Manuscript Studies

A Journal of the Schoenberg Institute for Manuscript Studies

The Schoenberg Institute
for Manuscript Studies
UNIVERSITY *of* PENNSYLVANIA LIBRARIES

MANUSCRIPT STUDIES

A Journal of the Schoenberg Institute for Manuscript Studies

Special Issue:
The Syriac Galen Palimpsest
Editors: William Noel and Ralph M. Rosen

Articles

Spectral Imaging Methods Applied to the Syriac Galen Palimpsest

ROGER L. EASTON, JR.
Chester F. Carlson Center for Imaging Science,
Rochester Institute of Technology

KEITH T. KNOX
Early Manuscripts Electronic Library, Kihei, HI

WILLIAM A. CHRISTENS-BARRY
Equipoise Imaging, Ellicott City, MD

KEN BOYDSTON
Megavision, Santa Barbara, CA

THE NOTION THAT IMAGES collected in bands of light that are invisible to the eye can convey important information about an object or scene is hardly new. In his book *The Pencil of Nature*, which was published in six segments between 1844 and 1846, William Henry Fox Talbot wrote about how photography could be used to image objects in "invisible" infrared wavelengths. He even suggested that the concept of imaging with invisible light could be the basis for a detective story with forensic overtones:

> Now, I would propose to separate invisible [infrared] rays from the
> rest [of the visible light], by suffering them to pass into an adjoining

apartment through an aperture in a wall or screen of partition. This apartment would thus become filled (we must not call it "illuminated") with invisible rays, which might be scattered in all directions by a convex lens placed behind the aperture. If there were a number of persons in the room, no one would see the other: and yet nevertheless if a camera were so placed as to point in the direction in which any one were standing, it would take his portrait, and reveal his actions.

For, to use a metaphor we have already employed, the eye of the camera would see plainly where the human eye would find nothing but darkness.

Alas! that this speculation is somewhat too refined to be introduced with effect into a modern novel or romance; for what a dénouement we should have, if we could suppose the secrets of the darkened chamber to be revealed by the testimony of the imprinted paper.[1]

In the intervening period of nearly 175 years, Talbot's proposal has been applied and adapted in a variety of applications—this is particularly true since digital imaging technologies became widely available during the past two decades or so, but was also performed successfully using analog methods as early as the 1890s by the well-known Prussian experimental physicist Ernst Pringsheim and his research partner, lawyer Otto Gradenwitz.[2] Their work was extended in the 1910s by Fr. Raphael (Gustav) Kögel to include fluorescence images generated under ultraviolet illumination.[3]

The inclusion of wavelength bands other than the infrared light suggested by Talbot and the use of interactions of light with the object other than mere "reflection" have increased the value to the results obtained from

---

1   William Henry Fox Talbot, *The Pencil of Nature* (London: Longman, Brown, Green and Longmans, 1844), 30.

2   E. Pringsheim and O. G. Gradenwitz, "Photographische Reconstruction von Palimpsesten," *Jahrbuch für Photographie und Reproduktionstechnik für das Jahr 1901* (1901): 52–56.

3   P. R. Kögel, "Die Palimpsestphotographie," *Sitzungsberichte der königlich Preussischen Akademie der Wissenschaften* (1914): 974–78.

the image processing, which "combines" two or more image bands to enhance the visibility of the feature(s) of interest, most often erased ink. Because the different image bands are combined, it is important for the individual images to be accurately aligned (or "registered"), which ensures that the corresponding pixels in different images "see" the same features on the original pages. The image collection conditions in this project made this more difficult compared with the conditions today.

It would be fair to say that the idea of imaging over a range of wavelengths invisible to the human eye now supplies answers to formerly unsolvable cases in the "detective stories" of history. Because it can reveal new and sometimes unexpected text from manuscripts in existing collections, multispectral technology has the promise of providing a growing canon of available writings to scholars and scientists.

Based on the positive results obtained in the imaging effort for the Archimedes Palimpsest, which spanned the years 2000–2008, the three authors who formed the imaging team for that project (Easton, Knox, and Christens-Berry) joined with the fourth author (Boydston) to image the Syriac Galen Palimpsest during the years 2008–10. The lessons learned during the Archimedes Palimpsest project were incorporated into the new task, though other changes were also implemented. That said, the relentless advance of technology since completion of the Syriac Galen project means that much of the imaging technology has now been superseded. For this reason, it may be useful to describe how the new advances are now implemented in imaging projects.

Imaging System

A Sinar-54 RGB camera had been used in the last Archimedes Palimpsest image collection in August 2007. It had a 22-megapixel sensor with a Bayer RGB screen to collect visible color imagery. Though not the original choice of the imaging team, primarily because the color filters over the individual pixel elements prevented the collection of "true" spectral imagery, this camera was a late substitute for the custom camera with a monochrome sensor that was not available in time for the scheduled imaging. Though the Bayer

screen did create some problems for the subsequent image processing, this "bug" turned out to be a useful "feature" for one of the undertexts in the Archimedes Palimpsest, which was recovered only by processing the RGB color fluorescence images under ultraviolet illumination.

In the Archimedes Palimpsest project, this camera was used to collect images at eleven different illumination bands using light-emitting diodes (LEDs) at wavelengths ranging from the near-ultraviolet end of the spectrum at λ = 365 nanometers (nm) through the visible wavelengths to the near infrared region out to λ = 870 nm. Additional images were collected using low-angle "raking" illumination in blue and infrared bands; it was hoped that these bands would provide useful information about the topography of the pages. The panels of LEDs were constructed by one team member (Christens-Barry).[4]

Unlike most common digital cameras that measure only one of the three additive primary colors (red, green, or blue) at each pixel and interpolate the other two, the sensor in the Sinar camera used in the 2007 Archimedes imaging could be shifted by half- and full-pixel increments relative to the image by a piezoelectric translator. This added the capability to record all three primary colors at each pixel in a stationary scene by collecting four images after full-pixel translations. The half-pixel translations also were used to collect additional sets of images of the same leaf that were combined to make 88-megapixel RGB images under each illumination. The value of these additional pixels proved to be very small, as spatial averaging of the fine features by the finite-sized pixels severely attenuated the modulation of sinusoidal components with the newly available high frequencies. Put another way, the 88-megapixel images contained very little additional information compared with the original 22-megapixel images without the half-pixel translations, and the time required to collect the additional twelve images at each wavelength increased the total imaging time by a factor of four. That said, the capability for full-pixel translations that enabled collection of images in the red, green, and blue bands proved to be essential for

4   W. A. Christens-Barry, K. Boydston, F. G. France, R. L. Easton Jr., K. T. Knox, and M. B. Toth, "Camera System for Multispectral Imaging of Documents," *Proc. SPIE* 7249-08 (2009).

recovering text of one of the undertexts in the Archimedes Palimpsest: a commentary on Aristotle's treatise entitled *Categories*. This led to the discovery that the color of the fluorescence can be very important for text recovery in some cases.[5]

With the experience of the 2007 Archimedes Palimpsest imaging in mind, the team returned to the original plan to use a monochrome sensor and no piezoelectric translation for imaging of the Syriac Galen manuscript. Two imaging sessions were held: a preliminary test in March 2009 and a complete session in March 2010. For the Syriac Galen Palimpsest project, images were collected in thirteen reflectance bands over the range from the ultraviolet region through the visible and into the infrared region (365 nm ≤ λ ≤ 940 nm). The 2010 session occurred after the value of the spectral fluorescence information had been demonstrated for the Aristotle commentary in the Archimedes Palimpsest, so images were collected in the same set of thirteen reflectance bands, four "raking" bands with low-angle illumination from the left and right sides at blue and infrared wavelengths, and three bands of spectral fluorescence through red, green, and blue filters under both blue and ultraviolet illumination. All told, twenty-three different image bands were collected for each leaf in the Syriac Galen Palimpsest.

Unfortunately, the only reflectance standard used during the imaging was a so-called Macbeth chart, which is commonly used for color correction of digital color photographs; a similar chart is now available from X-Rite as the Colorchecker. It includes twenty-four reference reflector "patches": six neutral gray reflectors and eighteen reference colors for calibrating the visual appearance of rendered color images. However, the reflectance of the white reference patch in the Colorchecker is significantly smaller at wavelengths outside of the visible range, so calibration of the image data in that scenario would require extrapolation of the measured reflectance. Had the images included a broadband reflectance standard made of Spectralon, which has a much "flatter" reflectance across the range of measured wavelengths including the ultraviolet and infrared regions, it would have been

---

5   K. Bloechl, H. Hamlin, and R. L. Easton Jr., "Text Recovery from the Ultraviolet-Fluorescence Spectrum for a Treatise in the Archimedes Palimpsest," *Proc. SPIE* 7531-08, 2010.

possible to extract reflectance spectra from the data by interpolation between measured values instead of the less-accurate extrapolation from smaller measured reflectance values. This would also have significantly simplified implementation of processing methods from one leaf to others without recalculating local image statistics.

The camera in both the 2009 and 2010 Syriac Galen imaging sessions incorporated the KAF-39000 39-megapixel sensor fabricated by the former sensor division of Eastman Kodak Company. In both collections, the commonly used medium-format Schneider 120-millimeter f/5.6 Apo Digitar lens was used, which is designed to correct chromatic aberrations and maintain focus only over the range of visible wavelengths from 400 to 700 nm. Since spectral images are collected at bands running from the ultraviolet through the visible and into the near-infrared region of the spectrum, this lens could not maintain focus at a single setting across the entire range of imaged bands. It would have been possible, though tedious and difficult, to refocus the camera at the different wavebands to maintain sharpness, but this strategy would have changed the magnification in the different bands and therefore would have significantly complicated the image registration process. At the time of the image collections, we considered using a 60-millimeter f/4 UV-VIS-IR lens designed to maintain focus over bands in the range 300 nm ≤ λ ≤ 1100 nm (from the ultraviolet to the near-infrared regions of the spectrum) for the 35-millimeter image format (36 × 24 mm), then manufactured by Coastal Optics, which now is a division of Jenoptik Group. Though this lens has received excellent reviews when used for imaging of cultural heritage, it is not designed for use with the larger format of the 39-megapixel sensor (49 × 36.8 mm, approximately 50 percent larger along each dimension). Use of the smaller-format lens would have resulted in degraded image quality near the edges of the larger sensor. Since the time of the Syriac Galen imaging, the same designer of the 60-millimeter UV-VIS-IR lens manufactures a 120-millimeter f/4.5 lens for the larger format with similar image quality specifications. Had this lens been available at the time of the image collection, it would have resulted in significantly improved image quality in the infrared bands.

Image Processing

Once collected, the sets of digital images for each leaf are "processed" by computer to combine the spectral bands and render an output image with enhanced visibility of the erased and overwritten text. This project again benefited from the experience of the Archimedes Palimpsest project, which pioneered pseudocolor rendering to improve the legibility of palimpsested text[6] and that also demonstrated success using processing based on the multiband statistics of the image, primarily the process known as principal component analysis (PCA).

The pseudocolor rendering method is based on the observation that the legibility of erased and palimpsested text is enhanced in images of light generated by fluorescence within the parchment. Short-wavelength light photons (most often at ultraviolet energies) are absorbed by the parchment, which then re-emits lower-energy photons with longer wavelengths in the visible spectrum. Because both the incident and re-emitted light are attenuated by remnants of the original ink, and the fluorescence of the parchment is also suppressed by these remnants, the contrast of the original writing relative to the background parchment is enhanced, sometimes to the point of legibility. The visibility of the text is usually maximized in the "blue" band of fluorescence generated by ultraviolet stimulation. To further enhance the visibility of the erased text, local variations in the background were attenuated by a local "contrast normalization," where the 12-bit gray value of each pixel was scaled based on the gray values of pixels in a neighborhood of selected size (usually 501 × 501). The calculation did a linear stretching of the gray value of the specified pixel, around the local mean of the neighborhood pixels, by a multiple of the standard deviation of the neighboring values. This had the effect of "normalizing" the background and minimized local variations in background contrast. Such images generally

---

6   William A. Christens-Barry, Roger L. Easton Jr., and Keith T. Knox, "Imaging and Image Processing Techniques," in *The Archimedes Palimpsest,* vol. 1: *Catalogue and Commentary* (Cambridge: Cambridge University Press, 2011), chapter 5.

showed both the overtext and undertext, and often with similar contrasts relative to the parchment.

The same normalization process was applied to a reflective band of the same leaf, usually a band collected at a red or infrared wavelength, where the undertext is nearly invisible but the overtext is readily apparent. The two normalized bands are then combined to form a "pseudocolor" rendering by inserting the normalized reflectance band into the red channel of a color image, while the normalized fluorescence band is inserted into the blue and green channels. The final image of the later overtext appears to be "dark" in all three colors, while the erased undertext is "dark" in the blue and green bands, but "light" in the red band. In other words, the overtext appears in a neutral "black" or "gray" tone, while the undertext is "reddish." The color cue of the origination of the text was demonstrated to be useful to the scholars working on the Archimedes Palimpsest. This "standard" pseudocolor rendering process was applied to the images of each leaf and significantly improved the legibility of much of the text. The improved legibility came from the addition of the color contrast of the difference of the two bands to the luminance contrast already present in the fluorescence band. The pseudocolor renderings were computed in custom image processing software written by one of the authors (Knox).

At the request of the late Robert Sharples of University College, London, the pseudocolor images were rendered as the monochromatic (black and white) difference of the two bands. This had the effect of "stripping" off the overtext because the gray-level difference of the two image bands in overtext areas very nearly matched that in parchment areas. Put another way, the overtext "disappears" into the parchment. Because the gray values of the undertext are significantly different in the two bands, this text appears with very different gray values in the difference image. These images have less contrast than the pseudocolor images, because they contain only the color difference contrast, without the luminance contrast from the fluorescence band. Even so, often there is character information visible in these images that does not appear in the pseudocolor images. As a result, these renderings have become a standard output of the image processing and quickly were dubbed "Sharpies" as a salute to the original requestor.

For some leaves, the pseudocolor renderings were not sufficient to transcribe the text. Customized processing based on the spectral statistics of the data was applied in those cases. In this case, each image from the 39-megapixel sensor has 7216 × 5412 pixels, with two bytes needed to store the recorded gray value. The N image bands for each leaf are combined to form a data "cube" or "stack" with coordinates $[x, y, \lambda_n]$, where n is the index of the spectral band ($1 \leq n \leq N$). Each pixel in the image cube is described by its N-dimensional vector of gray values recorded in each band. The statistics necessary to evaluate the N principal components are evaluated from the histogram of these N-dimensional pixel vectors, which describes the population of each such vector in the image. PCA transforms the data vectors to a new N-dimensional coordinate system with orthogonal axes that are sequenced so that the variance is maximized for the first component and decreases monotonically for the subsequent axes. In this way, PCA produces an equivalent set of N images that are mutually orthogonal and where the higher-order terms have the least variance. This means that random variations in pixel gray values due to noise mechanisms will be concentrated in the higher-order PCA bands and may be discarded. In practice, the PCA axes are evaluated from the $N \times N$ statistical covariance matrix evaluated from the N-dimensional data histogram. In the Syriac Galen project, the statistical processing was performed in the ENVI software package, which is now distributed by Harris Corporation. Though this software is intended for use in environmental remote sensing, it has many features that are useful for spectral processing of manuscript imagery.

In PCA processing of spectral data of palimpsests, the first principal component will be dominated by the large variance from the dark overtext atop the "bright" parchment. Because subsequent PC bands include variations that are orthogonal to the first band, the "dark" overtext pixels and "light" parchment pixels in the first band will map to the same gray value in the second component. In this way, the overtext "disappears" into the parchment, and the visible variation in the second component will be due to lower-contrast features, possibly enhancing the visibility of the fainter undertext. This was in fact the case for the RGB fluorescence images collected under UV illumination for the leaves of the Aristotle commentary on the Archimedes Palimpsest, and this observation led to the collection of

images under ultraviolet and blue illumination through red, green, and blue filters in the Syriac Galen project.

The segmentation of features by PCA rarely produces a single band containing only the undertext, but rather the undertext is generally visible in a small number of output bands. For this reason, pseudocolor images were usually fabricated from two or three output bands, and those renderings were often manipulated subsequently in imaging software, such as Adobe Photoshop.

Results

From the large number of possible examples, we have selected two bifolia for illustration, 214v–221r and 142r–147v. Figure 1 shows the visual appearance and the pseudocolor image with the ultraviolet band (λ = 365 nm) in the blue and green channels and a near-infrared band (λ = 780 nm) in the red channel. As described earlier, the later overtext is "dark" in both bands and appears with a dark neutral tone in the pseudocolor image, while the earlier undertext, which is much less visible in the infrared band, appears with a "reddish" tint.

The first of the two images in figure 2 is the grayscale difference "Sharpie" image generated from the two bands used in the pseudocolor image. The second is the pseudocolor rendering of two bands evaluated by principal component analysis evaluated from the set of six reflective bands with the shortest wavelengths (λ_1 = 365 nm through λ_6 = 570 nm), plus the red, green, and blue fluorescence bands under ultraviolet illumination. Note that the overtext is nearly invisible in the grayscale difference image and appears in a very different color in the PCA pseudocolor band.

The two images in figure 3 are the correspondents to those in figure 1 for bifolio 142r–147v, with the visual appearance on the left and the pseudocolor rendering on the right. This is a more difficult undertext that does not appear as prominently in the pseudocolor image. Figure 4 shows the "Sharpie" image on the left, which shows the undertext more clearly than the pseudocolor rendering. Single PC bands are shown from two different sets of input bands at the center and on the right of figure 4: that in the

FIGURE 1. Visual appearance of bifolio 214v–221r on the left and the pseudocolor rendering generated from the ultraviolet ($\lambda = 365$ nm) and a near-infrared band ($\lambda = 780$ nm) on the right.
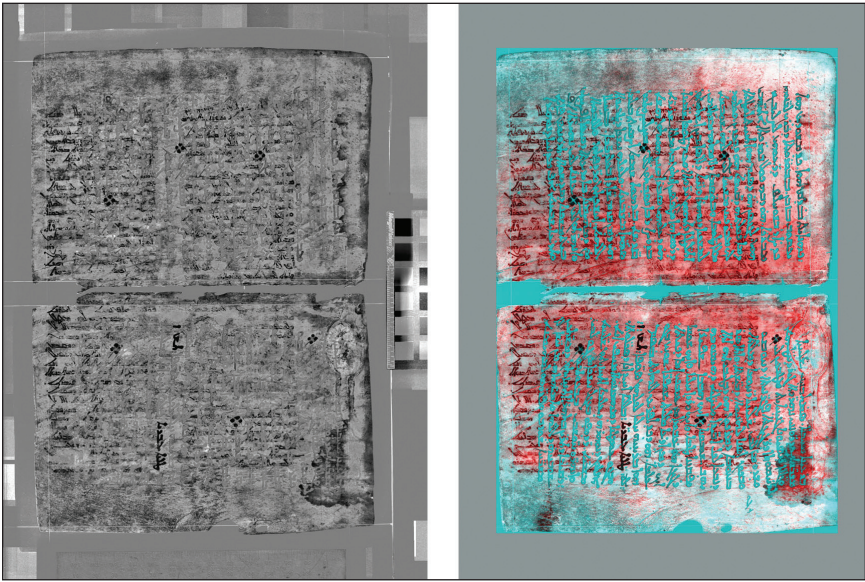

FIGURE 2. The "Sharpie" image (difference of normalized bands from the ultraviolet and near-infrared used in the pseudocolor of figure 1) appears on the left. The pseudocolor rendering on the right consists of two output bands from principal component analysis.

FIGURE 3. The visual appearance (left) and pseudocolor rendering (right) of the bifolio 142r–147v, where the undertext is more difficult to discern.
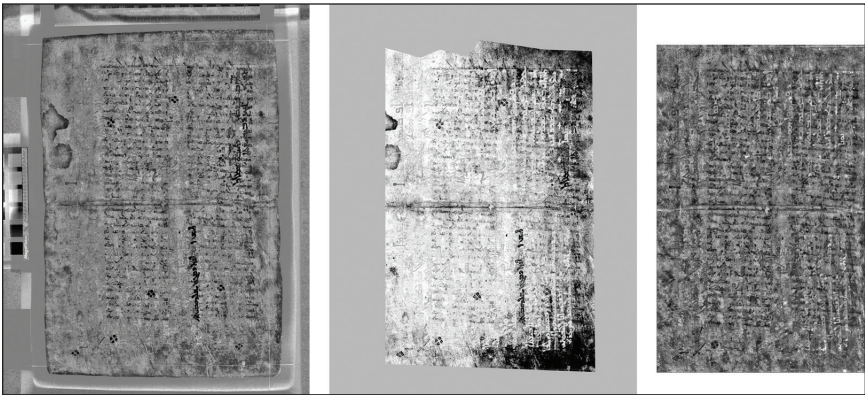

FIGURE 4. (left) The "Sharpie" image shows the undertext rather more clearly than the pseudocolor image in figure 3. Two different results from PCA are shown in the gutter and on the right, where the text may be more apparent yet.

center is PC band #2 from the same set used in figure 2, while that on the right is PC band #3 from a set of twelve bands. The text appears rather more clearly in this case.

Scholarly feedback generally indicates that comparison of several different output renderings is more useful for accurate transcription than a single output image. Not every character on a leaf responds in the same way. For that reason, we supplied the scholars with several different types of processed images.

Conclusions

The authors' experience in the Archimedes and the Syriac Galen projects has shown that no one method or algorithm will recover erased or damaged text in all cases. As can be seen in the figures in this paper, this is true even within different regions on a single individual leaf of the manuscript.

In cases where the spectral differences in the characters are visible in the individual wavelength bands, the pseudocolor method of combining a fluorescence band and a red or infrared reflectance band in a pseudocolor image can be an effective method to improve significantly the legibility of the erased text. The corresponding "Sharpie" difference image can also reveal some of these characters that are not enhanced by the pseudocolor process. These methods are easy to compute and can be applied in an unsupervised batch mode to every leaf.

Alternatively, when the contrast differences of the erased text in the different wavelength bands are not readily visible, it takes more sophisticated processing methods to reveal these very low contrast characters. Here is where the PCA method is valuable for separating character information from the background "noise" to make the erased characters more legible. Again, combining different-output PCA bands into a pseudocolor image can increase the legibility of the erased text through the addition of color contrast to otherwise black-and-white images. However, very powerful, supervised methods, like PCA, do require human supervision within the processing loop. As a result, these methods take more effort in

the processing step and are therefore applied more sparingly, only to the leaves that did not respond well to the unsupervised algorithms.

Given the state of the technology at the time, the imaging of the Syriac Galen Palimpsest arguably was rather successful. Since then, digital cameras have improved, more wavelengths have become available, and processing methods have become more sophisticated. As a result, given the relentless advance of technology, better results would likely be obtained if the manuscript were reimaged today.

Acknowledgments