6-1-2010

An Analysis of Binarization Ground Truthing

Elisa H. Barney Smith
Boise State University

An analysis of binarization ground truthing

Elisa H. Barney Smith
Boise State University
Boise, Idaho, USA
EBarneySmith@BoiseState.edu

ABSTRACT

The accuracy of a binarization algorithm is often calculated relative to a ground truth image. Except for synthetically generated images, no ground truth image exists. Evaluating binarization on real images is preferred. The ground truthing between and among different operators is compared. Four direct metrics were used. The variability of the results of five different automatic binarization algorithms were compared to that of manual ground truth results. Significant variability in the ground truth results was found.

Categories and Subject Descriptors

I.4.1 [**IMAGE PROCESSING AND COMPUTER VISION**]: Digitization and Image Capture; I.4.6 [**IMAGE PROCESSING AND COMPUTER VISION**]: Segmentation—*Pixel classification*

General Terms

Image Binarization, Ground Truthing

Keywords

Image Binarization, Ground Truthing

1. INTRODUCTION

Binarization is an important first step for many OCR algorithms. If this step is performed incorrectly it can affect segmentation of the page into zones, words and characters, and if these are incorrect, recognition will be poor. There are many algorithms for binarization. Trier and Taxt [12] compared 19 different binarization algorithms based on visual results, Trier and Jain [11] compared 11 algorithms based on the recognition performance on the binarized results. Sezgin and Sankur [9] presented a survey of binarization algorithms which covered 40 different algorithms. They used synthetically generated and degraded images and directly compared the binarization results to the source image. Stathis et al. [10] compared 30 algorithms using images created by combining synthetic text on backgrounds of real stained historical document. The studies listed above are often criticized for not evaluating the algorithms on real images or by not directly utilizing the binarized output images, but rather evaluating binarization performance based on visual appearance or an indirect measure such as recognition performance.

To address these criticisms as well as to offer an opportunity to evaluate the newest generation of binarization algorithms, the Document Image Binarization Contest[2], DIBCO 2009, held in conjunction with the 10th International Conference on Document Analysis and Binarization (ICDAR2009) presented an opportunity to compare several binarization algorithms directly on real images. At DIBCO 2009 there were 43 algorithms entered into the binarization contest. This shows that there is still a significant amount of active research and interest in the problem of document binarization. The images used in DIBCO 2009 were real degraded images with a selection of degradations including stains, show through, bleed through and fading. Four training images and ten test images were prepared, seven each of handwriting and machine print. This provided a diverse and realistic data set on which to test the algorithms. The organizers used a semi-automatic method based on [5] to prepare the ground truth images. While this ground truth was created with great care, it still contains a subjective component. The accuracy therefore is not guaranteed. This paper explores the variability that exists when images are ground truthed by humans and how this might affect the evaluation of automated binarization algorithms. Accuracy is evaluated with four metrics. The experiments attempt to answer the following six questions:

- How much difference do the two sets of carefully ground truthed images have?
- How much variabiility is there among ground truthers?
- Is variability among binarization algorithms greater or less than the variability among ground truthing efforts?
- How large of a range of difficulty is there in the DIBCO2009 data set?
- Which binarization evaluation metric exhibits the greatest variability?

- How does choice of ground truth affect binarization algorithm rankings?

In Section 2 the images used in this paper and the tool used to create the ground truth are presented. Section 3 describes the evaluation metrics used. The results of the experiments are presented in Section 4, and the paper concludes in Section 5.

2. IMAGE DATA

In the DIBCO 2009 contest[2] there were four training images and ten test images with ground truth provided to participants. These form the data used in this study. Some of the original images were provided in gray scale, and some were provided in color. Some contained show through, bleed through, staining of various degrees, paper edges, visible fibers and water marks. Half were handwritten documents and half were machine print. Thumbnails of these images are shown in Figure 1. A summary of the image characteristics is listed in Table 1.

Eric Saund et al. from PARC developed a GUI tool called PixLabeler to use for ground truthing images [7]. A screen shot of this tool in operation is shown in Figure 2. This allows each pixel to take a label as to its image content.

The tool provides a user friendly interface with which to do the binarization. In its current form it has two drawbacks: (1) The smallest cursor (or "brush") size is 2x2, and (2) stopping in the middle and restarting is not possible. The brush size constraint can be overcome by gradually working from the middle or one side of the character to the outside and "repainting" portions of the excess label to correct them making an effective 1x1 label. The restarting issue was found to be largely a timing issue making students either work a long concentrated period, or have to leave their machine on, avoiding other applications and locked when not in use.

This tool was installed at BSU and used to re-ground truth the DIBCO 2009 images. A research student was instructed in what image binarization was and asked to use the tool to re-do the binarization for all 14 DIBCO 2009 images. One image, H03, was selected from the set of 10. It was selected because it was representative of the images, but on the somewhat small side, so the time to do the ground truthing was estimated to be smaller (5 hours). Five other students in BSU's Signals Research Lab were asked to use the PixLabeler tool to ground truth this one image. They were instructed on the use of the tool as well as the ideology behind binarization but not given instructions as to where in the edge profile they should choose to separate text from background.

Ground truthing an image is ultimately a subjective process. As much as we as scientists like to consider the existence of a pure and perfect ground truth result, it likely does not exist in reality. Figure 3 shows a portion of one image in its original form as well as the ground truth from DIBCO 2009 and the results of binarizing that image portion by each member of the BSU team. The resulting binarizations are all different, and while a few can be labeled as lower quality, no one result is clearly the "best". The decision as to at which gray level to call the pixel foreground versus background depends on the local background level and the user's personal opinion as to whether the middle gray level, lightest non-background gray level, or sharpest change in gray levels should be the dividing point. These are the same questions that researchers must answer when developing their automated algorithms. This is biased by knowledge of the generation process, knowledge of the subsequent use of the binarized image, and plain old personal preference. Any head-to-head evaluation of binarization will then be biased toward the binarization algorithm that uses the same "definition" of binarization as was used in preparing the ground truth.

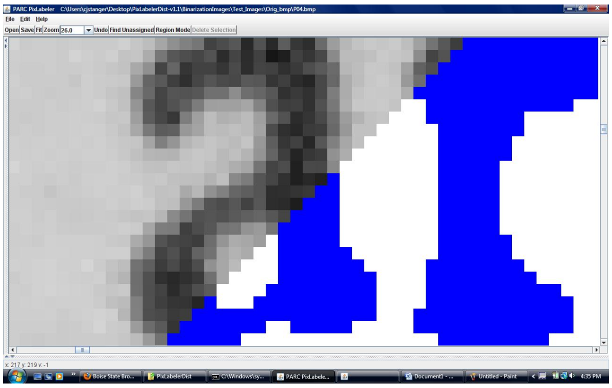
PARC PixLabeler C:\Users\cjstanger\Desktop\PixLabelerDist-v1.1\BinarizationImages\Test_Images\Orig_bmp\P04.bmp

Figure 2: A screen shot of the PixLabeler tool[7] used to create the ground truth images.

3. BINARIZATION EVALUATION METRICS

In the DIBCO 2009 contest there were 4 metrics used to evaluate the entries. Three of those are used in this analysis. Also the Normalized Cross Correlation metric is used. These are defined as follows.

F-Measure (FM)

These metrics are the same as used in information retrieval and were used as the primary metric for DIBCO[2]. A true positive (TP) is defined to occur when the image pixel is labeled as foreground (black) and the ground truth is also. A false positive (FP) occurs if the pixel is labeled foreground when the ground truth is background (white). A false negative (FN) occurs when the pixel is labeled background but the ground truth was foreground. The number of pixels in each image in each of these three categories, TP, FP and FN, are then combined to calculate

$$Recall = \frac{\#TP}{\#FN + \#TP} * 100, \tag{1}$$

and

$$Precision = \frac{\#TP}{\#FP + \#TP} * 100. \tag{2}$$

Precision and recall are combined into a single metric called an F-measure (FM)

$$FM = \frac{2 * Recall * Precision}{Recall + Precision}. \tag{3}$$

A higher F-measure indicates a better match.

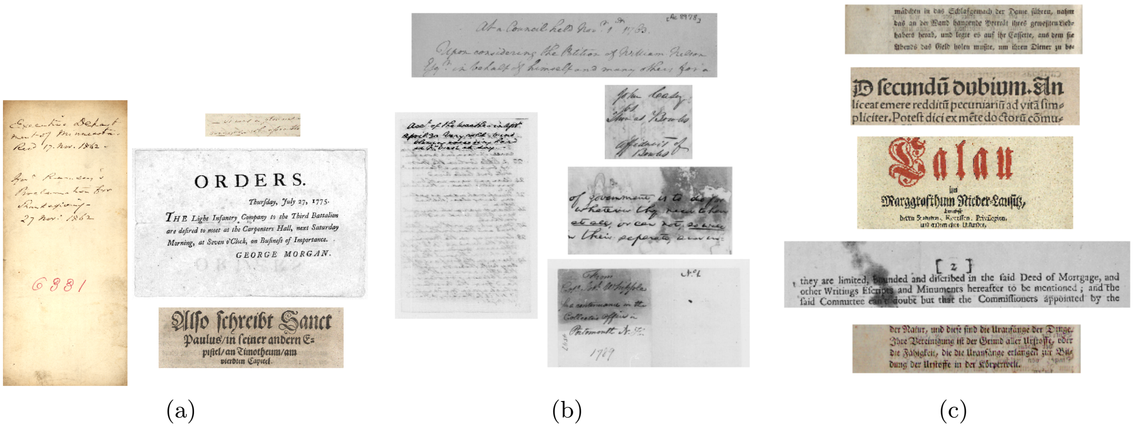

Figure 1: Thumbnails of DIBCO 2009 images used in this study. (a) training images HW* and PR*, (b) handwritten test images H0*, (c) printed test images P0*.

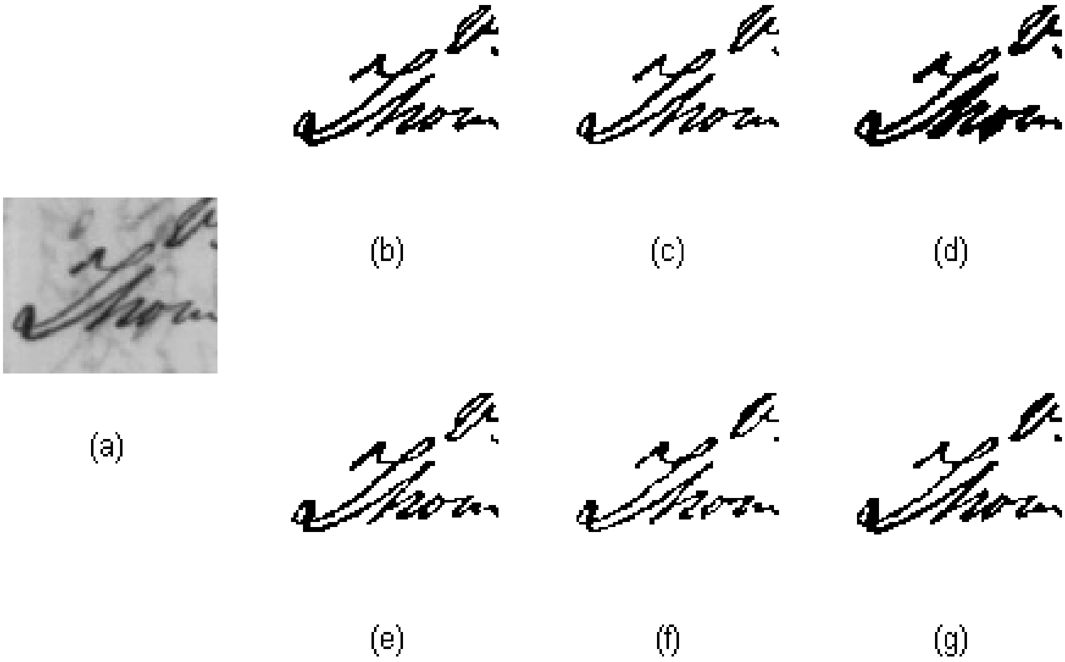

Figure 3: Samples of the images. (a) original image (b) DIBCO ground truth (c)-(g) new ground truths.

Table 1: Summary of DIBCO 2009 image characteristics

Image Reference Code	Image Size Color	Comments	Image Reference Code	Image Size Color	Comments
HW01	1078x2477 Color	uneven background, multi colored ink	PR01	2044x1308 Gray scale	show through folded paper and visible paper edge
HW02	824x201 Color	show through	PR02	1605x525 Color	show through fiber in paper some stain
H01	2025x426 Gray scale	variable ink density	P01	1268x263 Color	show through, non even background, some staining
H02	946x1366 Gray scale	show through and bleed through, paper edge visible	P02	1223x310 Color	show through
H03	582x492 Gray scale	show through	P03	1153x493 Color	multi colored ink, fiber in paper and watermark visible watermark visible
H04	1091x581 Gray scale	significant staining	P04	1849x357 Gray scale	significant staining, visible trace of pencil(?) underlining
H05	1341x713 Gray scale	significant staining, paper edge visible	P05	1218x259 Color	multiple types of show through (one might be ink from facing page) some staining

Negative Rate Metric (NRM)
The NRM uses the true positives, and false negatives as in F-Measure, but also uses the count of true negatives (TN) which occurs when both the image and the ground truth are labeled as background. As defined in DIBCO[2], the negative rate of false negatives,

$$NR_{FN} = \frac{\#FN}{\#FN + \#TP}, \tag{4}$$

and the negative rate of false positives,

$$NR_{FP} = \frac{\#FP}{\#FP + \#TN}, \tag{5}$$

are combined to form the Negative Rate Metric

$$NRM = \frac{NR_{FN} + NR_{FP}}{2}. \tag{6}$$

A lower NRM indicates a better match.

Peak SNR (PSNR)
A metric based more directly on the image difference is the Peak SNR. This is calculated by

$$PSNR = 10 * log_{10}(\frac{C^2}{MSE}), \tag{7}$$

where the Mean Square Error (MSE) is calculated from

$$MSE = \sum_{x=1}^{N}\sum_{y=1}^{M} \frac{(I_1(x,y) - I_2(x,y))^2)}{M * N} \tag{8}$$

and C is the difference between the foreground and background colors. Since all images were converted to a 0/1 scale for this study, $C = 1$. This metric was used in both DIBCO[2] and Stathis[10]. A higher PSNR indicates a better match.

Normalized Cross Correlation (NCC)
Another metric used often to compare images is the Normalized Cross Correlation (NCC). This was used in [1] to evaluate image binarization performance. This is defined as

$$NCC =$$

$$\sum_{x=1}^{N}\sum_{y=1}^{M} \frac{(I_1(x,y) - \bar{I_1})(I_2(x,y) - \bar{I_2})}{\sqrt{\sum_{x=1}^{N}\sum_{y=1}^{M}(I_1(x,y) - \bar{I_1})^2 \sum_{x=1}^{N}\sum_{y=1}^{M}(I_2(x,y) - \bar{I_2})^2}}. \tag{9}$$

A higher NCC indicates better a match.

4. EXPERIMENTS AND RESULTS

A series of experiments were run to compare the variability of the ground truthing. These are designed to explore the questions: How much difference do the two sets of carefully ground truthed images have? How much variability is there among ground truthers? Is variability among binarization algorithms greater or less than the variability among ground truthing efforts? How large of a range of difficulty is there in the DIBCO 2009 data set? Which binarization evaluation metric exhibits the greatest variability? How does choice of ground truth affect binarization algorithm rankings?

These questions were explored using the 14 DIBCO 2009 ground truth image, the 14 image DIBCO 2009 set re-ground truthed at BSU, one image from the DIBCO 2009 data set re-ground truthed at BSU by four other students and the application of five binarization algorithms to the DIBCO images.

(1) How much difference do the two sets of carefully ground truthed images have?

The full set of 14 images ground truthed at BSU were compared to the 14 DIBCO 2009 ground truth images to numerically quantify this variability. First the metric quantifying the difference between each BSU result and the corresponding DIBCO 2009 result was calculated. The resulting maximum, minimum and mean values and the variance among the measurements are shown in Table 2. While the two ground truths did not match and the scores look in total poor given that a 'perfect' score might be expected from a human, it was observed that for the metrics FM, NRM and PSNR that the BSU ground truth was on average similar in magnitude of closeness to the DIBCO ground truth as the best of the entries in the DIBCO contest. The matches were on average closer for the printed documents than the hand written documents.

Table 2: Maximum, minimum, mean and variance when comparing BSU ground truth and DIBCO ground truth across 14 DIBCO images. The desired response is bolded.

Comparison metric	FM	NRM	PSNR	NCC
Maximum	**94.6**	0.129	**23.5**	**0.94**
Minimum	84.9	**0.006**	15.9	0.85
Mean	89.3	0.049	18.4	0.89
Variance	2.6	0.040	2.55	0.025

(2) How much variability is there among ground truther?

This question was explored with one image across a set of 6 ground truth results. Figure 3 shows a portion of the image H03 with all the ground truths. In Figure 4 the average of that portion of image H03 over all six of the ground truth images, including both BSU and DIBCO 2009, can be seen. As expected most of the variability is along the edges of the strokes.

The images were compared to numerically quantify this variability. First the difference between each BSU result and the DIBCO 2009 result was calculated. The resulting mean and the variance among the measurements are shown in Table 3.

Then every pair of images including all 5 BSU results and the one DIBCO result were compared resulting in 6*5/2 comparisons. The results for these means and variances are also shown in Table 3. Overall the variance between ground truths is around 2% or less for all metrics except for NRM where the variance is 40%.

(3) Is variability among binarization algorithms greater or less than the variability among ground truthing efforts?

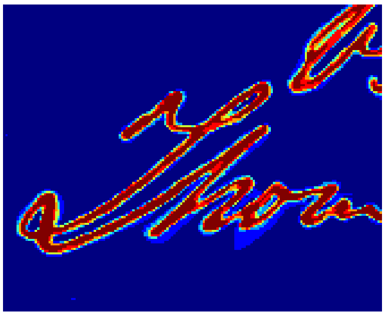

Figure 4: Average binarized ground truth image. Blue is 100% labeled background, red is 100% labeled foreground.

Table 3: Mean and variance when comparing BSU ground truth and DIBCO ground truth on image H03.

Test	Comparison metric	FM	NRM	PSNR	NCC
Many to one	Mean	84.9	0.070	15.8	0.83
Many to one	Variance	2.2	0.020	0.87	0.020
Many to many	Mean	84.7	0.093	15.6	0.83
Many to many	Variance	2.9	0.036	0.96	0.028

As stated in the introduction, many binarization algorithms have been developed. These are usually the target of the evaluation procedure. If the effectiveness of the binarization algorithm is desired, then it is desired to know whether the resulting difference between the binarization algorithm's result and the ground truth image is because of a algorithmic weakness or a difference in opinion when producing the ground truth image.

Five binarization algorithms have been selected for analysis to represent a broad sample of the many possible algorithms available. The binarization algorithms are (1) Otsu - a global thresholding algorithm[6], (2) Niblack - a common adaptive thresholding algorithm[4], (3) Sauvola - another adaptive binarization algorithm[8], (4) Gatos - An algorithm that is particularly suited to documents with uneven background from bleed through and stains[3] and (5) Background Estimation and Subtraction (BES) - an algorithm developed by the author that uses the Total Variation framework for image regularization [1].

Each of the 14 DIBCO images were binarized with each of the 5 binarization algorithms. The minimum, maximum, mean and variance of the algorithms between the DIBCO 2009 ground truths and the BSU ground truths are calculated for each of the four evaluation metrics. These results are shown in Table 4.

The variance between the binarized images and the ground truth was less than the variance between the two ground truth sets.

(4) How large of a range of difficulty is there in the DIBCO 2009 data set?

Table 4: Maximum, minimum, mean and variance among five binarization algorithms.

Comparison metric		FM	NRM	NCC	PSNR
DIBCO	Max	**88.6**	0.11	**16.8**	**0.88**
	Min	84.1	**0.04**	15.3	0.83
	Mean	86.42	0.08	15.9	0.85
	Var	1.5	0.028	0.57	0.016
BSU	Max	**87.3**	0.09	**16.6**	**0.86**
	Min	81.6	**0.041**	14.3	0.81
	Mean	84.9	0.07	15.8	0.84
	Var	2.2	0.020	0.87	0.02

While pure difficulty can not be easily quantified, if the results in Table 2 are considered information on this topic is available. The span between the maximum and minimum scores for the different metrics is quite wide. This indicates that some images have a clearer ground truth than others. This is also shown through the variance of the metrics.

This problem can also be considered by noticing that the range of average match between the five binarization algorithms across the 14 image data set is quite large.

(5) Which binarization evaluation metric exhibits the greatest variability?

Using the results from Tables 2, 3 and 4 the binizaration evaluation metrics can be compared. If the normalized variance is considered, the variance for FM, PSNR and NCC are each around 2% or less, whereas for NRM the variance is 40%. The results from the DIBCO contest [2] also show this trend. If the rankings of the entries are considered relative to the rankings of the NRM score, a lot of inconsistencies are seen. This could mean that NRM is not a good metric, or that it may pick up a totally different set of features than the other metrics do. The NRM is strongly affected by the number of pixels in the image, so smaller images and larger images with the same percentages of errors will have different scores. Further investigation as to the image qualities that lead to good and bad values of this metric need to be done if it will be used often.

(6) How does choice of ground truth affect binarization algorithm rankings?

The results from all five binarization algorithms on all 14 DIBCO images were compared first to the DIBCO ground truth. Then the same binarization output images were compared to the BSU ground truth. The average for each metric across the set of 14 images was calculated for each algorithm and the results were then ranked as appropriate for the metric. Even though the BSU and DIBCO ground truth images have differences, the rankings between them were the same except for one case: the fourth and fifth ranked algorithms with the PSNR metric were swapped. For this data set and this set of algorithms the results were pretty robust to the ground truth algorithm. However the five binarization algorithms that were evaluated were quite dissimilar and expected to have a wide range of abilities in binarization. Also of note is that the rankings between metrics were pretty similar among FM, NCC and PSNR, but very different when any of those three were compared to the rankings that resulted from NRM. This agrees with the earlier discussion on this topic.

5. CONCLUSIONS AND FUTURE WORK

The semi-automatic ground truthing method used to create the ground truth images for DIBCO 2009 was compared with some fully manual ground truthing results. The sets were not as close as might have been expected. For the single image that was ground truthed multiple times, an even larger variability was seen.

A special effort was made to not provide the students with too much information about the expected results to bias their ground truthing algorithm. This may have added to the variance.

Students were asked to share some of their comments on their thoughts of the process that they completed. Some of the comments include: “Sometimes it was difficult to decide where the boundary between text and background was placed.” “I used some knowledge of what the writer was intending.” “I didn’t like to have to use different gray level thresholds across the image, but it was the only way that made sense.” “For the color images it was hard to decide which color to use to make the decision.” These are largely the same issues that researchers contemplate and try to duplicate when designing their binarization algorithms.

Four direct evaluation metrics were used in this study. From the literature there are more available. These could be included in future studies. There is a need for a metric that is weighted by the distance from the edge of the ground truth image. DIBCO used such a metric but it was not coded for this set of experiments.

It would be interesting to reevaluate all 43 binarization algorithms submitted to DIBCO 2009 and compare their performance against each of the ground truth images produced for this study. How would that have changed the rankings? This is not so much important for the determination of a prize winner, but rather for knowing how robust the binarization metric is to different interpretations of the desired outcome as the set of binarization algorithms tested in DIBCO was much larger than the five used here, the sensitivity might be more evident.

It was noticed that the human results were on average comparable in closeness to the top DIBCO results. This may indicate that in a contest, no differentiation among algorithms can be made above a certain level of fit. A choice of a ‘winner’ might at some point become a decision of best mimicing the preferences of the contest organizer. This might lead to a return to evaluation based on OCR performance [11] and other indirect metrics where significantly larger data sets could be used.

Given the variability between operators, ultimately the creation of the best ground truth might take a couple of yet unseen forms. Some possibilities include the thresholded average of the result of multiple ground truthers operating on each image and taking the average of the results could be used. Calculating a median response on each pixel is another possibility. Or perhaps we will learn to live with the variability and approach the problem through using a 'fuzzy' ground truth based on the ground truth average.

Note to reviewers and editor:
(1) The images produced in this study are available for contribution to the DIA community if Gatos et al. agree the original image is available to the community. I don't know if the DIBCO test set is now openly and freely available to the DIA community.

6. ACKNOWLEDGMENTS

Special thanks go the students in BSU's Signal Processing Lab. Mr. C.J. Stanger, a research assistant, put in countless ours doing the bulk of the binarization and supervising the collection from other students. Thanks also go to the other research students who binarized a single image: Kris Burch, Will Grover, Chad Harper, Josh Johnson and Subramaniam Venkatraman.

7. REFERENCES

[1] E. H. Barney Smith, L. Likforman-Sulem, and J. Darbon. Effect of pre-processing on binarization. In *Proceedings SPIE Electronic Imaging Document Recognition and Retrieval*, San Jose, California, USA, January 2010.

[2] B. Gatos, K. Ntirogiannis, and I. Pratikakis. ICDAR 2009 document image binarization contest (DIBCO 2009). In *Proceedings of the Tenth International Conference on Document Analysis and Recognition, ICDAR-2009*, pages 1375–1382, Barcelona, Spain, 2009.

[3] B. Gatos, I. Pratikakis, and S. J. Perantonis. Adaptive degraded document image binarization. *Pattern Recognition*, 39:317–327, 2006.

[4] W. Niblack. *An Introduction to Digital Image Processing*. Prentice- Hall, Englewood Cliffs, NJ, 1986.

[5] K. Ntirogiannis, B. Gatos, and I. Pratikakis. An objective evaluation methodology for document image binarization techniques. In *Proceedings of the 8th International Workshop on Document Analysis Systems (DAS'08)*, pages 217–224, Nara Japan, September 2008.

[6] N. Otsu. A threshold selection method from gray level histograms. *IEEE Trans. Syst. Man Cybern.*, 9:62–66, 1979.

[7] E. Saund, J. Lind, and P. Sarkar. PixLabeler: User interface for pixel-level labeling of elements in document images. In *Proceedings of the Tenth International Conference on Document Analysis and Recognition (ICDAR'09)*, pages 646–650, Barcelona, Spain, 2009.

[8] J. Sauvola and M. Pietikäinen. Adaptive document image binarization. *Pattern Recognition*, 33:215–236, 2000.

[9] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168, 2004.

[10] P. Stathis, E. Kavallieratou, and N. Papamarkos. An evaluation technique for binarization algorithms. *Journal of Universal Computer Science*, 14(18):3011–3030, 2008.

[11] Øivand Due Trier and A. K. Jain. Goal-directed evaluation of binarization methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(12):1191–1201, 1995.

[12] Øivind Due Trier and T. Taxt. Evaluation of binarization methods for document images. *Transactions on Pattern Analysis and Machine Intelligence*, 17(3), March 1995.

Table 5: EXPANDED::Maximum, minimum, mean and standard deviation among five binarization algorithms.

DIBCO Ground Truth					
		Comparison Metric			
Binarization Algorithm		FM	NRM	NCC	PSNR
Otsu	max	97.2	0.12	0.97	25.23
	min	28.17	0.02	0.35	6.82
	mean	78.87	0.05	0.8	16.4
	std	23.37	0.03	0.2	5
Niblack	max	92.76	0.15	0.93	23.09
	min	57.09	0.05	0.58	10.44
	mean	81.77	0.09	0.81	16.06
	std	10.65	0.03	0.1	3.73
Sauvola	max	94.46	0.15	0.94	23.86
	min	51.18	0.04	0.54	9.44
	mean	82.01	0.07	0.81	16.19
	std	12.85	0.03	0.12	3.99
Gatos	max	94.67	0.24	0.94	23.9
	min	68.6	0.04	0.71	13.4
	mean	86.39	0.08	0.86	17.56
	std	7.13	0.05	0.07	3.3
BES	max	95	0.19	0.95	25.34
	min	76.54	0.04	0.76	11.84
	mean	87.86	0.07	0.87	17.92
	std	5.39	0.04	0.05	3.71
Total	max	97.2	0.24	0.97	25.34
	min	28.17	0.02	0.35	6.82
	mean	83.38	0.07	0.83	16.83
	std	13.85	0.04	0.12	4.06
BSU Ground Truth					
		Comparison Metric			
Binarization Algorithm		FM	NRM	NCC	PSNR
Otsu	max	96.34	0.14	0.95	23.8
	min	25.82	0.01	0.33	7.06
	mean	77.09	0.06	0.78	15.47
	std	21.72	0.04	0.19	4.69
Niblack	max	90.88	0.19	0.91	21.77
	min	54.09	0.02	0.52	9.72
	mean	79.72	0.1	0.79	15.51
	std	10.87	0.05	0.11	3.5
Sauvola	max	93.15	0.16	0.93	22.79
	min	47.8	0.01	0.51	9.13
	mean	79.71	0.08	0.79	15.48
	std	12.58	0.05	0.12	3.76
Gatos	max	92.4	0.23	0.92	22.26
	min	70.56	0.01	0.72	13.37
	mean	84.07	0.09	0.83	16.72
	std	6.59	0.06	0.06	3.02
BES	max	93.1	0.18	0.93	24.07
	min	72.77	0.01	0.72	12.06
	mean	85.34	0.08	0.84	17.05
	std	5.89	0.04	0.06	3.62
Total	max	96.34	0.23	0.95	24.07
	min	25.82	0.01	0.33	7.06
	mean	81.19	0.08	0.81	16.05
	std	13.21	0.05	0.12	3.82