 222–230, Mar. 2005.
[119] J. A. D. Oliveira, C. S. Oliveira, C. Justel, and N. M. M. D. Abreu, "Measures of irregularity of graphs," *Pesquisa Operacional*, vol. 33, no. 3, pp. 383–398, Nov. 2013.
[120] A. Frieze, R. J. Gould, M. Karoński, and F. Pfender, "On graph irregularity strength," *J. Graph Theory*, vol. 41, no. 2, pp. 120–137, Oct. 2002.
[121] S. Mukwembi, "On maximally irregular graphs," *Bull. Malaysian Math. Sci. Soc.*, vol. 36, no. 3, pp. 717–721, 2011.
[122] D. Rautenbach and L. Volkmann, "How local irregularity gets global in a graph," *J. Graph Theory*, vol. 41, no. 1, pp. 18–23, Sep. 2002.
[123] G. Ebert, J. Hemmeter, F. Lazebnik, and A. Woldar, "On the number of irregular assignments on a graph," *Discrete Math.*, vol. 93, nos. 2–3, pp. 131–142, Nov. 1991.
[124] S. Mukwembi, "A note on diameter and the degree sequence of a graph," *Appl. Math. Lett.*, vol. 25, no. 2, pp. 175–178, Feb. 2012.
[125] M. Bača, S. Jendroľ, M. Miller, and J. Ryan, "On irregular total labellings," *Discrete Math.*, vol. 307, nos. 11–12, pp. 1378–1388, May 2007.
[126] G. Chartrand, O. R. Oellermann, and M. Schultz, "Distance: A graphical tour," in *Proc. 2nd Quadrennial Int. Conf. Theory Appl. Graphs*. San Francisco, CA, USA: San Francisco State Univ., Jul. 1989, pp. 441–458.
[127] S. Piccard, *Sur les Ensembles de Distances des Ensembles de Points d'un Espace Euclidien*. Neuchâtel, Switzerland: Univ. Neuchâtel, 1939.
[128] T. Aste, *The Pursuit of Perfect Packing*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2008.
[129] C. Zong, "Packing, covering and tiling in two-dimensional spaces," *Expositiones Mathematicae*, vol. 32, no. 4, pp. 297–364, 2014.
[130] A. Okabe, B. Boots, W. Laurier, K. Sugihara, and S. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Chichester, U.K.: Wiley, 2000.
[131] S. M. Moser and P.-N. Chen, *A Student's Guide to Coding and Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[132] P. D. Grünwald, *The Minimum Description Length Principle*. Cambridge, MA, USA: MIT Press, 2007.
[133] R. Weatherford, *Philosophical Foundations of Probability Theory*. London, U.K.: Routledge & Kegan Paul, 1982.
[134] R. S. Nickerson, "The production and perception of randomness," *Psychol. Rev.*, vol. 109, no. 2, pp. 330–357, 2002.
[135] M. Challinor, "Change, chance and structure: Randomness and formalism in art," *Leonardo*, vol. 4, no. 1, pp. 1–11, 1971.
[136] Y. B. Sanderson, "Color charts, esthetics, and subjective randomness," *Cognit. Sci.*, vol. 36, no. 1, pp. 142–149, Jan. 2012.
[137] L. Lovász and K. Vesztergombi, "Geometric representations of graphs," in *Paul Erdős and His Mathematics*, vol. 2, G. Halász, L. Lovász, M. Simonovits, and V. T. Sós, Eds. Berlin, Germany: Springer, 2002, pp. 471–498.
[138] H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and Fractals: New Frontiers of Science*. New York, NY, USA: Springer, 1992.
[139] Wikipedia. (Jul. 1, 2017). *Sophie Piccard*. [Online]. Available: https://fr.wikipedia.org/wiki/Sophie_Piccard
[140] Y. Voegeli. (Aug. 2, 2017). *Suffrage Féminin. Dictionnaire Historique de la Suisse*. [Online]. Available: http://www.hls-dhs-dss.ch/textes/f/F10380.ph
[141] P. Erdös, "On sets of distances of n points," *Amer. Math. Monthly*, vol. 53, no. 5, pp. 248–250, May 1946.
[142] P. Erdös, "On sets of distances of *n* points," *Amer. Math. Monthly*, vol. 77, no. 7, pp. 739–740, 1970.
[143] P. Erdös, "Set theoretic, measure theoretic, combinatorial, and number theoretic problems concerning point sets in Euclidean space," *Real Anal. Exchange*, vol. 4, no. 2, pp. 113–138, 1978.
[144] E. Paul and R. K. Guy, "Distinct distances between lattice points," *Elemente Math.*, vol. 25, no. 6, pp. 121–123, 1970.
[145] E. J. Makai, J. Pach, and J. Spencer, "New results on the distribution of distances determined by separated point sets," in *Paul Erdős and His Mathematics*, vol. 2, G. Halász, L. Lovász, M. Simonovits, and V. T. Sós, Eds. Berlin, Germany: Springer, 2002, pp. 499–511.
[146] zbMATH. (Jul. 1, 2017). *Erdős, Paul. Zentralblatt MATH*. [Online]. Available: https://zbmath.org/authors/?s=0&q=au%3Apaul+erdo
[147] P. Erdős. (Aug. 2, 2017). *Paul Erdős' Papers*. [Online]. Available: http://login.math-inst.hu/~p_erdos/Erdos.htm
[148] L. Babai and J. Spencer, "Paul Erdős (1913–1996)," *Notices Amer. Math. Soc.*, vol. 45, no. 1, pp. 64–73, 1998.
[149] E. W. Weisstein. (Nov. 28, 2017). Discrepancy. MathWorld. [Online]. Available: http://mathworld.wolfram.com/Discrepancy.html
[150] E. W. Weisstein. (Nov. 28, 2017). Discrepancy theorem. MathWorld. [Online]. Available: http://mathworld.wolfram.com/DiscrepancyTheorem.html
[151] (Dec. 4, 2021). *Uniform Distribution Theory*. [Online]. Available: https://sciendo.com/journal/UDT
[152] B. Chazelle, *The Discrepancy Method: Randomness and Complexity*. Cambridge, U.K.: Cambridge Univ. Press, 2001.
[153] P. Hellekalek and G. Larcher, *Random and Quasi-Random Point Sets*. New York, NY, USA: Springer, 1998.
[154] M. Ghyka, *The Geometry of Art and Life*. New York, NY, USA: Dover, 1977.
[155] R. A. Dunlap, *The Golden Ratio and Fibonacci Numbers*. Singapore: World Scientific, 1998.
[156] N. J. A. Sloane. (Jul. 24, 2017). *A001622: Decimal Expansion of Golden Ratio. The On-Line Encyclopedia of Integer Sequences*. [Online]. Available: https://oeis.org/A00162
[157] M. Neveux and H. E. Huntley, *Le Nombre d'or. Radiographie d'un Mythe Suivi de La Divine Proportion*. Paris, France: Points, 2014.
[158] G. W. Berkhan and W. F. Meyer, "Neuere Dreiecksgeometie," in *Encyklopädie der Mathematischen Wissenschaften mit Einschluß Ihrer Anwendungen*, vol. III.1.2 AB 10, W. F. Meyer and H. Mohrmann, Eds. Leipzig, Germany: B. G. Teubner Verlag, 1914, pp. 1173–1276.
[159] P. Baptist, *Die Entwicklung der Neueren Dreiecksgeometrie*. Mannheim, Germany: BI-Wissenschaftsverlag, 1992.
[160] P. J. Davis, "The rise, the fall, and possible transfiguration of triangle geometry: A mini-history," *Amer. Math. Monthly*, vol. 32, no. 3, pp. 204–214, 1995.
[161] P. Romera-Lebret, "Die neue Dreiecksgeometrie: Der Übergang von der Mathematik der Amateure zur unterrichteten Mathematik," *Mathematische Semesterberichte*, vol. 59, no. 1, pp. 75–102, 2012.
[162] H. Fukagawa and T. Rothman, *Sacred Mathematics: Japanese Temple Geometry*. Princeton, NJ, USA: Princeton Univ. Press, 2008.
[163] P. Mehta and P. Majumder, *From Extractive to Abstractive Summarization: A Journey*. Singapore: Springer, 2019.
[164] J.-M. Torres-Moreno, *Automatic Text Summarization*. London, U.K.: Wiley, 2014.
[165] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.
[166] S. Le Moan, I. Farup, and J. Blahová, "Towards exploiting change blindness for image processing," *J. Vis. Commun. Image Represent.*, vol. 54, pp. 31–38, Jul. 2018.
[167] K. Berkner, E. L. Schwartz, and C. Marle, "SmartNails—Image and display dependent thumbnails," *Proc. SPIE*, vol. 5296, pp. 54–65, Jan. 2004.
[168] F. Brodbeck. (2011). *Cinemetrics*. Accessed: Jun. 9, 2021. [Online]. Available: http://cinemetrics.fredericbrodbeck.de
[169] M. A. Smith and T. Kanade, *Multimodal Video Characterization and Summarization*. New York, NY, USA: Kluwer, 2005.
[170] Z. Xiong, R. Radhakrishnan, A. Divakaran, Y. Rui, and T. S. Huang, *A Unified Framework for Video Summarization, Browsing & Retrieval: With Applications to Consumer and Surveillance Video*. San Diego, CA, USA: Elsevier, 2006.

[171] V. Atanasiu and R. Ingold, "Document towers: A MATLAB software implementing a three-dimensional architectural paradigm for the visual exploration of digital documents and libraries," *SoftwareX*, vol. 14, Jun. 2021, Art. no. 100684.
[172] D. S. Bloomberg and F. R. Chen, "Document image summarization without OCR," in *Proc. 3rd IEEE Int. Conf. Image Process.*, Lausanne, Switzerland, Sep. 1996, pp. 229–232.
[173] R. P. Futrelle, "Summarization of diagrams in documents," in *Advances in Automated Text Summarization*, I. Mani and M. T. Maybury, Eds. Cambridge, MA, USA: MIT Press, 1999, pp. 403–421.
[174] A. Brink, L. Schomaker, and M. Bulacu, "Towards explainable writer verification and identification using vantage writers," in *Proc. 9th Int. Conf. Document Anal. Recognit. (ICDAR)*, Curitiba, Brazil, Sep. 2007, pp. 824–828.
[175] K. McCall and K. Piersol, "Enhancing accuracy of jumping by incorporating interestingness estimates," U.S. Patent 20 070 180 355, Aug. 2, 2007.
[176] F. Matulic, "Automatic selection of visually attractive pages for thumbnail display in document list view," in *Proc. 3rd Int. Conf. Digit. Inf. Manage.*, London, U.K., Nov. 2008, pp. 221–226.
[177] R. N. Bracewell, *The Fourier Transform and Its Applications*, 3rd ed. Boston, MA, USA: McGraw-Hill, 2000.
[178] R. L. J. Easton, *Fourier Methods in Imaging*. Chichester, U.K.: Wiley, 2010.
[179] D. C. Champeney, *Fourier Transforms and Their Physical Applications*. London, U.K.: Academic, 1973.
[180] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, 2nd ed. San Diego, CA, USA: California Technical Publishing, 1999.
[181] E. W. Kamen and B. S. Heck, *Fundamentals of Signals and Systems Using the Web and MATLAB*, 3rd ed. Harlow, U.K.: Pearson, 2014.
[182] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
[183] D. Bradley and G. Roth, "Adapting thresholding using the integral image," *J. Graph. Tools*, vol. 12, no. 2, pp. 13–21, 2007.
[184] M. D. Fairchild, *Color Appearance Models*, 3rd ed. Chichester, U.K.: Wiley, 2013.
[185] C. E. Shannon, "A mathematical theory of cryptography," Bell Laboratories, Murray Hill, NJ, USA, Tech. Memorandum 45-110-92, Sep. 1945.
[186] N. J. A. Sloane and A. D. Wyner, Eds., *Claude Elwood Shannon Miscellaneous Writings*. Murray Hill, NJ, USA: Mathematical Science Research Center, AT&T Bell Laboratories, 2013.
[187] F. Attneave, "Stochastic composition processes," *J. Aesthetics Art Criticism*, vol. 17, no. 4, pp. 503–510, 1959.
[188] H. Meinhardt, *The Algorithmic Beauty of Sea Shells*, 4th ed. Berlin, Germany: Springer, 2009.
[189] J. Wagemans, Ed., *The Oxford Handbook of Perceptual Organization*. New York, NY, USA: Oxford Univ. Press, 2015.
[190] W. Prinz, "Quantitative Versuche über die Prägnanz von Punktmustern," Psychologische Forschung, vol. 29, no. 4, pp. 297–359, 1966.
[191] G. Jewell and M. E. McCourt, "Pseudoneglect: A review and meta-analysis of performance factors in line bisection tasks," *Neuropsychologia*, vol. 38, no. 1, pp. 93–110, Jan. 2000.
[192] H. Cartier-Bresson, *L'imaginaire d'après nature*. Paris, France: Fata Morgana, 1996.
[193] L. Pacioli, *De Divina Proportione*. Venice, Italy: A. Paganius Paganinus, 1509. [Online]. Available: https://archive.org/details/diuinaproportion00paci/
[194] B. Wolf, M. Jung, A. Fischer, G. Waeber, and V. Atanasiu, "Analyse 'automatisierte integration des alten sachkatalogs in alma,'" HES·SO//Fribourg, iCoSys, Fribourg, Switzerland, Tech. Rep., Dec. 2020.
[195] Wikipedia Contributors. (Jan. 31, 2021). *Tesseract (Software)*. [Online]. Available: https://en.wikipedia.org/wiki/Tesseract_(software)
[196] Google. (Jan. 31, 2021). *Vision AI*. [Online]. Available: https://cloud.google.com/vision/
[197] D. L. Wang and J. S. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, no. 4, pp. 619–681, Aug. 1982.
[198] K. Paliwal, K. Wójcicki, and B. Shannon, "The importance of phase in speech enhancement," *Speech Commun.*, vol. 53, no. 4, pp. 465–494, Apr. 2011.
[199] M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, no. 2, pp. 140–154, Apr. 1982.
[200] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE*, vol. 69, no. 5, pp. 529–541, May 1981.
[201] B. C. Hansen and R. F. Hess, "Structural sparseness and spatial phase alignment in natural scenes," *J. Opt. Soc. Amer.*, vol. 24, no. 7, pp. 1873–1885, 2007.
[202] P. Kovesi, "Image features from phase congruency," *Videre, J. Comput. Vis. Res.*, vol. 1, no. 3, pp. 2–26, 1999.
[203] A. W. Lohmann, D. Mendlovic, and G. Shabtay, "Significance of phase and amplitude in the Fourier domain," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 14, no. 11, pp. 2901–2904, 1997.

P12 – 338

VLAD ATANASIU passed the student admission examinations at the Faculty of Telecommunications, Polytechnic University, Timişoara, Romania, in 1989. He received the B.A. degree in middle eastern studies and the M.A. degree in Arabic linguistics from the University of Provence, Aix-en-Provence, France, in 1994 and 1995, respectively, and the Ph.D. degree in Arabic paleography from the École Pratique des Hautes Études, Paris, France, in 2003. He is currently pursuing the Ph.D. degree with the Informatics Department, University of Fribourg, Switzerland, with a focus on document analysis and visualization.

From 2003 to 2005, he was a Postdoctoral Fellow with the Architecture Department and the Brain and Cognitive Science Department, Massachusetts Institute of Technology, Cambridge, MA, USA. From 2006 to 2009, he was at the Visualization Commission, Austrian Academy of Sciences, Vienna, Austria, where he developed image processing methods for paper structure analysis and geographical information system tools to support paper history, and further co-coordinated the European computational paper history project "Bernstein." From 2010 to 2011, he developed handwriting analysis and visualization software for Télecom-ParisTech, Paris, and the back- and front-end of an online cartography atlas of Iran for the French National Centre for Scientific Research, Paris. He taught information visualization at the University of Graz, Austria, and the University of Fribourg, from 2011 to 2015. From 2019 to 2021, he was with the University of Basel, Switzerland, where he created software for the legibility enhancement of ancient papyri. Currently, he is doing data analysis of natural language datasets for the University of Bern, Switzerland. He does freelance typographical and graphic design work. He is the author of the books *On Letter Frequency and Its Influence on Arabic Calligraphy* (Paris: L'Harmattan, 1999) and *Expert Bytes: Computer Expertise in Forensic Documents* (Boca Raton: CRC Press, FL, 2014).

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 March 2023, accepted 24 July 2023, date of publication 28 July 2023, date of current version 16 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3299594

COMMENTS AND CORRECTIONS

Corrections to "The Structural Information Potential and Its Application to Document Triage"

VLAD ATANASIU
Department of Informatics, University of Fribourg, 1700 Fribourg, Switzerland
e-mail: atanasiu@alum.mit.edu

This work involved human subjects in its research. The author confirms that all human subject research procedures and protocols are exempt from review board approval.

In the above article [1], a few of the sections require corrections and some additional experiments to strengthen the work.

1) In the above article [1, p. 13109, col. 1, §3, Section II-A-5]," add a reference to the herein Fig. 1 in the sentence beginning with "The one domain that is almost pathologically fascinated by structural irregularity is that of the arts and architecture." The purpose of this correction is to illustrate the application of structural irregularity in the domain of architecture, with a real-world example.
2) In the above article [1, p. 13110, col. 1, §4, Section II-B]," add a reference to the herein Fig. 2 in the sentence beginning with "The problem becomes more critical as the number of documents to be processed increases." The purpose of this correction is to illustrate with a real-world example the application of triage to documents.

FIGURE 1. Example of an architecture based on the concept of redundancy minimization of edge lengths and texture—City archives, Berne, Switzerland; :mlzd architekten biel-bienne. (Credits: Vlad Atanasiu, 2022.)

FIGURE 2. Does document triage matter? An internationally famous case in point that would seem to answer in the affirmative is provided by the so-called *Pentagon Papers*, a classified history of the Vietnam War between 1945 and 1967 prepared by the U.S. Department of Defense [2]. Its publication revealed how multiple U.S. administrations misled the public on the reasons for and actions taken in the Vietnam War, and had consequential repercussions on the legal aspects of freedom of speech and journalism in the United States. The "papers" consisted of more than 7,000 pages of documents that were covertly photocopied by Daniel Ellsberg of the RAND Corporation and made available to the press in 1971. Among the major practical problems with which the journalists had to contend was that of rapidly identifying the most informative content and collating related pages so that articles could be finished before editorial deadlines [3, p. 392], [4]. Layout-based classification is a sensible strategy for both tasks under such circumstances. The difficulties posed by document triage for the evaluation of the Pentagon Papers were fictionalized in a memorable scene of the film *The Post* (2017), depicting the staff of the Washington Post at work (Credits: 20th Century Fox):
– Is that…?
– Yep.
– It's not the full report but it's over 4,000 pages of it.
– Are these in order?
– I don't think so.
– There are no page numbers.
– Yeah, that's where the "top secret" stamps were. My source had to cut 'em off.
– I was supposed to retire on Friday.
– Ben, how we supposed to comb through 4,000 pages of material?
– They're not even loosely organized?
– The Times had three months.
– There's no way we can possibly get this done.
– He's right. We've got less than eight hours.

3) In the above article [1, p. 13113, equation 9, Section III-A-5]," read "≈ 0.6942" instead of "0.6942."

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

4) In the above article [1, p. 13121, Section IV-A]," add after §1 of col. 2, the following section. The purpose of this correction is to strengthen the evaluation of the structural information potential (SIP) method with a series of psychophysical experiments.

3) PSYCHOPHYSICAL EXPERIMENTS

Our understanding of SIP and other pattern-ordering methods may be deepened beyond what is possible through the use of visual evidence and quantitative evaluations, on the basis of empirical insights obtained from psychophysical experiments. The primary goal of experiments such as those described below is to evaluate the agreement between computational and perceptual pattern orderings, which is especially important for SIP as a model for human behavior. The discovery of further insights into the perceptual, cognitive, and other practical aspects of SIP—for example, in the context of the triage task—represents a secondary goal of the experiments.

Experimental design—The participants in the experiments were 10 students in digital humanities from the University of Bern, Switzerland (4 females, $MD = 32$ years, $RG = 22$–43, $SD = 5.98$). The stimuli consisted of the 104 pages of the *New Yorker* magazine described above [1, Fig. 31], each reproduced on a 3-by-4.5 cm (1.18-by-1.77 in) piece of paper. This size ensured, on one hand, that all pages could be laid out within the confines of a small table and viewed at a glance; on the other hand, the small size helped to focus on the overall layout pattern, reducing the salience of semantic information such as text and picture content. To further emphasize the pattern and inhibit reading, the reproductions were placed upside-down.

P12 – 340

The concept of the uniform–clustered–regular pattern spectrum was explained to the participants with reference to representative page samples ([1, Fig. 10] a, b, d) and through discussion. To ensure that the concept had been effectively assimilated, a practical training exercise was conducted in which the participants were required to order 16 sample documents (uniformly selected from the library cards shown in Fig. 21), with corrective inputs from the experimenter.

Ordering experiment—This experiment aimed to evaluate the suitability of SIP as a model of human pattern ordering along the uniform–clustered–regular spectrum. To this end, the participants received a stack of randomly ordered stimuli and were required to order the page reproductions along the uniform–clustered–regular spectrum.

The differences between the perceptual and SIP orderings were quantified using the topological entropy method, and the statistical summary is presented in Fig. 3b. The median value of 0.57 indicates a sizable divergence of the perceptual ordering from the computational SIP ordering ($M = 0.57$, 95% confidence interval $[CI] = [0.54, 0.60]$). Nevertheless, the perceptual orderings are on average closer to SIP than are the other computational methods, with the median of the latter being 0.67 ($M = 0.66$, 95% $CI = [0.58, 0.74]$; Fig. 3a).

We were able to enhance our comprehension of the results by adopting a qualitative approach, based on spontaneous remarks made by the participants when commenting on the experiment, observations of their behavior, and post-experiment debriefings. First of all, participants commented on the tediousness of the task: in terms of the need to physically arrange so many stimuli, in cognitive terms (the difficulty associated with deciding on their proper order), and in motivational terms (due to the duration of the operation, ca. 20 to 30 minutes). Second, it was difficult for the participants to order the stimuli purely in terms of their layout without interpreting ink structures in terms of their semantic classes, such as pictures and titles, despite that some of these were identified as such precisely because of the layout that made them stand out. Third, the ordering of clustered patterns was more challenging than that of uniform and regular patterns; the latter could be intuitively comprehended as "mostly empty surfaces" and "homogeneous ink distributions," while the concept of scale-space clustering was less familiar. Perhaps the results would have been different with participants drawn from other demographics, such as document designers, artists, and computer scientists.

In conclusion, the foremost finding from this experiment is the difficulty of perceptual ordering for various practical and cognitive reasons. It also provides evidence for the poor performance of human pattern triage. Since the experiment was more relevant for revealing human factors in pattern ordering than for allowing an evaluation of SIP as a perceptual model, we designed a second experimental setup to address these issues.

Improvement experiment—To diminish the influence of factors affecting free perceptual ordering, in this experiment, the participants were presented with a fixed order of stimuli (the one produced by SIP) and asked to improve this ordering by shifting the page reproductions so as to create a transition from uniform to clustered to regular layouts that they deemed to be perceptually the most "smooth."

The numerical results, with median topological entropy of 0.38, now indicate a smaller, albeit non-trivial, divergence between computational and perceptual orderings for the improvement task relative to the case of the free ordering task ($M = 0.39$, 95% $CI = [0.33, 0.45]$; Fig. 3c).

The effect size of the participants may be better appreciated through a visual examination of the resulting patterns. Unlike the numerical assessment, this approach reveals that the improvements made were comparatively limited in terms of their number and the distance of pattern shifts along the spectrum. Fig. 3g shows a representative example of the improvement produced by participant MF, which is the closest to the median divergence from SIP; clearly, this ordering is quite similar to SIP (Fig. 3h) and far from the median ordering of other methods (Fig. 3e). In terms of qualitative results, the participants stated that this task was easier than the former, and, indeed, they completed the experiment faster (in ca. 15 to 20 minutes).

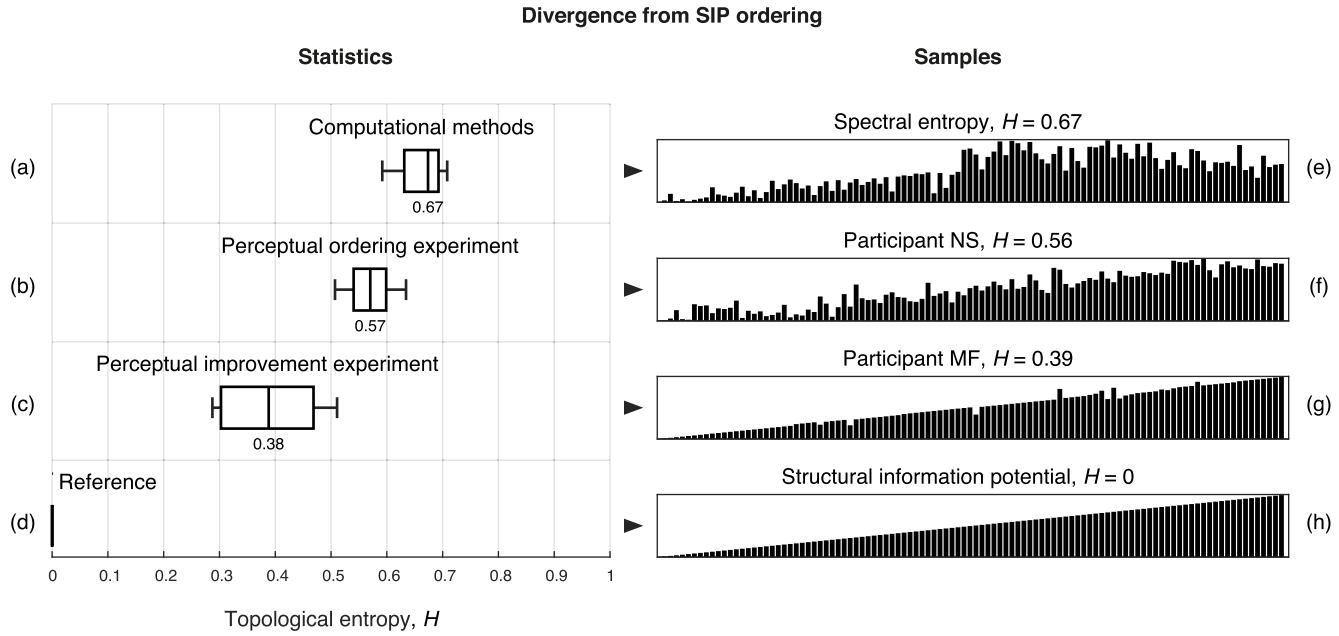


FIGURE 3. Results of the perceptual ordering and improvement experiments related to the divergence from the SIP-induced ordering. (a)–(d) Statistics on the distribution of topological entropy values, H, visualized as box-and-whisker plots: median, interquartile range, and range. "Computational methods" summarizes the data from [1, Fig. 19], while "Reference" refers to the value H of the reference SIP ordering. (e)–(h) Actual ordering samples from the corresponding statistical distributions, selected as the closest to the medians. The orderings of bars represent the permutations of the SIP ordering.

P12
–
341

It appears from this experiment that SIP-induced ordering is very much accepted by humans as producing a smooth transition between uniform, clustered, and regular patterns. It also greatly facilitates triage.

Rating experiment—In this last experiment, we intended to compare on a perceptual basis the smoothness of the pattern spectra induced by various computational methods, which we assessed from a formal quantitative point of view in the preceding section, "Quantitative Evaluation" [1, Fig. 19]. To do this, the participants were requested to rate on a five-level scale the smoothness of the five analyzed computational orderings, along with the physical ordering of the pages, based on printed copies of Figs. 31–36 from [1].

The results are presented in Fig. 4. The rating aggregation method employed was majority judgement, chosen because it is a non-parametric ranking method that favors consensual candidates and is based on median scores rather than means [5]. Clearly, not only does SIP produce a highly consistent and smooth uniform–clustered–regular pattern spectrum, but all other methods are also given lower consistency ratings.

General assessment—In conclusion, SIP is a good model of perceptual pattern ordering along the uniform–clustered–regular spectrum, as well as aiding greatly in pattern-based triage. The psychophysical experiments described herein confirm and expand, from an empirical point of view, the conclusions derived from visual evidence and formal quantitative evaluation. In the next section, we evaluate SIP on a much larger dataset containing 1.2 million stimuli.

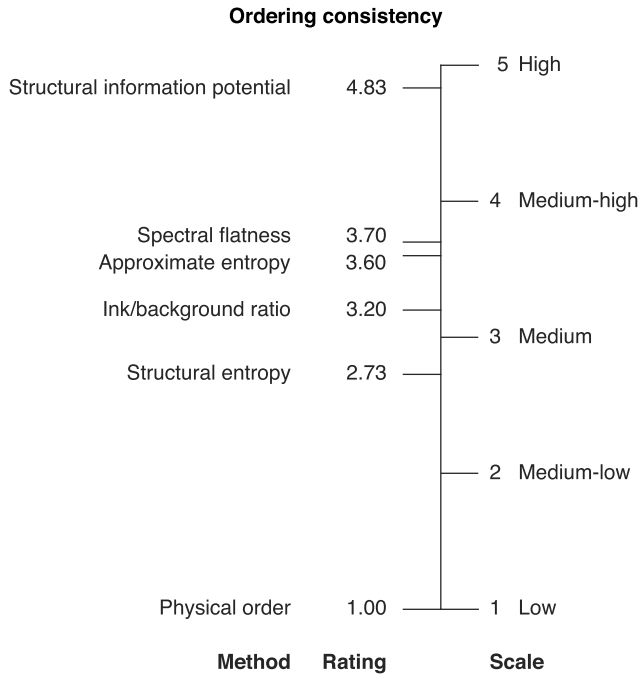


FIGURE 4. Results of the experiment on the perceptual rating of the computational ordering methods of [1, Fig. 19], as well as the physical order of pages in the document. The ratings are aggregated over 10 participants using the majority judgement method and a five-level rating scale.

5) In the above article [1, p. 13130, Section V-B]," add after §1 of col. 2, the following paragraphs and figure. The purpose of this addition is to provide an analytical explanation

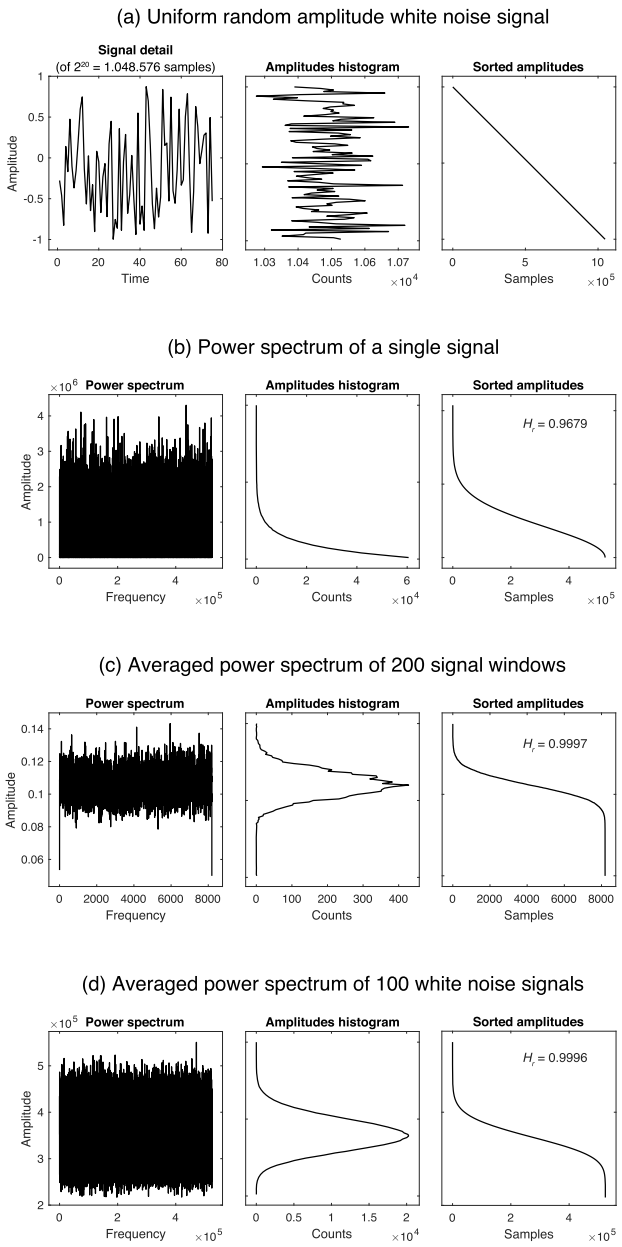


FIGURE 5. **Analysis of the power spectra of uniform random amplitude white noise. (a) Left: Signal detail. Center: Histogram of the signal amplitudes. The amplitude values are in the [0, 2] range, are aggregated in 100 bins, and are displayed on the y-axis; the counts per bin are on the x-axis. Right: Amplitudes sorted in descending order. The plot is linear, as expected from a uniform random distribution. (b) The power spectrum of the signal is shown in (a). Note that the amplitude distribution is not flat, as is expected from white noise. (c) Welch's power spectral density estimate of the signal is shown in (a), in which the signal is divided into 200 overlapping windows of equal length, filtered with a Hamming window [6], [7, pp. 415–417]. Note the noisy Gaussian distribution of the amplitudes and the flattening of their sorted plot, which leads to an increase of the Shannon relative entropy, H_r, with respect to that of the power spectrum of (b). (d) Averaged power spectrum of 100 white noise signals.**

of the randomness phenomena affecting the measurement of SIP.

In fact, the power spectrum of white noise with uniform random amplitude is theoretically flat, i.e. equal at all frequencies, and thus has maximal entropy [7, p. 385]. However, for a finite signal, the power spectrum distribution has a power law shape (Fig. 5 a, b). As in the case of run lengths, this is due to the unequal number of cycles per frequency available to compute the power spectrum. Various methods have been developed to overcome this limitation, such as averaging the power spectra of multiple random signals or averaging overlapping windows of the same signal (Fig. 5 c, d) [7, pp. 371–644], [8]. The distribution of the averaged amplitudes of the power spectra will now converge to a Gaussian, as per the central limit theorem. Increasing signal length also increases the distribution variance, with the power spectrum becoming uniform in the limit and its entropy maximal, as expected. The flattening of the power spectrum also characterizes the SIP method, since it comprises an averaging of the power spectra of all orientations of an image.

In conclusion, uniform random-amplitude white noise patterns of finite extent naturally exhibit substantial homogeneity, while their measured entropy is further increased through spectral averaging as part of the SIP method.

6) In the above article [1, p. 13132, col. 1, Section V-G]," replace §1 with the following paragraph. The purpose of this correction is to extend the discussion of SIP in relation to psychology.

The field of perceptual organization [9] studies from a psychological perspective many aspects related to SIP, among them structure [10], informativeness [11], entropy [12], uncertainty [13], redundancy [14], complexity [15], randomness [16], Gestalt Prägnanz [17], scene statistics [18], illusions [19], attractivity [20], and aesthetics [21]. Configurations are, however, difficult psychophysical research stimuli; this is more so the case for real data, such as documents, where semantics, aesthetics, and user experience (among many other factors) play an important role in their evaluation. As the pattern ordering resulting from the SIP is primarily intended to have humans as end users, it is nevertheless desirable to investigate in future work the relevance of modifying the SIP method to account for human pattern perception.

P12 – 342

REFERENCES

[1] V. Atanasiu, "The structural information potential and its application to document triage," *IEEE Access*, vol. 10, pp. 13103–13138, 2022.

[2] Wikipedia. (Feb. 15, 2022). *Pentagon Papers*. [Online]. Available: https://en.wikipedia.org/wiki/Pentagon_Papers

[3] D. Ellsberg, *Secrets: A Memoir of Vietnam and the Pentagon Papers*. New York, NY, USA: Penguin Books, 2003.

[4] B. Gallagher, J. Harlan, and J. Scott. (2021). *'We're Going to Publish' An Oral History of the Pentagon Papers*. New York, NY, USA: The New York Times. [Online]. Available: https://www.nytimes.com/interactive/2021/06/09/us/pentagon-papers-oral-history.html

[5] M. Balinski and R. Laraki, *Majority Judgment: Measuring, Ranking, and Electing*. Cambridge, MA, USA: MIT Press, 2010.

[6] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, no. 2, pp. 70–76, Jun. 1967.

[7] S. M. Alessio, *Digital Signal Processing and Spectral Analysis for Scientists: Concepts and Applications*. Cham, Switzerland: Springer, 2016.

[8] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.
[9] J. Wagemans, *The Oxford Handbook of Perceptual Organization*. New York, NY, USA: Oxford Univ. Press, 2015.
[10] E. Leeuwenberg and P. A. van der Helm, *Strutural Information Theory: The Simplicity of Visual Form*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
[11] J. Lauwereyns, *Brain and the Gaze: On the Active Boundaries of Vision*. Cambridge, MA, USA: MIT Press, 2004.
[12] R. D. Luce, "Whatever happened to information theory in psychology?" *Rev. Gen. Psychol.*, vol. 7, no. 2, pp. 183–188, Jun. 2003.
[13] W. R. Garner, *Uncertainty and Structure as Psychological Concepts*. New York, NY, USA: Wiley, 1962.
[14] F. Attneave, "Some informational aspects of visual perception," *Psychol. Rev.*, vol. 61, no. 3, pp. 183–193, May 1954.
[15] F. Attneave and M. D. Arnoult, "The quantitative study of shape and pattern perception," *Psychol. Bull.*, vol. 53, no. 6, pp. 452–471, Nov. 1956.
[16] R. Falk and C. Konold, "Making sense of randomness: Implicit encoding as a basis for judgment," *Psychol. Rev.*, vol. 104, no. 2, pp. 301–318, Apr. 1997.
[17] M. Wertheimer, "Untersuchungen zur Lehre von der Gestalt. II," *Psychologische Forschung*, vol. 4, no. 1, pp. 301–350, Jan. 1923.
[18] A. Desolneux, L. Moisan, and J.-M. Morel, *From Gestalt Theory to Image Analysis: A Probabilistic Approach*. New York, NY, USA: Springer, 2008.
[19] A. G. Shapiro and D. Todorovic, *The Oxford Compendium of Visual Illusions*. New York, NY, USA: Oxford Univ. Press, 2017.
[20] W. Prinz, "Quantitative versuche über die Prägnanz von punktmustern," *Psychologische Forschung*, vol. 29, no. 4, pp. 297–359, Dec. 1966.
[21] R. Arnheim, *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley, CA, USA: Univ. California Press, 1974.

Curriculum Vitae

Vlad Atanasiu

i value the fine grained texture of knowledge spun from blended perspectives and i enjoy playing at kindling ideas that make things move and sparkle

French & Romanian citizenships, Swiss residency (C)
Postgasse 38
3011 Bern, Switzerland
T +41 76 242 8118
atanasiu@alum.mit.edu
http://waqwaq.info

SKILLS

COMPUTER SCIENCE

image processing · visualization · handwriting and document analysis · geographical information systems · text analysis · statistics · data quality · data fusion · human visual perception · user experience · evaluation · robotics · web front- and back-end

HUMANITIES

digital humanities · science and technology studies · book history · paleography · typography · art history · history · linguistics · middle eastern studies

COGNITIVE SCIENCE

visual perception · somatosensory and motor systems

DESIGN

typography · graphic design · fonts · calligraphy

WRITING

creative · editing

FUNCTIONS

research · teach · program · plan · design · write

EXPERIENCE

languages · international work environments

Updated on 2024.05.22

EDUCATION

STUDENT INVOCATIONS OF CONFUCIUS

2012 - 2023	**Ph.D. Computer Science — University of Fribourg, Switzerland** thesis Human Factors as Constraints and Solutions in Document Processing advisor Rolf Ingold
2003 - 2005	**Post-Doc — Massachusetts Institute of Technology, Cambridge, MA, USA** topics Art History, Science and Technology Studies, Brain and Cognitive Sciences advisors Nasser Rabbat, Joseph Dumit, & Ruth Rosenholtz
1996 - 2003	**Ph.D. Art History & History — École Pratique des Hautes Études, Paris, France** thesis The calligraphic phenomenon in Mamluk times (Middle East, 13th-16th c.) advisor François Déroche
1995 - 1996	**M.A.Paleography — École Pratique des Hautes Études, Paris, France** **Institut National des Langues et Civilisations Orientales, Paris, France** thesis Of Letter Frequencies and Their Influence on Arabic Calligraphy advisor François Déroche
1993 - 1995	**M.A. Linguistics — University of Provence, Aix-en-Provence, France** thesis Lexical Fields in "Lisān al-'arab" & Semantic Unity of <gh**> roots advisors Claude-France Audebert & Pierre Larcher
1990 - 1993	**B.A. Middle Eastern Studies — University of Provence, Aix-en-Provence, France** thesis Meaning & Sound in "Maqāyīs al-lugha", a dictionary by Ibn Fāris (d. 395/1004) advisor Claude-France Audebert
1989	**Admission, Telecommunications Engineering — "Traian Vuia" Polytechnic University, Timișoara, Romania**
1989	**Baccalaureate, Mathematics & Physics — "Nikolaus Lenau" High School of German language, Timișoara, Romania**

LANGUAGES

RECOMBINATIONS

Natural Spoken/Written: Arabic (classical, Egyptian & Syrian dialects), English, French, German, Persian, Romanian — Notions: Hebrew, Italian, Japanese, Turkish; Ancient Near Eastern & Semitic languages and scripts (e.g., Akkadian, Egyptian, Pre-Islamic Arabic, South Arabic)

Formal Matlab, Java, Python, JavaScript, D3, Processing, XML, HTML, CSS, PHP, PhpMyAdmin, MySQL, Unix shell, Basic, MOS 6510 assembler — Systems & applications: Aseba for Thymio robots; ESRI ArcGIS Python, Web API, & desktop; Blender; Adobe InDesign IDML, scripting, & GUI; LaTex; FontLab

EMPLOYMENT

2021 - 2024 **Researcher**, Department of Digital Humanities, University of Bern, Switzerland — Full-stack web development of the Swiss historical phonebooks online application historic.localsearch.ch. / Entity names labelization of the Swiss phone book PTT-Archive using reinforcement learning / Quantitative data analysis of the Swiss National Library catalog for the exhibition "Political Representation" / Improved the OCR of historical archival documents through natural language and image post-processing / Historical information visualization / Graphic design

2022 **Teacher**, Department of Computer Science, University of Fribourg, Switzerland — Mentoring in scientific writing and visualization for PhD theses // Supervisor testimonial: "I had the feeling of reading a entirely new thesis!" Rolf Ingold

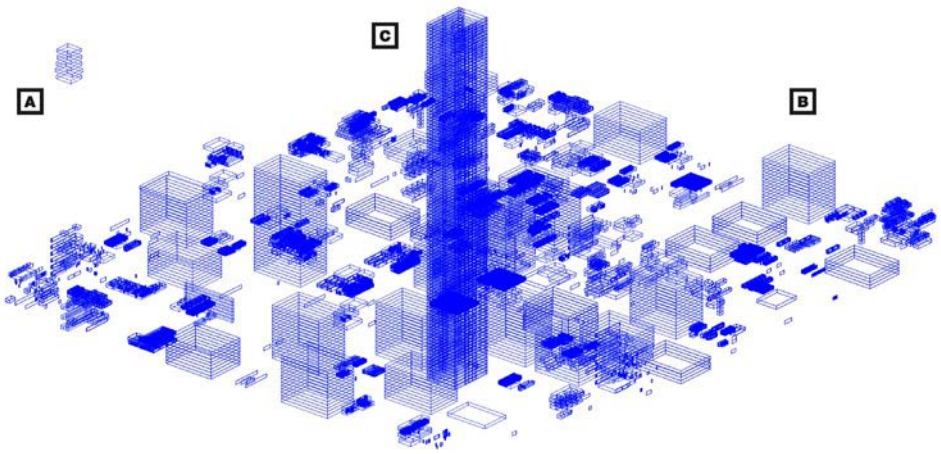
2019 - 2021 **Researcher**, Department of Ancient Studies, University of Basel, Switzerland — Software development for legibility enhancement of papyri using color processing and visual illusions // Testimonial: "Such a useful tool was incredibly needed in our papyrological community." Chiara D'Agostino

2017 - 2020 **Researcher & Planner**, School of Engineering and Architecture, Fribourg, Switzerland – Elaborated a mathematical theory of structural information potential / Automated the classification and visualization of 1.3 million scanned catalog cards of the Swiss National Library / Project architect for academia-industry-government collaborations on digital libraries (Swiss Federal Archives, Swiss National Library, Swiss Post, Sugarcube)

2016 - 2017 **Planner**, Chair of Digital Humanities, University of Vienna, Austria — Project architect for "Alice", Austrian–Flemish project on the use of computational logic for historical studies

2012 - 2015 **Researcher & Teacher**, Department of Informatics, University of Fribourg, Switzerland — How do you formulate queries to describe the unknown beyond the horizon? I conceived "Document Towers", a panoramic visualization based on an architectural metaphor for exploring digital libraries when precise linguistic queries are infeasible and speed is critical / Taught Information Representation for computer scientists (in English & French) // Student testimonial: "Because of that class [. . .] I have landed 2 contracts doing D3 development in America. I appreciate the skillset I developed in that class as it will pay off greatly!"

2011 - 2012 **Teacher**, Centre for Information Modeling / Austrian Centre for Digital Humanities, Graz University, Austria — Taught Information Visualization for Humanities (English & German)



This is a collection of ninety-nine digital articles and one book represented as a city: each document is a tower, each page a flat, each raster image a room; some are born digital documents, some are scanned, impacting keyword search quality.

2011	**Researcher**, National Institute for Applied Sciences (INSA), Lyon, France — Surveyed the state of the art in computational document forensics, a work published in book form
2010 - 2011	**Researcher**, Télécom ParisTech, Paris, France — Carried out research on automated handwriting analysis and developed "Rex", a Web application for document retrieval
2010 - 2011	**Programmer**, National Center for Scientific Research Paris, France — Back and front end programming of "Irancarto", a Web database for the statistical geography of Iran
2009	**Programmer & Designer**, Institute for Lexicography of Austrian Dialects and Names, Austrian Academy of Sciences, Vienna, Austria — Programed an interactive Web-based visualization of the geographical spread of Austrian dialects and designed a physical three-dimensional model thereof
2007	**Teacher**, National Library, Cairo / The Islamic Manuscript Association (TIMA) — trained library staff in paper studies (in Arabic)
2006 - 2010	**Researcher & Planner**, Commission for Scientific Visualization, Austrian Academy of Sciences, Vienna, Austria — Investigated the historiography of paper studies and its impact on its computational modeling / Created "Waqwaq", image processing, data quality control, georeferencing, & geographical information systems software suite for historical paper analysis / Project architect and co-coordinator of the European portal "Bernstein" for historical watermark and paper expertise
2004 - 2005	**Researcher**, Lighthouse International, New York, USA — Contributed to psychophysical and computational research on improving script legibility for people with low vision
2003 - 2005	**Researcher**, Commission for Scientific Visualization, Austrian Academy of Sciences, Vienna, Austria — Part-time research in image processing for Humanities
2003	**Teacher**, National Library of the Netherlands, The Hague / Istituto Centrale di Patologia del Libro, Rome — Installed and dispensed training for the AD751 software, resulting in automating the analysis of 21000 paper reproductions
2002	**Researcher & Programmer**, Centre National de la Recherche Scientifique / Multimedia Press, Paris, France — Developed analytical methods for dating papers based on their laid lines structure and implemented them in the AD751 software
1998	**Assistant**, Paris, France — Assisted visually impaired persons to browse, read, and manage text-based documents

A laser cutter is shown here engraving a plastic plate with a map representing by isolines the density of homonymous populated places in Austria. Lower densities appear at the country's borders, were German placenames coexist along toponyms influenced by Slavic and Romanic languages.

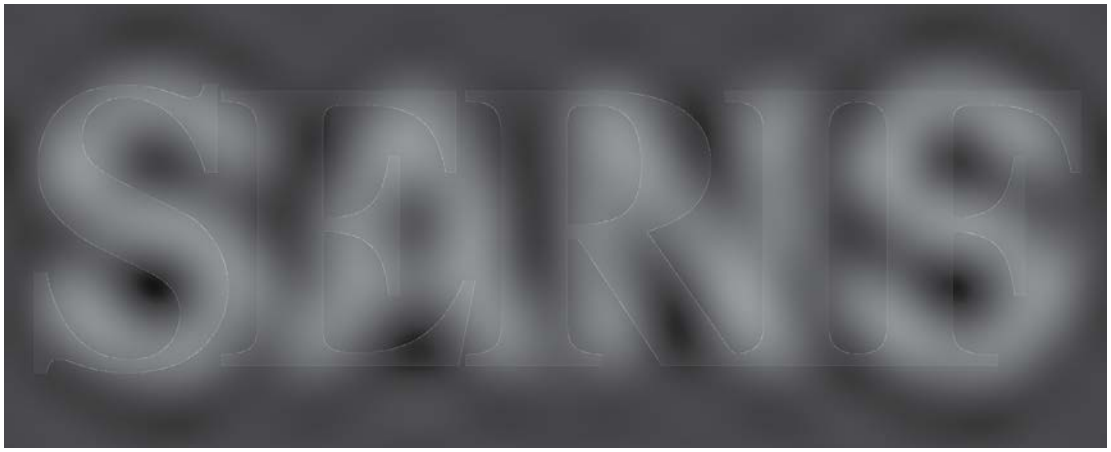
1997	**Designer**, Oxford University Press / American University in Cairo / Laboratory of Egyptology, École Pratique des Hautes Études, Paris — Created fonts for scientific transliteration
1996 - 2003	**Programmer & Researcher**, École Pratique des Hautes Études, Paris, France — Programmed "Kabikadj", the first Web database of manuscripts in Arabic script / Indexed first Hidjra century Qur'anic manuscripts
1994 - 1995	**Teacher**, University of Provence, Aix-en-Provence, France — Tutoring in Arabic language, Middle Eastern history & civilization (in French)
ongoing	**Designer**, freelance — Graphic art: book layout, covers, logos, business cards, Web sites

FIELD WORK

THE DESERT OF THE TARTARS

2001	**Iran**, French National Centre for Scientific Research (CNRS) mission for Project "Manuscript Heritage" (three months) — *also surprised a dragon fast asleep since Mongol times*
2000	**Iran**, French National Centre for Research (CNRS) mission for the digitization of the manuscript collection of the French Research Institute in Iran (three months) — *chased the rabbit in the moon and other tales of Chinese origin*
1999	**Iran**, French Institute of Research in Iran (IFRI), prospecting countrywide on manuscript collections, epigraphy, and contemporary calligraphers (three months) — *battled bandits by night en route through the Balkans*
——	**Syria**, French Institute for Arabic Studies in Damascus (IFEAD), prospecting on manuscript collections, epigraphy, & art history (two months) — *ate grapes from the gardens of King Salomon and chanted the Song of Songs on the Djebel Druze*
——	**Turkey**, Istanbul, prospecting on manuscript collections, epigraphy, & art history (two months) — *Grand Tour of Ottoman cemeteries and escape from the Great Earthquacke*
1998	**Egypt**, French Institute of Oriental Archaeology in Cairo (IFAO), study of Mamluk manuscripts and epigraphy (six months) — *rescued cats in distress and got help from the Shaykh al-Azhar and a blind librarian*
1993 - 1994	**Syria**, French Institute for Arabic Studies in Damascus (IFEAD), training in Middle Eastern Studies (one year) — *captured by soldiers on Mount Hermon while listening Bach and enjoying the vista of the shimmering Euphrates beyond the temples of Palmyra*
1991 - 1992	**Egypt**, Department of Arabic Studies in Cairo (DEAC), linguistic training center of the French Foreign Affairs Ministry, Arabic language training (one year) — *shipwrecked before Alexandria, then mined alabaster in the desert mountains with the seven dwarfs of Luxor*

Grants	European Commission, eContentPlus project "The Memory of Papers: Bernstein. Image based paper expertise and history" / Comprising an Internet portal and software, multilingual watermark database, research activities in humanities and computer science, publications, conferences, ten exhibitions, and an ongoing vibrant research network / A partnership between nine European institutions, including national libraries, with 3.000.000 € budget, between 2006 and 2009 / Role: co-coordinator
	Manuscript Heritage: Digital cataloguing and reproduction of manuscripts collections / École Pratique des Hautes Études (François Déroche), UMR 7528 Monde Iranien: Centre National de la Recherche Scientifique, Institut National des Langues et Civilisations Orientales, Université Paris III Sorbonne Nouvelle (Živa Vesel, Maria Szuppe, Bernard Hourcade, Francis Richard), 2001 to 2003 / Role: founder
Awards	Aga Khan Fellowship, Program for Islamic Architecture at the Massachusetts Institute of Technology (2003–2004)
	International Mobility Fellowship, French Ministry of Education & Research (1998)
Reviewer	Document Engineering (Association for Computing Machinery), International Conference on Document Analysis and Recognition (International Association for Pattern Recognition), Pattern Analysis and Applications, Heritage Science, Abstracta Iranica, Forensic Science International
Expert	European Commission, Horizon 2020 Framework Programme for Research and Innovation
Affiliations	Société Asiatique (France), History & Computing (Switzerland)
Licenses	Paragliding, parachuting, scuba diving



METAMORPHOSIS

Depending on the reading distance the above text says "SANS" or "SERIF". The illusion is created by filtering two words at different spatial frequencies and blending them into a single image. I created this font as a tangible metaphor of the polysemy of language.

REPRESENTATIVE PUBLICATIONS

Information Theory, Image Processing & Visualization

The Structural Information Potential and Its Application to Document Triage", IEEE Access, 2022 (10): 13103-13138.

Presents a theory and a method for the measurement of the information potential that extends the classical information theory of Shannon to structural data; develops a panoramic visual exploration paradigm of digital libraries based on an architectural metaphor as an alternative to keyword search.

History & Art History

Hypercalligraphy: The calligraphic phenomenon in Mamluk times (Middle East, 13th-16th c.); Ph.D. thesis, École Pratique des Hautes Études, Paris, 2003.

A study of cultural rivalry between the Great Powers of Egypt/Syria, Turkey, and Persia during the Eurasian Renaissance period. Showcases how a close technical reading of calligraphy and calligraphers reveals a crystaline hall of mirrors at the heart of socio-culture.

Computational Linguistics & Art History

Of Letter Frequencies and Their Influence on Arabic Calligraphy; book, Paris: L'Harmattan, 1999.

Introduction to the statistical foundations of Arabic calligraphy. Argues that the ortho-graphical rythm of the written language has shaped major calligraphic styles, in a context where the knowledge of letter frequencies gained from scribal practice was used in linguistics, theology, literature, cryptography, and further disparate domains.

Semantics & Lexicography

Meaning and Sound in "Maqāyīs al-lugha", a dictionnay by Ibn Fāris (d. 395/1004); B.A. thesis, University of Provence, Aix-en-Provence, 1993.

Identifies overlaping semantic and phonetic clusters of the Arabic lexicon.E.g., significant amounts of words containing the <gh> root denote something hidden, unclear, or negatively connotated: ghafala *forget*, ghâ'ib *absent*, ghurba *exile*, ghalaqa *close*, ghilâf *cover*, ghûl *death and demon*, gharar *danger*, ghalat *error*, ghair *without*, ghurûb *sunset*, ghaddan *tomorrow*, ghayn *cloud*, ghâba *forest*, ghubâr *dust*, ghanna *nasalisation*.

Human & Computational Geography

Geography & factors of placename homonymy; digital maps & laser engraved physical models, Commission for Visualization / Institute for Austrian Dialects and Name Lexica, Austrian Academy of Sciences, Vienna, Austria, 2010.

Welsh placenames are notoriously long, inhibiting reccurence, contrary to Japanese placenames. Besides social & linguistic factors, homonymity is indicative of cultural diversity, e.g., low in Romanian, German, and Hungarian-speaking Transylvania. It impacts the accuracy of civilian & military geographical information systems.

Psychology & Graphic Design

Ugraphia. The Utopia of the Perfectly Legible Script; article, Conference Proceedings: "Graphemics in the 21st century: From graphemes to knowledge", 14–15 June 2018, Brest, France, pp. 113–140.

Font and script designers use legibility improvement a marketing criterion, a claim backed by little historical and psychophysical evidence and an unattainable goal due to a variety of mathematical, cognitive, and cultural constraints. Nevertheless, legibility remains a compelling selling argument and its history makes a good story.

BOOKS

2024 **Ugraphia: The Pursuit of Perfect Legibility**, Brest: Fluxus Editions, 2024, 186 p., https://doi.org/10.36824/2024-atanasiu. [forthcoming]

2014 **Expert Bytes—Computer Expertise in Forensic Documents: Players, Needs, Resources, and Pitfalls**, Boca Raton, FL: CRC Press, 2014, 198 p., https://doi.org/10.1201/b15467.

1999 De la fréquence des lettres et son influence en calligraphie arabe [**Of Letter Frequencies and Their Influence on Arabic Calligraphy**], Paris: L'Harmattan, 1999, 192 p.

ARTICLES

2023 Vlad Atanasiu, Peter Fornaro, **On the utility of Colour in shape analysis: An introduction to Colour science via palaeographical case studies**, Heliyon, 2023, 9 (10): e20698, 25 p., https://doi.org/10.1016/j.heliyon.2023.e20698.

2022 **The Structural Information Potential and Its Application to Document Triage**, IEEE Access, 2022, 10: 13103–13138, https://doi.org/10.1109/ACCESS.2021.3133654.

— **Improving the Quality Control and Exploration of Digital Documents Using Visualization and Serendipity**, Journal of Information Architecture, 2022, 6(2): 7–81, https://doi.org/10.55135/1015060901/222.011/2.046.

— **Applications of the Document Towers method of representing document structures; Annex: Dreams, Architecture, and Information Systems: A brief review of the utility of document exploration**, Information Design Journal, 2022, 27(1): 1–47, https://doi.org/10.1075/idj.20026.ata.

— Wolfgang Eger (auth.), Heike Bazak, Christiane Sibille (interv.), Vlad Atanasiu (ed.), **Innovating for the Common Good**, in Vlad Atanasiu, Nicolas Chachereau, Enrico Natale, Christiane Sibille (eds), Framing Communication Infrastructures (1950–2020): Discours et imaginaires / Diskurse und Vorstellungen, Revue Histoire et Informatique / Zeitschrift Geschichte und Informatik, vol. 22 (2022), Zürich: Chronos, pp. 123–132.

2021 Vlad Atanasiu, Isabelle Marthot-Santaniello, **Legibility Enhancement of Papyri Using Color Processing and Visual Illusions: A Case Study in Critical Vision**, International Journal on

Document Analysis and Recognition (IJDAR) , 2022 (25): 129–160, https://doi.org/10.1007/s10032-021-00386-0.

—— Vlad Atanasiu, Rolf Ingold, **Document Towers: A MATLAB software implementing a three-dimensional architectural paradigm for the visual exploration of digital documents and libraries**, SoftwareX, 2021, 14: 100684 , 7 p., https://doi.org/10.1016/j.softx.2021.100684.

2018 **Ugraphia. The Utopia of the Perfectly Legible Script**, Extended Abstracts of the conference "Graphemics in the 21st century: From graphemes to knowledge", 14–15 June 2018, Brest, France, pp. 113–140.

2016 **Cuckoos among Your Data: A Quality Control Method to Retrieve Mislabeled Writer Identities from Handwriting Datasets**, IS&T International Symposium on Electronic Imaging, Document Recognition and Retrieval XXIII, 15–18 February 2016, San Francisco, CA, paper 056, 16 p,. https://doi.org/10.2352/ISSN.2470-1173.2016.17.DRR-056.

2011 **Forensic vs. Computing writing features as seen by Rex, the intuitive document retriever**, First International Workshop on Automated Forensic Handwriting Analysis, 17–18 September 2011, Beijing, China, pp. 16–20.

—— Vlad Atanasiu, Laurence Likforman-Sulem, Nicole Vincent, **Writer Retrieval—Exploration of a Novel Biometric Scenario Using Perceptual Features Derived from Script Orientation**, 11th International Conference on Document Analysis and Recognition, 18–21 September 2011, Beijing, China, pp. 628–632, https://doi.org/10.1109/ICDAR.2011.132.

—— **Rex Online**, Gazette du Livre Medieval, 2011 (56–57): 131–135.

—— Table ronde. Applications actuelles de l'informatique à la paléographie: Quelles méthodes pour quelles finalités ? [**Round Table. Paleographical applications of computer science: Which methods and to what aims?**], Gazette du Livre Medieval, 2011 (56–57): 122–123.

—— Bernard Hourcade, Amandine Lebugle, Emanuel Giraudet, Poupak Rafii-Nejad, Vlad Atanasiu, Irancarto : diffuser l'information spatiale pour analyser des questions controversées [**Irancarto: publishing spatial data to analyze controversial questions**], Colloque Fonder les Sciences du Territoire, 23–25 November 2011, Paris, France.

2008 **Using Cartography to reveal Historical Perspectives**, Electronic Cultural Atlas Initiative (ECAI) at Pacific Neighborhood Consortium (PNC) Conference, 4–6 December 2008, Hanoi, Vietnam.

—— Wenger Emanuel, Vlad Atanasiu, **Paper Expertise and Contextual Data Provided by the eContentPlus Project Bernstein**, 29th International Congress of Paper Historians, 27–31 May 2008, Stockholm, pp. 93–101.

2006 Wenger Emanuel, Vlad Atanasiu, eContentPlus Projekt Bernstein. Das Gedächtnis der Papiere [**eContentPlus Project Bernstein. The Memory of Papers**], Proceedings of Conference Electronic Information: the Visual Arts and Beyond, 2006, Berlin, pp. 65–69.

—— The President and the Calligrapher: Arabic Calligraphy and Its Political Use [**futility subverted. The instrumentalization of calligraphy in Middle Eastern politics**], Studies in Architecture, History and Culture. Papers by the 2003–2004 AKPIA@MIT visiting fellows, Aga Khan Program for Islamic Architecture at MIT, 2006, pp. 7–20.

—— Les allographes comme mesure biométrique et la synthèse du comportement [**Allographic biometrics and behavior synthesis**], Cahiers GUTenberg, 46–47 (2006): 5–15.

2004 **Assessing paper origin and quality through large-scale laid lines density measurements**, Proceedings of 26th Congress of the International Paper Historians Association, 30 August – 6 September 2002, Rome/Verona, pp. 172–184.

—— Les réalités subjectives d'un paléographe arabe du X^{e} siècle [**Subjective realities of a 10th century paleographer**], Gazette du Livre médiéval, 2004 (43): 14–22.

2003 **Allographic biometrics and behavior synthesis**, TUGboat, 2003 (24.3): 998–1002.

2002 Le traitement des images et son application à l'histoire du papier: La mesure à grande échelle de la densité de vergeures [**Image analysis and its application to the history of paper: Large-scale measurement of laid lines densities**], Gazette du Livre médiéval, 2002 (41): 31–41.

2001 5170 marches jusqu'à Dieu: Traditions musulmanes relatives à la Tour de Babel et à la confusion des langues [**5170 steps until God: Muslim traditions concerning the Tower of Babel and the confusion of languages**], Annales Islamologiques, 2001 (35): 1–16.

2000 Le Retroencrage: Analyse du ductus des écritures d'après le dégradé du coloris de l'encre [**Retroinking: finding the direction of writing from the intensity of ink**], Gazette du Livre médiéval, 2000 (37): 34–42.

1997 Calligraphie à 50Hz ou l'art de l'écriture influencée par la fréquence des lettres [**Calligraphy at 50 Hz or the art of writing influenced by the frequency of letters**], Proceedings of AFEMAM–EURAMES Conference, 4–7 July 1996, Aix-en-Provence, France.

THESES

2023 **Human Factors as Constraints and Solutions in Document Processing** — Ph.D. Computer science, University of Fribourg, Switzerland, advisor: Rolf Ingold

2003 Hypercalligraphie: Le phénomène calligraphique à l'époque du sultanat mamluk (Moyen-Orient, XIII^{e}–XVI^{e} siècle) [**Hypercalligraphy: The calligraphic phenomenon in Mamluk times (Middle East, 13th–16th century)**] — Ph.D. Art History & History, École Pratique des Hautes Études, Paris, France, advisor: François Déroche

1996 De la fréquence des letteres et de son influence en calligraphie arabe [**Of Letter Frequencies and Their Influence on Arabic Calligraphy**] — M.A. Paleography, École Pratique des Hautes Études, Paris, France, advisor: François Déroche

1995 Familles lexicales dans le "Lisān al-'arab" & Unité du sens des racines [**Lexical Fields in "Lisān al-'arab" & Semantic Unity of <gh**> roots**] — M.A. Linguistics, University of Provence, Aix-en-Provence, advisors: Claude-France Audebert and Pierre Larcher

1993 Les sens dans les sons: d'après "Maqāyīs al-lugha", un dictionnaire d'Ibn Fāris [**Meaning and Sound in "Maqāyīs al-lugha", a dictionary by Ibn Fāris (d. 395/1004)**] — B.A. Middle Eastern Studies, University of Provence, Aix-en-Provence, advisor: Claude-France Audebert

EDITED BOOKS

2024 Enrico Natale, Vlad Atanasiu (eds), **History of Forecasting**, Revue Histoire et Informatique / Zeitschrift Geschichte und Informatik, 2022 (23), Zürich: Chronos. [forthcoming]

2022 Vlad Atanasiu, Nicolas Chachereau, Enrico Natale, Christiane Sibille (eds), **Framing Communication Infrastructures (1950–2020):** [**Discourse and Imaginary**} Discours et imaginaires / Diskurse und Vorstellungen, Revue Histoire et Informatique / Zeitschrift Geschichte und Informatik, 2022 (22), Zürich: Chronos.

2019 Nicolas Chachereau, Christiane Sibille, Enrico Natale, Vlad Atanasiu (eds), Enjeux politiques du web: approches historiques / Das Web als politische Herausforderung: historische Perspektiven [**Political Challenges of the Web: Historical Perspectives**], Revue Histoire et Informatique / Zeitschrift Geschichte und Informatik, 2019 (20), Zürich: Chronos, DOI 10.33057/chronos.1466.

TECHNICAL REPORTS

2023 **Meaning Is Action: Document Analysis Methods for Accessing Switzerland's Historical Phonebooks**, Final Report, Hasler Project no. 22056.

2022	**Improving the matching accuracy of card images and electronic records of the Swiss National Library catalog**, School of Engineering and Architecture, HES·SO//Fribourg, iCoSys, Fribourg, Switzerland
2020	**CardOCR Visual Exploration**, in Beat Wolf, Michael Jungo, Andreas Fischer, Analyse "Automatisierte Integration des alten Sachkatalogs in Alma", School of Engineering and Architecture, HES·SO//Fribourg, iCoSys, Fribourg, Switzerland
2018	**Structural Information Potential: Theory, Measurement, Representation, Applications**, School of Engineering and Architecture, HES·SO//Fribourg, Fribourg, Switzerland
2016	**UHU: An interactive web-based software to visualize the graph edit distance**, Department of Informatics, University of Fribourg.
2013 – 2015	**Automation of EPUB Production: 1. How can InDesign help the tagging of PDFs?, 2. The Visualization of Document Structures**, Document Image and Video Analysis Group (DIVA), Department of Informatics, University of Fribourg / Sugarcube Sarl.
2011	**Computational Handwriting Analysis**, Télécom ParisTech, Paris.
2005	**Spatial proprieties of written information: measurements, encoding, applications**, Balkan Commission, Austrian Academy of Sciences, Vienna, Austria.
1992	**Ghagars of the Nile Delta. An Ethnographic and Linguistic Study** (private commission)

TALKS

2020	**Contextual Versus Generic Technologies: Lessons from selecting images among a large number of papyri enhancements**, IS&T DigiTips, 15 December 2020, online conference.
——	Vlad Atanasiu, Peter Fornaro, **On the Utility of Color for Shape Analysis**, Neo-Paleography conference, University of Basel, 27–29 January 2020, Basel, Switzerland.
2019	**Script Styles Panoramas by Computational Synthesis**, 29th International Congress of Papyrology, 28 July – 3 August, Lecce, Italy.
——	**Computational Graphonomics (d-scribes Project)**, Computerizing Handwritings: Current Approaches on Historical Documents, University of Basel, 27 June 2019, Basel, Switzerland.
——	**Digital Paleography as Odyseey**, Center for the Study of Manuscript Cultures, University of Hamburg, 23 Mai 2019, Hamburg, Germany.
——	Isabelle Marthot-Santaniello, Vlad Atanasiu, **The Digital Paleography of Greek and Coptic Papyri**, Digital Humanities Exchange Conference DHX2019, University of Basel, 21–22 February 2019, Basel, Switzerland.
2017	**See-thorugh Books: Visual Exploration of Vesalius' "Fabrica" (1543)**, Panel Hackathons als Zukunftslabor für die digitale Nachhaltigkeit, Conference Digital Humanities im deutschsprachigen Raum, 13–18 February 2018, Berne, Switzerland.
2015	**Document and library visualization**, Workshop on The state of the art in image recognition, Digital Cultural Heritage Forum, Association of European Research Libraries (Liber) 44th Conference, 24–26 June 2015, London, UK.
2008	**Reflets d'encre: une fantaisie typographique offerte par un paléographe**, Colloque international d'épigrqphie islamique, Hommage à Solange Ory, Maison méditerranéenne des sciences de l'homme (MMSH), 28–29 November 2008, Aix-en-Provence, France.
2008	**Visionary Scientists: What Scientific Visualization can learn form Graphic Arts**, Aesthetics and Science, Symposium on the Occasion of the Solemn Session of the Austrian Academy of Sciences, 13 Mai 2008, Vienna, Austria.
2006	**Bernstein: Collaborative systems for paper expertise and history**, The Islamic Manu-

script II: A Conference to Establish The Islamic Manuscript Association (TIMA), Cambridge University, 7–9 August 2006, Cambridge, UK.

2004 **Handwriting of the Dead: Advances in Script Identification and Synthesis**, Aga Khan Program for Islamic Architecture at the Massachusetts Institute of Technology, 10 May 2005, Cambridge, MA, USA.

2002 **Visible i'ajaz: Three graphic particularities in Qur'an copies**, International Conference on Manuscripts of the Qur'an, 26–28 September 2002, Bologna, Italy.

— **Type with Style!**, Symposium on The history of printing and publishing in the languages and countries of the Middle East, First Congress of Middle Eastern Studies, 9–12 September 2002, Mainz, Germany.

— **The Kabikadj database for manuscripts**, European Association of Middle East Librarians (MELCOM), 27–29 May 2002, Paris, France.

— **The Manuscript Heritage Project for Digital Manuscripts**, French National Centre for Research (CNRS), 23 May 2002, Villejuif, France.

2000 **Statistics, Language, & Calligraphy** (in Persian), Dramatic Arts Center of Iran, Tehran, Iran

— Structures linguistiques et structures graphiques : Problèmes et solutions de la calligraphie du Coran en écritures persanne nasta'liq [**Linguistic and graphical structures: Problems and solutions to writing the Qur'an in the Persian nasta'liq style**], French Institute of Research in Iran, 23 December 2000, Tehran, Iran.

— **Linguistic aspects of calligraphy**, 36th International Congress of Asian and North African Studies (ICANAS), 27 August – 2 September 2000, Montreal, Canada.

— **Cultural ramifications of letter frequencies in Arabic**, New York University Florence, Florence, Italy.

1998 **Sex, Calligraphy and Politics**, Mamluk Studies Conference / Conference of the Middle Eastern Studies Association, 3–6 December 1998, Chicago, IL, USA.

1997 Donne-moi ton regard, je t'offrirais l'immortalité – Apologie pour la sociologie du regard [**Give me your gaze, I will offer you Immortality: Apologue for a sociology of the gaze**], inaugural lecture, Société Asiatique, Palais de l'Institut de France, Paris, France.

— **Give me your gaze, I will offer you Immortality: Apologue for a sociology of the gaze**, 35th International Congress of Asian and North African Studies (ICANAS), Körösi Csoma Society & Eötvös Loránd University, 9–10 July 1997, Budapest, Hungary.

1995 5170 marches jusqu'à Dieu: Traditions musulmanes relatives à la Tour de Babel et à la confusion des langues [**5170 steps until God: Muslim traditions concerning the Tower of Babel and the confusion of languages**], Symposium on Atour de la Pentecôte : universalité du langage et pluralité des langues et cultures, 2 June 1995, Aix-en-Provence, France.

WEB PUBLICATIONS

URL http://waqwaq.info

2010 **Avatars in Real Life: Placename homonymy in Iran**

— **Inimitable handwriting. Graphical specificities of copying the Qur'an**

2009 **7385 Swabian bulls and 2 Viennese mermaids: Interpreting the Bernstein cartography of papers and watermarks**

— **The Score of Death. Plague Epidemics in Europe 1347–1600**

2004 **A Japanese Gaze. Using space to represent time and meaning in Japanese animation**

1999 **Devil in the Gaze. Depicting social non-conformism in Harīrī's Maqāmāt**

REVIEWS

2018 **Computer Networks Histories: Illustrated Review**, Lugano, 14–15 December 2017, 28 p., infoclio.ch

2015 **Data and History: Poetry & Cartoons Review**, Symposium, Berne, 16 September 2015, 20 p., infoclio.ch

— **Document Engineering (DocEng)**, Lausanne, 8–11 September 2015, infoclio.ch

2010 Yuka Kadoi, **Islamic Chinoiserie**, Edinburgh, Edinburgh University Press, 2010 / Abstracta Iranica, 32–3 (2009–10): 147.

2003 David J. Roxburgh, **Prefacing the Image: The Writing Art History in Sixteenth-Century Iran**, Leiden, E.J. Brill, 2000 / Abstracta Iranica, 24 (2001): 82–3.

— W. M. Thackston, **Album Prefaces and Other Documents on the History of Calligraphers and Painters**, Leiden, E.J. Brill, 2000 / Abstracta Iranica, 24 (2001): 84–4.

NOT YET PUBLISHED

2000 **Feathers of Angels: Homoeroticism and calligraphy in Renaissance Persia**

1998 Les Lampes de Déméter: Le symbolisme sacré des lampes à huile en terre cuite dans les bassin méditerranéen [**Demeter's lamps: Sacred symbolism of terracotta oil lamps around the Mediterranean**]

SOFTWARE

Open source code, Matlab File Exchange, MathWorks, https://ch.mathworks.com/matlabcentral/fileexchange/?q=atanasiu

Once you start thinking of impossible books you make them possible. — A book not written is impossible to read. — Every book not written increases the amount of impossible books. — Just-in-time books: when the author starts writing just before the book is shipped out. — The author might perpetually rewrite the entire book, except for the page the reader is in the process of reading. Nobody will notice.

— Impossible Books

SELECTION

ongoing	**Impossible Books**, The New Yorker (submitted)
ongoing	**Travel to Waqwaq** (photographical geo-fiction) — *À Constantinople, où les étoiles / ont le goût de vanille, / dans le Palais du Topkapi / les Marsupilamis / font la chasse à l'Okapi*
2003 - 2005	**Fresh Sushi News** (life at MIT) — *Leonardo da Vinci visits the MIT Nuclear Plant during IAP'04 · Dragon sets volcano ablaze, extinguished by MIT RainLab · Language obsolete! Brain space for rent*
1998	Le Moustique [**The fly**], Le Lien, Revue de l'Association Démocratique des Français à l'Étranger, 1998 (November): 14–15 — *Bzzz*
——	La chaise de Richelieu [**Richelieu's chair**] [the cat of the National Library of France] — *Chaque jour, avant sa tournée dans les galléries de la bibliothèque, et en vue de cette expédition, Richelieu avait pris l'habitude d'aller s'aiguiser les griffes sur un des pieds d'une certaine chaise de la grande salle des Imprimés…*
1995	Le talsiman de Si Sadeq [**Si Sadeq's Talisman**] [fictionalizing documents of the Archives nationales d'outre-mer, Aix-en-Provence] — *C'était un beau garçon. Il avait fait la guerre pour le Turc en 1874 contre les Persans et, de retour en Algérie, il affectait de porter à leur manière le turban, en laissant dépasser quelques mèches blondes. …*
1994	Virgule d'Orient ou comment fut construit Babel [**Oriental commas. Or how Babel was build**], in Anne Zali, Annie Berthier (eds), Les Écritures, Paris: Bibliothèque Nationale, 1997, p. 72. — *S'il y a beaucoup de mots que la langue arabe a prêté aux langues d'Europe, peu sont ceux qui lui ont été restitués. Il serait très étonnant de savoir que ceci s'est passé aussi avec les nobles litières dans lesquelles s'entassent les mots en voyage : les lettres elles-mêmes. …*
1990	Le divan bleu [**Blue Divan**] (novel, bound in rabbit fur) — *Locul îi de mult, / timpul-i de acum, / iar ei din nou au pornit, / prin mahalalei labirint, / ce îi scoate la liman, / limanul Alepului, / poarta Ierusalemului…*

JUSLES VERNE

SELECTION

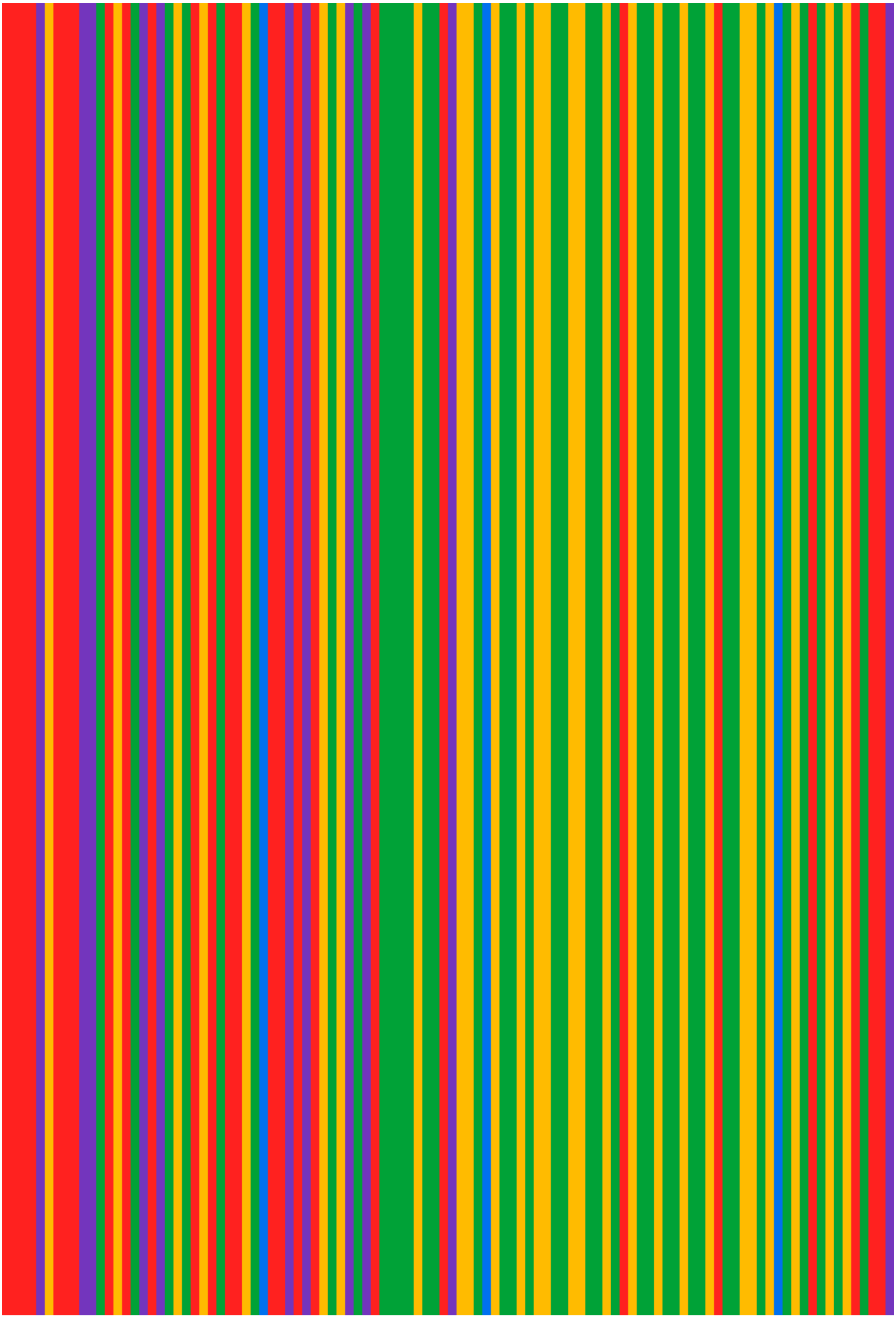
ongoing	**Photography**, Through my Eyes, A photographic journey
2018	**Font**, Metamorphosis
2015	**Layout**, Svizzera Open Access. Guida alle risorse online per le scienze storiche e sociali, infoclio.ch, Berne
2013	**Cover design**, Frank Peter, Sarah Dornhof, Elena Arigita, eds., Islam and the Politics of Culture in Europe: Memory, Aesthetics, Art, transcript Verlag, Bielefeld
2009	**Sonification**, Conversations in Deep Space, Les Grands Boulevards, The Alphabets
2008	**Font**, Sol, hommage to Solange Ory
——	**Logo**, Historians of Islamic Art Association
2006	**Layout**, Roger Delbarre, Communication et violence symbolique: Une autre approche de la société est-allemande, L'Harmattan, Paris
2003	**Impossible books**, objects, private collection Cristina Pallini, Italy/USA
2000	**Graphics**, "Art Slash Life", Art/Life, 2000 (5)
1997	**Fonts** for the transliteration of Oriental languages (Oxford University Press, American University in Cairo and Laboratory of Egyptology, EPHE4, Paris)
1996	**Web site**, Delia Dumitrașcu, painter

CALLIGRAPHY

1994	Student of calligrapher Adnān Sanaqnaqī in **Damascus**, Syria (nasta'līq calligraphic style)
1993	Follows the teaching of master Hassan Celebi in **Istanbul**, Turkey (ruq'a calligraphic style)
1992–3	Noviciate with Prof. Solange Ory at the University of Provence, **Aix-en-Provence**, France

Colophon

Text set in CP Mono by Tino Meinert for Liquitype (2009), a monospaced typeface inspired by British car plates; some adjustments were made by Vlad Atanasiu to make the page gray more even: the width of the “m” was increased (m > m) to avoid high density patches (“peaks”), and the width of spaces and punctuation decreased to avoid low density patches (“lakes”), Titles set in Dax by Hans Reichel for FontShop (1995).

Cover: This laser-engraved wooden tablet encodes, in a manner similar to the representation of sound wave amplitudes, the degree of layout regularity of double-pages in the author’s Expert Bytes book. The tactile document interface is a compact structural representation of documents, especially suitable for document navigation by those who are visually impaired. For such an application, the histogram could be embossed on book covers. The image illustrates the thesis’ idea that human factors create constraints, as well as suggest solutions to document processing. The tablet exudes a pleasant smell of burned wood. (Credits: V. Atanasiu)

Back cover: Color-coded representation of the structural information potential of an issue of the New Yorker cultural magazine, ranging between mostly empty pages and low informativeness (blue), clustered layout patterns and high informativeness (red), and homogeneous ink distribution and medium informativeness (green) (Credits: V. Atanasiu).