Recovery of handwritten text from the diaries and papers of David Livingstone

Keith T. Knox[a], Roger L. Easton, Jr.[b], William A. Christens-Barry[c], and Kenneth Boydston[d]
[a]Air Force Research Laboratory, 535 Lipoa Parkway, Kihei HI, 96753
[b]Chester F. Carlson Center for Imaging Science, Rochester Inst. of Tech., Rochester NY 14623
[c]Equipoise Imaging, LLC, 4009 St. Johns Lane, Ellicott City, MD 21042
[d]Megavision Inc., P.P. Box 60158, Santa Barbara CA 93160

ABSTRACT

During his explorations of Africa, David Livingstone kept a diary and wrote letters about his experiences. Near the end of his travels, he ran out of paper and ink and began recording his thoughts on leftover newspaper with ink made from local seeds. These writings suffer from fading, from interference with the printed text and from bleed through of the handwriting on the other side of the paper, making them hard to read. New image processing techniques have been developed to deal with these papers to make Livingstone's handwriting available to the scholars to read.

A scan of the David Livingstone's papers was made using a twelve-wavelength, multispectral imaging system. The wavelengths ranged from the ultraviolet to the near infrared. In these wavelengths, the three different types of writing behave differently, making them distinguishable from each other. So far, three methods have been used to recover Livingstone's handwriting. These include pseudocolor (to make the different writings distinguishable), spectral band ratios (to remove text that does not change), and principal components analysis (to separate the different writings). In initial trials, these techniques have been able to lift handwriting off printed text and have suppressed handwriting that has bled through from the other side of the paper.

Keywords: Palimpsests, spectral imaging, pseudocolor, principal component analysis

1. INTRODUCTION

David Livingstone (1813-1873), the African explorer, left a large volume of diaries and correspondence to the world after his death. Dr. Adrian Wisnicki, a scholar of Livingstone's writings, broadcast a message in the summer of 2009 requesting assistance in transcribing some of Livingstone's diaries. This request was noted by Dr. William Noel of the Walters Art Museum in Baltimore, who is the Director of the Archimedes Palimpsest project. At Dr. Noel's suggestion, the authors contacted Dr. Wisnicki to suggest the application of spectral imaging system and processing techniques developed during the Archimedes Palimpsest project[1]. This offer was accepted.

In preparation for imaging of the entire corpus of material, Dr. Wisnicki arranged to borrow a letter written by David Livingstone to his friend and future biographer Horace Waller from the collection of Peter and Nejma Beard. This letter, originally written in Bambarre, was successfully imaged as an adjunct to another palimpsest imaging effort in March 2010. The success with this effort contributed to a successful proposal to the National Endowment for the Humanities and the British Academy to publish a critical edition and spectral image database of the diary and letters Livingstone wrote in 1870-71

In the summer of 2010, the team traveled to the National Library of Scotland in Edinburgh to collect and process spectral images of pages of the Livingstone Manyema Field Diary, which consists of two halves: (1) the Bambarre Diary from 1870-71 and (2) the Nyangwe Diary from 1871; portions of these are in the collections of the David Livingstone Center in Blantyre, Scotland and in the National Library of Scotland. In addition, the team also imaged a few relevant letters from the same period also in the collection of the National Library of Scotland.

Today Livingstone's diary is in a fragile state: its pages, which were also subject to adverse environmental circumstances, are crumbling, and large portions of Livingstone's handwritten text have become illegible due fading, blotting, water damage, and other problems.

As a whole, the diaries and journals from Livingstone's final travels in Africa (1866-73) represent an under-researched field: to date most critical endeavor related to this period has focused on Livingstone's letters and Livingstone's Last Journals (1874), a highly-flawed, posthumously published text. The need for a fresh edition has been championed by many leading Livingstone scholars.

2. IMAGE COLLECTION

The images Livingstone's papers were collected with our standard multispectral imaging system[2] incorporating LED illumination at 12 wavelength bands from 365nm to 940nm and a 39-megapixel monochrome camera from Megavision, Inc. of Santa Barbara, CA. Images were collected at the National Library of Scotland in June-July, 2010.

The illumination is provided by two panels of LEDs that were built by Equipoise Imaging, LLC. Each panel contains seven banks of LEDs that emit in the ultraviolet (1 wavelength) and visible regions (6 wavelengths) of the spectrum and five additional clusters of LEDs that emit in the infrared region (5 wavelengths). The geometry of the LED light panels is designed to provide maximum uniformity of the illumination at the surface of a leaf being imaged. Each LED emits at a specific wavelength with a FWHM of approximately 25 nm.

A page from the Nyangwe Diary (1871) is shown in Figure 1. This diary provides Livingstone's original account of the notorious Arab-perpetrated massacre of Africans in the village of Nyangwe, which became an iconic rallying point for late-Victorian abolitionists. The diary also records the circumstances—including an uncensored record of Livingstone's deteriorating state of health—leading up to his famous meeting with Henry M. Stanley.

As he ran out of clean paper, Livingstone improvised by writing crosswise over the printed pages of the only texts available to him in Central Africa: in this case, on issues of The Standard. When his ink ran low, he resorted to making his own ink from the seeds of a local plant. The printed text makes the faint handwriting very hard to read.

A letter from the same time period is shown on the right side of Figure 1. In this case, the ink bled through both sides of the paper, making the handwriting difficult to read. The goal, in this effort, is to develop image processing methods to make the handwriting legible despite underlying printed text or bleed through from the opposite side of the page.

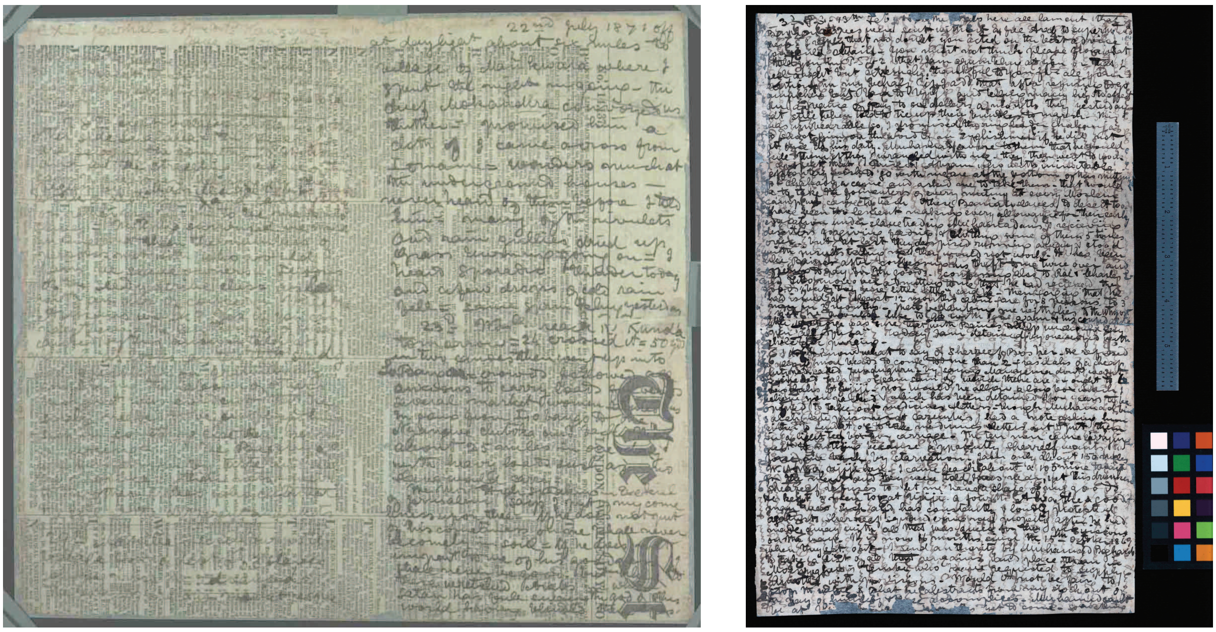

Figure 1: On the left, a page from David Livingstone's Nyangwe Diary, written in 1871. The text was written on top of a page from *The Standard.* Scholars have requested help in reading the faint text on the left side of the page. On the right, is a letter from the same time period. The writing is difficult to read as a result of ink bleeding through from the opposite side.

3. IMAGE PROCESSING

3.1 Pseudocolor Applied to Bleed Through

Pseudocolor was a technique that was used very successfully to reveal erased text on the Archimedes Palimpsest[3-6]. The erased text consisted of iron gall ink stains on parchment. The erased text was overwritten with a Christian prayer book, also written in iron gall ink. Under ultraviolet illumination, the stains from the erased text increase significantly in contrast, making them more visible. The difficulty is that the existing iron gall writing also increases in contrast. The result, under ultraviolet illumination, is two sets of writing of equal contrast, which is difficult to read.

Under visible illumination, the erased text had a significantly lower contrast. By combining the two renditions together, the ultraviolet and the visible images, the differing contrasts of the two inks (the existing ink and the erased ink) could be made into contrasting color, i.e. a neutral gray for the existing text and bright red for the erased writing. This emphasized the erased writing, allowing the scholars to easily read around the existing ink on top.

For the Livingstone papers, a similar procedure was tried. The difference is that for these papers, the underlying substrate is different, paper not parchment, and the inks are both visible, except that one is seen on the paper, while the other has bled through the paper. For the Livingstone papers, because one of the writings has bled through from the other side of the paper, the scans of the reverse side can be used to suppress the competing text.

In Figure 2, is an image of the letter by David Livingstone to Horace Waller shown in Figure 1. Part a) shows a portion of the recto side of the letter using 505 nm illumination. Part b) shows the reverse side of that same section, flipped horizontally. Part c) is a pseudocolor image with recto 505 nm in the red, recto 780 nm in the green and verso 505 nm in the blue. This combination displays the recto text in cyan-blue and the verso text in green-yellow. It can be seen in part c), that these color differences make the recto text much easier to read against the text from the other side.

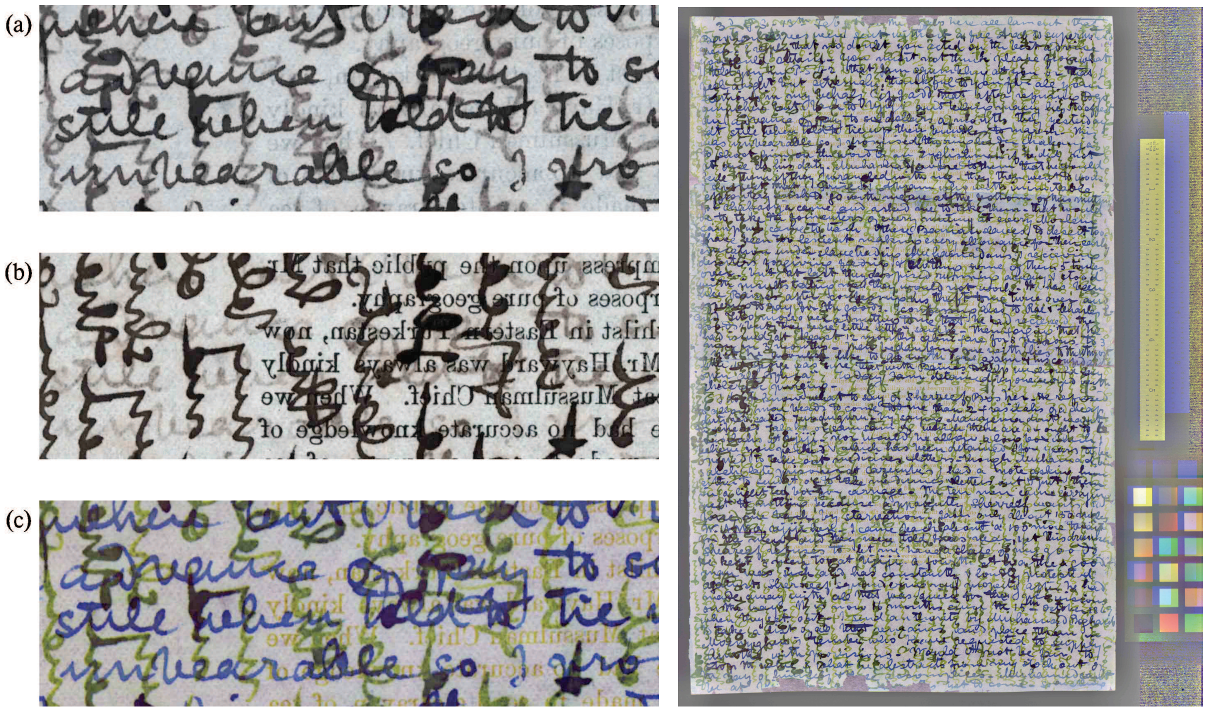

Figure 2: A letter written by Livingstone in 1871. On the right is a pseudocolor image that distinguishs the handwriting from the ink that bled through from the opposite side. On the left, (a) a section of the writing on the recto side; (b) the same section on the verso side, flipped horizontally to align with the bleed through; (c) the two sides combined in pseudocolor.

3.2 Principal Component Analysis Applied to Removal of Printed Text

The fact that the different layers of writing differ in color as well as contrast suggested the application of principal component analysis (PCA) to extract the handwriting from the printed text, a technique used successfully on the Archimedes Palimpsest.[7-8] PCA images were calculated using the ENVI software package from ITT Visual Information Solutions. Pixels outside the page (e.g., of the resolution and color targets) were masked from the statistics.

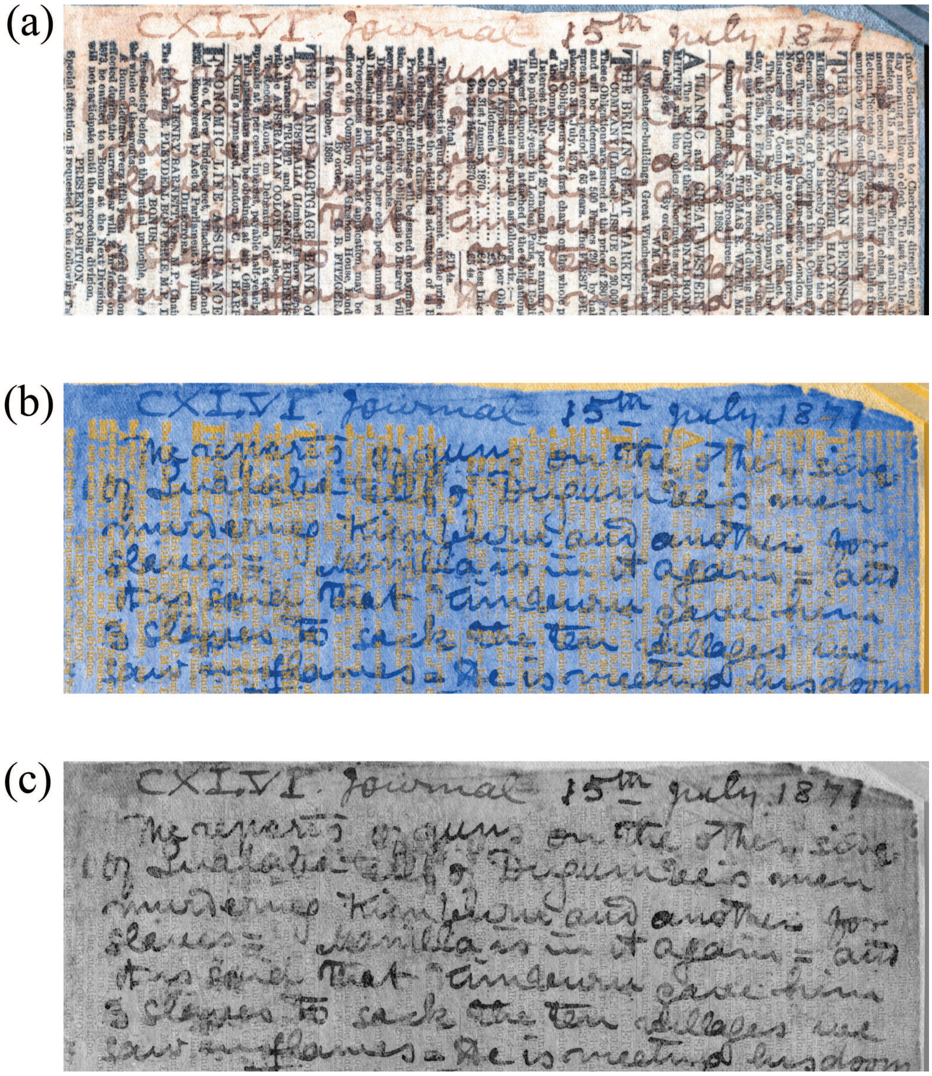

Figure 3: (a) Visual appearance of section of journal page for 15 July 1871 (DLC297b CXLVI f. 12r), showing fading of handwritten text and obscuration by printed text; (b) pseudocolor image created from first and second PCA bands with band 2 in red and blue channels and band 1 in green channel, after rotation of hue angle; (c) green channel of image in (b), showing near complete removal of printed text.

The 12 resulting component images were examined visually to determine which that might be most useful to the transcription (usually bands 2 and/or 3). Dr. Wisnicki prefers the printed text significantly suppressed but not eliminated, so that the handwriting is legible and its relationship to the printed text can still be seen. Only rarely did a PCA axis align with the histogram so that the printed text was significantly suppressed. One useful tool for image rendering was adapted from results of previous work on the Archimedes Palimpsest. Since the two writings exhibit distinct spectra, it should be possible to project the data onto a spectral axis where the handwritten text is approximately orthogonal to the printed text, so that mainly the former appears. Only rarely did an individual PCA band satisfy this requirement.

To better extract the handwritten text, pseudocolor images were created by inserting selected bands into the RGB channels of a color image; an example is shown in Figure 3b. The hue angle of the pseudocolor image was rotated until the visibility of the handwritten text was maximized in a particular color band, which was then extracted for use by the scholars in their transcription; an example shown in Figure 3c. Comparing Figure 3a with 3c, it can be seen that this method successfully removes the printed text, revealing the handwriting and making it legible.

3.3 Spectral Ratios Applied to Removal of Printed Text

A second method to remove the printed text, forming the ratio of pairs of spectral components, was tried and found to work well. It successfully suppresses the printed text, making the handwriting legible. It does not suppress the printed text as much as the PCA method does, but it can be implemented in batch mode, without operator intervention and therefore is a good first method to apply.

A section of the diary page from Figure 1 is shown in Figure 4. Parts a) through d) show the variation of the printed text and the handwriting with wavelength, 450 nm, 592 nm, 850 nm and 940 nm. In the blue region of the visible spectrum, the handwriting is dark. As the wavelength is increased, it almost completely disappears in the red, and does completely disappear in the near infrared. The printed text, on the other hand, is fairly constant across this spectrum.

Since the printed text is the only writing left at 940 nm, dividing the other wavelengths by the 940 nm separation effectively suppresses the printed text. The ratios are shown in the right-hand side of Figure 4. The color image, in Figure 4h, is a combination of the three ratios in Figures 4e, 4f and 4g, in the red, green and blue channels, respectively.

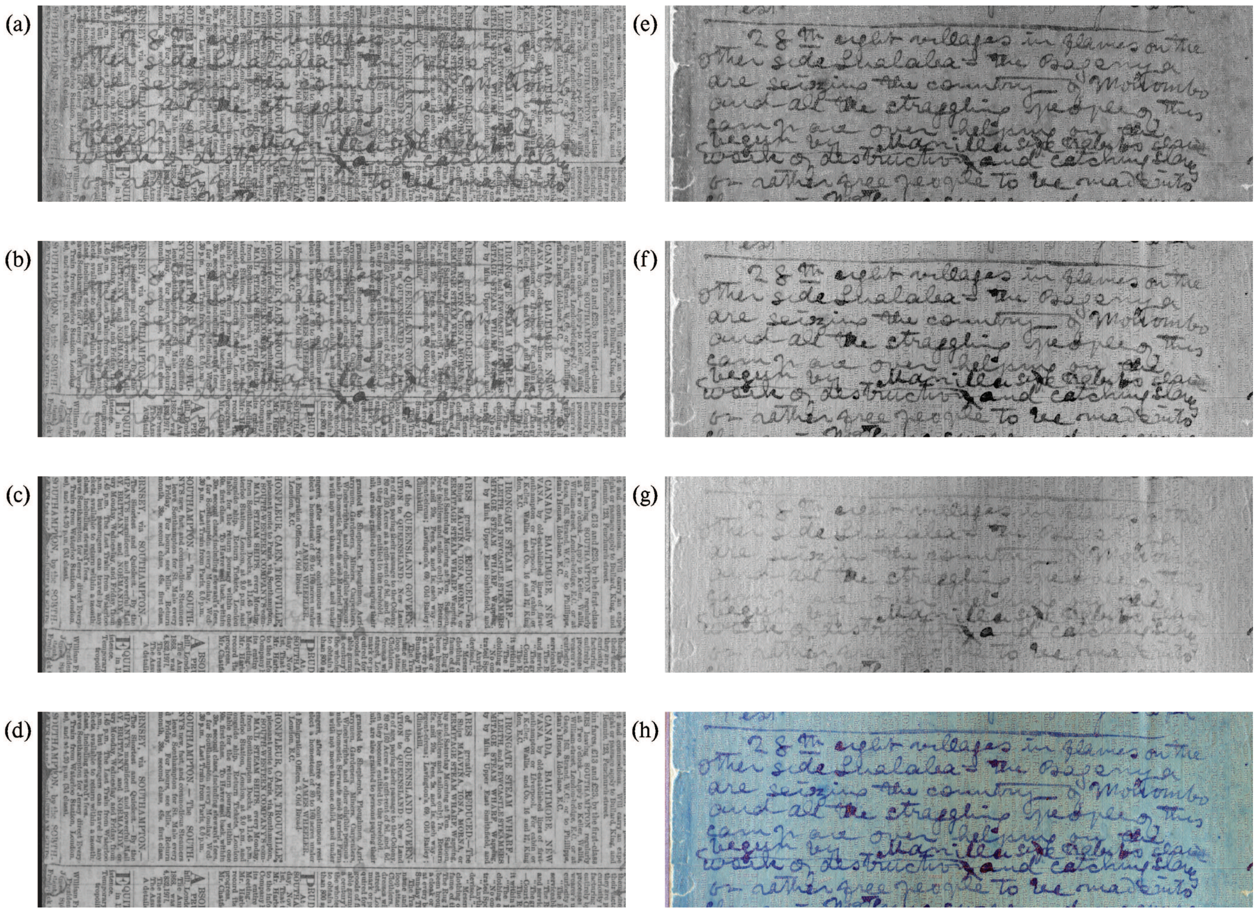

Figure 4: Section of the diary page in Figure 1. (a) 450 nm; (b) 592 nm; (c) 850 nm; (d) 940 nm; (e) ratio of 450 nm by 940 nm; (f) ratio of 592 nm by 940 nm; (g) ratio of 850 nm by 940 nm; (h) color image, (e) in red, (f) in green and (g) in blue.

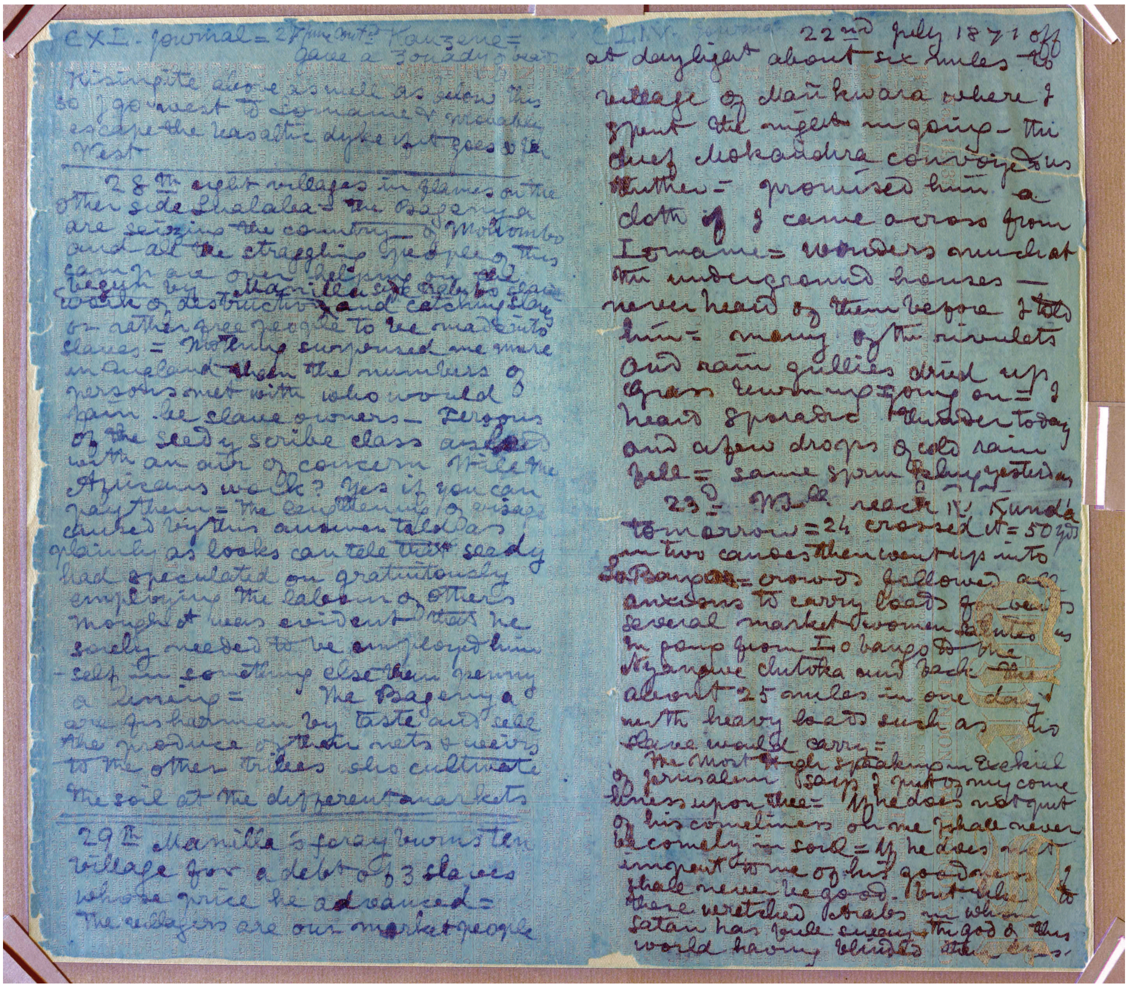

Figure 5: Processed image of the diary page from Figure 1. Three wavelengths, 450 nm, 592 nm and 850 nm were divided by the 940 nm separation and combined into a color image. The effect is to suppress the printed text and make the handwriting legible.

4. CONCLUSIONS

The image processing methods described here have effectively suppressed the interference of the printed text and the bleed through of the ink from the opposite side of the page. This effort is just beginning and further advances are expected. For example, the pseudocolor image that uses the images from both sides of the page, effectively colors both sets of handwriting, but does not separate them. Further efforts in applying color segmentation, or principal component analysis may be able to separate the two sets of writing. The two methods of suppressing the printed text, by principal components analysis and by spectral band ratios, greatly suppress, but do not eliminate the printed text. Further work is needed, including automation of the hue rotation selection. Future developments will be reported when available.

5. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the assistance of Dr. Adrian Wisnicki, Dr. William Noel, Michael B. Toth, Michael Phelps, and Doug Emery. Rab Jackson and Alison Metcalf from the National Library of Scotland in Edinburgh, Karen Carruthers and Anne Martin from the David Livingstone Centre, Suzanne Lamb from the National Trust of Scotland, and Kate Simpson from Edinburgh Napier University were essential to the successful data collection. Several RIT students, including Carrie Houston and Katie Lange, performed the PCA image processing. This work was supported by grants from the British Academy and the National Endowment for the Humanities. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily represent those of the National Endowment for the Humanities.

REFERENCES

[1] Netz, R., and Noel, W., [The Archimedes Codex], Da Capo Press, Philadelphia (2007).

[2] Easton, R.L., K.T. Knox, W.A. Christens-Barry, K. Boydston, M.B. Toth, D. Emery, and W. Noel, "Standardized system for multispectral imaging of palimpsests", Proc. SPIE 7531, pp. 75310D-1 - 75310D-11, (2010)

[3] Knox, K. T., "Enhancement of overwritten text in the Archimedes Palimpsest", Proc. SPIE 6810, pp.681004-1 – 681004-11, (2008).

[4] Bergmann, U., Knox, K. T., "Pseudo-color enhanced x-ray fluorescence imaging of the Archimedes Palimpsest", SPIE Proc. 7247, Document Recognition and Retrieval XVI, Kathrin Berkner and Laurence Likforman-Sulem, Editors, (2009).

[5] Knox, K. T., Easton Jr., R. L. and Christens-Barry, W., "Image Restoration of Damaged or Erased manuscripts", 16th European Signal Processing Conference, (2008).

[6] Easton Jr., R. L., Knox, K. T., Christens-Barry, W., "Multispectral Imaging of the Archimedes Palimpsest", Proc. 32nd Applied Imagery Pattern Recognition Workshop, AIPR'03, pp. 111-116, (2003)

[7] Bloechl, K., H. Hamlin, and R. L. Easton, Jr., "Text Recovery from the Ultraviolet-Fluorescent Spectrum for Treatises of the Archimedes Palimpsest," Proc. SPIE 7531, pp. 753109-1 -753109-9 (2010)

[8] Salerno, E., Tonazzini, A. and Bedini, L., "Digital image analysis to enhance underwritten text in the Archimedes Palimpsest ", Intl. J. on Document Analysis and Recognition, vol. 9, pp 79-87, (2007).