Analysis of the spectra of degraded documents

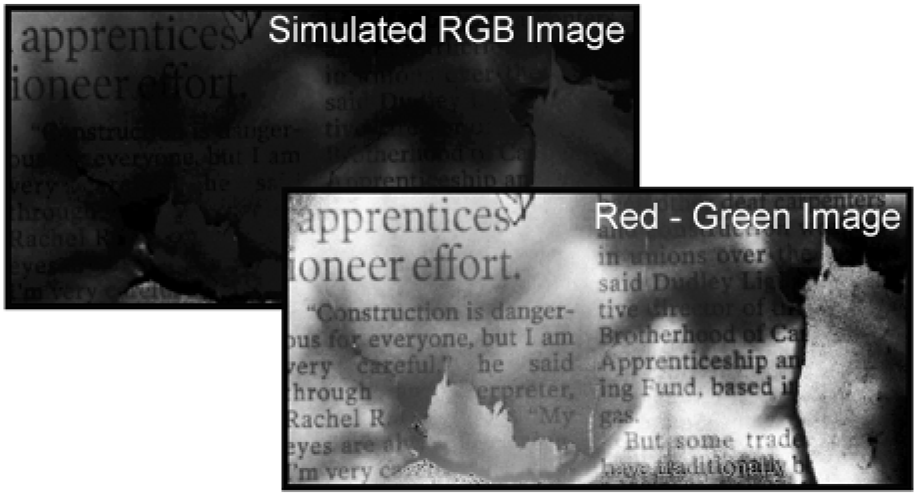
Ryan Swift

SIMG-503

Senior Research

Analysis of the Spectra of Degraded Documents

Final Report



Ryan Swift
Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology
May 2001

Table of Contents

Analysis of the Spectra of Degraded Documents

Ryan Swift

Table of Contents

Analysis of the Spectra of Degraded Documents

Ryan Swift

Abstract

The intent of this research was to examine the spectral reflectance of degraded documents to determine whether information is available in regions visible to the human eye. Ultimately we attempt to determine the practicality of manipulating visible (RGB) images to obtain similar information commonly obtained by using infrafred (IR) imaging methods. Charred newspaper was tested as an example of a degraded document. Spectral reflectance of the degraded documents was measured using a fiber optic grating spectrometer. In addition, images were taken of the documents in both the visible (RGB) and infrared (IR) regions of the spectrum.

By examining the spectral reflectance at different wavelengths in different areas of the documents, we were able to locate information (i.e., characters) on the documents, and determine the wavelengths where the greatest contrast exists between the text and the charred paper. The areas of interest on the charred pages were imaged in several wavelength bands of the spectrum, including the near-ultraviolet and near-infrared regions. Useful information was observed in the IR images, and we attempted to obtain similar results by digitally manipulating visible light (RGB) images. Our hypothesis, that manipulation of the RGB images (specifically subtraction of channels) will find similar contrasts between text and parchment as found in the IR images, was supported by the research.

Table of Contents   Copyright

Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology
Rochester, NY 14623-5604

This report is accepted in partial fulfillment of the requirements of the course SIMG –503 Senior Research.

Title: Analysis of the Spectra of Degraded Documents
Author: Ryan Swift
Project Advisor: Dr. Roger Easton
SIMG 503 Instructor: Dr. Anthony Vodacek

Table of Contents   Introduction

Analysis of the Spectra of Degraded Documents

Ryan Swift

Introduction

Images of degraded documents are often processed to enhance readability and to reveal new information. In addition, these images can be distributed to many users simulateously, and with greater ease, than can be done with original manuscrpts. Most common imaging equipment is limited in sensitivity to the visible regime of the spectrum. However, additional information is often obtained from images taken at other wavelengths, specifically in the infrared region. The focus of this research is to determine whether some of the information that exists in the infrared range can be obtained by manipulating visible light images.

Table of Contents

Background

Ancient parchments have often been photographed to better analyze their contents. From black-and-white imaging to color to multispectral digital imaging, many steps have been made to improve the readability of the parchments. Because of the unstable nature of both the documents material, (usually animal skin or papyrus) and inks (specifically Iron Gall inks (1)) the parchments may be very difficult to read with the visible eye (2). Because of the deterioration of such parchments, the original information on the texts is no longer is visible to the naked eye or in traditional color photographs. It has been established that imaging in the infrared region of the spectrum (800nm $< \lambda <$ 1200nm) can provide greater visibility to the text written on the parchments. This is largely due to the fact that the materials that make up the parchment reflect a great deal of energy in the infrared range of the spectrum, while the ink is a general broadband absorber. Since the parchment reflects better in the IR than in the visible range, while the ink absorption remains constant, increased contrast is observed between the ink and parchment (3).

Since the reflectance and absorbance of the parchments vary across the spectrum, techniques that measure reflectance in many wavelengths are useful. Multispectral imaging (MSI) is a class of procedures that have proved useful in a number of fields, such as geology, astronomy, and remote sensing. The technique of MSI is relatively straightforward; an imaging spectrometer is used to capture the reflectance of various sections of a target over a range of wavelengths. When this series of readings is complete, the researcher can construct a reflectance spectrum for that specific section of the parchment. Since sections of the parchment will have unique reflectance spectra according to their material, these spectra can be used to identify useful sections of the document. For example, a section with ink on degraded parchment will have a different spectrum than an undamaged section or a degraded section with no text. Therefore, MSI can be used to locate sections of the parchment that contain ink, even if the sections are badly degraded or faded (2).

In the course of imaging studies of the Dead Sea Scrolls, Knox, et al., found that significantly enhanced contrast in a monochrome image generated one of the three channels of the "YES" color space defined by Xerox Corp (4). In the "YES" color space, the luminance channel, Y, is a weighted sum of the RGB values, while the chroma channels are spectral differences: The signal in the E channel is proportional to the difference of the red and green color channels, while the S color channel is proportional to yellow minus blue. The matrix is:

$$\begin{pmatrix} 0.253 & 0.684 & 0.065 \\ 0.5 & -0.5 & 0 \\ 0.25 & 0.25 & -0.5 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} Y \\ E \\ S \end{pmatrix}$$

Xerox Systems Institute, 1989

Figure 1: RGB - YES Conversion Matrix (5)

In consideration of the success in the "YES" color space, the intent of the research is to explore the possibility that manipulation of visible RGB color data can be an ample substitute for IR imagery. Should such a procedure be established and proven, existing RGB images of degraded parchments could be reanalyzed and additional information could be uncovered. In addition, the need for expensive infrared filters, cameras and film could be eliminated, allowing for more economical research to be performed.

Table of Contents

Methods

Spectral Analysis

The first step in the research was to measure the spectral reflectance of various sections of a document, which in this case was charred newspaper. Different sections included those which are text-free, both degraded and intact, and those which contain text. The intact text-free regions are analyzed in order to provide baseline for comparison. These measurements were obtained for a single document.

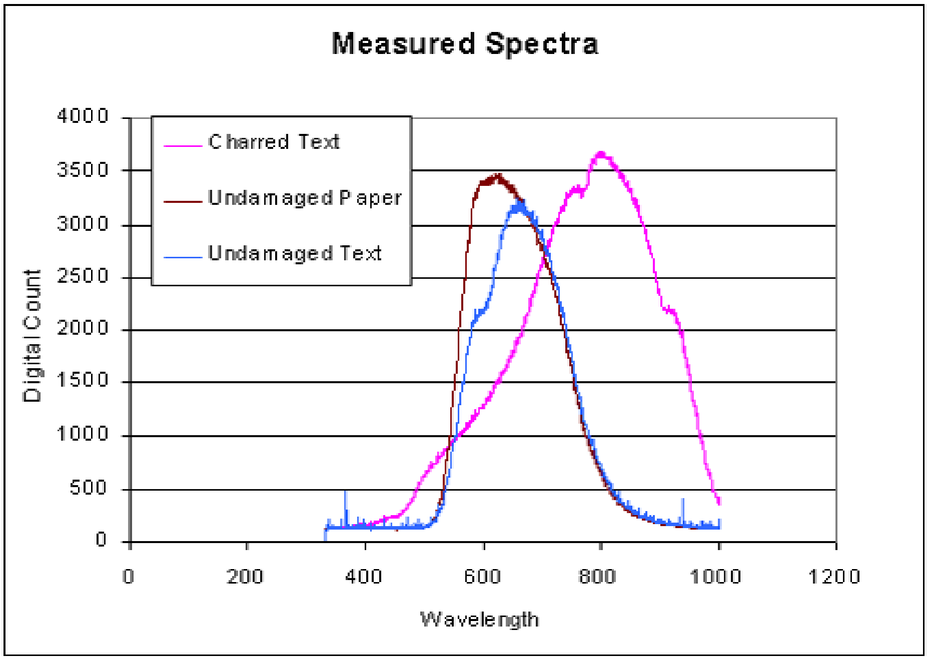


Figure 2: Reflectance Spectra of Various Regions of Charred Newspaper

As can be seen from the measurements, the different sections of the parchment have distinct reflectance spectra. The spectra were further analyzed to determine specific areas of interest. As can be seen in Figure 2, the spectra for the undamaged text and paper peak in the 600nm range.

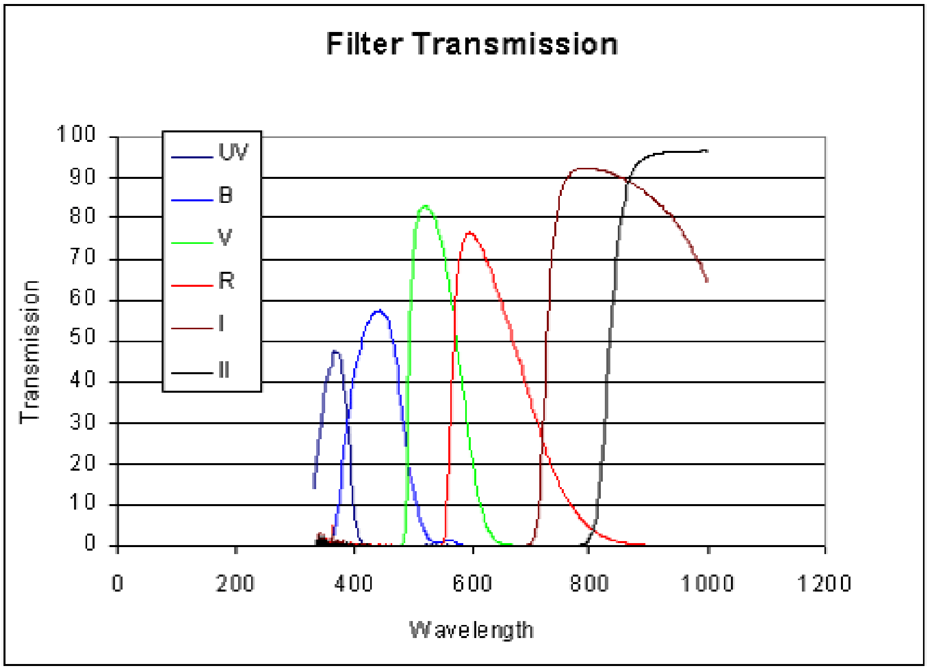


Figure 3: Transmission Spectra for Digital Camera Filters

The transmission spectra for the filters used with the digital camera were then analyzed. When compared to the spectra in Figure 2, it appeared that the paper peaked close to the peak of the green filter, and the text (ink) peak is shifted slightly towards longer wavelengths.

The spectra of the various regions were then calculated through each of the camera's filters.

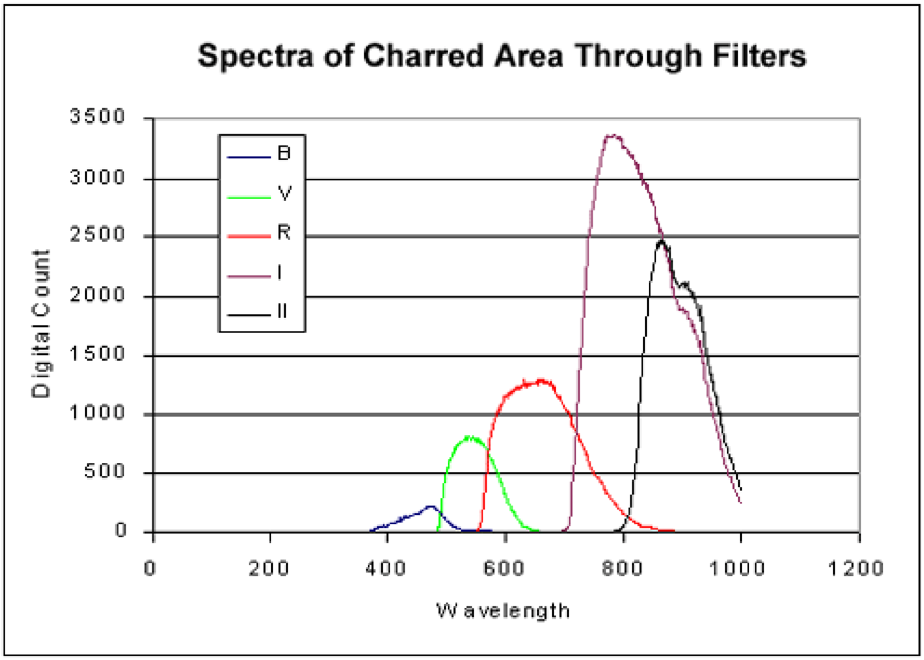


Figure 4: Reflectance Spectra of Charred Region through Camera Filters

As can be seen in Figure 4, there appears to be information in the red and green areas of the spectra of the charred section of the newspaper, even though the peak is shifted greatly into the IR region.

Image Aquisition & Manipulation

After completing the spectral analysis, images of the document were obtained using the Photometrics SenSys™ digital camera. Images were taken through clear, red, green, blue, and infrared filters.

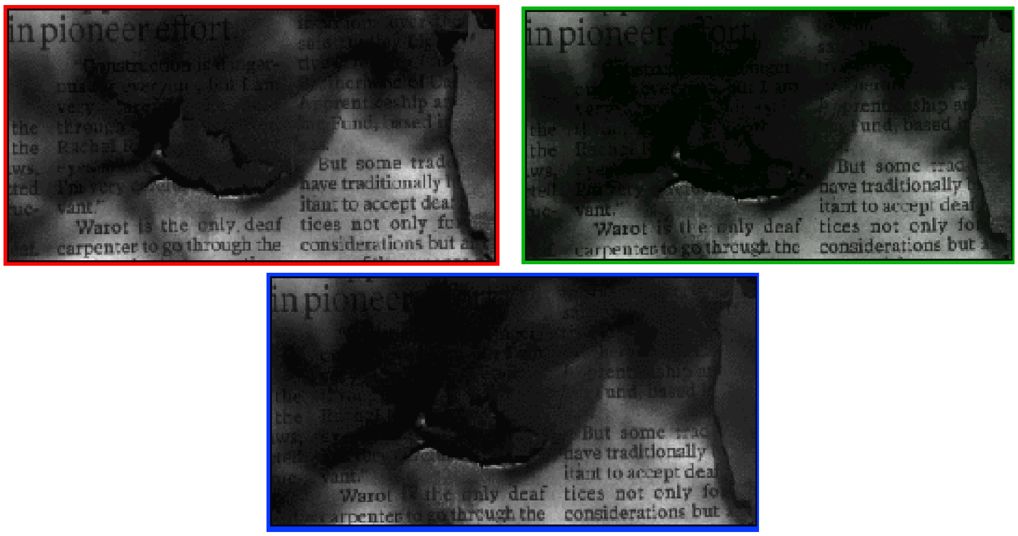


Figure 5: Newspaper Imaged With Red, Green, & Blue Filters

The red, green and blue images were combined in the ratio of 0.35, 0.55, and 0.1, respectively, to obtain a typical RGB image.

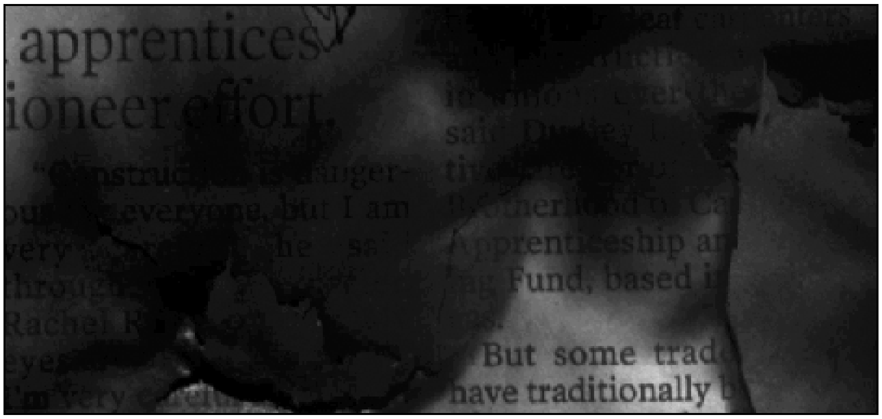


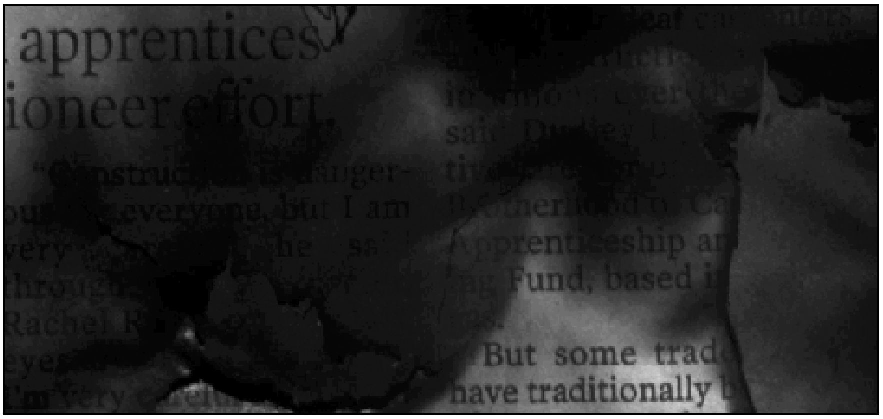
Figure 6: RGB Sum Image Simulating Monochrome Image

IDL®
code was then written to subtract the green image from the red image. Before image subtraction was performed, the images were slightly translated, as they were not exactly aligned. This was possibly due to the camera filters being tilted, or a slight shift in the camera's position due to movement involved with changing the filters. After image subtraction, the IDL histogram equalization function HIST_EQUAL was implemented to enhance the contrast of the images. The IDL function WRITE_JPEG was used to write the 8-bit images included in the following figures.

Table of Contents

Results

As expected, the image quality of the infrared image was significantly greater than that of the RGB image.



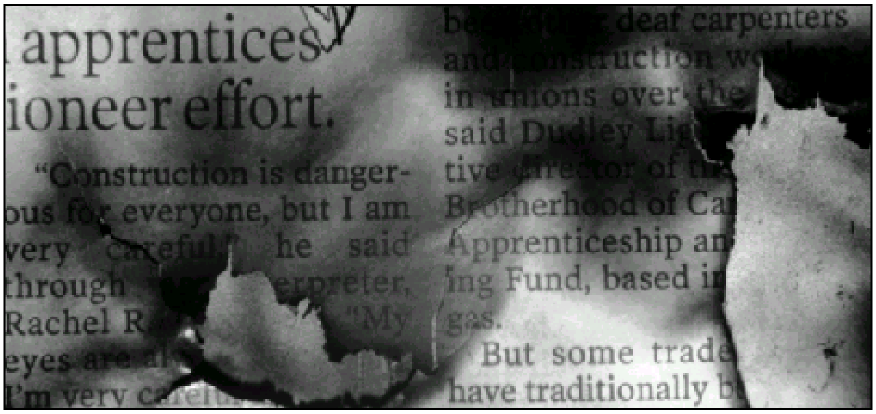


Figure 7: RGB Image and IR Image

Likewise, the Red - Green image also showed improvement compared to the RGB image.

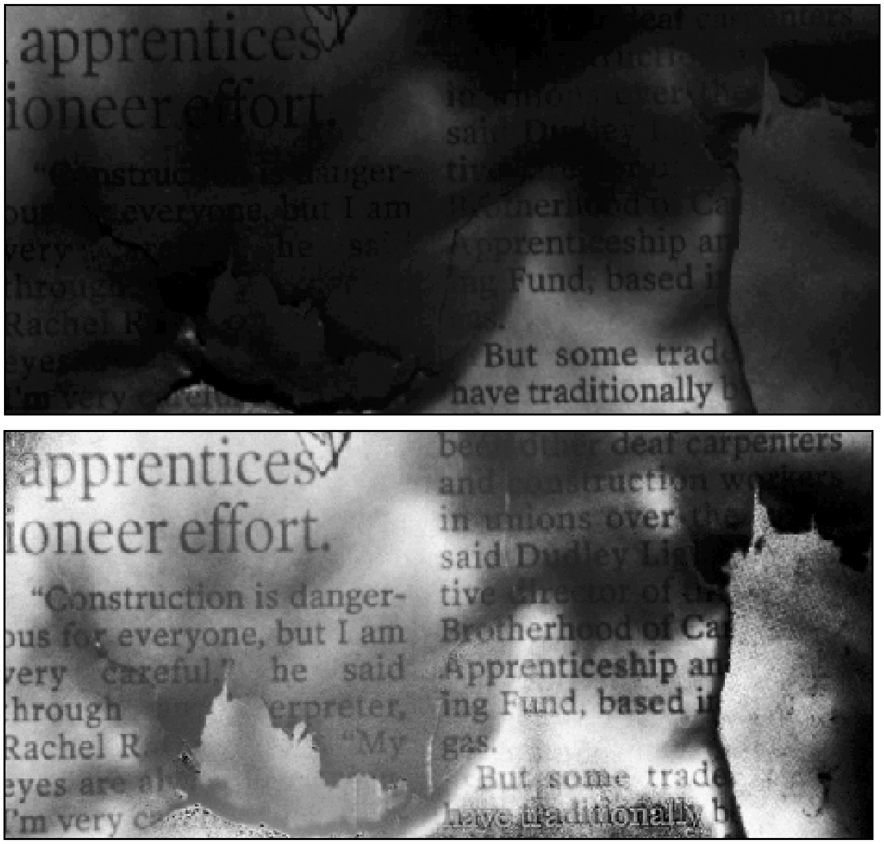


Figure 8: RGB Image and R-G Image

When compared directly to one another, the IR and the R-G images appear to have similar contrast, and much of the information recovered in the IR image is also apparent in the R-G image. In addition, certain areas of the images suggest that the subtracted image has displayed the text somewhat more clearly than the IR image.

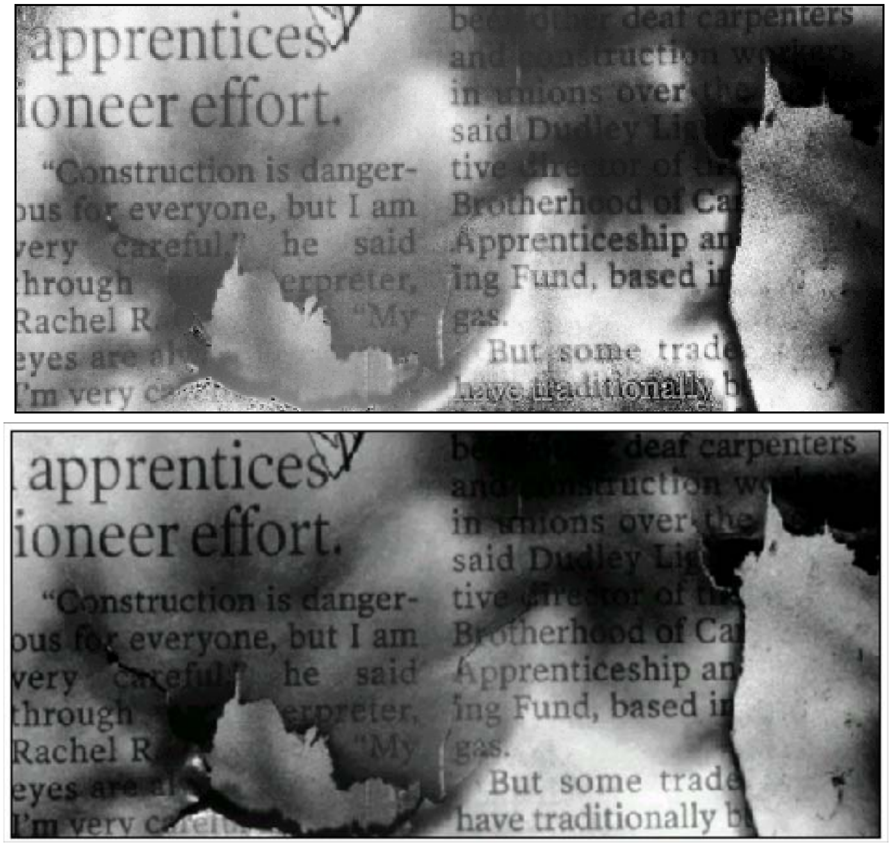


Figure 9: R-G Image and IR Image

Table of Contents

Discussion

Analysis of the IR and R-G images shows that image subtraction yields comparable contrast to an IR image. Information that was obscured by charring to the eye and to traditional RGB imaging was recovered adequately by both infrared imaging and visible image subtraction. However, although the same histogram equalization routine was run on all final images, including the RGB image, the question still arises as to how much of an effect that process has on the improvement of the images. The HIST_EQUAL function used to process the images equalized the images independantly. Thus the image histograms were not identically "stretched." Other histogram-stretching routines should be explored as well. Despite the inequivalent equalization, the evidence still supports the theory that image subtraction is a viable substitute for infrared imaging. Implementing the same technique on other damaged subtrates may also yeild similar results. Further research incorporating other subtrates and inks would prove to be a useful exercise.

Table of Contents

Conclusions

This research was intended to determine the plausibilty of using image subtraction within the visible range of the spectrum as a viable substitute for infrared imaging in the analysis of degraded documents. The difference of the red and green channels conveys similar information to that available in infrared images of degraded newspaper. The details of the text were clearly visible in both R-G and IR images, indicating that the technique of image subtraction may be a viable substitute for infrared imaging in similar circumstances.

Table of Contents References

Analysis of the Spectra of Degraded Documents

Ryan Swift

References

1. Elmer Eusman and Birgit Reißland, The Iron Gall Ink Corrosion Website, http://www.knaw.nl/ecpa/ink/.

2. G.H. Bearman and S.I. Spiro, "Archaeological Applications of Advanced Imaging Techniques," Biblical Archaeologist, 59:1: 56-66 (1996).

3. Peter Tyson, "Deciphering the Dead Sea Scrolls," Technology Review, Vol. 98: 22-23 (April 1995).

4. K.T. Knox, R.H. Johnston, and R.L. Easton, "Imaging the Dead Sea Scrolls," Optics and Photonics News, Vol. 8, No. 8: 31-34 (August 1997).

5. Xerox Systems Institute, Color Encoding Standard, Xerox Corp., Palo Alto, Calif.: p.3-1 (1989).

Use "Back" button to return to thesis. Table of Contents