

# Robust Binarization of Degraded Document Images Using Heuristics

Jon Parker<sup>a,b\*</sup> and, Ophir Frieder<sup>a</sup>, Gideon Frieder<sup>a</sup>

<sup>a</sup> Dept. of Computer Science, Georgetown University, Washington DC, USA;

<sup>b</sup> Dept. of Emergency Medicine, Johns Hopkins University, Baltimore, MD, USA

## ABSTRACT

Historically significant documents are often discovered with defects that make them difficult to read and analyze. This fact is particularly troublesome if the defects prevent software from performing an automated analysis. Image enhancement methods are used to remove or minimize document defects, improve software performance, and generally make images more legible. We describe an automated, image enhancement method that is input page independent and requires no training data. The approach applies to color or greyscale images with hand written script, typewritten text, images, and mixtures thereof. We evaluated the image enhancement method against the test images provided by the 2011 Document Image Binarization Contest (DIBCO). Our method outperforms all 2011 DIBCO entrants in terms of average F1 measure – doing so with a significantly lower variance than top contest entrants. The capability of the proposed method is also illustrated using select images from a collection of historic documents stored at Yad Vashem Holocaust Memorial in Israel.

**Keywords:** readability enhancement, historic document processing, document degradation

## 1. INTRODUCTION

Affordable computer hardware and ubiquitous Internet access has made large scale digitization of document collections ever more worthwhile. The New York Review of Books reported in April 2013 that Google Books' database contained over 30 million volumes [1]. The IMPACT project [2] seeks to archive 18th century books to preserve European cultural heritage. Digitization efforts like these rely on techniques such as optical character recognition (OCR) and automatic data extraction. These techniques perform better when supplied with crisp black and white input images.

Unfortunately, original documents do not always fit that description. Some documents - particularly historic documents - can exhibit defects that inhibit OCR and data extraction. Consequently, converting a color image to a clean black and white image - i.e., binarizing the image - has become a well-studied task. Due to the complexity and importance of image binarization, a Document Image Binarization contest (DIBCO) was held during the International Conference on Document Analysis and Recognition (ICDAR) in 2009 and 2011. The 2009 contest compared 43 different binarization algorithms while the 2011 contest compared 18 algorithms. The contest summary papers [3, 4] provide short descriptions of the submitted algorithms as well as a description of the results.

We present a novel method to enhance and binarize a color image. The image enhancement method is designed to minimize the effect of defects in the original image. The method is novel because it:

- Has better average performance than other known methods;
- Has lower variance than other known methods – and thus more consistently provides better results;
- Requires no parameter setting or other human interaction. This technique can, therefore, be integrated into a high throughput workflow;
- Requires no training data. Consequently, this technique can be applied to any image corpus and parallelized easily due to its input page independence.

We evaluated our method by applying it to the set of test images provided by the 2011 Document Image Binarization Contest. We also demonstrated our method on a selection of images from a collection of historical document images.

---

\* Further author information: (Send correspondence to Jon Parker)  
E-mails: {jon, ophir, gideon}@cs.georgetown.edu

## 2. RELATED WORK

In 1979 Otsu [5] proposed an elegantly simple - and surprisingly effective - binarization method based on optimal clustering. This early method is sometimes referred to as a global thresholding method because it computes one threshold for the entire image for which darker pixels are set to black and lighter pixels are set to white. Niblack [6] proposed a thresholding method that considered local pixels and consequently enabled the threshold to change for different parts of the image. This method generates results that are significantly better than Otsu's method on some images but notably worse on other images. Nevertheless, the insight behind Niblack's method was commendable, and Sauvola [7] proposed an extension that overcame the main problem with Niblack's method (performance when no text was within the window of local pixels).

Beyond these early binarization methods [5-7] significant research was done on document image processing. Gatos, et al. [8] improved upon the adaptive thresholding in [7]. Multiple recent methods are based on Markov random fields [9-11]. One of these methods [9] was the winner of the 2011 DIBCO competition. Work also focused on foreground-background separation [12] as well as background estimation [13]. The method from [13] placed 2nd in the 2011 DIBCO competition. There are machine learning based methods such as [14] and [15] as well as works that examine processing images of "document-like" images. For example, [16] extracts information from images of engraved cemetery headstones, [17] enhances pictures of whiteboards, and [18] automatically extracts data from census tables.

The image enhancement and binarization method presented here is a descendant of - and significant improvement upon - an earlier method by Parker, et. al [19]. Both methods are based on the guiding principles that: (1) writing should be darker than nearby non-writing and (2) writing should generate a detectable edge. The algorithm introduced in [19] relies on parameters that are automatically set using heuristics. The parameter setting process has two notable downsides. First, the process is time consuming because it is governed by statistics that can only be computed when multiple fully rendered images are available for comparison. In other words, the parameter setting process merely crowns one output image as the (likely) best image from a collection of output images. The second problem is that the heuristics guiding parameter selection have not been well studied.

The improved method presented herein renders the entire parameter setting process unnecessary because it more robustly identifies pixels that are "near" an edge. This revised method also adds two post processing image cleaning steps. The first cleaning removed stray pixels and the second cleaning removes "white islands" from output images. White islands are a specific type of undesirable artifact that can appear in images produced by the prior method. They are discussed in detail in Section 3.5.2 where a process that removes them is introduced

## 3. METHODOLOGY

The image enhancement method described herein converts a color document image to a strictly black and white document image. The method highlights prominent aspects of an image while reducing the effect of degradation found in the original color image. This method is based on these two guiding principles:

1. "Writing" pixels should be darker than nearby "non-writing" pixels.
2. "Writing" should generate a detectable edge.

This improved method is summarized in Figure 1; each of the intermediate steps is discussed in detail in the section below.

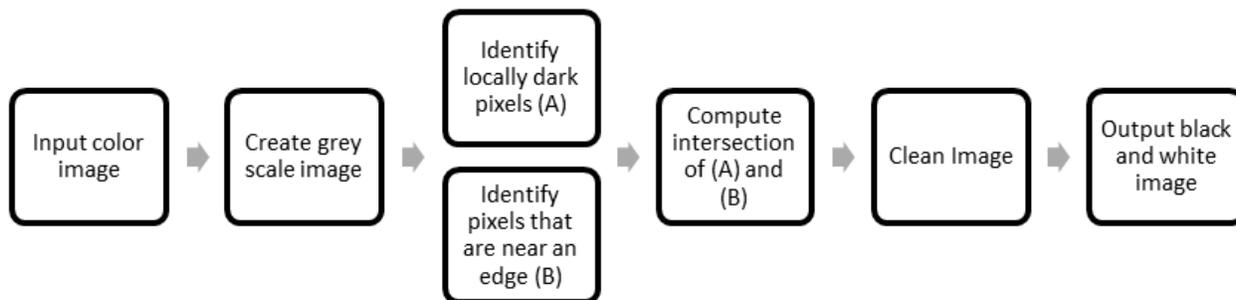


Figure 1. A high level view of the image enhancement method.

### 3.1 Creating a Greyscale Image

The image enhancement process begins by creating a greyscale image from the input color image. We use principal component analysis (PCA) to convert the input color image (which can be viewed as collection of 3-dimensional RGB values) to a greyscale image (which can be viewed as collection of 1-dimensional greyscale values). This step can be skipped if the input image is already in greyscale. However, applying PCA to a greyscale image does not alter the greyscale image.

### 3.2 Identifying Locally Dark Pixels

Identifying locally dark pixels is the first of two steps using the greyscale image as input. Each pixel in the greyscale image is analyzed separately by applying Otsu's method to a snippet of local pixels extracted from the greyscale image (similar to the method from [20]). A pixel is added to the set of "locally dark" pixels if the pixel is set to black when Otsu's method is applied to its corresponding snippet of local pixels. If a pixel is to be black in the final output image it must be flagged as "locally dark" in this step (excluding two corner cases discussed in Section 3.5). This requirement means that the output of this step (Figure 2, right panel) can be viewed as a filter each pixel must pass to be included in the final output.

Figure 2 illustrates a typical output of this step. As shown, this filter generates a sharp outline around text found in the input image. However, this filter is highly ineffective when applied to the background portion of the image. When the filter is applied to the background it merely highlights noise from the greyscale image (similar to Niblack's method). Section 3.3 introduces another filter that, when intersected with this filter, clarifies the background portion of the image.

The snippets of local pixels are created by extracting an  $n$  pixel by  $n$  pixel region where the region is centered on the pixel being analyzed. We require  $n$  to be odd so that there is always a single pixel at the exact center of the  $n$  by  $n$  region. The results illustrated were obtained when  $n$  was set to 21.

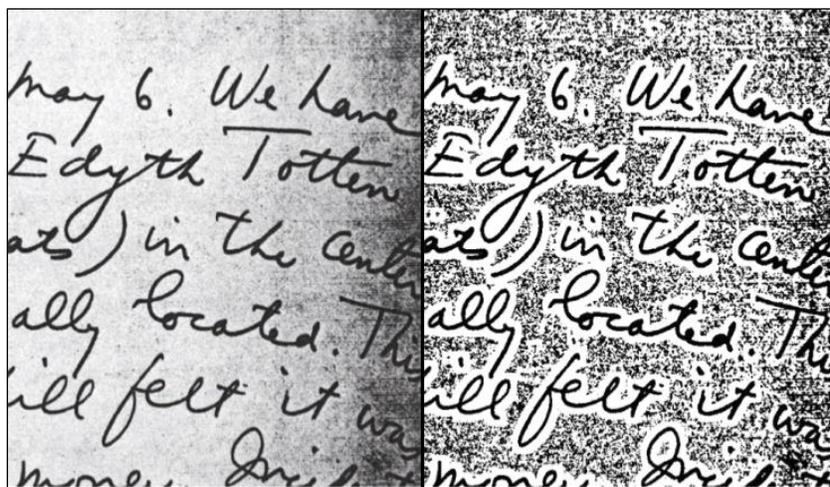


Figure 2. A typical input (left) and output (right) pair for identifying locally dark pixels using a window size of  $n = 21$ .

### 3.3 Identifying Pixels that are Near an Edge

Next, we identify pixels that are near an edge. This is the second of two steps using the greyscale image as input. This step executes the process shown below in Figure 3.



Figure 3. The process used to identify pixels that are near an edge.

This process begins by running Sobel edge detection on the greyscale image. The image produced by the Sobel operator (Figure 4, left panel) is then gently smoothed using a bilateral filter [21] that reduces the effect of image noise. Next, we compute the standard deviation of the greyscale values found in each snippet of local pixels extracted from the smoothed edge detection image. These standard deviations are then normalized so that they range from 0 to 255. The result is treated as a greyscale image (Figure 4, center panel) upon which Otsu’s method can be applied. Applying Otsu’s method to the image of local color standard deviations identifies regions of an image that are near an edge. A typical result is shown in the last panel of Figure 4.

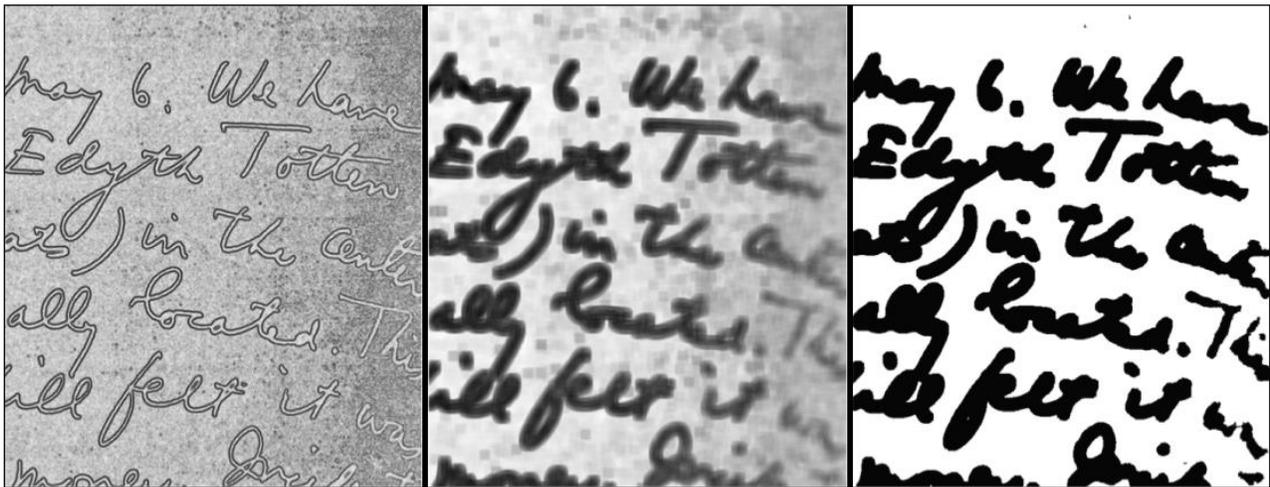


Figure 4. Typical results from applying edge detection (left), computing standard deviation of local color greyscale values with  $n = 15$  (center), and applying Otsu’s method (right). The result obtained from applying the bilateral filter is not shown.

### 3.4 Computing the Intersection

The results obtained when identifying locally dark pixels and pixels that are near an edge can be viewed as filters based on one of the guiding principles from the beginning of Section 3. The intersection of these filters generates a strikingly good black and white version of the input image – see Figure 5 – provided the snippets used to compute the standard deviations of local greyscale values are smaller than the snippets used to identify locally dark pixels. If this condition is not met the intersection will contain a noticeable artifact. The artifact is caused because the black regions produced when identifying pixels that are near an edge are too big with respect to the white outlines produced when identifying locally dark pixels. When this occurs a small portion of the noisy region from the output of Section 3.2 is seen in the final intersection. The artifact generates a speckled halo effect around all text. To prevent this artifact the snippet size  $n$  is set to 15 in Section 3.3. Empirical results show that keeping snippet sizes in a ratio 3:4 produce good results.

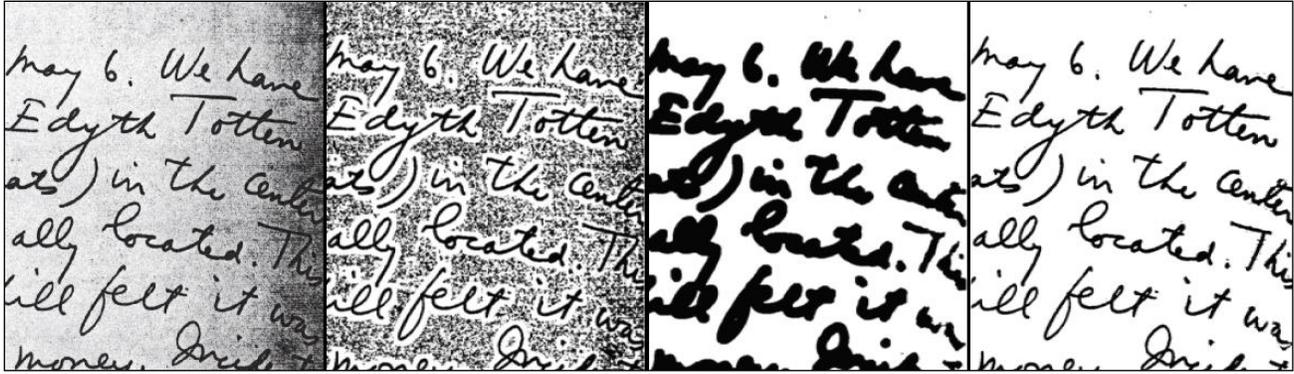


Figure 5. The input image (far left) and the output intersection (far right) of the locally dark pixels identified in Section 3.2 (left-center) and the pixels that are near an edge identified in Section 3.3 (right-center).

### 3.5 Cleaning the Image

The intersection obtained in Section 3.4, generally speaking, is a good result. Yet, two image improving corrections can be made. The first correction fixes what are likely to be erroneously classified pixels. The second correction fixes what are likely to be erroneously classified white regions.

#### 3.5.1 Stray pixel correction

The first cleaning step looks for stray pixels in the intersection computed in Section 3.4. The presumption underlying this step is that a black (or white) pixel does not typically appear by itself or nearly by itself. To remove stray pixels, we examine each pixel along with its eight neighbors. When the pixel at the center of a group of 9 pixels is outnumbered by pixels of the opposite color by 1-to-8 or 2-to-7, it is flipped to the locally dominant color. We do not flip pixels that are outnumbered by 3-to-6 because doing so would destroy a fine line of pixels (see the bottom right panel of Figure 6). Figure 6 illustrates 4 examples in which the center pixel would be corrected and 2 examples in which the center pixel would be left unchanged.

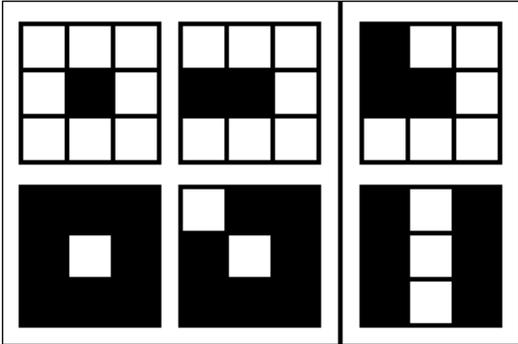


Figure 6. Example pixels arrangements that are changed (columns 1 and 2) and not changed (column 3).

#### 3.5.2 White island correction

The method as described from Section 3.1 to Section 3.4 would not correctly handle large regions that should be classified black in the output image. The problem occurs because the pixels in the center of a large black region may not be included in the set of pixels that are “near an edge”. A document with unusually large font may exhibit this error as shown in Figure 7. We refer to these incorrectly classified regions as “white islands”.

Correcting white islands begins by finding contiguous regions of white that are surrounded by a single contiguous region of black – i.e., a black boarder. When a (white island, black board) pair is identified in the intersection computed in Section 3.4 we must refer back to the greyscale image produced in Section 3.1 to determine if a correction should be

made. If a correction is indicated then all the pixels in the white island are set to black thus “plugging a hole” in the black boarder.

We assume that a correctly classified white island will have pixels with a statistically different mean greyscale value than the pixels within the black boarder. Consequently, to determine if a white island was incorrectly classified we perform a two sample z-test to see if pixels corresponding to those in the white island (but selected from the greyscale image) are statistically different from pixels corresponding to those from the black boarder (but selected from the greyscale image). If the pixels found in these regions are not statistically different, we flip the color of all the pixels in the white island to be black.

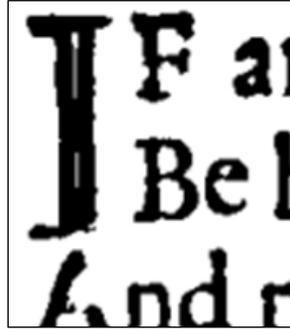


Figure 7. This letter exhibits 2 white islands because the pixels at the center of the letter are not identified as pixels that are “near an edge”

### 3.6 Method Summary

There are four aspects of the method described above that are worth emphasizing explicitly. First and foremost, this method requires no training data; therefore, this method can be applied to any image dataset. Second, this method is trivial to parallelize because: every sub-step is itself parallelizable, and the method as a whole is input page independent. Third, this method is simple to implement. There are no EM calculations to perform nor are there any other complex operations. Finally, this method requires no human (or non-human) interaction to set parameters. Thus, this method can easily be incorporated into a high throughput image processing/preprocessing workflow.

## 4. EXPERIMENTAL RESULTS

Two document image datasets are used to evaluate the proposed method. This first image dataset comes from the 2011 a Document Image Binarization Contest (DIBCO) that was held at the International Conference on Document Analysis and Recognition (ICDAR). The images within the DIBCO corpus were selected because they reflect various types of document degradation including bleed through, blotching, and faint script. The DIBCO corpus contains 8 images of hand written script as well as 8 images of printed script. It is available online at: <http://utopia.duth.gr/~ipratika/DIBCO2011/>. Importantly, this corpus permits a methodical evaluation of an image binarization algorithm because the corpus contains one hand-created strictly black and white “ground truth” image for each of the 16 test images within the collection. The second dataset contains scans of historic documents that are currently stored at Yad Vashem Holocaust Memorial in Israel. The input-output results shown from this corpus are selected both to illustrate the versatility of the proposed method as well as to illustrate the variety of test cases available in this corpus. Pages from this corpus contain typewritten text, handwritten script, pictures, multiple languages, and combinations of these.

### 4.1 DIBCO Results

As discussed above, each of the 16 DIBCO test images comes with a corresponding hand-created black and white ground truth image. These ground truth images permit each output image to be compared against the “ideal result”. One metric used to judge the DIBCO competition is F1 measure. Figure 8 and Table 1 shows how our proposed method compares to its predecessor [19] as well as the 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> place methods [9, 13, and 22] (out of 18) from the 2011 DIBCO competition.

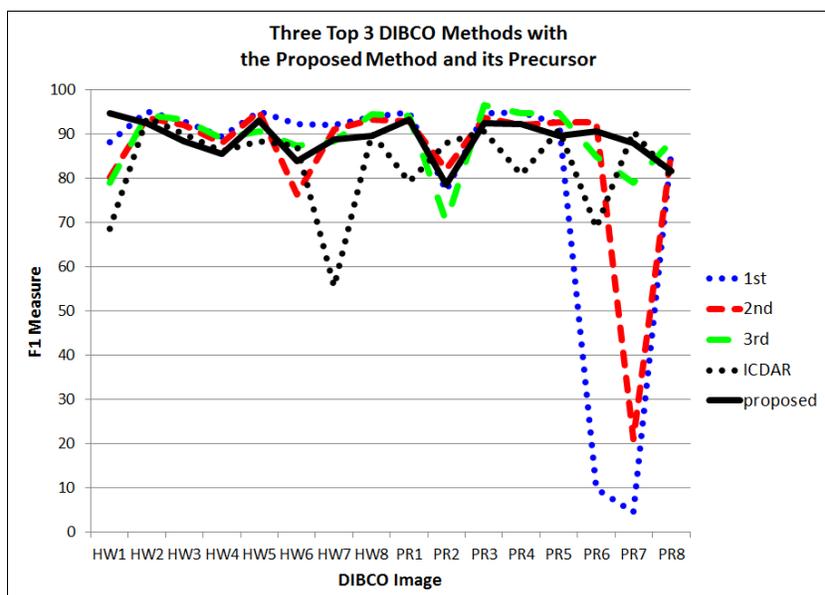


Figure 8. This chart shows the F1 measure each of top 3 methods from the 2011 DIBCO competition as well as the proposed method and its predecessor achieved on the 16 images in the 2011 DIBCO dataset.

Table 1. This table contains summary statistics for the F1 measures each of top 3 methods from the 2011 DIBCO competition as well as the proposed method and its predecessor achieved.

	1 <sup>st</sup> place 2011 DIBCO	2 <sup>nd</sup> place 2011 DIBCO	3 <sup>rd</sup> place 2011 DIBCO	Predecessor	Proposed Method
Mean F1	80.9	85.2	88.7	83.1	88.9
Median F1	92.3	92.1	90.6	88.0	89.7
Variance of F1	794.9	302.5	49.8	104.5	19.2

As shown in Table 1, the proposed method has a higher mean F1 measure than any method entered in the competition. The proposed method also has a significantly lower variance than any of the top 3 methods. The dramatic difference in variance is due to the top 2 methods failing catastrophically on image PR6, or both PR6 and PR7. It is worth noting that these two methods also had a slightly higher median F1 measure than the proposed method. Figure 9-12 show examples of DIBCO test images and the corresponding black and white output image.

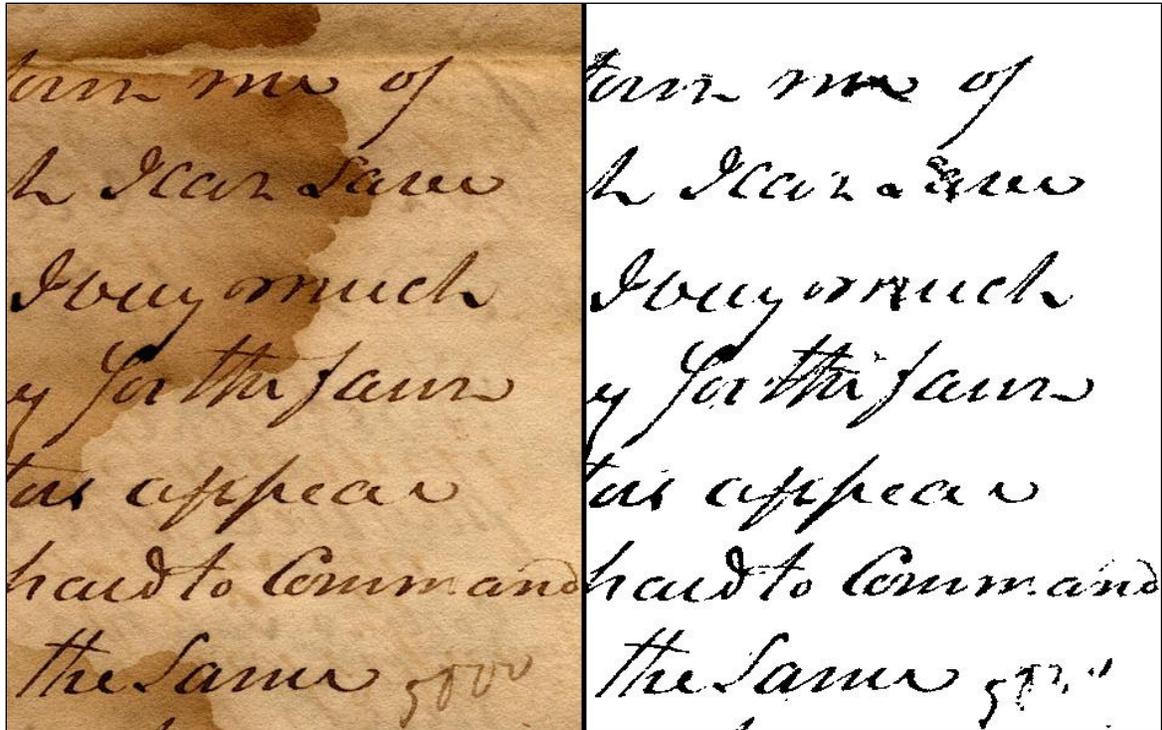


Figure 9. DIBCO image HW4 and its corresponding black and white image.

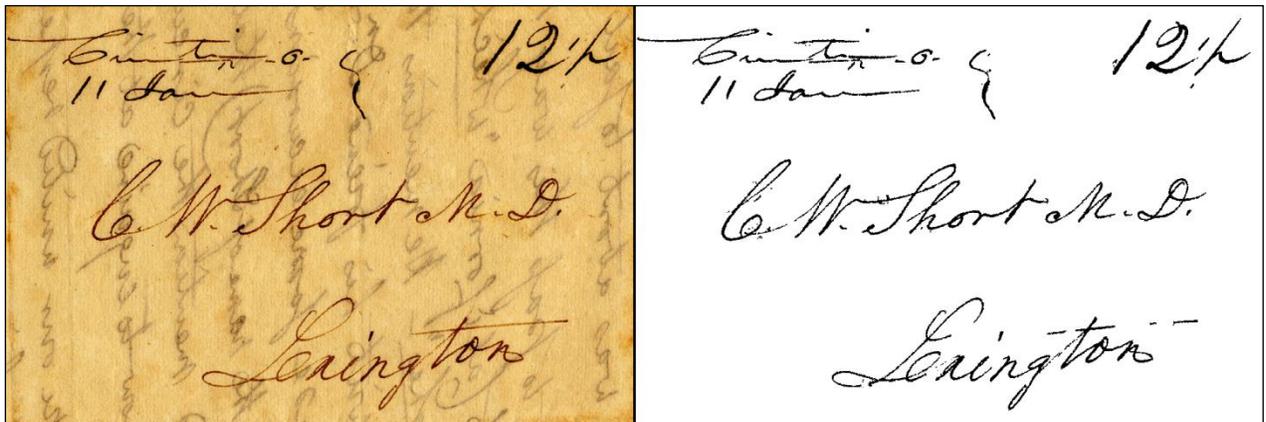


Figure 10. DIBCO image HW7 and its corresponding enhanced image.

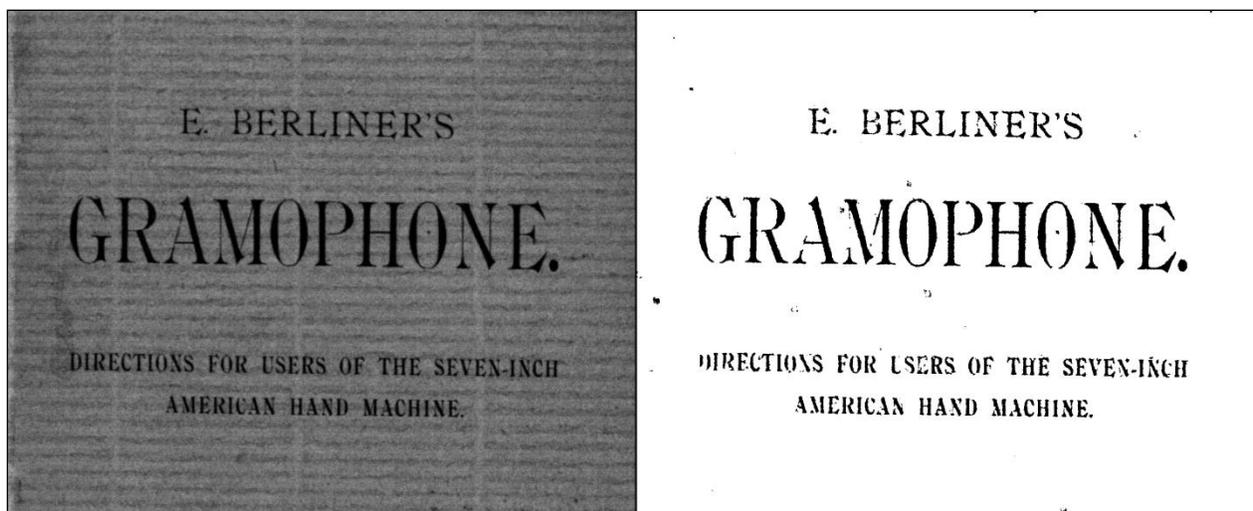


Figure 11. DIBCO image PR6 and its corresponding enhanced image.

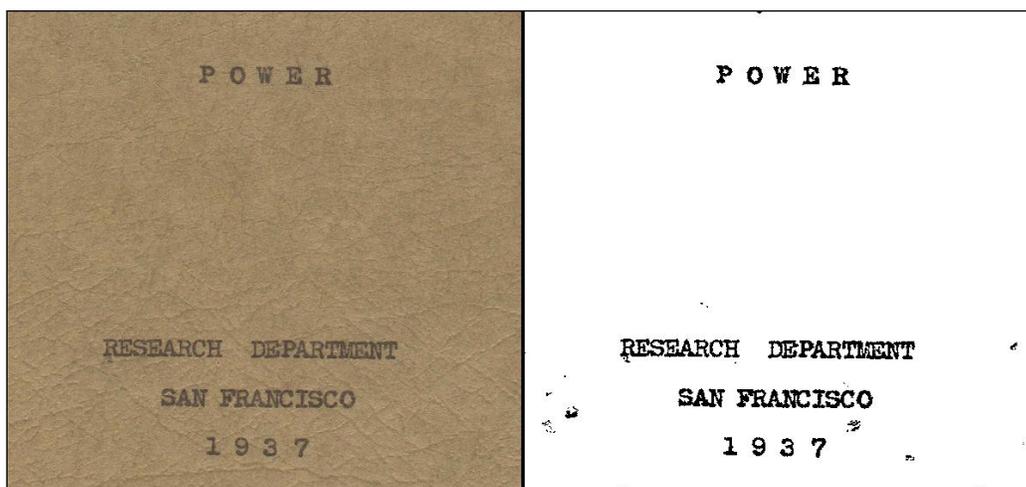


Figure 12. DIBCO image PR7 and its corresponding enhanced image.

## 4.2 Examples From the Frieder Dairies

The Yad Vashem Armin Frieder Diaries image corpus contains approximately 900 high resolution scans of historically significant documents [23]. Most of the originals were authored from the late 1930s to the middle of the 1940s. Due to a variety of causes some of the original documents were better preserved than others. Figures 13 and 14 show how the proposed method performed when it was applied to 4 images from this corpus. The pair of pages in Figure 13 show different types defects. The excerpt on the left side of Figure 13 has a two-tone effect while the excerpt on the right side contains bleed-through (where the typewritten script from the reverse side is visible). The pages shown in Figure 14 are both difficult to read to the naked eye. The blotching in the left-hand excerpt and the faint text on the right-hand excerpt render the script difficult to understand. Both Figure 13 and 14 contain enhanced version of the diary excerpts. The enhanced versions are noticeably clearer. It is worth noting that the enhanced version of M.5\_193\_95 (Fig. 14, left side) has no blotching even though the dotted i's, commas, and accent marks were retained.

Slávny Okresný Úrad !

**OKRESNÝ ÚRAD**  
**OVÉ MESTO NOVÉ MĚSTO**  
 číslo: 28 J. 1940

Podpísaná Miestna skupina cionistického svä-  
 zu pre Slovensko v Novom Meste nad Váhom zastúpená Arminom Frie-  
 derom hlavným rabinom ako predsedom a Ing. Deziderom Koložvárim  
 ako tajomníkom týmto úctive žiadame o povolenie založenia uvede-  
 ného spolku, poťažne povolenie účinkovania .

Túto našu žiadosť podporujeme nasledovnými  
 deklarámi :

1./ Dľa pod 1./ pripojených stanov Ústredného sväzu cionistické-  
 ho pre Slovensko, je tento sväz ako taký povolený a stanovky jeho  
 sú schválené.

92

-yc-

**NEKUDAH MUKAČEVO**

BANKVERBINDUNG: JÜDISCHE KREDITGENOSSENSCHAFT MUKAČEVO

BUCHDRUCKEREI UND VERLAGSGENOSSENSCHAFT TELEFON No. 183 TELGR.: NEKUDAH MUKAČEVO

ABTEILUNG: VERLAG Herrn

Oberrabbiner Armin Frieder  
 Nové Město n/V

17. Juni 1938 G./D 21. Juni 1938

Wir danken Ihnen sehr für Ihre ausführliche Informationen.  
 Wir möchten folgendes resumieren:  
 Wir nehmen gerne zur Kenntnis, dass Sie die innerliche und

Slávny Okresný Úrad !

**OKRESNÝ ÚRAD**  
**OVÉ MESTO NOVÉ MĚSTO**  
 číslo: 28 J. 1940

Podpísaná Miestna skupina cionistického svä-  
 zu pre Slovensko v Novom Meste nad Váhom zastúpená Arminom Frie-  
 derom hlavným rabinom ako predsedom a Ing. Deziderom Koložvárim  
 ako tajomníkom týmto úctive žiadame o povolenie založenia uvede-  
 ného spolku, poťažne povolenie účinkovania .

Túto našu žiadosť podporujeme nasledovnými  
 deklarámi :

1./ Dľa pod 1./ pripojených stanov Ústredného sväzu cionistické-  
 ho pre Slovensko, je tento sväz ako taký povolený a stanovky jeho  
 sú schválené.

92

-yc-

**NEKUDAH MUKAČEVO**

BANKVERBINDUNG: JÜDISCHE KREDITGENOSSENSCHAFT MUKAČEVO

BUCHDRUCKEREI UND VERLAGSGENOSSENSCHAFT TELEFON No. 183 TELGR.: NEKUDAH MUKAČEVO

ABTEILUNG: VERLAG Herrn

Oberrabbiner Armin Frieder  
 Nové Město n/V

17. Juni 1938 G./D 21. Juni 1938

Wir danken Ihnen sehr für Ihre ausführliche Informationen.  
 Wir möchten folgendes resumieren:  
 Wir nehmen gerne zur Kenntnis, dass Sie die innerliche und

Figure 13. Excerpts from diary images M.5\_192\_61 and M.5\_192\_92: each excerpt is paired with its enhanced version.

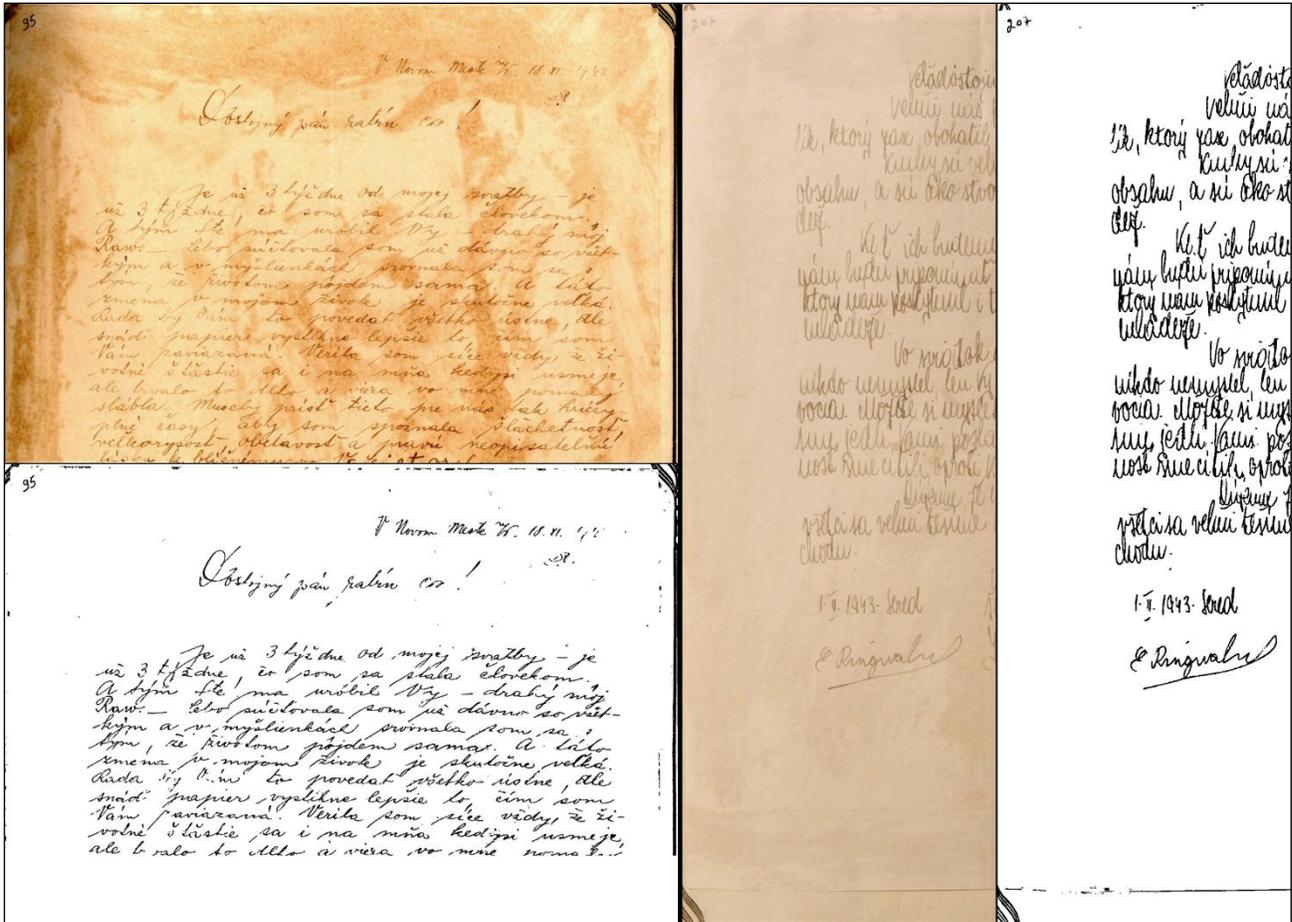


Figure 14. Excerpts from diary images M.5\_193\_95 and M.5\_193\_207: each excerpt is paired with its enhanced version.

## 5. CONCLUSION

The image enhancement and binarization method presented herein was evaluated by applying the method to the test corpus distributed as part of the 2011 Document Image Binarization contest. The proposed method has a higher average performance than any entrant into the 2011 Document Image Binarization contest. Moreover, it has a significantly lower variance than all of the top entrants into the competition. Consequently, the proposed method returns high quality results more consistently than other image binarization methods.

The proposed method was also applied to select images of pages found in the Yad Vashem's Armin Frieder Diaries – a real world corpus of historically significant documents with corresponding images. The diaries themselves are stored at Yad Vashem Holocaust Memorial in Israel. When the proposed method was applied to the diary images containing a variety of defects the results showed no sign of the defects that occluded the original document.

The proposed method has other benefits in addition to its robust high quality output. The method is input page independent. This input independence means the method can be applied without training data, incorporated into automated high throughput (i.e., no human interaction) workflows, and parallelized with ease. An ancillary benefit of this algorithm is that it simple to implement and intuitively understood.

## REFERENCES

1. <http://www.nybooks.com/articles/archives/2013/apr/25/national-digital-public-library-launched/?page=1>
2. de Does, Jesse, and Katrien Depuydt. "Lexicon-supported OCR of eighteenth century Dutch books: a case study." *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013.

3. B. Gatos, K. Ntirogiannis, I. Pratikakis, "ICDAR 2009 Document Image Binarization Contest (DIBCO 2009)," *Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on*, pp.1375-1382, 26-29 July 2009
4. I. Pratikakis, B. Gatos, K. Ntirogiannis, "ICDAR 2011 Document Image Binarization Contest (DIBCO 2011)," *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pp.1506-1510, 18-21 Sept. 2011
5. Otsu, N., "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics* (1979).
6. Niblack, W., [An Introduction to digital image processing], Prentice Hall, Englewood Cliffs, NJ, USA (1986).
7. Sauvola, J. and Pietikainen, M., "Adaptive document image binarization," *Pattern Recognition* (2000)
8. Gatos, B., Pratikakis, I., and Perantonis, S. J., "Adaptive degraded document image binarisation," *Pattern Recognition* 39(3), 317–327 (2006).
9. Lelore, T. and Bouchara, F., "Document image binarisation using markov field model," *Proc. IEEE 10<sup>th</sup> ICDAR*, 551–555 (July 2009).
10. Lettner, M. and Sablatnig, R., "Spatial and spectral based segmentation of text in multispectral images of ancient documents," *Proc. IEEE 10th ICDAR*, 813 –817 (July 2009).
11. Kuk, J. G. and Cho, N. I., "Feature based binarization of document images degraded by uneven light condition," *Proc. IEEE 10th ICDAR*, 748–752 (July 2009).
12. G. Agam, G. Bal, G. Frieder, O. Frieder, "Degraded document image enhancement", *Proc. SPIE 6500*, pp. C1–11, 2007
13. Lu, S., Su, B., & Tan, C. L. (2010). Document image binarization using background estimation and stroke edges. *International Journal on Document Analysis and Recognition (IJ DAR)*, 13(4), 303-314.
14. Obafemi-Ajayi, T., Agam, G., and Frieder, O., "Ensemble lut classification for degraded document enhancement," in [Document Recognition and Retrieval XV], Yanikoglu, B. and Berkner, K., eds., *Proc. SPIE 6815*, 681509 – 681509–9 (2008)
15. Esser, D., Schuster, D., Muthmann, K., Berger, M., & Schill, A. (2012, January). Automatic indexing of scanned documents: a layout-based approach. In *IS&T/SPIE Electronic Imaging* (pp. 82970H-82970H). International Society for Optics and Photonics.
16. Christiansen, C. S., & Barrett, W. A. (2013, February). Data acquisition from cemetery headstones. In *IS&T/SPIE Electronic Imaging* (pp. 86580I-86580I). International Society for Optics and Photonics.
17. He, Y., Sun, J., Naoi, S., Minagawa, A., & Hotta, Y. (2010, January). Enhancement of camera-based whiteboard images. In *IS&T/SPIE Electronic Imaging* (pp. 75340G-75340G). International Society for Optics and Photonics.
18. Clawson, R., Bauer, K., Chidester, G., Pohontsch, M., Kennard, D., Ryu, J., & Barrett, W. (2013, February). Automated recognition and extraction of tabular fields for the indexing of census records. In *IS&T/SPIE Electronic Imaging* (pp. 86580J-86580J). International Society for Optics and Photonics.
19. Parker, J., Frieder, O., Frieder, G. "Automatic Enhancement and Binarization of Degraded Document Images" *Document Analysis and Recognition (ICDAR), 2013 International Conference on*, In Press
20. R. F. Moghaddam, M. Cheriet, AdOtsu: An adaptive and parameterless generalization of Otsu's method for document image binarization, *Pattern Recognition*, Volume 45, Issue 6, June 2012, Pages 2419-2431, ISSN 0031-3203, <http://dx.doi.org/10.1016/j.patcog.2011.12.013>.
21. C. Tomasi, R. Manduchi, "Bilateral filtering for gray and color images," *Computer Vision, 1998. Sixth International Conference on*, pp.839-846, 4-7 Jan 1998
22. N. R. Howe, "Document binarization with automatic parameter tuning." *International Journal on Document Analysis and Recognition (IJ DAR)* (2012): 1-12.
23. [http://ir.cs.georgetown.edu/collections/frieder\\_diaries/browse/The%20Diaries%20of%20Rav%20Frieder.txt](http://ir.cs.georgetown.edu/collections/frieder_diaries/browse/The%20Diaries%20of%20Rav%20Frieder.txt)