

Enhancement of overwritten text in the Archimedes Palimpsest

Keith T. Knox

Boeing LTS, 535 Lipoa Parkway, Kihei, HI, USA 96753

ABSTRACT

The Archimedes Palimpsest is a thousand-year old overwritten parchment manuscript, containing several treatises by Archimedes. Eight hundred years ago, it was erased, overwritten and bound into a prayer book. In the middle of the twentieth century, a few pages were painted over with forged Byzantine icons. Today, a team of imagers, scholars and conservators is recovering and interpreting the erased Archimedes writings. Two different methods have been used to reveal the erased undertext. Spectral information is obtained by illuminating the manuscript with narrow-band light from the ultraviolet, through the visible wavebands and into the near-infrared wavelengths. Characters are extracted by combining pairs of spectral bands or by spectral unmixing techniques adapted from remote sensing. Lastly, since all of the text was written with iron gall ink, X-Ray fluorescence has been used to expose the ink underneath the painted icons. This paper describes the use of color to enhance the erased text in the processed images and to make it visible to the scholars. Special pseudocolor techniques have been developed that significantly increase the contrast of the erased text and make it possible to be read by the scholars despite the presence of the obscuring, overlaid text.

Keywords: enhancement, pseudocolor, manuscripts, visualization, remote sensing

1. INTRODUCTION

Pseudocolor is a standard technique that uses color in an image to represent non-color information, hence the name pseudo- or false-color. Typically, pseudocolor is used to represent grayscale details in an image that extend over a dynamic range larger than what can be seen by the human visual system. This technique results in sharp color contours that delineate arbitrary and not-always meaningful grayscale boundaries in the original image.

We use pseudocolor as a means to enhance the visibility of erased text in the digital images of a very old parchment manuscript. The erased text is not in a different dynamic range from the non-erased text on the page, but rather is in a different spectral range. The two texts, erased and not erased, can be differentiated by spectral classification techniques. The pseudocolor technique is used to visually distinguish the two texts in the processed images, so that a scholar can easily read the erased writing. Because the color contours in the pseudocolor image align with the boundaries of the erased text, the color changes significantly enhance the readability of the erased text for the scholars. The pseudocolor technique described here will be illustrated using images from the Archimedes Palimpsest. Unfortunately, due the nature of color, the black & white hardcopy version of this paper will not present this technique in its best light.

A palimpsest is a manuscript leaf that has been overwritten by being “scraped again” to clean off the first writing, before the second layer of writing is applied. The Archimedes Palimpsest is a copy of several treatises by Archimedes that was erased and overwritten 800 years ago. The manuscript was purchased at auction by a private collector in 1998.

Shortly after the purchase of the Archimedes Palimpsest, the new owner contracted the Walters Art Museum of Baltimore, Maryland to conserve the manuscript and to make a transcription available to scholars. An imaging team was organized consisting of the author of this paper and his two colleagues Roger Easton, Jr. of the Center for Imaging Science at the Rochester Institute of Technology and William Christens-Barry of Equipose, Inc. Since 2000, the team has imaged the manuscript both optically and with X-Ray fluorescence. During this time, several scholars of classical mathematics and ancient manuscripts have been engaged in transcribing and interpreting the processed images from the imaging team. The final results of this study will be published near the end of 2008. More information about the manuscript and the attempts to recover the original text of Archimedes can be found in a recent book¹ just published by the Curator of Manuscripts at the Walters and the project’s lead scholar of classical mathematics.

2. ARCHIMEDES PALIMPSEST

The Archimedes Palimpsest consists of 177 leaves that were bound in book form. Each leaf was written on both the front and back sides of the leaf. Two leaves from the manuscript are shown in Figure 1. The one on the right-hand side of the figure is in reasonable shape, except for some staining around the edges. The leaf on the left-hand side, on the other hand, has had much rougher treatment. The upper half of the left leaf has several moldy sections where the text has been eaten away. The bottom half of that leaf was separated from the top half, erased again and a fake Byzantine icon was painted over it. It is thought that a previous owner might have added the icon with the misguided idea that it would increase the value of the manuscript.



Fig. 1. Two leaves from the Archimedes Palimpsest, 28 verso – 21 recto on the left and 93 verso – 92 recto on the right. The text oriented vertically on the two leaves is the overlying, non-erased text that is part of the prayer book. The erased Archimedes writing that the scholars want to read is oriented horizontally in this figure. The bottom half of the left leaf was erased a second time and a forged Byzantine icon was painted on the leaf obscuring both layers of text. The copyrights of all the images in this paper reside with the owner of Archimedes Palimpsest.

Multispectral imaging works very well for revealing the erased text on the leaves that are similar to the leaf on the right-hand side of Figure 1. Images were taken over a range of wavelengths from ultraviolet through the near infrared. The erased text is most visible in the ultraviolet, and it is least visible in the infrared.

The moldy sections are the hardest to restore and little has helped to recover text in those sections. The painted icons, on the other hand, have responded very well to X-Ray fluorescence. All of the text in the manuscript was written using iron gall ink. Since X-Ray fluorescence is able to distinguish different elements in the material, the iron map from the X-Ray fluorescence is able to see through the paintings and recover the text underneath.

We have found that our pseudocolor technique was able to significantly enhance the visibility of the erased text for both the multispectral and the X-Ray fluorescence images. Fortunately, the pseudocolor technique developed for one image type worked very well for the other.

3. MULTISPECTRAL IMAGING

During the month of August 2007, the Archimedes Palimpsest was completely re-imaged. Images were taken at 833 dpi using a Sinar 54H digital camera. Since the Sinar has a Bayer color filter on its sensor, the camera was set to take four separate images, while the sensor was piezo-electrically moved by one pixel width or height between exposures. The resultant four images were combined to create a single image with three full-color separations.

The actual size of the sensor was only 4,080 x 5,440 pixels, but the sensor could also be moved one-half pixel width or height, thereby doubling the resolution of the final image. The total process of increasing the resolution and removing the color filter pattern meant that 16 separate images were taken at each location for each illumination. The resultant images contain 8,160 x 10,880 full-color, RGB pixels.

Each leaf of the manuscript was illuminated with a series of LEDs at 365, 445, 470, 505, 530, 570, 617, 625, 700, 735, and 870 nanometers. Two LED panels were designed and built by W. Christens-Barry and operated under computer control. The panels were cycled through all the wavelengths and 16 exposures taken for each illumination. Given that the parchment leaf did not move during this cycle, all of the spectral images for a given leaf are in perfect registration, enabling us to apply multispectral unmixing techniques to the data.

3.1 Spectral response

The images shown in Figure 2 illustrate the spectral behavior of the two inks. The erased text is more sensitive to wavelength changes and disappears under red light while the non-erased text is still visible. This makes sense, since both the erased text and the parchment are reddish in color.

Both types of text are very visible under ultraviolet illumination. This is expected given the reddish color of the iron gall ink. They may be an additional effect, however, due to the ultraviolet illumination. The parchment fluoresces under ultraviolet illumination and emits visible light in the blue end of the spectrum.

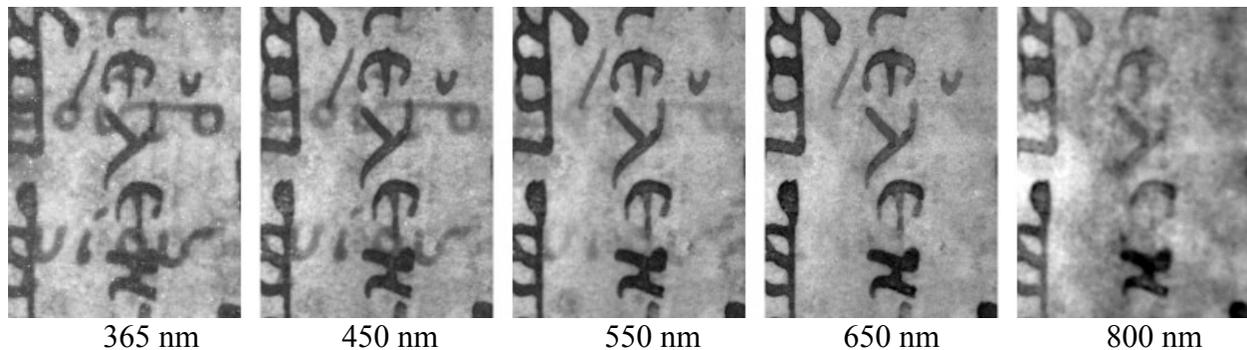


Fig. 2. Images of the manuscript in 5 different wavelengths. The erased text is horizontally oriented and the non-erased text runs vertically. Both texts lose contrast as the wavelength increases, but the erased text disappears more quickly and is almost gone under red illumination.

3.2 Difference image

The images in Figure 2 suggest an immediate method for recovering the erased text. Since the ultraviolet image shows both texts and the red image shows only the unwanted text, it seems that it should be possible to subtract the unwanted text from the ultraviolet image, leaving an image of the desired, erased text.

To take the difference between two images, they first must be adjusted so that the parts that should cancel are approximately the same value. In Figure 3, a section of 93 verso – 92 recto is shown. The original color image is on the top. The second image was taken in ultraviolet light and the third from the top is the red separation of the color image.

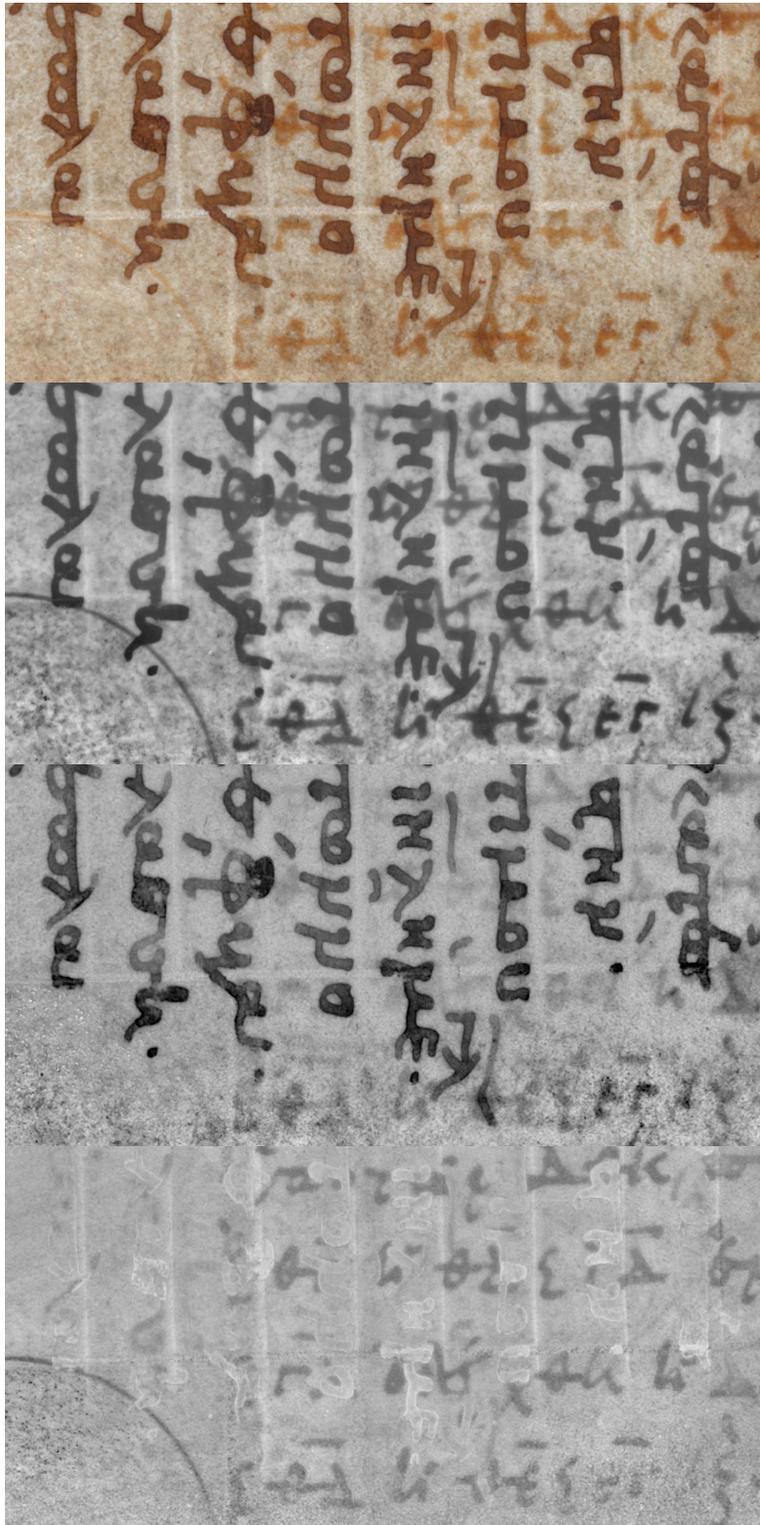


Fig. 3. Difference of two images. The top image is the original color image. The second image was taken in ultraviolet light, which shows both texts. The third image down is the red separation of the color image, which shows only the unwanted text. The bottom image is the difference of the second and third images and shows the desired erased text.

The bottom image is the difference of the ultraviolet and red images, after they have been adjusted to be of the same average contrast. The contrast was adjusted by sliding a window across each image (401 pixels square), and changing the center pixel in the window according to the variance and mean of the image data within the window. The mean was moved to mid-gray and the contrast was adjusted so that 6 standard deviations stretched between black and white.

The difference image that results shows the erased text as black text on a gray background. A piece of a circle from a diagram is also visible in the lower left corner. The text along the bottom of the image is very legible. The text in the main body of the image is visible, except where the text was overlaid by the unwanted text. If the balancing was done right, the overlap regions will become the same gray value as the background parchment.

While this approach appeals to the scientist, the scholars were bothered by the difference images and preferred to read the erased text from the ultraviolet images. Their reasoning was that a gap in the difference image could be caused by either a region of overlap or a region where there is no text information and they could not tell the difference. They could more easily read around the overwritten text in the ultraviolet image.

3.3 Pseudocolor

To accommodate the scholars, we moved to pseudocolor in order to show both the difference image and the two overlapping texts. The pseudocolor image in Figure 4 was created by taking the two contrast adjusted images from Figure 3 and putting the red image into the red separation and the ultraviolet image in the green and blue separations.

The result in the pseudocolor image is that the erased text shows up as red text and the unwanted text is seen as neutral gray or black. The contrast of the erased text is therefore much higher. It has the same black & white contrast as it does in the ultraviolet image and it has an additional color contrast equal to its contrast in the difference image. This effect will not be seen in the black & white hardcopy version of this paper.

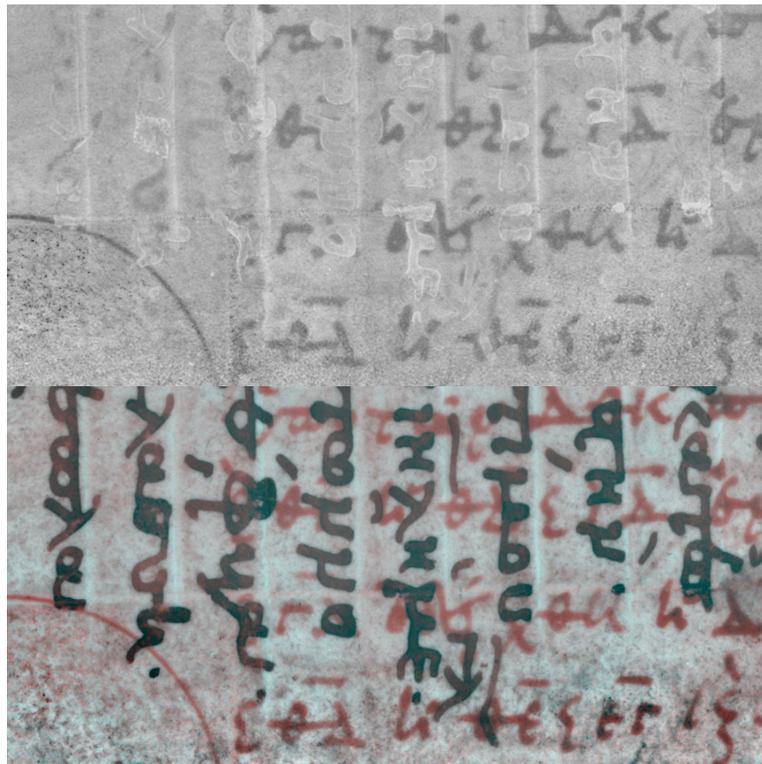


Fig. 4. Pseudocolor. The top image is the difference image showing the desired text. The bottom image is the pseudocolor image where the differences are shown in color. The combination of the black & white character information with the differences in color significantly increases the contrast of the desired text in the pseudocolor image.

4. SPECTRAL UNMIXING

The erased text on the majority of the leaves in the manuscript is sufficiently revealed using the simple multispectral method with just two wavelength bands, as described in the previous section. In a few places, there is very little left for the scholars to read and the ultraviolet illumination does not produce high contrast text. In those cases, we are trying to apply more sophisticated processing techniques that have been developed in the remote sensing field.

4.1 Unconstrained spectral unmixing

The previous method only used two wavebands of the multispectral image. There are 11 bands in the data that we collected on the manuscript, so more sophisticated techniques can be applied. The linear spectral unmixing method², is an unconstrained least squares optimization of a linear component model.

The assumption is that a given spectral response is a linear sum of several spectral components, called signatures. If the individual signatures can be determined independently, then a least squares fit of those signatures to the spectral response can be computed to determine the fraction of each signature that is contained in the response.

The spectral responses are contained in the collection of images at the 11 different wavelengths. Since each image is exactly registered with each other image in the set for a given leaf, there are 11 known spectral values for each pixel.

A signature is determined by identifying specific pixels that belong to known classes. For example, consider the moldy section of 28 verso that is shown in Figure 5. There are places where exposed erased Archimedes text can be identified. Pixels from those regions can be used to determine a signature for the erased Archimedes ink. Similarly, regions are visible that can be used to determine signatures for plain parchment, moldy parchment and non-erased text.

The goal is to express every pixel as a linear sum of the four identified signatures. The first step is to make this a linear problem. Since the reflectance of ink on parchment, or on another ink, is a multiplication of the corresponding transmissions, the problem can be made linear by taking the logarithm of the spectral values at each pixel.

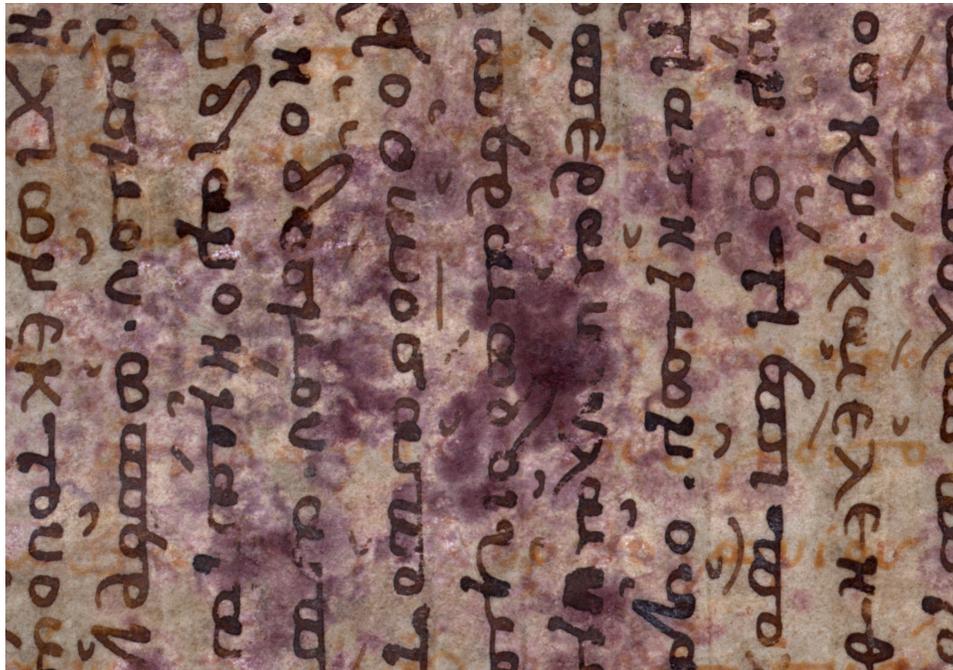


Fig. 5. Moldy section of 28 verso. Parts of this leaf have been partially or completely eaten by mold. This image contains regions where sections of erased text, non-erased text, plain parchment and moldy parchment can be identified.

Once the signatures are determined, the least squares solution for the fraction of each signature at a given pixel is given by multiplying the spectrum at that pixel by the Moore-Penrose pseudoinverse of the matrix containing the spectral signatures. This is a standard solution that is thoroughly described in reference 2.

The result of the least squares solution is a fraction map for each signature component. At each pixel, a fraction value is determined and when these fractions are plotted as an image, the places where that component is located are revealed. For text, it is simply a matter of reading the characters. Examples of several fraction maps are shown in Figure 7.

These fractions should have values between 0 and 1. A zero value says that there is no trace of that signature in that pixel and a unity value says that there is only that signature present. Another constraint, therefore, is that all of the fractions should sum to unity.

The problem is that the outputs from an unconstrained least squares solution do not have such nice values. To accommodate noise across the spectral range, a fraction map may have very large values, either positive or negative. These are spurious values and they can be ignored if the fraction map is viewed by itself.

These aberrant values in the fraction map are a significant problem, though, if the fraction maps are to be combined together. The fraction maps look very much like the difference images from Section 3. That means that a pseudocolor image could be created from the fraction maps, if they can be combined together. To do that, a constrained least squares solution, that limits the values of the fraction map, is necessary.

4.2 Constrained least squares

Two problems are addressed with the constrained least squares solution described here. The first problem is that noisy values resulting from the unconstrained least squares solution, can also exceed the expected bounds of zero and unity. The second problem is that the spectral responses of the inks and the parchment are not uniform over the leaf. The noise problem was solved by using a constrained least squares³ algorithm and the non-uniformity of the spectral signatures was addressed using a stepwise linear unmixing⁴ method. The combined method is shown below in Figure 6.

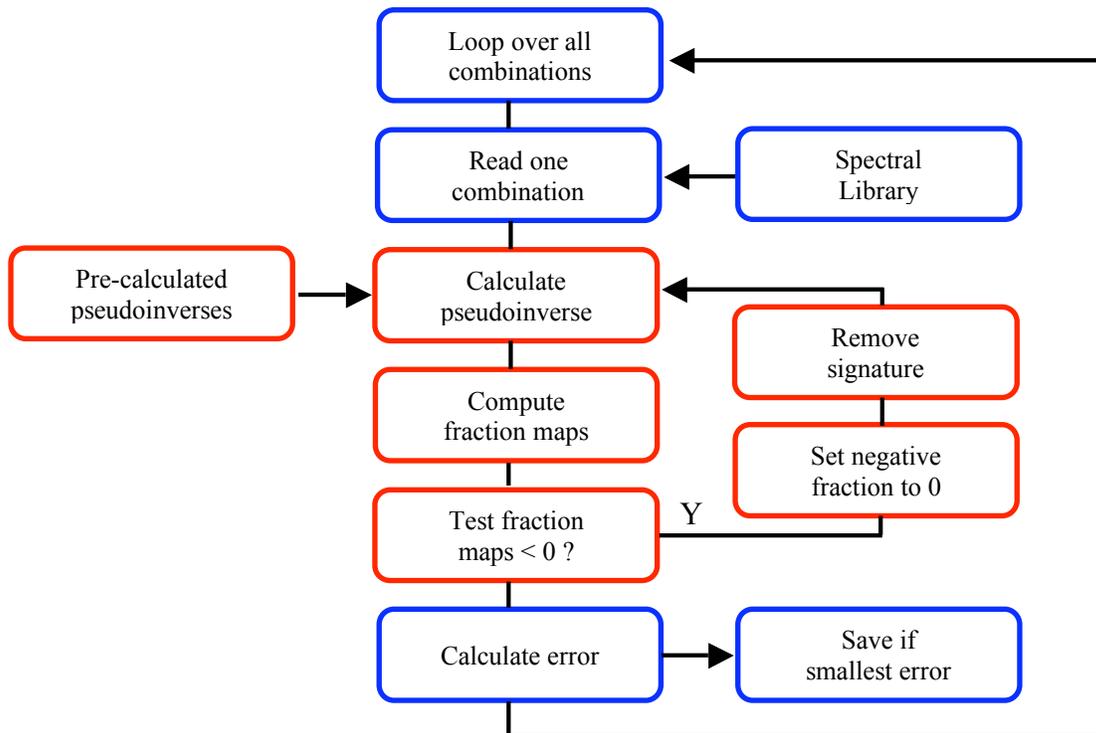


Fig. 6. Flow chart of combined constrained least squares solution. In the inner loop, a trial inverse is tested for the existence of a negative fraction value. In the outer loop, successful inverse solutions are compared for the best fit.

Before the algorithm starts testing solutions, a library of all possible combinations of signatures/solutions is created. For example, there is only one solution which uses 4 signatures, there are 4 solutions which use only 1 signature, and there are 6 solutions that use 2 signatures. In addition, if there are N pixels chosen to represent parchment across the leaf, then the number of solutions is multiplied by N . With several representative choices for each signature type, the number of possible solutions to test increases significantly.

In the outer loop, one set of 4 values is read from the spectral library. In the inner loop, all combinations of those 4 values are tested against the spectral data for a given pixel. If any combination produces a negative fraction for a signature, then that signature is dropped and other combinations are tried. For each combination that produces a non-negative result, the outer loop tests the goodness of that fit against all other trials at this pixel. After the inner and outer loops have tried all possible cases, there will be one case which gives no negative fraction values and the smallest error. If any fraction value is missing at this location, then its value is set to zero. Figure 7 shows the fraction maps for this algorithm applied to the image in Figure 5.

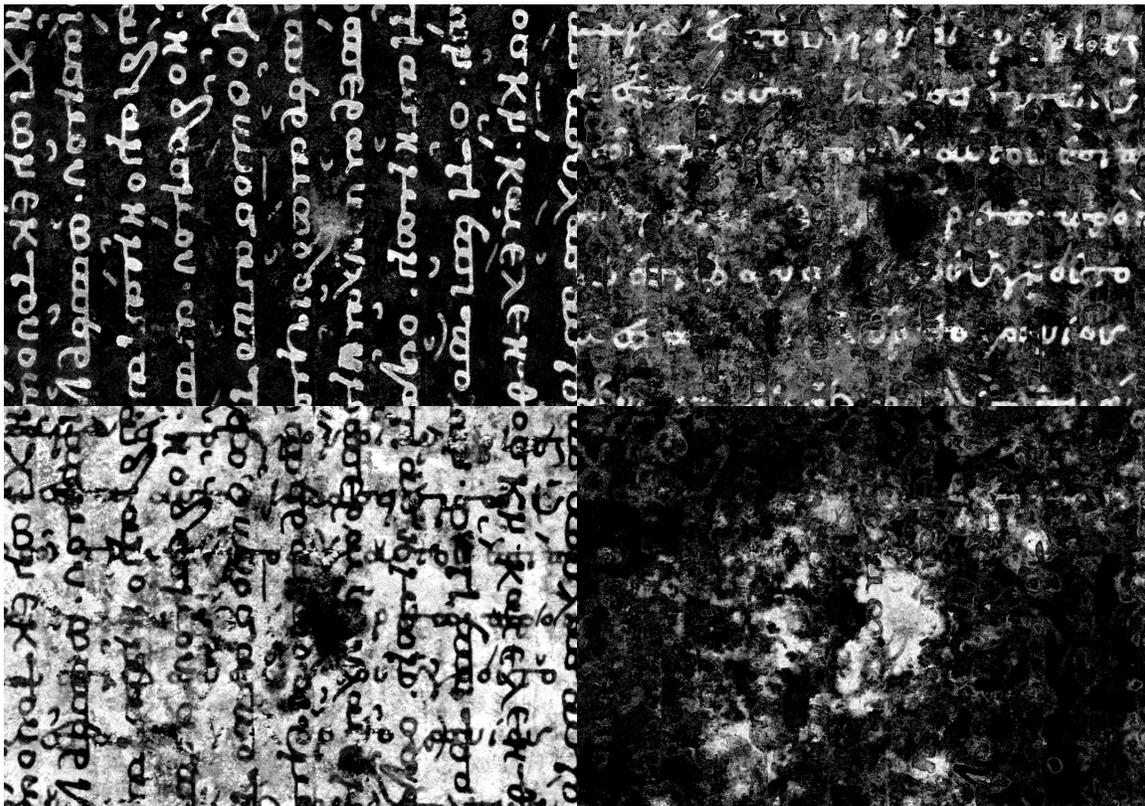


Fig. 7. Fraction maps using constrained least squares solution. Upper left, non-erased text; upper right, erased text; lower left, parchment; lower right, mold.

The presence of a component in the fraction maps in Figure 7 is shown with white. Black indicates that none of that component exists at that pixel. The non-erased text is well separated, because it is already visible. The parchment shows the location of all text characters and the mold as black. The white areas in the mold fraction map correspond to the moldy regions on the leaf. The erased Archimedes text is clear in many locations, but it very difficult to read in the moldy regions.

The fraction maps produced by this method are significantly less noisy than those produced by the unconstrained algorithm. In addition, they have the advantage that they can be combined together in order to create a pseudocolor image that enhances the erased text.

4.3 Pseudocolor

The interpretation of the fraction maps can be aided by using pseudocolor. Analogous with the two-band case, the color separations of the pseudocolor image are designed to create one separation that has both texts visible and the other separation showing only one text. That combination will create an image in which the desired text will be shown in color and the unwanted text will show as neutral gray.

The separation which shows both texts is created by adding the parchment and the mold fraction maps from Figure 7. To remove the erased text, the other separation is the sum of the parchment, the mold and the Archimedes text fraction maps. To make the pseudocolor image, the latter sum is used for the red separation and the former sum for the other two separations. The result is shown in Figure 8.

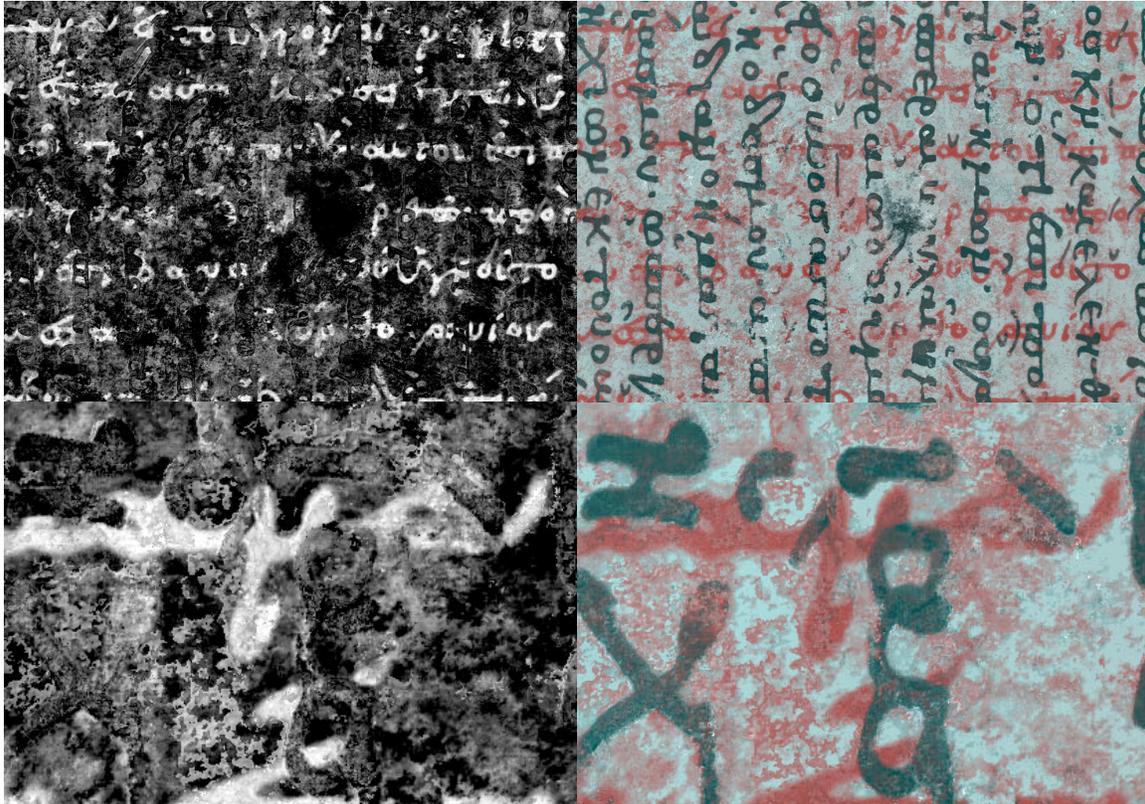


Fig. 8. Pseudocolor using fraction maps. Upper left is the erased text fraction map. Upper right is the pseudocolor image constructed from the fraction maps. The lower images are blowups of the upper left corner. In the blowups, it is clear that some partially overwritten characters, recognizable in the pseudocolor, are unclear in the fraction map by itself.

Some characters can be recognized in the pseudocolor image that are not clear in the fraction map. For example, the alpha in the upper left corner of the image is sufficiently obscured, so that it is not recognizable in the fraction map alone. The same principle in the difference image, that made gaps ambiguous and confused the scholars, applies to the fraction maps, as well. That ambiguity is removed by combining the fraction maps and showing the two layers of overlapping text together.

5. X-RAY FLUORESCENCE

The Archimedes Palimpsest has a handful of leaves that were erased a second time and a Byzantine forgery painted on the leaf. An example is shown in Figure 1. Some of the paint contains gold and most of it is opaque to optical wavelengths. X-Ray fluorescence was used to read the characters underneath the paint without removing the forgeries.

Three trips were made to Stanford University to the Stanford Linear Accelerator Center (SLAC) to use the synchrotron. The forgery leaves were scanned using X-Ray fluorescence to search for traces of iron in the parchment. Dr. Uwe Bergmann of SLAC sponsored the visits and prepared the site and equipment to accommodate the delicate leaves in the normal operating environment at SLAC.

Although it seems simple to pass X-Rays through the paint and parchment to detect the iron underneath, it presents an interesting complication. The original Archimedes leaves were written on both sides of the parchment. When the leaves were included in the prayer book, they were erased and overwritten on both sides. That means that the X-Ray images have 4 layers of writing visible, two from each side of the parchment. The optical multispectral images only have two layers of images to interpret. The overlapping four layers make the interpretation significantly more difficult.

An example X-Ray fluorescence image is shown in Figure 9, together with the corresponding optical image. In the X-Ray fluorescence image, a jumble of characters appears. There is writing from both sides of the parchment and this single image does not indicate which characters belong on which layer or on which side. Still, being able to read the characters at all from underneath the painting was exciting to the scholars. What we found, however, was that the pseudocolor technique could help differentiate between the layers from the two sides.

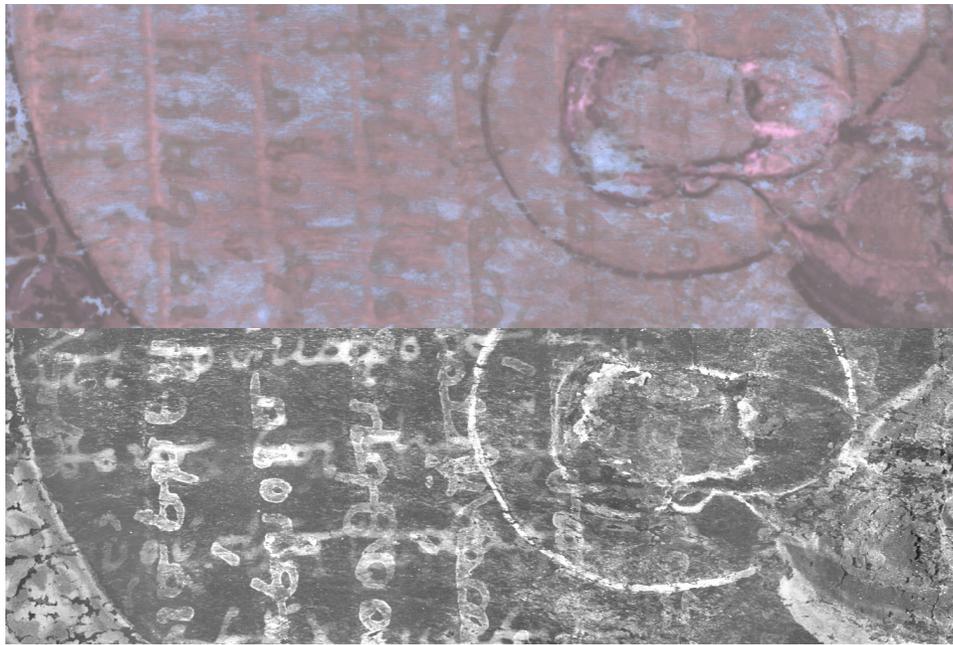


Fig. 9. X-Ray fluorescence image. Upper image is an optical image in ultraviolet. A few characters are visible underneath the paint. The lower image is the iron map of an X-Ray fluorescence image. In this image, characters appear that were not previously visible. There are 4 layers of characters from both sides of the parchment.

5.1 Pseudocolor

To create a pseudocolor image, there need to be two images which show different information. These differences are what are emphasized in color and provide the enhancement to the information that is being read.

For the X-Ray fluorescence images, the two different images were created by using two X-Ray detectors on the two sides of the parchment leaf. One detector was on the same side as the X-Ray source and the other was on the back side of the parchment.

The X-Ray beam would strike the parchment at one location and a reading was taken from both detectors. After the reading was completed, the leaf was moved (under computer control) to the neighboring location and the measurement was repeated. In this way, the whole leaf was scanned. This process took 24 hours to complete a scan of a single leaf.

The fortunate occurrence was that the signal received from an iron deposit at a given detector was stronger if the writing was on the same side of the parchment as was the detector. In other words, the text on the front side of the parchment gave a stronger signal in the detector on the front side than it did in the detector on the back side of the parchment.

This is just the kind of difference that is needed to create a pseudocolor image. The interesting difference between the X-Ray images and the optical images is that both sets of text, front and back, are strong in one detector and weak in the other. For the optical case, one set of text was always strong and the other was always weak. This symmetry in the X-Ray images means that the two sets of text from the two sides of the parchment will be manifested in complementary colors, rather than color and neutral gray.

The pseudocolor image for the above image is shown in Figure 10. In this figure, the text from the front side (legible) is shown in yellow. The text from the back side (seen here in reverse) is shown in blue.

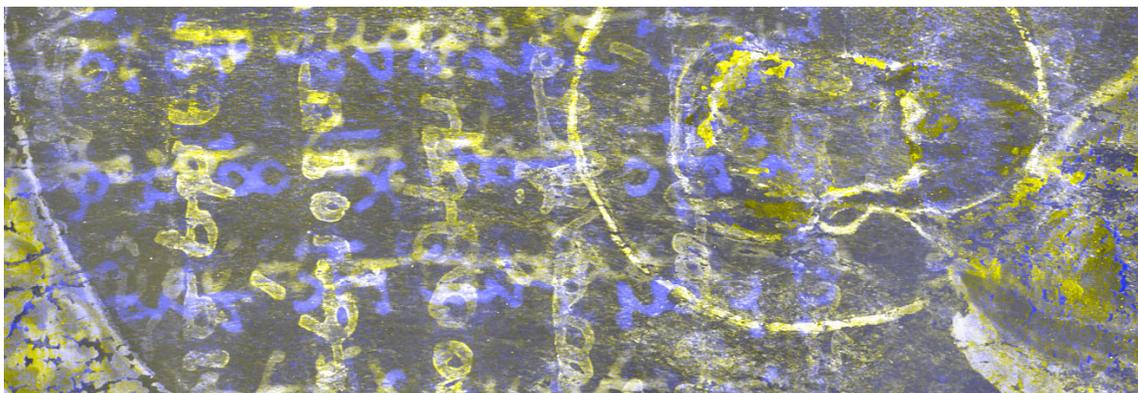


Fig. 10. X-Ray fluorescence pseudocolor. The layers of text from the two sides of the parchment show up in this pseudocolor image in complementary colors, yellow and blue.

The erased Archimedes text is also distinguishable from the non-erased text. The erased text is seen in the X-Ray fluorescence images as soft text with fuzzy edges. The non-erased text, which has ink still on the parchment, has very sharp edges. Using these sharpness differences and the color differences between the two sides, the scholars can separate out all 4 layers from these pseudocolor X-Ray images.

6. CONCLUSIONS

Pseudocolor introduces sharp color contours into an image. Ordinarily those color boundaries are not very meaningful. By aligning the color contours with meaningful text edge information, individual types of text can be differentiated in color. This color alignment adds significant contrast to the aligned text, thereby greatly enhancing its visibility.

REFERENCES

- ¹ R. Netz and W. Noel, *The Archimedes Codex*, Da Capo Press, Perseus Book Group, 2007.
- ² Robert Schowengerdt, *Remote Sensing*, Academic Press, San Diego, CA, 1997, p. 447.
- ³ Shimabukuro, Y. E. and Smith, J. A., "The Least-Squares Mixing Models to Generate Fraction Images Derived From Remote Sensing Multispectral Data", *IEEE Trans. On Geoscience and Remote Sensing*, vol. 29, no. 1, 16-20, 1991.
- ⁴ Gross, H.N., "An Image Fusion Algorithm for Spatially Enhancing Spectral Mixture Maps," Doctoral Dissertation, RIT, 1996.