Legibility Enhancement of Papyri Using Color Processing and Visual Illusions: A Case Study in Critical Vision

Vlad Atanasiu [1,2] and Isabelle Marthot-Santaniello [1]

University of Basel [1] and University of Fribourg [2], Switzerland, atanasiu@alum.mit.edu, i.marthot-santaniello@unibas.ch

Abstract

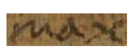
Purpose: This article develops theoretical, algorithmic, perceptual, and interaction aspects of script legibility enhancement in the visible light spectrum for the purpose of scholarly editing of papyri texts. — *Methods:* Novel legibility enhancement algorithms based on color processing and visual illusions are proposed and compared to classic methods. A user experience experiment was carried out to evaluate the solutions and better understand the problem on an empirical basis. — *Results:* (1) The proposed methods outperformed the comparison methods. (2) The methods that most successfully enhanced script legibility were those that leverage human perception. (3) Users exhibited a broad behavioral spectrum of text-deciphering strategies, 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 method satisfied all factor configurations. Therefore, using synergetically a range of enhancement methods and interaction modalities is suggested for optimal results and user satisfaction. (4) A paradigm of legibility enhancement for critical applications is outlined, comprising the following criteria: interpreting images skeptically; approaching enhancement as a system problem; considering all image structures as potential information; deriving interpretations from connections across distinct spatial locations; 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 [48: 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: would you trust money or a passport that has been "enhanced"? 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 mini-experiment also illustrates how enhancement quality depends on more factors than algorithms alone. These two principles frame the work being presented here.

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, 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.

Relevance — Papyri make up an extensive but underexploited cultural heritage. Digital enhancement reduces the time and uncertainty associated with reading these difficult texts from low-quality reproductions. More broadly, image enhancement is a ubiquitous stage in image processing.

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.

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.

Mathematics, algorithms, software, and hardware are never neutral when used by humans, especially so when the

Article accepted with minor revisions by the International Journal on Document Analysis and Recognition (IJDAR) on 2021.03.11.

machine and the human form such a closely coupled system as in the case of legibility enhancement. Therefore, the article present from various points-of-view how psychological and sociological factors are codeterminant in the design of legibility enhancement methods. This context-aware approach also contributes to develop more inclusive technologies.

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.

Organization — Section 2, "Topic", introduces papyrology and scholarly text editing, then defines legibility enhancement trough 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 6, "Paradigm", discusses a type of image enhancement that makes uncertainty explicit, and summarizes the criteria of a computer-aided critical vision system. Section 7, "Implementation", shows how the paradigm and empirical findings were translated in software.

2 Topic

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. Figs. 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" [115: 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 can be manufactured, that is suitable for writing [24]. 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

Figure 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.

sensitivity to horizontal and vertical gratings [18: 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, the presence of holes, and document fragmentation.

2.3 Task

Papyrology is the study of papyri, although it also encompasses inscribed potsherds, wax tablets, metal foils, and other materials [10: 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 [101: 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 [101: 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,

Figure 2. The papyri images used in the evaluation. **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 in 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.

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.

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 *histech* or *adapthistech* 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 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. The strategy fits well with models of visual search [65] and information seeking [63]. In-depth user studies and theoretical models of papyrological reading strategies implemented in expert systems are discussed in [38, 96, 110, 111].

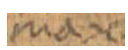
3 Related work

3.1 Papyrological practice

Three approaches to script enhancement characterize common papyrological practice: interactive image manipulation, spectral band decorrelation, and imaging techniques [88, 25].

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.

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 [46, 88]. DStretch decorrelates spectral bands by using either eigenvectors or principal components [5, 44]. 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 pseudo-colors. The method yields excellent results for images with a Gaussian distribution, which is characteristic of inscriptions such as 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. A comparison of DStretch with the *rgb* and *lbk* options (left) and the novel methods *vividness* and *lsv* (right) is shown below:

,  vs , .

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 [114, 40, 2]. With the advent of commercial narrow-band hyperspectral scanners, large-scale digitization has also become possible [58, 57, 30]. 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 [88]. X-ray and terahertz-based imaging have been used for the non-destructive reading of subsurface inscriptions in papyri and mummy cartonnage [66, 42]. 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 [4, 71].

3.2 Computer science research

The authors have found very few articles specifically examining the enhancement of papyri legibility from visible light images [109]. The wider document image literature must be surveyed to identify techniques of potential interest.

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: bleed-through [32], foxing [102], shadows [59], termite bites [94], cross-outs [20], photocopied low-contrast carbon copies [26], low resolution raster images [87], and background texture interference [82] (for a history of image denoising, see [70]). 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) [16]. 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 [85]. Advances in image quality measurement [54, 123] have benefited from research into *visual perception* and *neuroscience* [119, 12], as well as models of *scene statistics* [120]. The role of *tasks* in image enhancement is of particular interest in cultural heritage applications [113]. 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 [37]. 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 [99, 15].

Apart from methods, prominent work areas are datasets, benchmarks, and groundtruthing [72]. Visual confirmation of the attenuation of conspicuous artifacts such as mentioned above is a typical mean of comparing methods, supplemented with numerical characterization if feasible (availability of reference images or appropriateness of reference-free image quality measurements); user evaluations are rare.

4 Methods

4.1 Justification

Color image processing is a highly interdisciplinary area of computer science, owing to the multifarious apsects of color, as diverse as electromagnetic, perceptual, genetic, and

Figure 3. Enhancement results on papyri PSI XIII 1298 (15a) r 1 (top) and P.Oxy.XXII 2309 (bottom). From left to right are 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 the papyrus texture (lower row).

linguistic [41, 60]. 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 [53, 108, 117, 118]. Fig. 3 illustrates the outcomes of the proposed and comparison methods, highlighting interactions between color and shape.

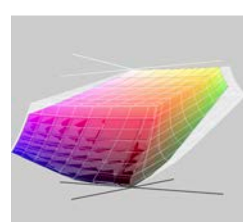
One far-reaching tenet of color research [52, 116] is that the technological, cognitive, and perceptual modalities through which visual reality is apprehended are not necessarily consistent with one another. In the negative polarity method, for example, physically and numerically identical gray-level 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 [90, 103]). 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 straightforward*, and considered by the experimental participants useful. Researchers also have noticed that simplicity is beneficial to legibility enhancement quality [112: 404, 416], and the likelihood of implementation of methods in software and their use by the end-users [62: 920].

4.2 Chroma contrasting by gamut expansion

In the above, 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.

Specifically, the embedded profile of the left image is sRGB IEC 61966-2.1, a standard designed for devices with a limited amount of reproducible color, as were typical in the 1990s. The space defined by these colors is called a color *gamut*. The right image has the larger Adobe RGB (1998) gamut, designed to meet the needs of photographers and high-fidelity reproduction. Humans perceive far more colors than are present in either of these gamuts (compared in the figure below within the CIELAB color space).



Numerical values alone do not define colors; a coordinate system and a space topology are also necessary. When a native 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 is that it does not require any numerical manipulation from software developers.

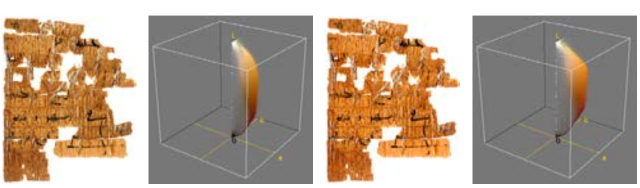
The underlying psychophysics is based on stronger chroma increasing the perceived level of luminance (*brightness*), which is known as the Helmholtz–Kohlrausch effect [36, 31]. This effect is leveraged as an unintended application of gamut mapping to improve the legibility of papyri.

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 work of converting numerical values to visible color is left to 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 [95, 104]. 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 below exhibit the change in chroma and color distribution. On the left is the original image with sRGB profile and its color distribution, and on the right the image after conversion to Adobe RGB (1998).



Papyrus image credits: Sorbonne Université, Institut de Papyrologie.

4.3 Lightness contrasting by stretching

Lightness refers to perceived reflectance, and is an achromatic component of color, together with brightness [98: 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

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

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 [35: 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 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 [105; 35: 201–210]; 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 [29]. D65 (noon daylight, 6504 K) provided in practice the greatest improvement to papyri legibility.

4.4 Lightness contrasting by negative polarity

The characters above, 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, non-linear focus/surround contrast gain, and spatial frequency [97; 98; 18: 354–356, 358–359, 370–371]. It belongs to the broader category of *color constancy*, the neural and ecological basis of which has been a matter of debate since antiquity [43, 1]. Furthermore, the *irradiation effect*, well-known to astronomers, makes light figures in dark surrounds (e.g. stars in the night sky) appear larger [122, 89].

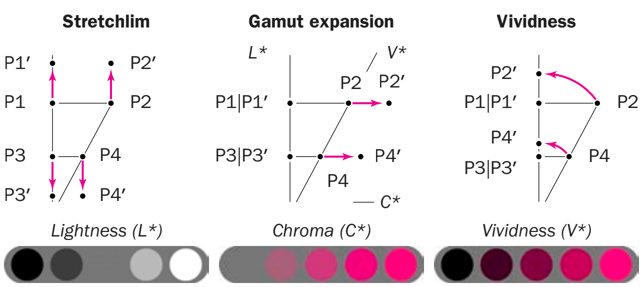
These phenomena allow us posit that reversing the polarity of the image (i.e., so that the script appears lighter), will improve legibility. The empirical testing of this hypothesis as applied to papyri images is reported in Section 5, "Experiment". The formula for negative polarity (*neg*), $L^{*\prime}$,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

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

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).

4.5 Selective contrasting by vividness colorization



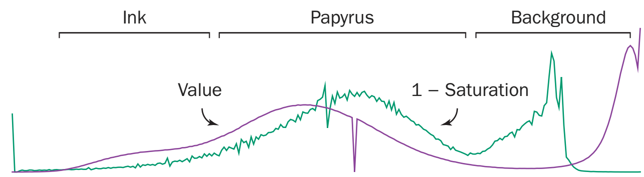
Chroma is a dimension of the CIELAB color space that is orthogonal to lightness and hue. Chroma modification effects a change between achromatic and chromatic color. Another dimension is *vividness*, which defines a concomitant change in lightness and chroma [14]. Vividness can be used as a model of 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 non-linear 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

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

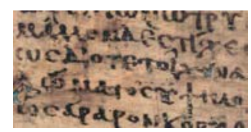
where L^* is the lightness, and a^* and b^* are the chromatic components in CIELAB ($L^{*\prime}$ values exceeding 100 are clipped). The mathematical expression of vividness corresponds to the l^2-norm of the CIELAB image values.

This method amounts to a heuristic component analysis. In the presence of interfering objects in the image, such as color charts, the principal component analysis would be biased in its identification of the ink–papyrus color shift axis. Vividness offers a more robust and simple solution.

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 [106]. Nonetheless, in HSV the distributions of saturation and value in papyri images roughly overlap (see example above). On the basis of a simple difference between them, it is possible to attenuate the background and 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. Before and after images are presented below.

 vs  .

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:

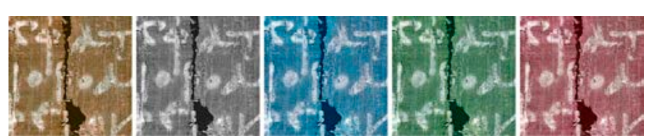
$$L^{*\prime} = 100\,{}_0\big[1 - ({}_0[100 - L^*]^1 + {}_0[V + S - 1]^1)/2\,\big]^1 \,, \tag{4}$$

$$V = \max(R, G, B) \,, \tag{5}$$

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

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 ${}_0[\,\cdot\,]^1$ denotes normalization to the [0, 1] range.

4.7 Hue contrasting by hue shift



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 [18: 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 above images 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 [34: 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 [36]. Moreover, the visual field for blue is broader than that for both green and red, thereby allowing for better comparison between focus and surround [18: 108–109]. A blue text background has also been found to improve performance in analytic tasks [75]. Blue is furthermore robust to the most common colorblindness types [84].

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 [78: 133–169].

Achromatic vision can resolve finer details than chromatic vision [79]. 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 [45, 19].

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

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

where h_i and $h_i{}'$ are the input and output hue in the cylindrical CIELAB color space, with $h = \tan^{-1}(b^*/a^*)$ [35: 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 [33].

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 below:

 vs  .

4.8 Dynamic range increase with CIELAB retinex

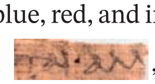
The retinex theory provides an explanation for the constancy of color perception across a wide range of illumination conditions. Essentially, it expresses lightness as the product of ratios of spatial neighborhoods. As an image enhancement technique, it reveals details in dark and light areas. Many variants have been proposed since Land's and McCann's original work, and the outcomes differ according to image quality, content, and intent [67, 74]. 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 of lightness. Pursuing our thinking within the framework of color processing, we find that performing retinex in a perceptual color space (i.e., 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 the illustration below, 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.

a. Input b. RGB Retinex c. CIELAB Retinex d. Blue Negative

Methodology — Conversion from RGB to CIELAB in the sRGB color space. Multiscale retinex with color restoration (MSRCR) [50, 86, 100] 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 the (*b*) RGB and (*c*) V^* (vividness), a^* (blueish–yellowish), and b^* (greenish–reddish) dimensions. Output normalization of (*c*) to the [0, 100], [−128, 127], and [−128, 127] ranges, respectively, and back-conversion to RGB. (*d*) Blue negative applied after CIELAB retinex, with $V^{*\prime} = 100 - V^*$, $a^{*\prime} = -a^*$, $b^{*\prime} = -b^*$.

4.9 Extension to multispectral images

The proposed methods are designed for color images; and stretchlim, negative, and retinex are also for intensity images. All methods can provide benefits even for multispectral images, if these achromatic images, perhaps obtained from the fusion of multiple bands, are considered as CIELAB lightness L^*, and combined with the chromatic channels of a color image (i.e., CIELAB a^* and b^*). Next is shown a papyrus in visible light; in infrared; the infrared "colorized" with the chromatic channels of the visible light image and enhanced with the gamut expansion and stretchlim methods; vividness enhancement of the same; its negative; MSRCR-RGB retinex of the colorization; its negative blue shift; and the decorrelation stretch of the blue, red, and infrared bands:

, , , ,

, , , .

5 Experiment

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.

5.1 Setup

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.

Stimuli — The stimuli consisted of a training and an evaluation set of papyri images (Fig. 1). 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) [92]. 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.

Algorithms — Four novel enhancement algorithms (*vividness, lsv, negvividness, and neglsv*) and five comparison algorithms (*stretchlim*, *histeq*, *adapthisteq*, *locallapfilt*, and *retinex*) were used. All were implemented in Matlab (R2020a), except for retinex, which was produced with ImageJ. 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 grounded in theories of human vision (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 whitepoint and saved them to sRGB to expand the color gamut. Lightness was obtained from CIELAB. — The novel algorithms were those described in Section 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* [73] expands the dynamic range of an image's CIELAB lightness to its bounds, [0, 100]. — *Histeq* [49, 73] 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* (CLAHE) [125, 73] 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* [83, 73] 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) [47].

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.

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 used as demographic sample, 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 "field study" enabled *in situ* simulation of the use of legibility enhancement software under 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, and the display size and number are only some of the factors that would have biased the results if participants had to work the unfamiliar laboratory settings.

The decision not to anonymize the image file names was also purposeful so that the participants' reactions to this information could be observed. The data analysis revealed interesting benefits, particularly with respect to the status of the "original" image among papyrologists. The differences between the results of the enhancement methods in the experiment were sufficiently large to make the methods identifiable after some familiarization. Furthermore, this experimental setup simulates the real-life situation in which users select methods in the interface of the enhancement software by name.

Frame — The experiment measured the preferences for legibility enhancement methods in the context of specific documents and available hardware and software, as developed by participants as they gained expertise in using these methods in real-life 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 another design) transcription errors or duration of task completion that is measured to judge the utility of methods for enhancing legibility, but rather their utility as manifested in the actions of rating and ranking by papyrologists.

Table 1. Ranking and rating of legibility enhancement methods.

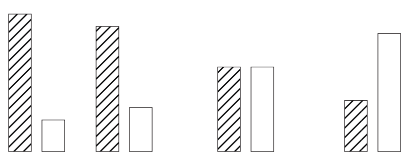
	AW ♂ 2 displays	**FG** ♀ 3 displays	**FRG** ♂ 1 display	**LG** ♂ 2 displays	**OC** ♂ 2 displays	**OR** ♀ 1 display	**SR** ♂ 2 displays	**SN** ♂ 1 display
columbia.apis. p367.f.0.600	A adapthisteq **neglsv** retinex B **negvividness** **vividness** stretchlim **lsv** *original* histeq N locallapfilt	A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

Note: "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



	X	A	B	N
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

5.2 Results

Table 1 presents the ratings and rankings of the experiment's image processing methods. Providing the data *in extenso* is essential to make visibile the statistical features that are quantified, summarized, and discusssed in the next section.

5.3 Analysis

5.3.1 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).

5.3.2 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. 3) 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

Table 4. Intra-class 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	**neglsv**	**neglsv**	retinex	**lsv**	**lsv**
vividness	**vividness**	histeq	*original*	stretchlim	**neglsv**	**neglsv**	*original*
adapthisteq	**neglsv**	**neglsv**	**vividness**	**lsv**	**negvividness**	adapthisteq	**vividness**
neglsv	**negvividness**	*original*	**lsv**	adapthisteq	locallapfilt	**vividness**	stretchlim
negvividness	stretchlim	**lsv**	stretchlim	locallapfilt	**lsv**	stretchlim	**neglsv**
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
neglsv	*original*	*original*	**vividness**	**vividness**	*original*	**vividness**	*original*	**neglsv**
original	**negvividness**	**vividness**	*original*	*original*	**vividness**	**neglsv**	**lsv**	**lsv**
vividness	stretchlim	**negvividness**	**negvividness**	stretchlim	**negvividness**	*original*	**neglsv**	*original*
negvividness	retinex	**neglsv**	**neglsv**	**neglsv**	stretchlim	**negvividness**	**vividness**	**vividness**
lsv	adapthisteq	stretchlim	retinex	retinex	retinex	**lsv**	**negvividness**	**negvividness**
adapthisteq	**vividness**	locallapfilt	**lsv**	**negvividness**	**neglsv**	retinex	retinex	adapthisteq
stretchlim	histeq	retinex	stretchlim	**lsv**	adapthisteq	stretchlim	stretchlim	retinex
locallapfilt	**lsv**	**lsv**	adapthisteq	adapthisteq	**lsv**	adapthisteq	adapthisteq	stretchlim
retinex	**neglsv**	adapthisteq	histeq	histeq	locallapfilt	histeq	histeq	locallapfilt
histeq	locallapfilt	histeq	locallapfilt	locallapfilt	histeq	locallapfilt	locallapfilt	histeq

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.

5.3.3 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.

Methodology — Ranking is treated extensively in statistics [6], but the methods usesd here were found in operations research, marketing, sports, and voting theory. These methods are non-parametric, and thus facilitate handling data with unknown distributions. They take categorical and ordinal data as input, which corresponds to the experimental ratings and rankings. They differ in their perspectives, with the mean-based methods being competitive and rewarding the individual performance of the enhancement methods, while the median-based method is consensual and satisfies the majority of users. The methods are briefly described below; the code was published in [8].

The *Centroids* and *Ratio of Offense and Defense* (ROD) ranking methods are both based on matrix analysis [68: 127–129, 176–197]. This is obtained in the first case by summing the ratings of individual enhancement methods and then calculating the differential between all pairs; here, the row-wise average gives the ratings for the Centroids method (hence its name). In the second case, the approach is more akin to graph analysis, as known from e.g. website ranking methods. Here, negative differentials are set to zero, so that the matrix reflects the dominance of each enhancement method. The ratio of row averages (the "offensive" strength of a method) and column averages ("defensive" strength) yields the final method ratings. (The terminology comes from the world of sports, where these ranking methods originate.)

Majority Judgment is a state-of-the-art consensual rating and ranking method [11] (Table 3B). Its main features are that voters rate each candidate individually, that their aggregated rating is given by the median, and that the method is robust to manipulation. The algorithm consists in removing the median value from the ratings of each candidate, followed by the next median, which is appended lexicographically to the first, and so on, for all ratings. By sorting the resulting scalars, called "majority values", the aggregated ranking of candidates is obtained, where the first median determines the enhancement methods' utility class. "Majority rating" is an extension of the method, contributed by author V.A. It adds the positive or negative value of the so-called "majority gauge" to the rank of a rating class. These values represent the percentage of votes above and below a candidate's majority value.

Empirical evidence suggests that aggregated opinions may outperform individual ones, an approach formalized in *ensemble methods* [64]. When comparing the results of the Centroids and Majority Judgment methods, differences can be seen in ranking and rating, reflecting the methods' outlook. For a balanced result, the two rating sets are combined by considering them as scores in favor of and against a given method; these scores are then normalized and fed into the ROD method. The new ranking is identical to the Majority Judgment and is not modified by *Kemenization*. Kemenization is an optimization method that involves switching consecutive ranks in the results if one method is dominant in most of the rankings to be aggregated [55: 9–23; 68: 175–176].

5.3.4 Are the evaluated methods sufficient?

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 [77, 107]). 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.

5.3.5 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. 4). Various distributions are present: unimodal, bimodal, and uniform (Fig. 4A). 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 [42: 1]. In the 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 [78: 3–5, 107–109, 194]), the linguistic dimension (preserving the names given to colors [78: 218– 219]), or the semantic content (e.g., skin [78: 216]).

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

Methodology — Because the rating and ranking of enhancement methods are not normally distributed (Fig. 4), non-parametric methods were used to analyze the rating and ranking *agreement* and *spread*.

Kendall's coefficient of concordance W is a statistic of agreement between multiple rankings [56: 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:

$$W = 12\,S \,/\, [m^2 (n^3 - n) + 2 - m \sum (t^3 - t)]\,, \tag{9}$$

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:

$$\chi^2 = 12\,S \,/\, [m\,(n^2 + n) + 2 - \sum (t^3 - t)/(n-1)]\,. \tag{10}$$

Spatial flatness SF is the complement of the spectral flatness [124; 21: 104–105], 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:

$$SF = 1 - n\,(\prod |\,\mathcal{F}(x)\,|\,)^{1/n} \,/\, \sum |\,\mathcal{F}(x)\,|\,. \tag{11}$$

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 [51: 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* of a range of possible values by actual values in a vector (e.g., to how many different ranks experimental

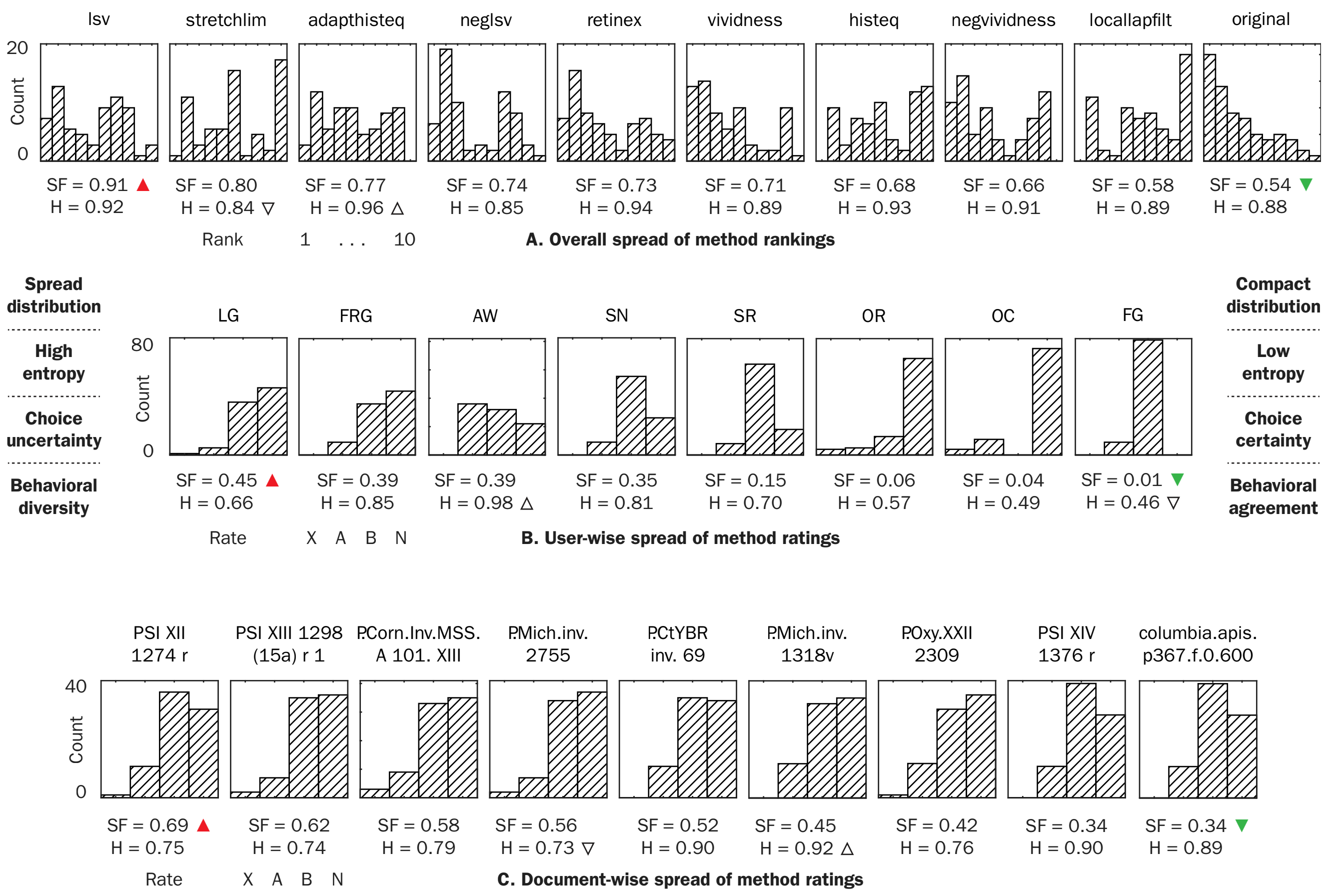


Figure 4. Spread of enhancement method rankings and ratings as indicative of behavioral diversity, measured with spatial flatness *SF* and Shannon entropy *H*.

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. 4 gives the Shannon entropy, *H*, of the ranking histograms. The "peakiest" histogram in Fig. 4A has the lowest *SF* value, but the lowest *H* value is that of a histogram wherein the values are fairly evenly distributed.≠

5.3.6 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.

5.3.7 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 TOOLS 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.

5.3.8 Contextual system optimization is preferable

Image interpretation in scholarly transcription, and even more so in forensic or medical contexts, can have significant repercussions on the advancement of knowledge and people's lives (for example, consider the direct and indirect consequences of the transcription of the Rosetta Stone by J.-F. Champollion). 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 having the system adapt to the behavior of a given user.

5.3.9 The primacy of the "original" image

Overall, papyrologists ranked original images as the most useful for transcription. Experimental 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. 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 made the counterintuitive finding that an 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 [81: 149–151].

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.

5.3.10 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) [22, 93, 69, 61, 18: 206–207]. 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 [18: 244–243] or disease [69]. Also, aeronautical displays with negative polarity help to maintain the adaptation of the visual system to a dark environment during night flight [91]. 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 [13]. 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 [76]; the word "God" calligraphed in gold on a green eight-meter-high panel in the Aya Sofia mosque of Istanbul [17: 504]; and the 12th-century "Blue Sutra" of Japan, in gold on blue paper [39: 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.

5.3.11 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, above all, a state of mental flow during task completion.

5.3.12 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 [27, 3]). For example, females have been found to have greater chromatic discrimination ability, whereas males are more sensitive to lightness variation [121]. 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.

6 Paradigm

Here, is provided a paradigm of legibility enhancement for critical applications, as a stage towards developing computer-assisted critical vision systems. This paradigm is a requirement for 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.

6.1 Explicitation of uncertainty

Requirements — Because vision is purposeful, we might sometimes stare at things without seeing them when there is no reason to seek them out. A converse problem is to believe what we see. Generally speaking, the ideal enhancement system would inform about the uncertainty in both images and their interpretation. Uncertainty is important and its presence should be made explicit. The system should thus aid in its detection and analysis, represent it visually and numerically, and preserve it for verification. *The management of uncertainty should be at the core of the enhancement system and its components, from the algorithms to the interface.*

Examples — An analogy is the use of white material for pottery reconstruction in archeology to distinguish between missing and original parts. A classic example of script ambiguity and the role of context in reading is the figure to the right, wherein the central shape is interpreted as either a letter or a number depending on the direction in which it is read. Mundane cases of visual ambiguity include faded maps or letters, and frightening shadows at night. A critical vision system should make the analyst aware of the potential for misinterpretation.

Rationale — A paradigm called *conservative preprocessing* has been proposed by Chen, Lopresti, and Nagy in a rare work on the theory of image enhancement [28]. It consists in preserving the original by reversible transformations and by 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, or was not scanned, or belongs to a missing second or third page [113]. Such invasive enhancement can thus destroy historical, forensic, and conservation-related information in documents and reproductions [80, 40]. Problems of this kind are common in digital library projects. Technical reasons can also make enhancement problematic: for example, because band-pass filtering is known to frequently lead to the suppression of script parts [82: 445], should the user not be made aware of the possibility of erroneous readings?

Formalization — An information theoretical approach may help formalize the explicitation of uncertainty. Specifically, optimal enhancement is obtained for a certain equilibrium of the entropy of both the image structure and content between minimal (= no alternative) and maximal (= quandary). The heterogeneity of user behavior in this experiment suggests that this optimum is contextually determined.

Operationalization — The need to support critical vision entails a system with a visual representation of uncertainty, numerical methods driving the enhancement process, and human–machine interaction tools to assist users in image interpretation. The result is much more than an image enhancement algorithm: it is a *computer-assisted critical vision system*. Such systems are under development, e.g., to support radiographic diagnosis in clinical medicine [99: 359–414].

6.2 Criteria

Our work resulted in 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 — Artificial support for critical vision is realized by making alternative image interpretations explicit. This is optimal for a certain equilibrium between minimal and maximal signal entropy. It is manifested visually by the enhanced image, numerically in enhancement methods, and in human–machine interaction.

4. SYSTEMS APPROACH — Legibility enhancement for critical applications is a system problem. Optimization depends on tasks, data, users, tools, and their interactions.

7 Implementation

The research findings were implemented in the Open Source software Hierax [9]. To account for the heterogeneity of user behavior, a broad range of processing methods were offered: the novel methods, adapthisteq, and retinex. Since interaction with the images was found to affect the usefulness of the enhanced images, the software includes an image viewer. One interface tab presents an overview of the enhanced images, and a second allows for detailed comparison. Zooming, panning, and rotation are synchronized. The arrow keys are used to scroll through the images, decoupling vision from motor control. Rapid switching between images helps user discover differences. These facilities were appreciated by the paleographers who beta-tested Hierax. Hierax has already been used in a real-world setting, to enhance the 70 papyri of the University of Basel Library, in view of an online catalog.

8 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* and *contextualization*, instead of searching for an overall optimum.

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

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).

Acknowledgments — 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; as well as Gorge Nagy and Andreas Fischer, for a constructive 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.

References

1. Adelson EH (2000). "Lightness Perception and Lightness Illusions", in Gazzaniga MS, ed., *The New Cognitive Neurosciences*, Cambridge, MA: MIT Press, 2nd ed., pp. 339-351.
2. Alexopoulou AA, Kaminari AA, Panagopoulos A, Pöhlmann E (2013). "Multispectral documentation and image processing analysis of the papyrus of tomb II at Daphne, Greece", *J. of Archaeological Science*, 40 (2): 1242–1249.
3. Allberry CRC *et al.* (1938). *Actes du Ve Congrès international de papyrologie, Oxford, 30 Août – 3 Septembre 1937*, Bruxelles: Fondation égyptologique Reine Élisabeth.
4. Allegra D, Ciliberto E, Ciliberto P, Milotta FLM, Petrillo G, Stanco F, Trombatore C (2015). "Virtual Unrolling Using X-ray Computed Tomography", *Proc. 23rd European Signal Processing Conf. (EUSIPCO), 31 August – 4 September 2015, Nice, France*, pp. 2864–2868.
5. Alley RE (1996). *Algorithm Theoretical Basis Document for Decorrelation Stretch*, ASTER-AST06, Pasadena, CA: Jet Propulsion Laboratory, http://www.dstretch.com/DecorrelationStretch.pdf.
6. Alvo M, Yu PhLH (2014). *Statistical Methods for Ranking Data*, New York, NY: Springer.
7. Asano Y, Fairchild MD, Blondé L (2016). "Individual Colorimetric Observer Model", *PLoS ONE*, 11 (2): e0145671.
8. Atanasiu V (2020). "Non-parametric rating and ranking functions", *MATLAB Central File Exchange*, https://www.mathworks.com/matlabcentral/fileexchange/78384.
9. Atanasiu V, Marthot-Santaniello I (2020). *Hierax Legibility Enhancement Software*, https://hierax.ch.
10. Bagnall RS (2009). *The Oxford Handbook of Papyrology*, Oxford: Oxford Univ. Press.
11. Balinski M, Rida L (2010). *Majority Judgment Measuring, Ranking, and Electing*, Cambridge, MA: MIT Press.
12. Barten PGJ (1999). *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*, Bellingham, WA: SPIE Optical Engineering Press.
13. Bauer D, Cavonius C (1980). "Improving the legibility of visual display units through contrast reversal", in Grandjean E and Vigliani E (eds.), *Ergonomic aspects of visual display terminals*, London: Taylor & Francis, pp. 137–142.
14. Berns RS (2014). "Extending CIELAB: Vividness, V^*_{ab}, Depth, D^*_{ab}, and Clarity, T^*_{ab}", *Color Research and Application*, 39 (4): 322–330.
15. Beutel J, Van Metter RL, Kundel HL, eds (2000). *Handbook of Medical Imaging*, Bellingham, WA: SPIE, vol. 1.
16. Bhowmik Sh., Sarkar R, Nasipuri M, Doermann D (2018). "Text and non-text separation in offline document images: a survey", *Intl. J. on Document Analysis and Recognition*, 21 (1–2): 1–20.
17. Blair Sh (2006). *Islamic Calligraphy*, Edinburgh: Edinburgh Univ. Press.
18. Boff KR, Lincoln JE, eds. (1988). *Engineering Data Compendium: Human Perception and Performance*. Ohio, OH: Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, vol. 1.
19. Breuil C, Jennings BJ, Barthelmé S, Guyader N, Kingdom FAA (2019). "Color improves edge classification in human vision", *PLoS Computational Biology*, 15 (10): e1007398.
20. Brink A, van der Klauw H, Schomaker L (2008). "Automatic removal of crossed-out handwritten text and the effect on writer verification and identification", *Proc. SPIE–IS&T Electronic Imaging Conf., Document Recognition and Retrieval XV, 29–31 January 2008, San Jose, CA, USA*, vol. 6815: 68150A.
21. Broersen PMT (2006). *Automatic Autocorrelation and Spectral Analysis*, London: Springer.
22. Buchner A, Mayr S, Brandt M (2009). "The advantage of positive text-background polarity is due to high display luminance", *Ergonomics*, 52 (7): 882–886.
23. Buckle CE, Udawatta V, Straus CM (2013). "Now You See It, Now You Don't: Visual Illusions in Radiology", *RadioGraphics*, 33 (7): 2087–2102.
24. Bülow-Jacobsen A (2009). "Writing Materials in the Ancient World", in Bagnall RS, *The Oxford Handbook of Papyrology*, Oxford: Oxford Univ. Press, pp. 3–29.
25. Bülow-Jacobsen A (2021). "Photography of Papyri and Ostraca", in Caputo C, Lougovaya J, eds., *Using Ostraca in the Ancient World. New Discoveries and Methodologies*, Berlin: De Gruyter, pp. 59–83.
26. Cannon M, Hochberg J, Kelly P (1999). "Quality assessment and restoration of typewritten document images", *Intl. J. on Document Analysis and Recognition*, 2 (2–3): 80–88.
27. Capasso M, Davoli P (2019). *Abstracts of the 29th Intl. Cong. of Papyrology, 28 July – 8 August 2019, Lecce, Italy*, Lecce: Università del Salento.
28. Chen J, Lopresti D, Nagy G (2016). "Conservative preprocessing of document images", *Intl. J. on Document Analysis and Recognition*, 19 (4): 321–333.
29. CIE (2004). *Colorimetry*, TR 15:2004, Vienna: CIE.
30. Ciortan I, Deborah H, George S, Hardeberg JY (2015). "Color and Hyperspectral Image Segmentation for Historical Documents", *Proc. 2015 Digital Heritage Conf., 28 September – 2 October 2015, Granada, Spain*, pp. 199–206.
31. Donofrio RL (2011). "Review Paper: The Helmholtz-Kohlrausch Effect", *J. of the Society for Information Display*, 19 (10): 658–664.
32. Drira F, LeBourgeois F (2017). "Mean-Shift segmentation and PDE-based nonlinear diffusion: toward a common variational framework for foreground/background document image segmentation", *Intl. J. on Document Analysis and Recognition*, 20 (3): 201–222.
33. Elliot AJ, Fairchild MD, Franklin A, eds. (2015). *Handbook of Color Psychology*, Cambridge: Cambridge Univ. Press.
34. Fairchild MD (1996). "Refinement of the RLAB Color Space", *Color Research and Application*, 21 (5): 338–346.
35. Fairchild MD (2013). *Color Appearance Models*. Chichester: Wiley, 3rd ed.
36. Fairchild MD, Pirrotta E (1991). "Predicting the lightness of chromatic object colors using CIELAB", *Color Research and Application*, 16 (6): 385–393.
37. Fitts PM, ed. (1947). *Psychological Research on Equipment Design*, Report no. 19, Army Air Forces: Washington D.C.
38. de la Flor G, Luff P, Jirotka M, Pybus J, Kirkham R, Carusi A (2010). "The Case of the Disappearing Ox: Seeing Through Digital Images to an Analysis of Ancient Texts", *Proc. Conf. on Human Factors in Computing Systems (CHI 2010), 10–15 April 2010, Atlanta, GA, USA*, pp. 473–482.
39. Fu Sh, Lowry GD, Yonemura A (1986). *From Concept to Context:*

Approaches to Asian and Islamic Calligraphy, Washington, DC: Smithsonian Institution.
40. Gargano M, Bertani D, Greco M, Cupitt J, Gadia D, Rizzi A (2015). "A perceptual approach to the fusion of visible and NIR images in the examination of ancient documents", *J. of Cultural Heritage*, 16 (4): 518–525.
41. Gevers Th, Gijsenij A, van de Weijer J, Geusebroek J-M (2012). *Color in Computer Vision: Fundamentals and Applications*. Chichester: Wiley.
42. Gibson A, Piquette KE, Bergmann U, Christens-Barry W, Davis G, Endrizzi M, Fan Sh, *et al.* (2018). "An assessment of multimodal imaging of subsurface text in mummy cartonnage using surrogate papyrus phantoms", *Heritage Science*, 6: 7.
43. Gilchrist A (2006). *Seeing Black and White*, Oxford: Oxford Univ. Press.
44. Gillespie AR, Kahle AB, Walker, RE (1986). "Color enhancement of highly correlated images: I. Decorrelation and HSI contrast stretches", *Remote Sensing of Environment*, 20: 209–235.
45. Hansen Th, Gegenfurtner KR (2009). "Independence of color and luminance edges in natural scenes", *Visual Neuroscience*, 26: 35–49.
46. Harman J (2008). "Using Decorrelation Stretch to Enhance Rock Art Images", *DStretch: Rock Art Digital Enhancement*, website, http://www.dstretch.com/AlgorithmDescription.html.
47. Hernández F-J (2010). *Retinex for ImageJ*, https://web.archive.org/web/20130323053256/http://www.dentistry.bham.ac.uk/landinig/software/retinex/retinex.html.
48. Homer (1990). *The Iliad*, trans. Robert Fagles, New York, NY: Penguin.
49. Hummel R (1977). "Image enhancement by histogram transformation", *Computer Graphics and Image Processing*, 6: 184–195.
50. Jobson DJ, Rahman Z, Woodell GA (1997). "A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes", *IEEE Trans. on Image Processing*, 6 (7): 965–976.
51. Hyoung-Gook K, Moreau N, Sikora Th (2005). *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*, Chichester: Wiley.
52. Kambara A (2017). "Effects of Experiencing Visual Illusions and Susceptibility to Biases in One's Social Judgments", *Perceptual and Motor Skills*, 7 (4): 1–6.
53. Karatzas D, Antonacopoulos A (2007). "Colour text segmentation in web images based on human perception", *Image and Vision Computing*, 25: 564–577.
54. Keelan BW (2002). *Handbook of Image Quality: Characterization and Prediction*, New York, NY: Marcel Dekker.
55. Kemeny JG, Snell JL (1972). *Mathematical Models in the Social Sciences, Cambridge*, MA: MIT Press.
56. Kendall MG (1970). *Rank Correlation Methods*, New York, NY: McGraw-Hill, 4th ed.
57. Kim SJ, Deng F, Brown MS (2011). "Visual enhancement of old documents with hyperspectral imaging", *Pattern Recognition*, 44: 1461–1469.
58. Kim SJ, Zhuo Sh., Deng F, Fu Ch-W, Brown MS (2010). "Interactive Visualization of Hyperspectral Images of Historical Documents", *IEEE Trans. on Visualization and Computer Graphics*, 16 (6): 1441–1448.
59. Kligler N, Katz S, Tal A (2018). "Document Enhancement using Visibility Detection", *Proc. 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 18–23 June 2018, Salt Lake City, UT, USA*, pp. 2374–2382.
60. Koschan A, Abidi M (2008). *Digital Image Processing*, Chichester: Wiley.
61. Kowalik G, Nielek R (2016). "Senior Programmers: Characteristics of Elderly Users from Stack Overflow", in Spiro E, Ahn Y.-Y., eds., *Social Informatics: 8th Intl. Conf., SocInfo 2016, Bellevue, WA, USA, November 11–14, 2016, Proceedings, Part II*, pp. 87–96.
62. Krupinski EA (2009). "Practical Applications of Perceptual Research", in Beutel J, Van Metter RL, Kundel HL, eds., *Handbook of Medical Imaging, Bellingham*, WA: SPIE, vol. 1, pp. 895–929.
63. Kuhlthau CC (2004). *Seeking Meaning: A Process Approach to Library and Information Services*, Westport, CT: Libraries Unlimited.
64. Kuncheva LI (2014). *Combining pattern classifiers : methods and algorithms*, New Jersey, NJ: Wiley, 2nd ed.
65. Kundel HL (2000). "Visual Search in Medical Images", in Beutel J, Van Metter RL, Kundel HL, eds., *Handbook of Medical Imaging, Bellingham*, WA: SPIE, vol. 1, pp. 837–858.
66. Labaune J, Jackson JB, Pagès-Camagna S, Menu M, Mourou GA (2010). "Terahertz investigation of Egyptian artifacts", *Proc. 35th Intl. Conf. on Infrared, Millimeter, and Terahertz Waves, 5–10 September 2010, Rome, Italy*, pp. 1–3.
67. Land EH (1977). *The Retinex Theory of Color Vision*, Scientific American, 237 (6): 108–128.
68. Langville AN, Meyer CD (2012). *Who's #1?: The Science of Rating and Ranking*, Princeton, NJ: Princeton Univ. Press.
69. Legge GE, Rubin GS, Schleske MM (1987). "Contrast-polarity effects in low-vision reading", in Woo GC, ed., *Low Vision: Principles and Applications*, New York, NY: Springer, pp. 288–307.
70. Li X (2013). "Image Denoising: Past, Present, and Future", Gunturk BK, Li X, eds., *Image Restoration: Fundamentals and Advances*, Boca Raton, FL: CRC Press, pp. 1–19.
71. Lin Y, Seales WB (2005). "Opaque Document Imaging: Building Images of Inaccessible Texts", *Proc. 10th IEEE Intl. Conf. on Computer Vision (ICCV'05), 17–21 Oct. 2005, Beijing, China*, vol. 1, pp. 662–669.
72. Marinai S, Karatzas D (2011). "Report from the AND 2009 working group on noisy text datasets", *Intl. J. on Document Analysis and Recognition*, 14 (2): 113–116.
73. MathWorks (2020). "Contrast Adjustment", *Matlab: Image Processing Toolbox*, https://www.mathworks.com/help/images/contrast-adjustment.html.
74. McCann JJ (2017). "Retinex at 50: color theory and spatial algorithms, a review," *J. of Electronic Imaging*, 26 (3): 031204.
75. Mehta R, Zhu R (2009). "Blue or Red? Exploring the Effect of Color on Cognitive Task Performances", *Science*, 323 (5918): 1226–1229.
76. Miller E, Lee NJ, Uprichard K, Daniels V (2000). "The Examination and Conservation of the Rosetta Stone at the British Museum", in Roy A, Smith P, eds., *Tradition and Innovation: Advances in Conservation*, The Intl. Institute for Conservation of Historic and Artistic Works, pp. 128–132.
77. Mocella V, Brun E, Ferrero C, Delattre D (2015). "Revealing letters in rolled Herculaneum papyri by X-ray phase-contrast imaging", *Nature Communications*, 6:5895.
78. Morovič J (2008). *Color Gamut Mapping*, Chichester: Wiley.
79. Mullen KT (1985). "The contrast sensitivity of human colour vision red-green and blue-yellow chromatic gratings", *The J. of Physiology*, 359: 381–400.
80. Mussell J (2012). *The Nineteenth-Century Press in the Digital Age*, London: Palgrave Macmillan.
81. Oppenheimer DM (2006). "Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly", *Applied Cognitive Psychology*, 20: 139–156.
82. Pan X-B, Brady M, Bowman AK, Crowther Ch, Tomlin RSO (2004). "Enhancement and feature extraction for images of incised and ink texts", *Image and Vision Computing*, 22: 443–451.
83. Paris S, Hasinoff SW, Kautz J (2015). "Local Laplacian Filters: Edge-Aware Image Processing with a Laplacian Pyramid", *Communications of the ACM*, 58 (3): 81–91.
84. Parry NRA (2015). "Color vision deficiencies", in Elliot AJ, Fairchild MD, Franklin A, eds., *Handbook of Color Psychology*, Cambridge: Cambridge University Press, pp. 216–242.

85. Perry S (2018). “Image and Video Noise: An Industry Perspective”, in Bertalmío M, ed., *Denoising of Photographic Images and Video: Fundamentals, Open Challenges and New Trends*, Cham: Springer, pp. 217–230.
86. Petro AB, Sbert C, Morel J-M (2014). “Multiscale Retinex”, *Image Processing On Line*, 4: 71–88.
87. Pham Th-A, Delalandre M (2017). “Post-processing coding artefacts for JPEG documents”, *Intl. J. on Document Analysis and Recognition*, 20 (3): 189–200.
88. Piquette KE (2018). “Revealing the Material World of Ancient Writing: Digital Techniques and Theoretical Considerations”, in Hoogendijk FA J, van Gompel SM 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*, Leiden: Brill, pp. 94–118.
89. Plateau J (1839). *Mémoire sur l’irradiation*, Bruxelles: Hayez.
90. Da Pos O, Zambianchi E (1996). *Visual Illusions and Effects: A Collection*, Milano: Guerini.
91. Poston AM (1974). *A literature review of cockpit lighting*, technical memorandum 10-64, Aberdeen Proving Ground, MD: US Army Human Engineering Laboratory.
92. Pratikakis I, Zagoris K, Karagiannis X, Tsochatzidis L, Mondal T, Marthot-Santaniello I (2019). “ICDAR 2019 Competition on Document Image Binarization (DIBCO 2019)”, *Proc. 2019 Intl. Conf. on Document Analysis and Recognition (ICDAR), 20–25 September 2019, Sydney, Australia*, pp. 1547–1556.
93. Pons C, Mazade R, Jin J, Dul MW, Zaidi Q, Alonso J–M, (2017). “Neuronal mechanisms underlying differences in spatial resolution between darks and lights in human vision”, *J. of Vision*, 17 (14): 5, 1–24.
94. Raj V, Arunkumar C (2014). “Content Restoration of Degraded Termite Bitten Document Images”, *Intl. J. of Engineering Research and Applications*, 4 (5): 151–155.
95. Rodney A (2005). *Color Management for Photographers*, Oxford: Focal Point.
96. Roued-Cunliffe H (2011). *A decision support system for the reading of ancient documents*, Ph.D. thesis, Faculty of Classics, Univ. of Oxford, Oxford.
97. Rudd ME (2017). “Lightness computation by the human visual system”, *J. of Electronic Imaging*, 26 (3): 031209.
98. Rudd ME, Popa D (2007). “Stevens’s brightness law, contrast gain control, and edge integration in achromatic color perception: a unified model”, *J. Optical Soc. of America*, A/24 (9): 2766–2782.
99. Samei E, Krupinski EA, eds. (2019). *The Handbook of Medical Image Perception and Techniques*, Cambridge: Cambridge Univ. Press, 2nd ed.
100. Schivre G (2020). “Multiscale Retinex”, *MATLAB Central File Exchange*, https://www.mathworks.com/matlabcentral/fileexchange/71386-multiscale-retinex.
101. Schubert P (2009). “Editing a Papyrus”, in Bagnall RS, ed., *The Oxford Handbook of Papyrology*, Oxford: Oxford Univ. Press, pp. 197–215.
102. Seuret M, Chen K, Eichenberger N, Liwicki M, Ingold R (2015). “Gradient-domain degradations for improving historical documents images layout analysis”, *Proc. 13th Intl. Conf. on Document Analysis and Recognition (ICDAR), 23–26 August 2015, Nancy, France*, pp. 1006–1010.
103. Shapiro A, Todorović D, eds. (2017). *The Oxford Compendium of Visual Illusions*, Oxford: Oxford Univ. Press.
104. Sharma A (2018). *Understanding Color Management*, Hoboken, NJ: Wiley, 2nd ed.
105. Sharma G, Rodríguez-Pardo CE (2012). “The Dark Side of CIELAB”, *Proc. SPIE–IS&T Electronic Imaging, Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications*, vol. 8292: 82920D.
106. Smith AR (1978). “Color gamut transform pairs”, *SIGGRAPH ‘78: Proc. 5th Annual Conf. on Computer Graphics and Interactive Techniques, 23–25 August 1978, Atlanta, GA, USA*, pp. 12–19.
107. Sider D (2005). *The Library of the Villa Dei Papiri at Herculaneum*, Los Angelse, CA: Getty Publications.
108. Sobottka K, Kronenberg H, Perroud T, Bunke H (2000). “Text extraction from colored book and journal covers”, *Intl. J. on Document Analysis and Recognition*, 2 (4): 163–176.
109. Sparavigna AC (2009). “Digital Restoration of Ancient Papyri”, *Computing Research Repository arXiv.org*, http://arxiv.org/abs/0903.5045.
110. Tarte S (2011). “Papyrological investigations: transferring perception and interpretation into the digital world”, *Literary and Linguistic Computing*, 26 (2): 233–247.
111. Terras MM (2006). *Image to Interpretation: An Intelligent System to Aid Historians in Reading the Vindolanda Texts*, Oxford: Oxford Univ. Press.
112. Tonazzini A (2010). “Color Space Transformations for Analysis and Enhancement of Ancient Degraded Manuscripts”, *Pattern Recognition and Image Analysis*, 20 (3): 404–417.
113. Tonazzini A, Gerace I, Martinelli F (2012). “Document Image Restoration and Analysis as Separation of Mixtures of Patterns: From Linear to Nonlinear Models”, in Gunturk BK, Li X, eds., *Image Restoration: Fundamentals and Advances*, Boca Raton, FL: CRC Press, pp. 285–310.
114. Tournié A, Fleischer K, Bukreeva I, Palermo F, Perino M, Cedola A, Andraud C, Ranocchia G (2019). “Ancient Greek text concealed on the back of unrolled papyrus revealed through shortwave-infrared hyperspectral imaging”, *Science Advances*, 5 (10) eaav8936.
115. Van Minnen P (2009). “The Future of Papyrology”, in Bagnall RS, ed., *The Oxford Handbook of Papyrology*, Oxford: Oxford Univ. Press, pp. 644–660.
116. VanderPlas S, Hofmann H (2015). “Signs of the Sine Illusion — Why We Need to Care”, *J. of Computational and Graphical Statistics*, 24 (4): 1170–1190.
117. Velázquez A, Levachkine S (2004). “Text/Graphics Separation and Recognition in Raster-Scanned Color Cartographic Maps”, *Proc. Intl. Workshop on Graphics Recognition, GREC 2003: Graphics Recognition. Recent Advances and Perspectives, Lecture Notes in Computer Science*, vol 3088, pp. 63–74.
118. Ventzas D, Ntogas N, Ventza MM (2012). “Digital Restoration by Denoising and Binarization of Historical Manuscripts Images”, in Ventzas D, ed., *Advanced Image Acquisition, Processing Techniques and Applications*, Rijeka: InTech, pp. 73–108.
119. Wang Zh, Bovik AC (2006). *Modern Image Quality Assessment*, San Rafael, CA: Morgan & Claypool.
120. Wang Z, Simoncelli EP (2005). “Reduced-reference image quality assessment using a wavelet-domain natural image statistic model”, *Proc. SPIE–IS&T Electronic Imaging Conf., Human Vision and Electronic Imaging X, 18 March 2005, San Jose, CA, USA*, pp. 149–159.
121. Webster MA (2015). "Individual differences in color vision", in Elliot AJ, Fairchild MD, Franklin A, eds., *Handbook of Color Psychology*, Cambridge: Cambridge Univ. Press, pp. 197–215.
122. Westheimer G (2007). “Irradiation, border location and the shifted-chessboard pattern”, *Perception*, 36 (4): 483–494.
123. Wu HR, Rao KR, eds. (2006). *Digital Video Image Quality and Perceptual Coding*, Boca Raton, FL: CRC Press.
124. Ziyaee T (2014). “Unsupervised Denoising via Wiener Entropy Masking in the STFT Domain”, *Proc. 2014 IEEE Military Communications Conf., 6–8 October 2014, Baltimore, MD, USA*, pp. 467–472.
125. Zuiderveld K (1994). “Contrast Limited Adaptive Histograph Equalization”, *Graphic Gems IV*, San Diego, CA: Academic Press Professional, pp. 474–485.

Supplementary Material

1 Enhancement algorithm

A papyri legibility enhancement algorithm is presented here. The vividness method is taken as an example, with gamut expansion, dynamic range stretching, negative polarity, and blue shift also included. This accounts for most of the proposed methods. For retinex methods, the reader is kindly asked to refer to the source code [8].

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)
$a^{*\prime} = -a^*$ and $b^{*\prime} = -b^*$

▷ 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