Digital Image Restoration Technology as Applied to Ancient Degraded Textual Material Using Color Imaging Systems

Robert H. Johnston and Roger Easton
Rochester Institute of Technology, Rochester, New York
Keith Knox and Reiner Eschbach
Xerox Corporation, Webster, New York
Joseph Tusinski and Mihai Cuciurean Zapan
Rochester Institute of Technology, Rochester, New York

Abstract

A team of research scientists from academia and industry has been applying digital image processing techniques for the enhancement of degraded ancient textual documents including Dead Sea Scrolls and burned papyrus scrolls from Petra, Jordan. The team consists of scholars from the Rochester Institute of Technology, the University of Rochester and the Xerox Digital Imaging Technology Center, Webster, New York. Using remote sensing, network transmission, digital enhancement, color manipulation and current storage and output technology, the investigators are working to provide maximum image information for the accurate translation of the degraded documents.

Introduction

Textual scholars worldwide deal with a wealth of images that have survived the ravages of time. These images suffer from a broad range of degradation. Some survive in totally readable condition. Others are damaged to the point that scholars need assistance in bringing forth as much usable text as possible. Details of ancient characters, accent marks, overwritten lines, and the shape and form of letters are sought. Erasures, in lines of text, are explored to attempt to discover what has been erased and perhaps why. The discovery of the Dead Sea Scrolls in 1947 brought to light ancient documents relating to the Bible from approximately 250 BCE to 135 CE. These documents are among the oldest biblical texts by 1000 years. Many of these writings were in fragments and many were damaged making accurate translation extremely difficult. Documents found were written on clay, leather, and papyrus. Some stored in clay jars were in remarkable condition while others found on cave floors under as much as nine feet of debris were in a deplorable state. Scholars have been working for over 50 years studying this rich find. Recently, a number of papyrus scrolls were uncovered at Petra, Jordan. These papyri were burned in antiquity. Experts were brought in to separate the charred fragments. These fragments have been photographed and the negatives have been sent to us for scanning and digital enhancement. We are dealing with dark text, written in ancient Greek, on black charred background. There has been and there is a dire need for help from scientists to assist these dedicated scholars by making visible as much material as possible using current imaging science technology. Scholars, according to Dr. Emanuel Tov, Editor-in-Chief of the Dead Sea Scroll Publishing Project, pursue the study of biblical text as an auxiliary science to biblical exegesis. Such scholars, with their studies in orthography and morphology, investigate each minuscule detail of text. It is the purpose of this research to apply elements of the imaging chain and to develop techniques that will apply to the broad based needs of degraded image enhancement. Much of our work has direct forensic application as well.

Methodology

Our research has progressed on two tiers. Initially we conducted our study using standard Macintosh equipment. We applied our enhancement techniques using readily available software. Adobe Photoshop played a key role in its several versions. Our rationale was that most enhancement applications would be used by non-imaging science professionals who needed access to standard and easily available technology. Our second level of research uses very sophisticated hardware and software available at the Carlson Center for Imaging Science and at the Kodak and Xerox research facilities. Dr. Keith Knox made the Xerox image processing software available to us. The system is UNIX based and includes a broad range of imaging applications. We will be describing this two-tiered system during this presentation and will present some of our results.

The enhancement of the Targum of Job is an example of color subtraction using Photoshop tools (See Figures 1 and 2).

Figure 1. Original Targum of Job

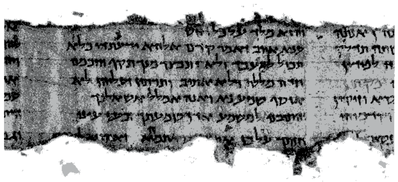


Figure 2. Enhanced Image

Our early image capture procedure used a flat-bed scanner for digitizing photographic images. The files were stored on optical discs at a resolution of 300-600 dpi. Later a drum scanner was used to scan in documents with significantly better resolution. We also captured images of three-dimensional objects using a digital video camera. Once captured, the images were enhanced using a broad spectrum of Photoshop™ applications. Some images were captured using a DCS-100 camera system at the Kodak facility and both IR and UV illumination. Eventually some material was scanned at the Kodak facility and stored on CDs as RGB images with a resolution of 288 pixels/inch. Results were converted into TIFF files and transmitted over the Internet to the Xerox Webster Research Center. The images were loaded into the Xerox image processing system for more sophisticated enhancements. During this time, we were in regular contact with textual scholars in England, Israel, and the United States through our Internet connection. This made the exchange of ideas and information easily available. Our initial images were of text written by ancient scribes on leather or papyrus. Later we also worked with documents written on clay. Each presented us with unique enhancement problems. The material came from the Dead Sea Scroll archives and from a new find at Petra, Jordan. The Petra material, written in ancient Greek on charred material, presented some unique research challenges.

Since the papyrus from Petra had been burned, it is difficult to distinguish the characters from the background. Also the fibrous structure presents an additional problem. The texture of the papyrus generates edges throughout the document, which complicates digital enhancement.

The enhancement of the images of the papyrus was done with the Adobe Photoshop™ software package. Within Photoshop, various operations were utilized to process the papyrus. To begin, each piece of papyrus is selected from the image and processed without the "white" background. Next, the selections are filtered with kernels that produce a result with larger mean values and contrast. By applying successive kernels, we can obtain images containing the texture of the papyrus in varying amounts.

Then the image is subject to the "Levels" command, which allows for the compression/expansion of the range of brightness values via the "Output Levels" option. To increase the visibility of the characters, the "Selective Color" command is utilized, which allows for the percentage of CMYK to be altered for each color. Specifically, we have modified the Neutral and Black levels. The modification of the Neutral levels increases the dc values of the background and the Black levels are modified to decrease the dc values of the text. This last procedure provides optimum results if the image is transformed to the CMYK color space. (See Figures 3 and 4) A similar method was attempted in the frequency domain, but the non-uniform distribution of the texture presented a difficult obstacle. Another very useful tool for enhancing the papyrus was the "Unsharp Mask Filter" that can sharpen only the edges in an image or it can sharpen any portion of an image according to our exact specification.

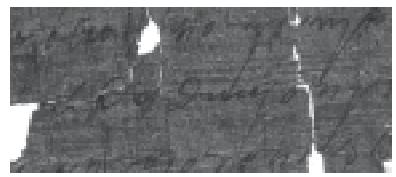


Figure 3. Original Papyrus

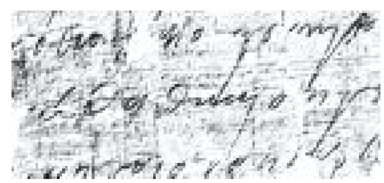


Figure 4. Enhanced Papyrus

The Temple Scroll was written in ink on parchment. Its image was obtained by scanning a color transparency taken in white light by a Scanmaster™ D4000 drum scanner at 600 dpi and stored as a RGB image. The rolled scroll was exposed to moist debris before its discovery, and contains regions of unreadable or degraded text that were exposed to elements, interspersed with regions where the text is virtually pristine. Some regions of the scroll are missing completely. In the degraded regions, the darkening of the color of the background obscured the lettering. These color changes suggested that the first efforts to enhance the readability of the degraded regions should be directed at examining the individual RGB images and projecting the data onto different axes in the color space. A transformation to the "E" channel in the Xerox "YES" space improved the readability of some of the characters.(See Figures 5 and 6) As "E" is a combination of the "R" and "G" channels, other weighting combinations of "R" and "G" were tried, eventually leading to projections onto axes in the R-G plane at different azimuth angles.

The image information was examined further by computing the various two-dimensional histograms of the RGB data. (See Figures 7 and 8) The histogram of the R-G data exhibited two extended clusters in the histogram corresponding roughly to the degraded and non

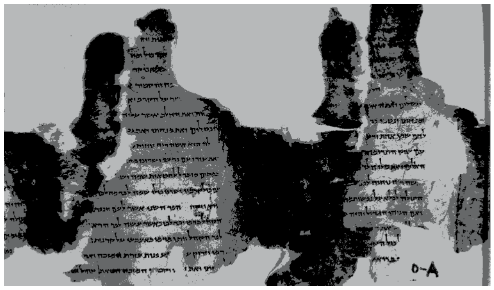


Figure 5. Original Temple Scroll

Figure 6. Scroll after YES Color Space Transformation

degraded text. The respective slopes of these clusters are approximately 2 and 1.3, respectively. The regions of the image that correspond to different regions of the histogram were identified by masking different portions of the 2-D histograms.

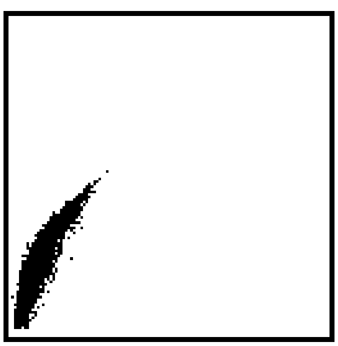


Figure 7. Two-dimensional Histogram of a Degraded Text Region of the Temple Scroll

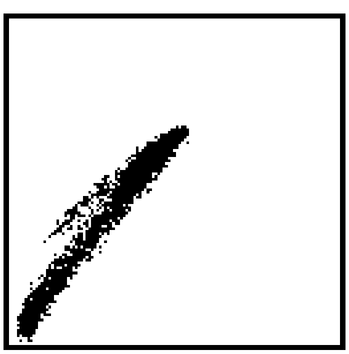
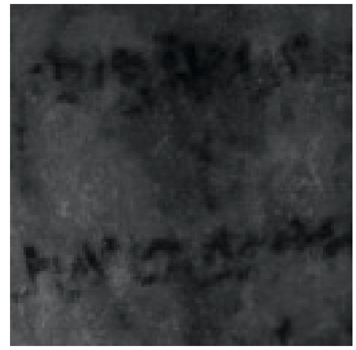
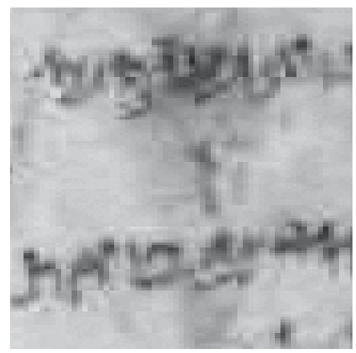


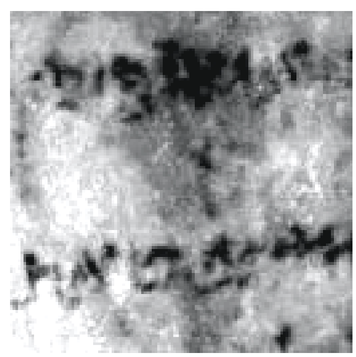
Figure 8. Two-dimensional Histogram of the Non Degraded Text Region of the Temple Scroll



(a)



(b)



(c)

Figure 9. (a) Original Portion of Degraded Text in the Temple Scroll, (b) YES Color Space Transformation of Same Region and (c) First Principal Component of Same Region

The three-dimensional RGB histograms of the image were examined by using the AVS visualization environment running on an SGI stereographic workstation. It was apparent that the contrast of the degraded text could be maximized by projecting the data onto the “long axis” of the histogram cluster, i.e. the first principal component of the data set. The principal components are the eigenvectors of the sample covariance matrix and have the property that the projection of the data onto a subset of the largest principal components is a minimum-error representation of the data. Principal components often are used in pattern recognition, such as in environmental remote sensing. The principal components of the image of the Temple scroll were computed in the ERDAS software package (Earth Resources Data Analysis Software). The projection of the first principal component onto the B-axis is small compared to the projections onto the R- and G-axes. In other words, the first principal component lies approximately in the RG plane, and in fact, approximately along the E-axis in the YES color space. The images were processed by computing the principal components for subsections of the image with degraded text projecting the data onto the first principal component, and enhancing the contrast of that one projection. Results are shown in Figures 9.

Conclusion

Our research continues in developing hardware and software applications to assist translators of ancient degraded textual material by providing them with all available material gleaned from the ultra violet, infrared and the visible spectra. We will continue to develop algorithms, kernels and plug-ins and make them applicable to our mission. Our applied research has successfully provided much needed additional information to translators here and abroad. Scholars have been able to read sections of previously unreadable text. We have successfully begun to apply our results to forensic applications working with local law enforcement agencies. Our research in applying ultraviolet florescence has provided access to seriously degraded text on papyrus material. In addition, the results of the color space work has provided scholars with additional characters that were previously not seen. Our future plans involve going into the field both here and abroad and capturing sections of degraded text with our DCS camera system at high resolution and bringing that material back to our laboratory for enhancement treatment.

We are grateful to New York State for supporting us with a Center for Advanced Technology grant and for the support of the Xerox corporation in providing us space, proprietary enhancement systems, advanced printing output and the expertise of two outstanding Xerox scientists, Dr. Keith Knox and Dr. Reiner Eschbach. Special thanks to Dr. Bruce Zuckerman, of the West Semitic Research Project, for providing images of the Temple Scroll. Dr. Edward Przybylowicz, Director of the Carlson Center for Imaging Science, has been supportive of our research by providing constant encouragement, space, equipment and technical expertise. Without the support mentioned our work could not have succeeded to this point.