Interactive Enhancement of Handwritten Text through Multi-resolution Gaussian

Oliver A. Nina
Department of Electrical Engineering and Computer Science
University of Central Florida
Florida, USA
onina@eecs.ucf.edu

***Abstract*—Handwritten text found in historical documents is often difficult to read due to issues such as contrast, noise and degradation. There has been much work on how to process such documents including improvements on binarization of these images. Despite the different advances in this area, improving the quality and readability of these documents is still an open research area.**

In this paper a novel approach is proposed to improve the text of historical documents through interactive stroke enhancement. This approach utilizes user interaction to indicate parts in the image where stroke enhancement is needed. The algorithm uses a difference of multi-resolution Gaussians to detect text at different scales and to modulate the amount of enhancement needed. This approach could be used for manually restoring text images or for improving readability of the text. Results are given in this paper that show the effectiveness of the proposed method.

***Keywords*-handwritting; enhancement; binarization; user interaction;**

I. INTRODUCTION

Handwritten text from historical documents has been a subject of study and research for many years. Reading text from historical documents through optical character recognition software or through human efforts can be cumbersome, especially in documents which contain significant noise and low contrast. There are different methods currently employed to improve the visibility of text in historical text documents. These methods include histogram equalization, bilateral filter, binarization and others.

In this paper a method to improve readability of historical documents through manually enhancing text strokes is proposed. Feedback from the user is recorded based on click and drag events to enhance parts of the text while simultaneously limiting the amount of noise in other parts of the image.

This method uses the mouse motion events to indicate where the strokes are. The strokes are constantly discovered as the mouse cursor revisits the areas that need more enhancing. Strokes emerge continuously through the difference of the original image to the continuous approximation of the background leveraged by a multi-resolution Gaussian filter. The proposed method assumes that the user is not an image processing expert.

II. BACKGROUND

One method to enhance images where the text is difficult to read due to its background is histogram equalization. Histogram equalization maps an input histogram to a desired output histogram by choosing the transformation T that minimizes :

$$|p_1((T(x)) - p_0(x)|$$

where p_0 is the probability density function (pdf) of the original image and p_1 is the pdf of the output histogram.

Histogram equalization often works for images which contain a distinct background and foreground. Nonetheless, it does not perform well when the background is non-uniform (i.e. noisy) as seen in figure 2.

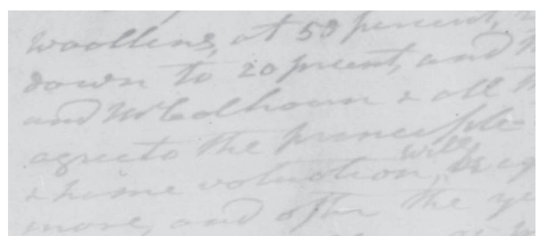


Figure 1: Original Image

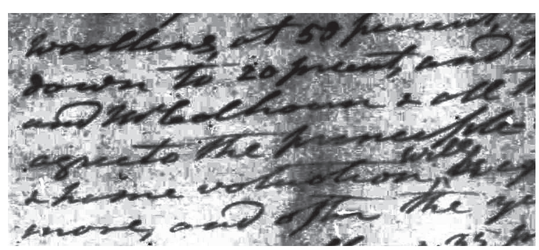


Figure 2: Histogram Equalized Image

There are other employable methods such as adaptive histogram equalization [1] that could perform better. However, choosing the correct local parameters for every image could be difficult and time-consuming.

Using a bilateral filter, such as the one described by Tomasi [2], is another way to improve readability of these

978-0-7695-4774-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICFHR.2012.222

types of text documents. A bilateral filter smooths an image while preserving strong edges. Additionally, a bilateral filter uses weighted spatial and range or intensity information within a window to smooth pixels in the center of the window relative to the rest of pixels in that region. A center pixel in the window is computed by the following bilateral function:

$$I_p' = \frac{\sum f(p-s)g(I_p - I_s)I_p}{\sum f(p-s)g(I_p - I_s)}$$

where I_p' is the new value for the center, f and g are the spatial and range Gaussian functions, I_p is the intensity of the center of the window, p and s are the coordinate positions of the center and other pixels in the window respectively.

For bilateral filtering to work, we need the standard deviations for the spatial and range functions which would be cumbersome to set for every image. In an ideal case, where the range and spatial parameters are known for every image, a bilateral filter does not provide any enhancement on the image but usually reduces local contrast because of the blurring involved. Figure 3 shows the results of using a bilateral filter on the original image on figure 1

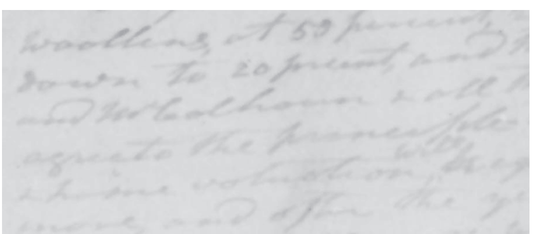


Figure 3: Bilateral Filtering (s= 3, r= 0.1)

Binarization is another way to separate the text from the background in these types of images. Recently, there has been great improvements on binarization methods due to the existence of competitions such as DIBCO [3] [4] and HDIBCO [5]. One of the methods that showed significant binarization results at DIBCO 2009 was Lu [6]. Later on, Su [7] at HDIBCO 2010 showed an augmentation of about 1 percent over Lu. Despite such improvements on binarization methods, there remains a strict classification between foreground and background with these methods which can make the text difficult to read as evidenced in figure 4.

In this paper, an innovative approach employing guided feedback from user interaction is proposed that enhances only the user-specified parts of the image. This method uses multi-resolution Gaussian approximation of background to enhance potential stroke pixels while discarding others.

This approach is not to be confused with binarization with local feedback. Binarization produces a binary image with two different set of values (i.e. 1, 0) where one of them defines the values for the foreground (1) and the other for the background (0). The result of the proposed method is a grayscale image (0-255) but with localized contrast enhancement based on the user interactive input. The next section describes this method.

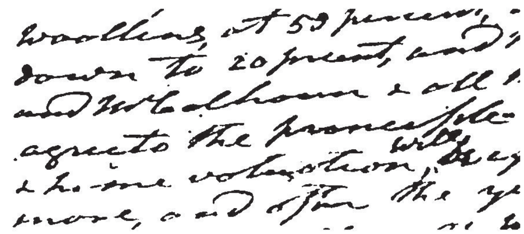


Figure 4: Lu algorithm

III. Method

The method proposed in this paper uses an interactive multi-resolution Gaussian approach to approximate the background of the image and using that approximation to find different edges or text strokes at different scales. In order to accomplish this, the following steps are performed:

A. Mouse Interaction Map

A user interaction map is created of the same size and resolution of the image in order to record the click and drag events that the user made. The more drag events recorded in a certain area, the higher the score for that area, which indicates the size of kernel to be used for the Gaussian filter.

B. Multi-resolution Difference of Gaussian

This concept is based on the multi-scale Difference of Gaussians (DOG) in Lowes scale invariant feature transform (SIFT) [2]. DOG is a faster way to calculate edges at different scales or resolution in the image. DOG performance is also comparable to that of the Laplacian of Gaussian (LOG) method as explained in [8]. DOG was used by Lowe to find matching keypoints of similar objects in two different images. In contrast with Lowes implementation, only the size of the Gaussian kernel is changed by a small kernel augmentation for the multi-resolution approach instead of the size of the image. Also, we subtract the original image from the Gaussians at every scale instead of performing Gaussian subtraction at every DOG level. Using the information from the interaction map, the kernel size for performing Gaussian filtering is determined for a certain area. At every pixel in the interaction map a different kernel size is obtained based on how many click and drag events the area received. By increasing the Gaussian kernel, more sampling area is attained which will help obtain a better approximation of the background and thus, better strokes.

C. Stroke Enhancement

The last kernel size recorded in the interaction map is kept as the approximated background to subtract from the original image. This process will leave a stroke map to enhance the strokes in the original image based on certain thresholds. Because there is noise around the text similar in intensity to that of the text pixels, it is necessary to only enhance those that belong to the text. In order to differentiate text pixels from background pixels, a stroke map is used that unveils stronger strokes as the scale increases. In order to improve visibility and avoid noise around the edges, the mean and standard deviation are calculated around the window being used. These values are used to calculate the amount of threshold to use to enhance a certain window area weighted by the stroke map. In this way, the edge map pixels are multiplied in the following way:

$$x' = ws\frac{\sigma}{h - \mu}$$

where x' is the enhanced intensity, s is the pixel in the stroke map, σ and μ are the standard deviation and mean of the area within the window, h is the highest intensity in the range (255) plus one (to avoid division by zero) and w is a weighting parameter determined by the stroke map at the same location. This enhancement function indicates that the new pixel intensity is proportional to the standard deviation of the area and inversely proportional to the distance between the mean and the highest range. Hence, if we are in an area where the pixels are similar and closer to the lowest intensity, the enhancement proportion will be low. If the opposite occurs the enhancement will be higher. The more passes the user does with the mouse on a certain area, the more enhancement will be performed.

Using the x' values for the window, we perform:

$$e = x' - x$$

to obtain the enhanced value e based on the original pixels x. Also, only positive values of e are considered, negative values are set to 0. Finally, at every mouse event, the enhanced image redraws only pixels where intensities are lower compared to the previous state of the image.

IV. Results

In this section the results of the proposed method are discussed. The results obtained by an implementation similar to the one described in Lu's paper [6] are compared with the proposed method. Although the original implementation of the Lu algorithm provided by the authors in their website [9] was used at first, such implementation failed to produce any results for figures 5 and 6. Moreover, Lu's implementation used in this paper also failed to produce any significant results on figure 5. However, it did produce results for figure 6. Figure 6 shows the original image with enhancements only for the words “the” and “and’. Figure 8 shows enhancements only in the strokes of the bottom line.

The results shown in this section were produced using human interaction such as click and drag events. Time spent in completing each task and ergonometric variables were not measured. However, the tasks the user performs with the proposed method are much simpler and faster than stroked enhancement in a pixel by pixel basis. A more extensive validation of the method could be done in a user study as explained in the future work section.

V. Conclusion

In this paper a novel method has been introduced that uses multi-resolution Gaussian approximation of the background to help enhance text images with low contrast or noisy backgrounds through user guided interaction. The initial results of the method as shown in this paper are promising.

An implementation of this method is available at code.google.com/p/enhancetext/

VI. Future Work

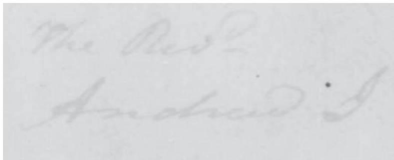
Although this method is promising, there is still more research to be done regarding the method. For instance, there are two parameters that still need to be set by the user in order for the algorithm to work. One of them is the window size which depends on the stroke width of the text being analyzed. An improvement for avoiding manually setting this parameter would be to automatically detect the stroke width in order to determine the window size. Another parameter that needs initial setup is the increment of the kernel size of the Gaussian on the interaction map. Currently, in cases where text size and resolution do not differ much between images, these two parameters need to be set only once before processing the images.

Also, speeding up the implementation with integral images for calculating the mean and standard deviation of the images would be another improvement that would make it more interactive. Although the method works in real time on images of small resolution, it slows down as the resolution increases because of the number of calculations needed with bigger windows. Other issues such as mouse input per image resolution need to be further addressed as well.

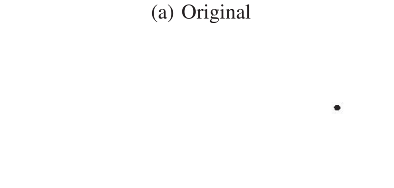
Finally, although the results shown in this paper seem successful, assessing the improvement of the text by a scoring system would be most helpful to better determine the extend of the success of this method. This could be done by a user study in which readability, easiness, ergonometric, time spent by the user and other variables could be measured to better validate the proposed approach.

Acknowledgment

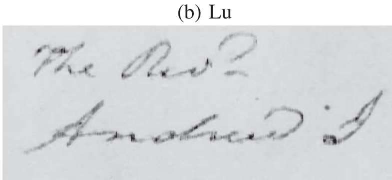
The author would like to thank Dr. William Barrett, Dr. Mubarak Shah, Intel Corporation and the National GEM consortium for their support and the reviewers for their comments.



(a) Original

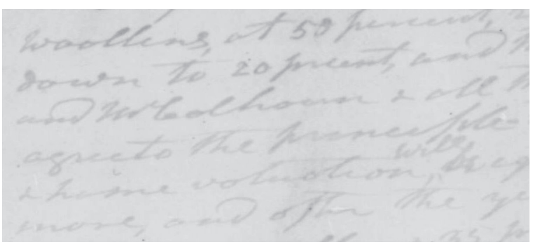


(b) Lu

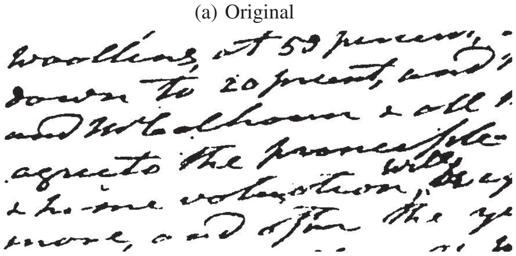


(c) Proposed

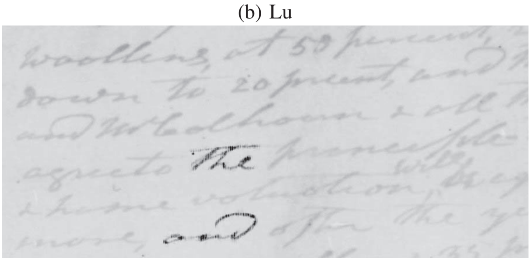
Figure 5: All characters enhanced



(a) Original

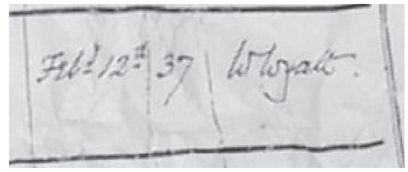


(b) Lu

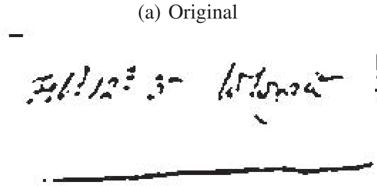


(c) Proposed

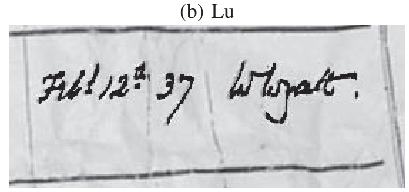
Figure 6: Only the words “the” and “and” were enhanced



(a) Original

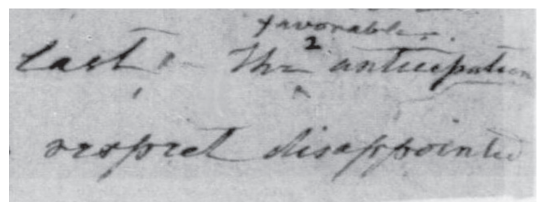


(b) Lu

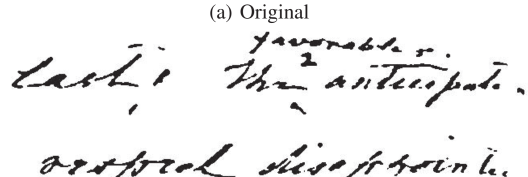


(c) Proposed

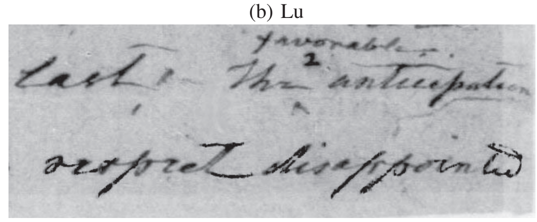
Figure 7: Only text characters were enhanced



(a) Original



(b) Lu



(c) Proposed

Figure 8: Only some of the characters in the bottom line were enhanced

References

[1] S. M. Pizer, "Adaptive histogram equalization and its variations," *Computer Vision, Graphics and Image Processing*, vol. 39, pp. 355–368, 1987.

[2] C. Tomasi, "Bilateral filtering for gray and color images," in *Proceedings of the 1998 IEEE International Conference on Computer Vision*, 1998.

[3] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "ICDAR 2009 document image binarization contest (DIBCO 2009)," in *Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on*, July 2009, pp. 1375 –1382.

[4] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "ICDAR 2011 document image binarization contest (DIBCO 2011)," in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, Sept. 2011, pp. 1506 –1510.

[5] ——, "H-DIBCO 2010 - handwritten document image binarization competition," in *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*, Nov. 2010, pp. 727 –732.

[6] S. Lu, S. Bolan, and C. L. Tan, "Document image binarization using background estimation and stroke edges," in *International Journal on Document Analysis and Recognition*, 2010.

[7] B. Su, S. Lu, and C. Tan, "Binarization of historical document images using the local maximum and minimum," in *International Workshop on Document Analysis Systems*, 2010.

[8] T. Lindeberg, "Scale-space theory: A basic tool for analyzing structures at different scales." *Journal of Applied Statistics*, pp. 224–270, 1994.

[9] "http://www.comp.nus.edu.sg/ subolan/bin/bin.html."