Article

Statistical Processing of Spectral Imagery to Recover Writings from Erased or Damaged Manuscripts

Roger L. Easton and David Kelbe | Rochester

Abstract

Imaging techniques for recovering historical writings that have been deliberately erased to make palimpsests or otherwise damaged have a long history. These writings often may be recovered from images collected under different illuminations over a range of wavelength bands. These spectral images are combined in subsequent image processing to enhance the visibility of the desired text. This paper considers the application of several processing algorithms based on the spectral statistics of the image data. These methods may be tailored to the specific condition of the manuscript and may enhance erased or damaged writings better than other techniques.

1. Introduction

Many historical manuscripts have been damaged by a variety of mechanisms: the ravages of natural processes over time, inappropriate storage, and deliberate action. Inks fade, manuscripts may have been scorched or charred by fire or damaged by animal excretions, parchments were often deliberately erased and reused to make palimpsests, etc. These faded, erased, or otherwise damaged manuscripts may be impossible to read without artificial aids to human vision. Moreover, simple devices (such as optical magnifiers to enlarge the text) and more complicated tools (e.g. ultraviolet lights to enhance the contrast of erased text relative to the background parchment) only improve the appearance to the human eye. The ultimate limit to the success of these visual aids is the human vision system itself, which is limited by the capability of the eye lens, the photochemical mechanism for sensing light, and the perception of images by the brain. One example of the constraint of the visual system is its narrow range of spectral sensitivity, which is limited to wavelengths between approximately 400 nm (blue) and 700 nm (red). Significant information is often present at wavelengths outside this range.

To obtain even better readings from the manuscripts, the limitations of the human visual system in terms of wavelength and sensitivity must be overcome. In other words, it is necessary to render subjectively imperceptible information so that it becomes visible to the eye. The need for better renderings has long been recognized and led to attempts to enhance the legibility of manuscripts by applying the very earliest imaging technologies. Chiara di Sarzana has presented an excellent history of photography applied to manuscripts.[1] In 1840, only one year after the accepted date of the first photograph, Jean Baptiste Biot reported that William Henry Fox Talbot had made images on sensitized paper of a Hebrew psalm, a Persian newspaper, and a 13th-century charter in Latin.[2] Apart from this early work, the use of photography to document historical writings did not become common until the technology of monochrome emulsion imaging matured in the 1880s, when flexible cellulose film substrates became available, thus eliminating the requirement for fragile glass plates. James Rendel Harris was an early imaging experimenter, having photographed manuscripts at St. Catherine's Monastery in Sinai in 1888.[3] Harris also assisted the Scottish twin sisters Agnes Smith Lewis and Margaret Dunlop Smith Gibson when they imaged the *Codex Sinaiticus Syriacus* on a subsequent visit to St. Catherine's in 1892.[4] It is interesting to note that the thousand or so film sheets exposed during that trip to the monastery were not processed until the travelers returned to England several months later, which prevented them from

---

[1] Di Sarzana 2006.

[2] Biot 1839-1840, *Chronique of Bibliothèque de l'École des Chartes* I, p. 408.

[3] Harris and Harris 1891.

[4] Gibson 1893.

assessing the value of their images until it was much too late to make any adjustments.

These first efforts at photographing manuscripts were not directed at improving the visibility of the text, but rather at creating copies of the manuscripts that could be studied intensively at a more convenient location. The first attempts to apply scientific technology to enhance the visibility of texts, rather than just to document them, apparently also took place in the mid-1890s, when Ernst Pringsheim and Otto Gradenwitz pioneered the use of multiple photographic images to enhance the erased 'undertext' of a palimpsested manuscript relative to the later text. They created a 'sandwich' of positive and negative photographic transparencies that had been exposed and processed specifically to make the later 'overtext' less visible than the original 'undertext' in the final image.[5] This was probably the first use of 'image processing' to enhance a manuscript that was difficult to read. The single documented result was quite successful.[6]

This early work set the stage for a long history of imaging and image processing for recovery of erased and damaged texts. In the early 1900s, the Benedictine monk Raphael (Gustav) Kögel photographed fluorescence emission from manuscripts after illumination with ultraviolet light generated by a variety of sources then available, including electric arc, mercury vapor, and metal wire lamps. As bandpass filters, he used various liquids in glass cuvettes placed in front of the camera lens. Kögel's goal had been to use fluorescence photography to enhance the visibility of erased text and therefore to replace chemical treatments of palimpsests with tincture of oak galls[7] or Gioberti's tincture[8] to enhance the original texts. Such treatments sometimes rendered the metallic ink more legible for a short time, but left severe damage in their wake. Kögel's goal was to improve the readability without inflicting any lasting damage. In the dedication to his detailed monograph on palimpsest photography in 1920,[9] Kögel acknowledged the pioneering contributions of Pringsheim and Gradenwitz as his inspirations. Though difficult and tedious to implement and requiring long exposures (up to 24 hours), the potential

capability of Kögel's method was evident, as he reported improvements in readability over previous results by up to 50%.[10] Kögel also helped initiate the establishment of the Institut für Palimpsestphotographie at the Archabbey of Beuron. In 1913, the Institute published a book of 152 black-and-white plates of the leaves of Codex Sangallensis 193 that were produced using the techniques developed by both Kögel and by Pringsheim and Gradenwitz.[11] This volume was intended to be the first of a series that would document the undertexts of palimpsests, but the effort was interrupted by World War I and apparently not resumed thereafter. Nonetheless, Kögel's work provided the first examples of what has become the basis of modern imaging methods used to recover text from palimpsests. An apparently similar technique was developed independently in Italy by Luigi Pampaloni.[12]

The methods used to image historical manuscripts advanced relatively little over most of the rest of the twentieth century. Color and infrared emulsions were first invented in the 1920s and were used later to assist the reading of manuscripts,[13] but the technology and capability of emulsion photography changed only in minor details from that time forward. For this reason, imaging methods to assist in the reading of manuscripts changed little until late in the 20th century, when there was a revolution in imaging and illumination hardware and processing software, which itself had been driven by the revolution in computing technologies. This advance in technology has resulted in new 'lights', 'eyes', and 'brains' that may now be applied to the task of recovering damaged or erased writings.

Several generations of the new imaging technologies have been applied to assist scholarly readings of manuscripts. The first arguably was the combination of broadband illumination and discrete bandpass filters to collect sets of spectral images, as had been used in the early experiments on the *Archimedes Palimpsest*.[14] The Forth-Photonics 'MUSIS' camera exemplified a later generation; it used a tunable optical filter to obtain images at a number of spectral bands from ultraviolet to near-infrared with a spatial resolution of

---

[5] Pringsheim and Gradenwitz 1894; Schnauss 1900.

[6] Pringsheim and Gradenwitz 1901.

[7] Edmonds 1998.

[8] Albrecht 2012.

[9] Kögel 1920.

[10] Kögel 1914.

[11] Dold 1913.

[12] Rostagno 1915.

[13] Haselden 1935.

[14] Netz, Noel, Tchernetska, and Wilson 2011.

1280 x 960 pixels (approximately 1.2 megapixels), which is quite coarse by today's standards.[15] Cameras belonging to the newest generation have much better spatial resolution (up to 50 megapixels) and use different illumination wavelengths generated by the new 'lights' to collect spectral image sets. These lights are made from light-emitting diodes (LEDs), which emit radiation generated by changes in electronic states rather than as a byproduct of heat, which makes them much cooler and safer to use for imaging of historical artifacts. Another useful feature of LEDs is their narrow emission bandwidth, which is typically between 10 nm and 50 nm. The illumination may be configured to interact with the manuscript in the usual reflection mode, but also in transmissive mode and in fluorescence, where the ink and substrate absorb incident radiation and emit longer wavelengths that are characteristic of the material. To ensure accurate measurements of the reflectance, transmittance, or fluorescence over the range of available wavelength bands, which is necessary for statistical analysis, standard reflectance and transmittance targets in the field of view are used to calibrate the collected spectral images.

The photoelectronic receptors in the sensors of the new eyes can 'see' light that is invisible to the human visual system, including visible light that is too faint to be perceived and wavelengths outside the human's range of sensitivity. For example, the silicon charge-coupled device (CCD) detector is usefully sensitive over the range of wavelengths from approximately 350 nm (in the near-ultraviolet region) to 1100 nm (near infrared), which is much broader than the range of human vision from approximately 400 nm (blue) to 700 nm (red). The 'invisible' bands of energy that are both shorter and longer than the range of human vision convey useful information about a manuscript. This extended range of vision is useful by itself, but when combined with LED illumination, the new 'eyes' can see a larger number of 'colors' than the three discrete color sensors of the human eye. If additional optical bandpass filters are incorporated in the optical path when using short-wavelength illumination (ultraviolet or blue), spectral images of the fluorescence may be collected, which have proven to be useful for some manuscripts, such as one undertext in the *Archimedes Palimpsest*, a commentary on Aristotle's treatise entitled *Categories*.[16]

The spectral images used in these projects typically include 12 reflective bands, as many as eight fluorescence bands, and often several transmittance bands (particularly if the subject is a palimpsest or a paper watermark that might benefit from enhanced visibility in this mode). The obvious negative aspect of transmissive illumination is that text on both sides of the leaf is often visible in the images, but this may be segmented by subsequent statistical processing.

It is important to recognize that the camera lens in the imaging system must transmit light and maintain focus over the wider range of wavelengths in spectral imaging from the near-ultraviolet to the near-infrared. This means that standard photographic lenses designed for visible light are not satisfactory. Because the glass components of standard lenses severely attenuate any ultraviolet reflection or emission from the object being examined, it is necessary to fabricate elements from crystalline quartz, which is transparent at these wavelengths. Also, if the lens is focused on an object at wavelengths in the middle of the visible range, wavelengths at the extrema of the transmitted ultraviolet and infrared bands will be unfocused. Fortunately, such 'UV-VIS-IR spectral lenses' have already been designed and are available on the market,[17] although they are significantly more costly than otherwise-similar standard lenses designed for visible light.

The camera system that forms the new 'eye' used for collecting these spectral images was constructed by Megavision, Inc., and has been documented elsewhere.[18] The rapid access to the imagery available from such a camera is a far more desirable state of affairs than that faced by the Smith sisters during their trips to Sinai. This imaging system has been used in several important projects, including the study of the *Archimedes Palimpsest* and the current project to image the palimpsests in the 1975 'New Finds' at St. Catherine's Monastery in Sinai.

The images collected at the different wavelength bands and using the different imaging modes are analyzed and processed in the new computer 'brains', with the goal of creating 'new' images that allow subtle differences in reflectance and color of the features to be distinguished. The choice of processing algorithm to enhance the desired feature depends on the specific situation. For example, it is occasionally possible for text that is invisible to the

---

[15] Rapantzikos and Balas 2005.

[16] Bloechl, Hamlin, and Easton 2010.

[17] E.g., http://www.jenoptik-inc.com/coastalopt-standard-lenses/uv-vis-nir-60mm-slr-lens-mainmenu-155.html.

[18] Easton et al. 2010.

Fig. 1: Visual appearance (on left) compared to pseudocolor rendering (on right) of a section of an *Archimedes Palimpsest* leaf (f. 94r–91v of the *Euchologion* overtext) including gutter. The visibility of the undertext relative to the parchment is enhanced because of the rendering in a contrasting color.

human eye to be read directly from a single image at a single wavelength; this is most often true for manuscripts that were scorched or carbonized, in which case the writing may be visible in an image collected under near-infrared illumination. More often, it is necessary to combine images collected in different bands to enhance subtle variations in the spectral reflectance of the text of interest. The algorithms for combining image bands may be rather loosely classified as 'deterministic', where the same combinations of bands are used for more than one leaf, or as 'statistical', where the band combinations are calculated from the statistics of the gray values in the ensemble of spectral bands for each class of object. In either case, the condition of the leaf (hair or flesh side) and variations in the degradation across the leaf ensure that the optimum choice of bands and processing method generally mean that the processing method also changes.

2. Deterministic processing methods

Although the focus of this paper is on statistical methods, a brief introduction to deterministic renderings may be useful to the reader. Perhaps the simplest technique of this kind is the evaluation of the difference in gray values of two spectral bands, so that regions on the leaf with similar reflectances subtract to small numerical values, while features with different gray values take on an extremum value – either positive or negative – that may be rendered as a more visible feature. Another useful example of deterministic processing is pseudocolor rendering of the image, where monochrome image bands generated under different conditions of illumination are inserted into the red, green, and blue channels of a visual color image. If judiciously chosen, the visibility of the text of interest may be enhanced in a particular combination of bands. This was the primary

means for rendering image data that was used in the study of the *Archimedes Palimpsest*, where images through a blue filter under ultraviolet illumination and through a red filter under tungsten lights were combined to render the overtext in neutral gray or black and the undertext in a reddish tint.[19] The resulting color 'cue' helps the reader distinguish between the two texts.

Because they do not require evaluation of spectral statistics, deterministic methods may usually be implemented quickly, which is a very distinct advantage in large projects with many leaves to be transcribed. For this reason, they are often used productively as a 'first pass' in image processing. If the resulting images are sufficient for scholarly transcription, no further processing is necessary. For those images that are not readable from images processed by deterministic methods, a second pass involving more computationally intensive statistical methods is applied.

3. Statistical processing methods

Statistical image-processing methods analyze the ensemble of gray values at each pixel over the range of spectral bands with the same goal as deterministic processing: to find linear combinations (weighted sums) of spectral bands that enhance the desired text. Numerous techniques of this kind exist; many were originally developed for military purposes (such as camouflage detection) or for environmental applications (such as characterizing ground conditions or assessing the health of crops). The same methods are directly applicable to the goal of enhancing subtle differences in reflectance spectra of the different features on a manuscript.

Consider a set of images collected under N distinct conditions of illumination that may include reflective, transmissive, and fluorescent modes. The integer gray values of a specific pixel measured under the N conditions form a vector with N components. The ensemble of N-dimensional vectors may be plotted, at least in theory, as an N-dimensional histogram, which will exhibit 'clusters' of pixels belonging to the same object class. For example, the gray values of a pixel in images of the same manuscript under green and red light form a two-dimensional vector, and the ensemble of such vectors from all the pixels in the image forms a statistical probability distribution that may be analyzed to look for correlations among object features. If a third illumination is added, then the histogram is formed from the
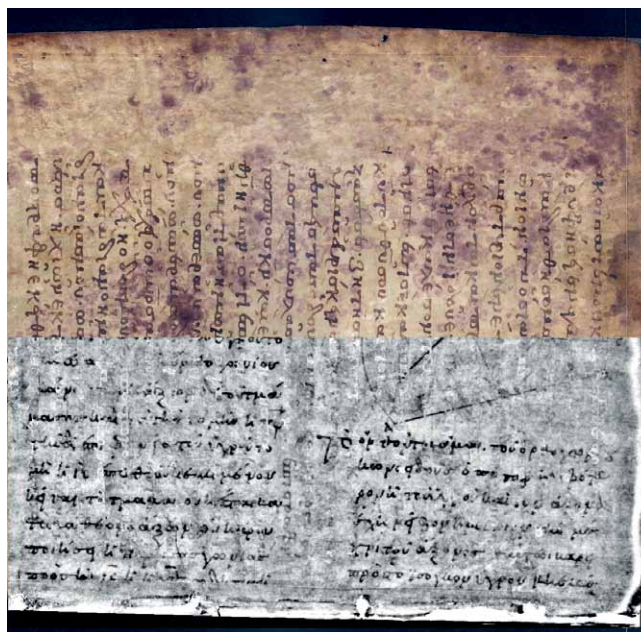

Fig. 2: Visual appearance (top half) compared to result of spectral unmixing (bottom half) of contiguous halves of f. 28v of the *Archimedes Palimpsest*. The unmixing process effectively removes the overtext. In particular, the diagram is much easier to see in the output image.

three-dimensional vectors. Of course, the dimensionality of the vector is equal to the number of bands. A rule of thumb is that the number of bands should equal or exceed the number of features to be distinguished. In these projects, it is common to include a dozen or more spectral bands, although it is often useful to use subsets of the images based on observations of the text's visibility.

4. Spectral unmixing

One method that has proven useful in text analysis that requires significant user interaction is 'spectral unmixing', where pixel regions belonging to specific object classes are identified by the user first, e.g. parchment, mold, overtext, erased text, etc. The algorithm then calculates the class membership of each pixel in the image based on the similarity of its spectrum to each of the specified classes. Although it is intensive both in terms of human interaction and subsequent computation time, this method was applied with success during the early experiments on the *Archimedes Palimpsest* and also to spectral image data collected with a MUSIS camera.[20] Spectral unmixing of manuscript imagery deserves additional study, particularly in situations where the feature spectra are known or can be determined a priori.

---

[19] Netz and Noel 2007.

[20] Knox et al. 2001.

In our applications, we have most frequently employed two similarly named statistical methods for enhancing the visibility of erased or damaged text: principal component analysis (PCA) and independent component analysis (ICA). PCA 'rearranges' the data in a set of N spectral images to create a different and equivalent set of N images that satisfy two properties: (1) the derived images are uncorrelated and (2) they are arranged in descending order of variance. Interested readers are advised to consult the very accessible introduction to PCA by Schlens.[21] ICA also rearranges the set of N spectral images to make a different set of N images, but the output bands are distinguished by statistical 'independence', which will be discussed shortly. Hyvärinen and Oja have written a useful introduction to ICA.[22]

5. Principal Component Analysis

PCA is implemented by evaluating the difference in the gray value of each pixel for each combination of two spectral bands and then evaluating the expectation value for these differences over all pixels in the image. The result is the real-valued and symmetric covariance matrix. The eigenvectors of this matrix are the orthogonal axes of the principal components. The eigenvectors are ordered in sequence based on the magnitudes of the corresponding eigenvalues, with the eigenvector associated with the largest eigenvalues first. The implementation of principal component analysis may be viewed as 'projecting' the image pixels of an N-band image onto each of these N orthogonal axes, followed by rendering each pixel as a gray value based upon its location on the particular axis. In other words, the act of data 'projection' onto each axis evaluates a weighted sum of the original N images. The end result of the projections onto the N orthogonal axes is a set of N 'new' monochrome images that are equivalent to the original N image bands.

The first PC band results from the projection of the data in the N-dimensional histogram onto the axis that spans the widest possible range of variation of the image data, so that the first PC image exhibits the widest possible range of variance of the statistics or the widest range in 'contrast' of image features. In the application to manuscript imaging, the range of gray values of the first PC band is determined by the pixels in the areas of the image that are brightest and darkest overall, such as the light parchment and darkest overtext characters.

The second PC image is the projection of the N bands of data onto the axis of the N-dimensional histogram with the largest possible range of variation in a direction that is orthogonal to the axis used for the first PC image, so that this second band exhibits a smaller range of variation than the first. The second PC band is also a weighted sum of the original N spectral images, but the fact that the two axes are orthogonal means that the first and second PC images are 'uncorrelated.'

The process of determining the orthogonal axis with the next largest variation and projecting the data from the histogram onto that axis is repeated to generate a total of N mutually orthogonal PC bands. The monotonic decrease in the sequence of eigenvalues corresponding to each axis means that the low-order bands (evaluated first) are dominated by large-scale variations in the original N bands of data, while the high-order bands (evaluated last) are dominated by small-scale variations in the original data, which may be random fluctuations ('noise') or image features with very little contrast (such as erased undertext).

It is important to recognize that the projected data values are floating-point numbers that must be mapped to integer gray or color values for display, usually in an 8-bit format with $2^8 = 256$ possible values per color that can be displayed on a computer screen. The process requires selection of the 'lightest' and 'darkest' values to be 'quantized' to the extrema of the available integer values. This process of quantizing the floating-point value to an 8-bit integer means that a range of different numerical values will be mapped to a single integer gray value. This means in turn that different features may be rendered at one gray value or over a small range of gray values, so the features may not be distinguishable for a particular choice of the limiting light and dark pixels. For this reason, it is essential to have the option to change the range of values for the rendering of the image. Similar problems appear in other applications, such as medical X-ray computed tomography, where the radiologist often changes the limits of the grayscale rendering to visualize subtle features of the pathology. The importance of the choice of rendering is easy to overlook.

In the best possible result of PCA processing, each feature class in the scene (e.g., parchment, overtext, undertext, etc.) would appear exclusively in a single specific band in the new set. In this case, each class of feature would dominate the range of gray scale in the specific PC band, while pixels belonging to the other classes of feature would exhibit the same gray value and thus 'disappear' into the background in that band. In fact, this happy occurrence is rare; traces

---

[21] Shlens 2009.

[22] Hyvärinen and Oja 2000.

Red
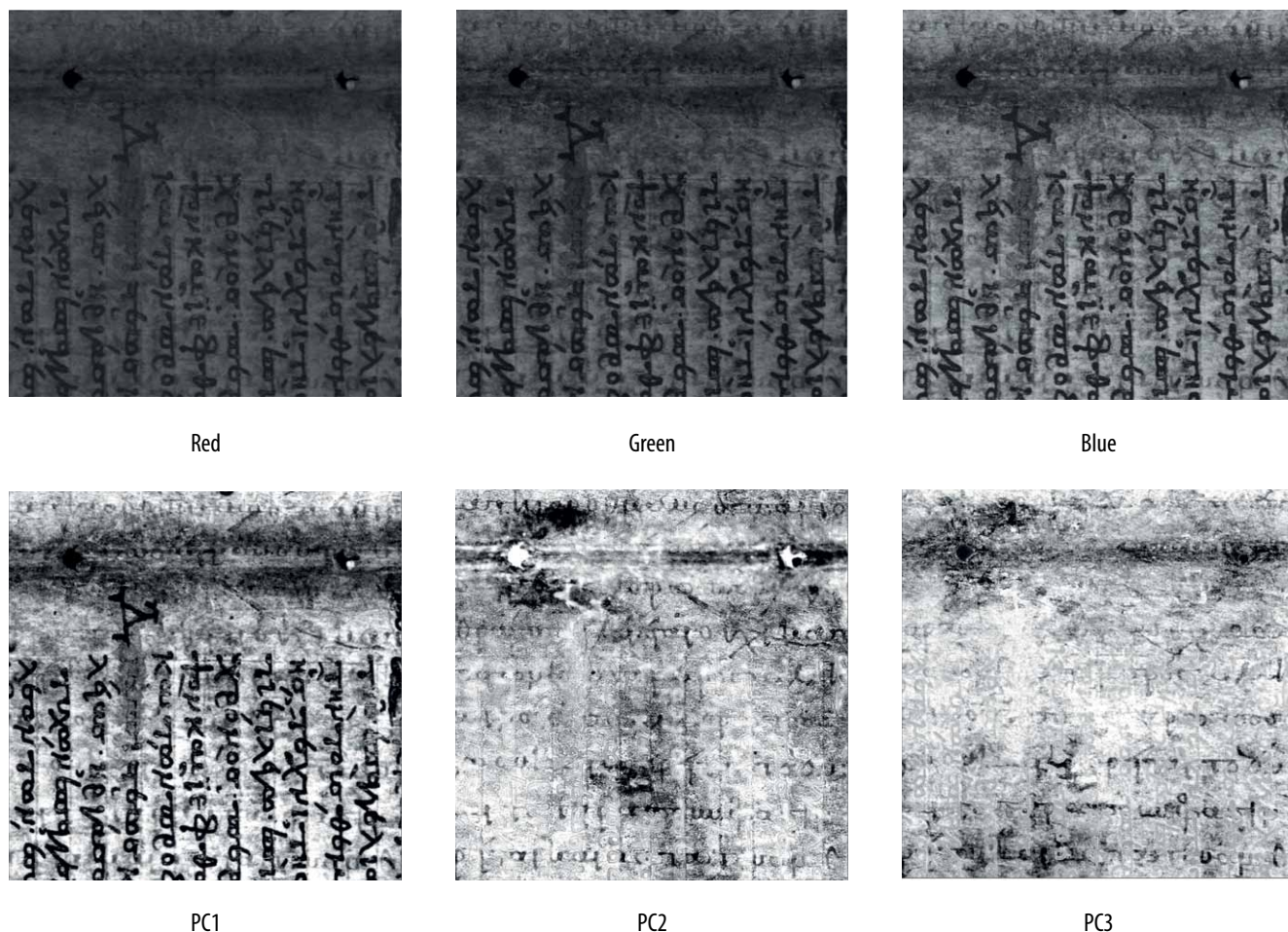

Green

Blue

PC1

PC2

PC3

Fig. 3: The top row shows the red, green, and blue color channels of an image of fluorescence generated by ultraviolet light. The images in the bottom row are the corresponding principal components. Note that the first principal component closely resembles the blue channel, which is considerably lighter than the red or green channels. The dark overtext that is so visible in the first principal component is rendered with approximately the same gray value as the parchment in the second and third principal components, so the overtext 'disappears' into the parchment and the erased undertext is far more visible in these bands.

of overtext usually appear in those bands where the erased undertext is most visible, for example, and the variation in statistics across the scene (e.g. due to variations in erasures) ensures that the desired erased text often appears in more than one PC band. This last observation results in the strategy of inserting two or three bands into the channels of a pseudocolor image, and the rendering is altered interactively by the observer, as will be described.

Software to implement principal component analysis is widely available, including versions in open-source imaging software such as ImageJ.[23] The processing implemented here was performed with the ENVI software toolkit that was written for environmental remote sensing by Exelis Visual Information Systems.

Because the 'rearrangement' of the original N input image bands is based only on the statistics evaluated from the N-dimensional histogram, there are no selectable parameters other than the choice of the specific original bands and of the region in the scene where the statistics are evaluated (both of which are important). This independence from input parameters means that PCA is widely applicable for many types of data sets (not just images), but it also means that it may not succeed in separating the features unless the statistics of the different features are truly orthogonal or readily distinguishable.

An example of PCA processing of a leaf of the *Archimedes Palimpsest* is shown in fig. 3. The undertext is a commentary on Aristotle rather than one of the treatises by Archimedes that occupy most of the leaves. This text was particularly difficult to read in the deterministic renderings, but was made quite

---

[23] ImageJ is available from the National Institutes of Health at http://imagej.nih.gov/ij/.
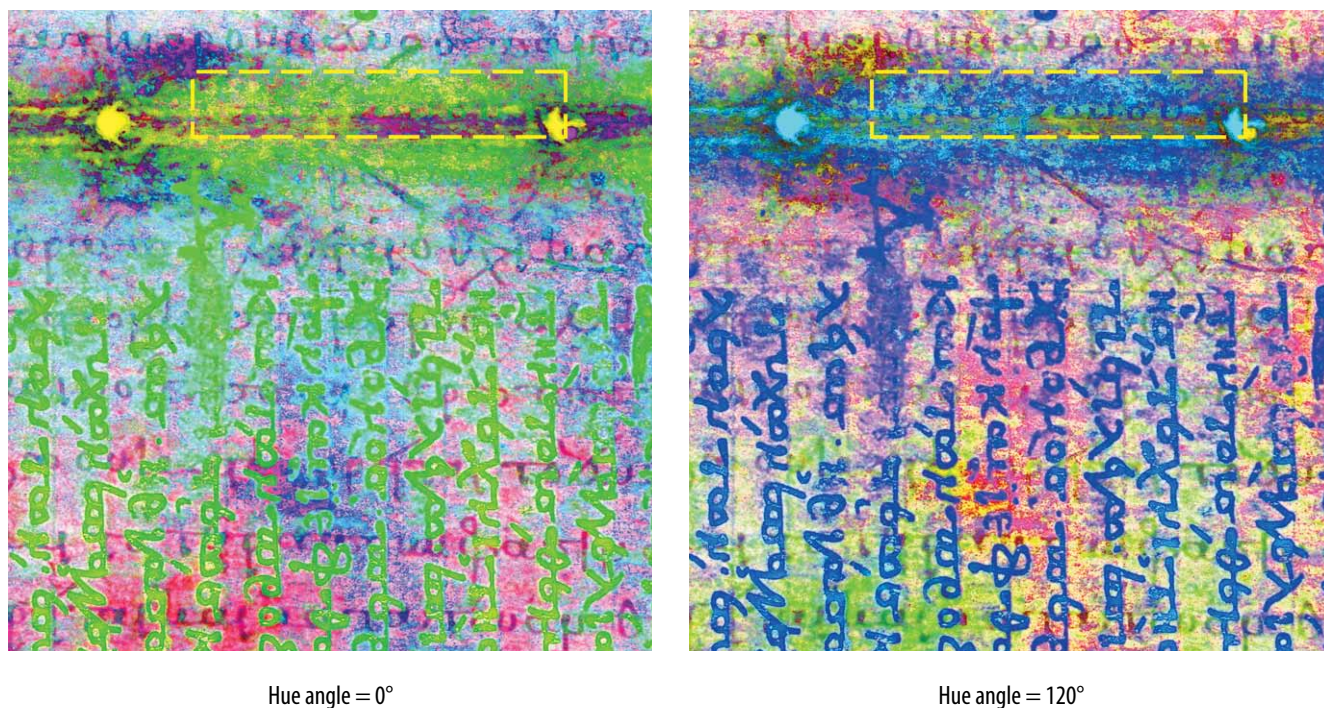
Hue angle = 0°   Hue angle = 120°

Fig. 4: Pseudocolor rendering of the principal components in Fig. 3 at two different hue angles. Undertext in the gutter region outlined by yellow dashed line is barely visible, if at all, in the image with hue angle equal to 0°. The visibility of the text in the gutter is much improved when rendered with a hue angle equal to 120°.

legible by PCA processing. The manuscript was illuminated by a narrow band of ultraviolet light centered approximately about $\lambda = 365$ nm. The ultraviolet illumination generated visible fluorescence in the manuscript, which was imaged by a sensor whose pixels were covered by visible color filters in the common Bayer pattern, so that each pixel in the array measures the amount of light in one of the three primary colors (red, green, and blue); half of the pixels in the sensor measure green light, and one quarter each measure red light and blue light. The camera had the capability to translate the sensor piezoelectrically relative to the image by increments of the pixel separation, so the amount of red, green, and blue light at each pixel could be measured from exposures at the different sensor positions. Three principal components are generated, matching the number of input bands. The contrast between the overtext and parchment dominates the first PC band because it spans the widest range of variation in gray level in the three-band image. Because the second and third PC bands are orthogonal to the first, the three-dimensional vectors of gray value of the overtext and parchment are projected onto the same location on the projection axis. In other words, the parchment and overtext are rendered in the same small range of gray values in PC bands #2 and #3, so that the overtext 'disappears' into the parchment. When rendered as black-and-white images, the much smaller range

of gray values determined by the undertext and background parchment is rendered to the full range of eight available bits of gray value, so that the contrast between the undertext and parchment dominates these higher-order principal component bands.

Although the goal of PCA is to construct orthogonal renderings that will segment the feature classes, the typical overlap of class statistics means that the desired feature appears in more than one output principal component band. In a palimpsest, the variation of the erasure of the original text across the leaf generally means that the histogram of the image varies with the position on the leaf, as was the case for the Aristotle commentary. This means, in turn, that the undertext appears most clearly in different locations of the three principal-component images and that pseudocolor rendering of PCA bands is often useful. Three PC bands are selected and inserted into the red, green, and blue color channels of a pseudocolor image. This means that small variations in grey value may appear as large changes in color, which may improve the readability of the erased text. The user may also actively change the pseudocolor rendering by varying the hue angle (formally, this changes the mapping of the color tones without varying the saturation and luminance values at each pixel). In practice, perceived changes in luminance and saturation accompany hue rotation; it is this

triad of rendering adjustments that is exploited to enhance the visibility of the undertext at different locations on the leaf. A comparison of pseudocolor renderings on one leaf of the Aristotle commentary in the *Archimedes Palimpsest* for two different hue angles is shown in fig. 4. The visibility of the text in the gutter enclosed by the dashed square is noticeably improved in the second example where the hue angle has been rotated by 120º. The variation in color rendering is most valuable when performed interactively by the transcriber, who can often see features more clearly in a dynamic rendering than in a static image.

Hue-angle rotation proved to be essential in the different problem of recovering information from an illuminated armorial on the manuscript of the French epic poem *Les Echéz d'Amour* at the Saxon State Library in Dresden. The manuscript was a victim of water damage after the Allied bombing of the city in 1945, and the armorial had been so smudged and tarnished that almost no structure in the seal remained visible. Spectral images of the leaf were collected and analyzed by PCA in the standard manner. Rotation of the hue angle of a subset of three PCA bands rendered in pseudocolor resulted in the image in fig. 5, where the seal clearly shows two unicorns, the second one being on a shield. After a short search of an online archive, this image made it possible to identify that the book had been owned by the Waldenfels family in Bavaria.

6. Independent Component Analysis

The second similarly named statistical processing algorithm of 'independent component analysis' (ICA) also rearranges the set of N spectral images to make a different set of N images, but in this case the output bands are distinguished by statistical 'independence,' rather than the PCA criteria of being 'uncorrelated' and 'orthogonal.' The method is an example of 'blind source separation' applied to a mixture of input signals with the goal of determining the original independent components from different measurements of the combinations. A common example is the so-called 'cocktail party problem,' where independent conversations occur simultaneously at different locations in a crowded room. A single microphone positioned in the center of the room records the sum of independent conversations with weightings determined by the distances of each one from the microphone. The experience of the reader probably validates the difficulty of segmenting individual conversations from a single recording of this kind. Nonetheless, conversations
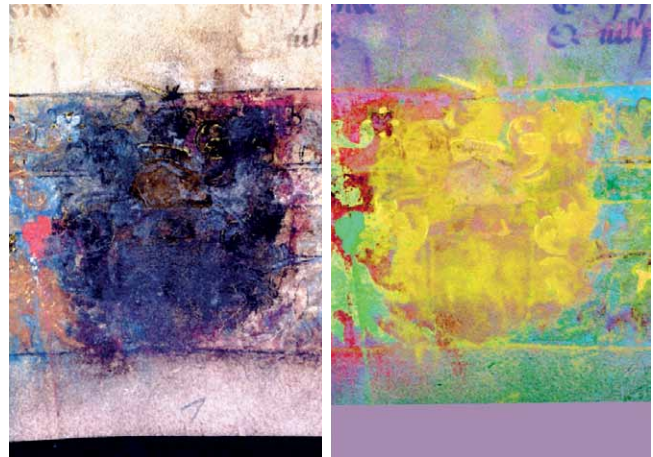

Fig. 5: Pseudocolor rendering of principal components of Armorial Achievement from f.1r of *Les Echéz d'Amour*, (OC66 from Saxon State Library, Dresden). The visual appearance on the left shows just a hint of a unicorn horn at the top. That unicorn is much more easily seen in the PCA pseudocolor on the right after rotation of the hue angle, which also shows a second unicorn on a shield and hints of a rampant lion or other animal on the left side of the shield. This image enabled identification of the family that owned the book.

could be separated from a recording by a single microphone if the frequency ranges of the voices were disjoint; a high-pitched voice could be separated from a simultaneous low-pitched voice in the same recording by applying the appropriate filter that passes one frequency range and blocks the other. The analogous situation for a manuscript would be the separation of two texts written in different colors of ink from images collected through two different bandpass filters. In the more realistic model of simultaneous conversations with overlapping frequency ranges, the required process for separating the components is less obvious; the corresponding imaging analogy is that the reflectance spectra of the different component writings will overlap.

The simultaneous voices may be recorded by multiple microphones placed around the room so that each one measures a different weighted sum of the conversations. The signals from the ensemble of microphones are analyzed and compared to segment the voices. At this point, it is useful to make an observation about the statistics of a meaningful conversation. The histogram of the spectrum of disordered random 'noise' tends to be uniform, with approximately equal populations at each frequency. The contrapositive statement is that histograms of 'ordered' components generally exhibit 'peaks' or 'clusters' at different frequencies, a feature that characterizes the statistics of a conversation as 'structured' or 'ordered.' From this observation, the process of independent

(a)   (b)   (c)   (d)
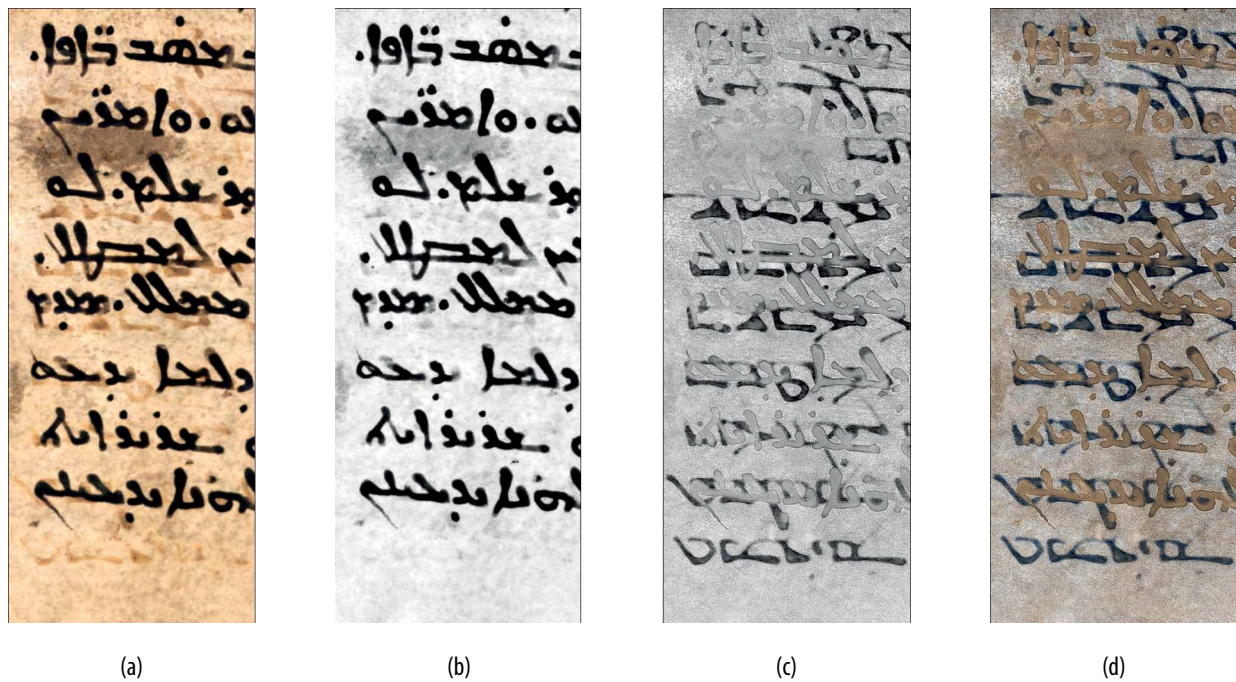
Fig. 6: Example of ICA processing applied to Syriac 2A from St. Catherine's Monastery; the visual appearance is shown in (a) for reference. ICA rearranges the spectral image sets into uncorrelated output bands which effectively separate the overtext (b) and undertext (c). These output IC bands can then be combined into a pseudocolor rendering (d) that shows both texts.

component analysis is the search for a set of component conversations such that the histogram of each component is as 'clustered' as possible (with minimum 'disorder') and that the weighted sums of the hypothetical conversations match those of the individual recordings. The fact that the component signals are assumed to be independent means that the joint probability of all signals is the product of the probabilities of the individual components. This requirement for statistical independence is a stronger condition than the assumption in PCA that the signals are not correlated. The stronger condition invoked in ICA means that signals that are not well segmented by principal component analysis may be separable by independent component analysis.

When applying ICA to spectral images, the different feature classes (e.g. parchment, overtext, erased text, mold, etc.) are analogous to the individual simultaneous conversations, and the different spectral bands correspond to the individual recordings from the different microphones. Just as each sample of the recording from a microphone is a weighted sum of contributions from the different conversations, each pixel in a spectral image is a weighted sum of contributions from the individual object classes. ICA uses the multiband statistics of each pixel in the set of spectral images in an attempt to estimate the contributions of each object class at

each pixel. The process is often combined with pseudocolor rendering and has been quite successful in recovering text from some palimpsests in the New Finds at St. Catherine's Monastery. The ICA tool available at ENVI was used in this analysis, although the algorithm is also available in other packages. An example of ICA applied to an image from St. Catherine's Monastery is shown in fig. 6.

Further, and often dramatic, improvements may be obtained from weighted combinations of images, including processed results from ICA and PCA and possibly original image bands. The choice of bands and the combination depends on the condition of the desired text. For example, erased text appears different under transmissive and fluorescence illumination; the process of scraping the text may thin the parchment so that the erased text is brighter than the surrounding parchment in transmission, while the erased text is darker than the surrounding parchment in fluorescence images. Statistical processing thus often produces different images that render different regions of the leaf better than other regions. It is often advantageous to combine the feature content from the two modes into a single image for viewing, while minimizing the visibility of the overtext. This can be done by combining multiple grayscale images into an intermediate pseudocolor image. Text features are thus
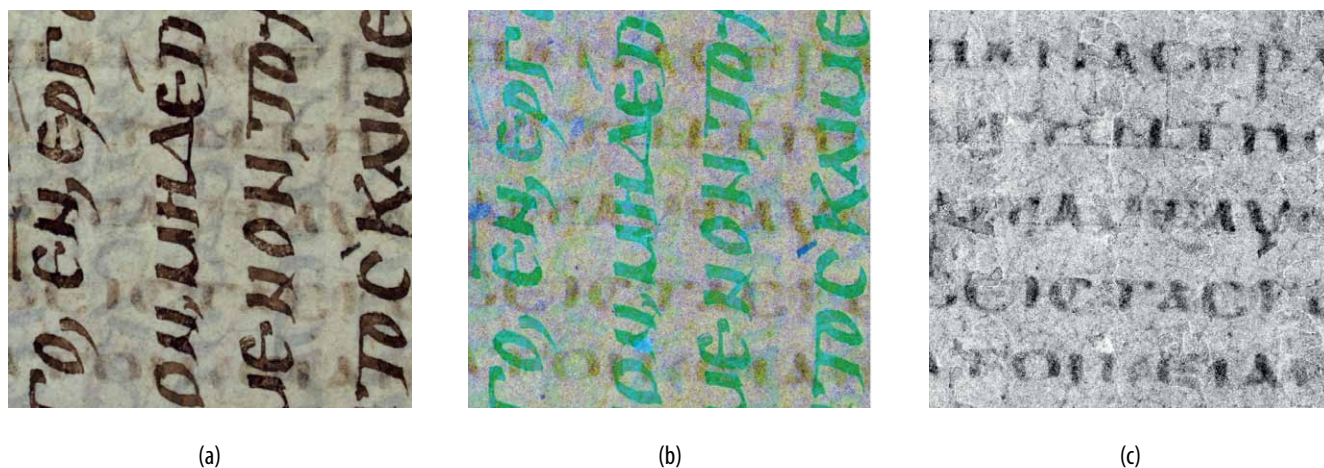
| (a) | (b) | (c) |

Fig. 7: ICA processing applied to Greek New Finds MG14 from St. Catherine's Monastery after subsequent processing to attenuate the visibility of the overtext: (a) visual appearance, (b) intermediate pseudocolor image showing different text features separated by color, (c) output grayscale image after modulating the brightness of individual colors in order to reduce the visibility of the overtext.

differentiated by color in the pseudocolor image, allowing precise control over the lightness and contrast of individual text features. In other words, color effectively separates different features and serves as a proxy for modulating the 'weight' of individual text components. The various features can be adjusted to optimize the visibility of the undertext in an output grayscale image. One such result is compared to the visual appearance and a pseudocolor PCA image in fig. 7.

It is important to note that all methods of statistical processing require an observer to make important selections from the image data, including the bands to be processed, the region(s) where the statistics are evaluated, the bands to be rendered in the final image, and the type and settings of any post-processing of the rendered image (such as the angle of hue rotation of a pseudocolor rendering). This observation indicates that the skill of the experienced observer remains an important factor in the success of the final result.

7. Conclusion

In summary, statistical processing techniques applied to spectral images of historical writings have been successfully applied to the task of distinguishing features that are not otherwise visible in images of manuscripts. These methods may be applied to a wide range of imagery. It seems quite certain that new developments in hardware and processing software in the near future will further enhance the value of statistical processing methods for spectral imaging. New sensors capable of imaging over a wider range of wavelengths may be anticipated, for example, and these techniques may be applied directly to that imagery.

REFERENCES

Albrecht, Felix (2012), 'Between boon and band: the use of chemical reagents in palimpsest research in the nineteenth century', *Care and Conservation of Manuscripts,* 13: 147–165.

Biot, J. B. (1840), reported in *Chronique* of Bibliothèque de l'École des Chartes I p. 408 (available from http://gallica.bnf.fr/ark:/12148/bpt6k123741.image).

Bloechl, K., Hamlin, H., and Easton, R. L., Jr. (2010), 'Text recovery from the ultraviolet-fluorescence spectrum for a treatise in the Archimedes palimpsest', *Proceedings of SPIE,* 7531-09.

Di Sarzana, Chiara Faraggiana (2006), 'La fotografia applicata a manoscritti greci di difficile lettura: origini ed evoluzione di uno strumento di ricerca e i principi metodologici che ne regolano l'uso', in Angél Escobar (ed.), *El palimpsesto grecolatino como fenómeno librario y textual*, 65-80 (available from http://ifc.dpz.es/recursos/publicaciones/26/54/_ebook.pdf).

Dold, Alban (1913), *Codex Sangallensis 193, Spicilegium Palimpsestorum Volumen I* (Leipzig: Harrassowitz).

Easton, R. L. Jr., Knox, K. T., Christens-Barry, W. A., Boydston, K., Toth, M. B., Emery, D., Noel, W. (2010), 'Standardized system for multispectral imaging of palimpsests', *Proceedings of SPIE*, 7531-12.

Edmonds, Tony E. (1998), 'An indicator of its time: two millennia of the iron-gall-nut test', *The Analyst*, 123: 2909–2914.

Gibson, Margaret D. (1893), *How the Codex Was Found* (Cambridge: MacMillan and Bowes) (available from https://archive.org/details/howcodexwasfound00lewi).

Harris, Helen B., and Harris, James R. (1891), *The Newly Recovered Apology of Aristides*, London: Hodder and Stoughton (available from https://archive.org/details/newlyrecoveredap00harr).

Haselden, Reginald Berti (1935), *Scientific Aids for the Study of Manuscripts* (Oxford: Oxford University Press for the Bibliographic Society).

Hyvärinen, Aapo and Oja, Erkki (2000), 'Independent Component Analysis: Algorithms and Applications', *Neural Networks*, 13: 411–430 (available from http://www.cs.helsinki.fi/u/ahyvarin/papers/NN00new.pdf).

Knox, K., Dickinson, C., Easton, R., Wei, L., and Johnston, R. (2001), 'Multispectral imaging of the Archimedes Palimpsest', *Proc. 54th IS&T Ann. Conf. PICS*, Montreal, 206–210.

Kögel, Raphael Gustav (1920), 'Die Palimpsestphotographie', *Enzyklopädie der Photographie*, 95 (available from http://goobipr2.uni-weimar.de/viewer/resolver?urn=urn:-nbn:de:gbv:wim2-g-2965569).

Kögel, Raphael Gustav (1914), 'Die Palimpsestphotographie', *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, 27: 974–978 (available from https://archive.org/details/sitzungsberichte1914deut).

Netz, Reviel, and Noel, William (2007), *The Archimedes Codex*, (London: Weidenfeld and Nicolson).

——, Noel, W., Tchernetska, N., and Wilson, N. (eds.) (2011), *The Archimedes Palimpsest* (Cambridge University Press).

Pringsheim, E., and Gradenwitz, O. (1894), 'Photographische Reconstruction von Palimpsesten', *Verhandlungen der Physikalischen Gesellschaft zu Berlin im Jahre 1893*, 58–60 (available from http://catalog.hathitrust.org/Record/000544812).

——, and Gradenwitz, O. (1901), 'Photographische Reconstruction von Palimpsesten', *Jahrbuch für Photographie und Reproduktionstechnik für das Jahr 1901*, 52–55 (available from https://archive.org/details/jahrbuchfrphoto04edergoog).

Rapantzikos, K., and Balas, C. (2005), 'Hyperspectral imaging: potential in non-destructive analysis of palimpsests', *IEEE International Conference on Image Processing (ICIP)*, 0-7803-9134-9/05/.

Rostagno, Enrico (1915), 'Della Riproduzione de' Palinsesti e d'un nuovo sistema italiano ad essa applicato', *Rivista Delle Biblioteche,* 26, 58–67 (available from http://babel.hathitrust.org/cgi/pt?id=njp.32101073752162;view=1up;seq=68).

Schnauss, H. (1900), 'Photography as an Aid to Palaeography', *The American Amateur Photographer*, 12: 504–506 (available from Google Books).

Shlens, Jonathon (2009), *Tutorial on Principal Component Analysis* (http://www.snl.salk.edu/~shlens/pca.pdf).