ICDAR 2011 Document Image Binarization Contest (DIBCO 2011)

Ioannis Pratikakis [1], Basilis Gatos [2] and Konstantinos Ntirogiannis [2,3]

[1]Department of Electrical and Computer Engineering, Democritus University of Thrace, GR-67100 Xanthi, Greece
e-mail: ipratika@ee.duth.gr

[2]Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, NCSR "Demokritos", GR-15310 Agia Paraskevi, Athens, Greece, e-mail: {bgat, kntir}@iit.demokritos.gr

[3]Department of Informatics and Telecommunications, National and Kapodistrian University of Athens GR-15784 Panepistimioupoli, Ilissia, Athens, Greece
e-mail: kntir@di.uoa.gr

Abstract - DIBCO 2011 is the International Document Image Binarization Contest organized in the context of ICDAR 2011 conference. The general objective of the contest is to identify current advances in document image binarization for both machine-printed and handwritten document images using evaluation performance measures that conform to document image analysis and recognition. This paper describes the contest details including the evaluation measures used as well as the performance of the 18 submitted methods along with a short description of each method.

I. INTRODUCTION

Document image binarization is of great importance in the document image analysis and recognition pipeline since it affects further stages of the recognition process. The evaluation of a binarization method aids in verifying its effectiveness and studying its algorithmic behaviour. To this end, following the success of DIBCO 2009 [1] organized in conjunction with ICDAR'09 as well as of H-DIBCO 2010 [2] organized in conjunction with ICFHR 2010, the follow-up of these contests in the framework of ICDAR 2011 was organized. In this contest, we focused on the evaluation of document image binarization methods using a variety of scanned machine-printed and handwritten documents for which we created the binary image ground truth following a semi-automatic procedure based on [3]. The authors of submitted methods registered in the competition and downloaded representative samples along with the corresponding ground truth. At a next step, all registered participants were required to submit their binarization executable. After the evaluation of all candidate methods, the testing dataset (8 machine-printed and 8 handwritten images with the associated ground truth) along with the evaluation software became publicly available (http://utopia.duth.gr/~ipratika/DIBCO2011/benchmark).

II. METHODS AND PARTICIPANTS

Sixteen (17) research groups have participated in the competition with eighteen (18) different algorithms (one participant submitted two algorithms). Brief descriptions of the methods are given in the following (The order of appearance is based upon the order of submission of the algorithm).

1) Qatar University, Qatar and Northumbria University, UK (*Abdelâali Hassaïne, Larbi Boubchir Somaya Al-Maadeed and Ahmed Bouridane*): This team has submitted two variations of an algorithm that are presented in the following.
(a) The method classifies each pixel as foreground or background according to its global k-means and Otsu segmentation, the values of its neighbors in these two segmented images, the values of the basic morphological operations and gaussian filters of several sizes. Standard edge detection techniques including Sobel and Laplace are also used as classifiers. All these classifiers are combined using a logistic regression to perform the classification. A second "cleaning" step is performed on the resulting image using the same classifiers in order to remove the pixels which are badly classified.
(b) A second variation of the aforementioned method, keeps only the most discriminant descriptors for the training stage.

2) Houari Boumediene University of Sciences and Technologies, Algeria (*M. Zayed*): The idea of this method is to detect different classes of pixels in the document image (background, foreground and others). Then, a local binarization technique is used to separate background and foreground in a local area so that a better performance is achieved. Finally, a post-processing is applied to remove small connected components and fill holes.

3) Illinois Institute of Technology, USA (T. Obafemi-Ajayi and G. Agam): The algorithm is based on a multi-resolution framework using an adaptive window selection for effective binarization of historical documents. An unsupervised learning technique is applied (Linear Discriminant Analysis via Otsu method [4]) to cluster the image pixels as foreground or background. A complete description of the method can be found in [5].

4) Kobe University, Japan (*N. Tanaka*): The method comprises three distinct steps. In step 1, it is

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.299

identified whether a character region is black region or white region by the corresponding ratio in the gray-level histogram. Step 2 removes the gradual fluctuation of background (since they cause "ghost objects") by the mathematical morphology top-hat operation. In step 3, the output image is obtained by following conditional dilation. In the latter operation, the condition image is the binarized image which is obtained by Otsu method**.**

5) Institute for Language and Speech Processing (ILSP) of Athena - Research and Innovation Center in Information, Communication and Knowledge Technologies, Greece, & National Technical University of Athens (NTUA), Greece (*V. Papavassiliou and F. Simistira*): In this method, the gray image is first filtered by applying the top-hat by reconstruction technique with a large structuring element (e.g. a disk with radius of 25 pixels). Next, a global threshold T is calculated by using the Otsu's method. Considering seed points, the pixels with values lower than 0.95*T, we grow the initial regions by examining the values of neighboring pixels of seed points and determining whether the pixel neighbors should be added to the region. This process results in an initial binary image (BW1). The Otsu's method and the region growing technique are also applied on the second derivative of the filtered image to produce a second binary image (BW2). Finally, the morphological reconstruction of the image marker (BW1 $\cap$ BW2) under the image mask (BW1 $\cup$ BW2) results in the final binary image.

6) University of Tunis, Tunisia & Technische Universitaet Braunschweig, Germany (*I. Ben Messaoud, H. Amiri, H. El Abed Haikal, V. Märgner*): The proposed method is decomposed into 5 parts. First, the regions of interest are detected using Canny's edges or connected components while the other regions of the image are considered as background. Second, a Wiener filter is applied. Then, the edges are detected using a combination between the high contrast images and the Canny's edge. A local threshold is calculated according to pixel intensities within a specific window. Finally, a post-processing method is performed. A detailed description of the method can be found in [6].

7) Federal University of Pernambuco, Brazil (*R. Neves, C.A.B. Mello*): The algorithm is mainly dedicated to grayscale handwritten documents and it is divided into three phases: The first is responsible to indentify the main objects of the image; the second phase divides the image into sub-images according to the previous identification; and the latest phase evaluates a local threshold for each sub-image and proceeds with the binarization of each region.

8) National University of Singapore and Institute for Infocomm Research, Singapore (*B. Su, S. Lu and C-L. Tan*): This method comprises four main steps. First, local image contrast which is evaluated by local maximum and minimum and local image gradient are combined based on the local mean intensity and variation to select the edge point candidates. Second, the stroke edges which are extracted using Canny's method are employed to produce a better edge map. Third, the document image is binarized by a local threshold is decided based on the constructed edge map and estimated stroke width. At last, some post-processing work is applied to produce better results.

9) Concordia University, Canada (*T.H. Ngan Le, T.D.Bui and C.Y. Suen*) : In this method, a novel adaptive binarization algorithm using ternary entropy-based approach is used. Given an input image, the contrast of intensity is first estimated by a grayscale morphological closing operator. A double-threshold is generated by our Shannon entropy-based method to classify pixels into text, near-text, and non-text regions. The pixels in the second region are relabeled by the local mean and the standard deviation values. The proposed method classifies noise into two categories which are processed by binary morphological operators, shrink and swell filters, and graph searching strategy.

10) South University of Toulon-Var, France, (*T. Lelore and F. Bouchara*): The method first use a median filtering of the input image and then upscale it using linear interpolation. Then, the method proposed by the authors in the H-DIBCO 2010 contest in the framework of ICFHR [2] with a low threshold to produce a noisy version of the final result. The text pixels from this image are then mixed into another temporary three-valued image (Text, Background, Unknown) obtained using correct threshold estimation. The noisy pixels are then easily detected and removed to produce the final image. This image is resized back to the correct resolution using bicubic interpolation. The final black and white image is obtained using a global threshold set to 70.

11) Smith College, USA (*N. Howe*): The method minimizes a global energy function, formulated as a graph cut problem for efficient exact computation. The Laplacian of the original image determines the local likelihood of foreground and background labels for each pixel, granting the method invariance to intensity shifts caused by illumination, stains, etc. without the need for thresholding. An energy penalty between neighboring pixels with different labels serves to enforce smoothness. The penalty does not apply if Canny edge detection identifies a likely discontinuity between the two neighbors, thus allowing the solution to conform to the natural boundaries in the image.

12) SAIC-Frederick,Inc., USA (*I. Filippov*): The proposed approach is using a simple modification to Otsu algorithm. Two thresholds are calculated - one is the mid-point between two intensity peaks computed by global Otsu method. The other is a weighted average (75% of lighter peak, 25% of darker peak) computed within a 25x25 window. The global threshold is taken if the difference between local peaks is smaller than a predefined value and the local threshold is higher than 1.1 * global_threshold. Otherwise, the local threshold is used. The window size and the condition for preferring global threshold vs. local are the parameters of the model.

13) Vienna University of Technology, Austria and Fraunhofer-Institute for Production Systems and Design Technology (IPK), Germany (F. Kleber, M. Diem, R. Sablatnig): The algorithm uses the saturation channel of the IHLS color space [7] to enhance colored text with a low contrast. To handle noisy images, a foreground-estimation is applied. Integral Images are used for an efficient calculation of morphological operations, the mean and the standard deviation.

14) Vienna University of Technology, Austria (F. Kleber, M. Diem, R. Sablatnig): The proposed algorithm is based on the toggle mapping operator [8]. The image is first mapped on the corresponding morphological erosion and dilation. Then, if the pixel value is closer to the erosion, it is marked as background otherwise it is marked as foreground. To avoid salt and pepper noise, pixels whose erosion and dilation are too close, are excluded from the analysis. Pixels are then classified into three classes: foreground, background and homogeneous. Finally, homogeneous regions are assigned to foreground or background according to the class of their boundaries. A hysteresis threshold is also used in order to reduce the critical effect of the threshold parameter.

15) Indian Institute of Technology Kharagpur, India (*S. Bag, P. Behera, and P. Bhowmick*): This is an adaptive-cum-interpolative binarization method for degraded document images. A multi-scale framework is added to an adaptive version of Otsu's method [4]. To convert Otsu's method to an adaptive model, the local threshold value is calculated for each pixel by observing the intensity behavior of its neighbor pixels. The function for this adaptive version of the Otsu's method is taken from Moghaddam and Cheriet method [9]. Like other adaptive methods, we use two parameters (k1 and k2) which act as threshold multipliers to make the threshold value more effective for image binarization.

16) University of Guadalajara, Mexico and Freie Universität Berlin, Germany (M.A. Ramírez-Ortegón, E. Cuevas, R. Rojas): The main idea is to compute transition values using pixel-intensity differences in a neighborhood around the pixel of interest [10]. Two subsets are considered in the neighborhood corresponding to high positive and negative transition values, called transition sets. These sets are refined by morphological operators in the transition image [11]. The binarization threshold is computed over the pixels in the transition sets using a statistical model to generate a preliminary binary image [12]. Finally, stains are removed using several morphological operators while erroneous connected components are detected and removed using contextual rules.

17) University of São Paulo, Brazil (*W.A. Luz Alves, A. Morimitsu and R.F. Hashimoto*): The proposed method is based on a morphological operator. First, an image contrast is extracted from the input image by ultimate attribute opening (UAO) of height attribute. Then, a binary image is obtained by the application of thresholdings based on the toggle mapping operator. Finally, in order to eliminate small connected components an area-opening operator is applied to the image and heuristics are used to recover lost foreground pixels.

III. EVALUATION MEASURES

For the evaluation, the measures used comprise an ensemble of measures that are suitable for evaluation purposes in the context of document analysis and recognition. These measures consist of (i) F-Measure; (ii) PSNR; (iii) Distance Reciprocal Distortion Metric and (iv) Misclassification Penalty Metric.

A. F-Measure

$$FMeasure = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{1}$$

where $\text{Recall} = \frac{TP}{TP + FN}$, $\text{Precision} = \frac{TP}{TP + FP}$

TP, FP, FN denote the True positive, False positive and False Negative values, respectively.

B. PSNR

$$PSNR = 10\log\left(\frac{C^2}{MSE}\right) \tag{2}$$

where $MSE = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} (I(x,y) - I'(x,y))^2}{MN}$

PSNR is a measure of how close is an image to another. Therefore, the higher the value of PSNR, the higher the similarity of the two images. We consider that the difference between foreground and background equals to C.

C. Distance Reciprocal Distortion Metric (DRD)

The Distance Reciprocal Distortion Metric (DRD) has been used before to measure the visual distortion in binary document images [13]. It properly correlates with the human visual perception and it measures the distortion for all the $\boldsymbol{S}$ flipped pixels as follows:

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN} \qquad (3)$$

where DRD_k is the distortion of the k-th flipped pixel and it is calculated using a 5x5 normalized weight matrix $\boldsymbol{W_{Nm}}$ as defined in [13]. DRD_k equals to the weighted sum of the pixels in the 5x5 block of the Ground Truth $\boldsymbol{GT}$ that differ from the centered kth flipped pixel at (x,y) in the Binarization result image $\boldsymbol{B}$ (Eq. 4).

$$DRD_k = \sum_{i=-2}^{2} \sum_{j=-2}^{2} |GT_k(i,j) - B_k(x,y)| \times W_{Nm}(i,j) \quad (4)$$

Finally, NUBN is the number of the non-uniform (not all black or white pixels) 8x8 blocks in the GT image.

D. Misclassification penalty metric (MPM)

The Misclassification penalty metric MPM evaluates the prediction against the Ground Truth (GT) on an object-by-object basis. Misclassification pixels are penalized by their distance from the ground truth object's border.

$$MPM = \frac{MP_{FN} + MP_{FP}}{2} \qquad (5)$$

where $MP_{FN} = \frac{\sum_{i=1}^{N_{FN}} d_{FN}^i}{D}$, $MP_{FP} = \frac{\sum_{j=1}^{N_{FP}} d_{FP}^j}{D}$

d_{FN}^i and d_{FP}^j denote the distance of the ith false negative and the jth false positive pixel from the contour of the GT segmentation. The normalization factor D is the sum over all the pixel-to-contour distances of the GT object. A low MPM score denotes that the algorithm is good at identifying an object's boundary.

IV. EXPERIMENTAL RESULTS

The DIBCO 2011 testing dataset consists of 8 machine-printed and 8 handwritten images resulting in a total of 16 images for which the associated ground truth was built for the evaluation. A representative example of the dataset is shown in Fig. 1(a),(c). The documents of this dataset originate from the collections of the following libraries: The Goettingen State and University Library (UGOE), The Bavarian State Library, the British Library and the Library of Congress. The selection of the images in the dataset was made so that should contain representative degradations.

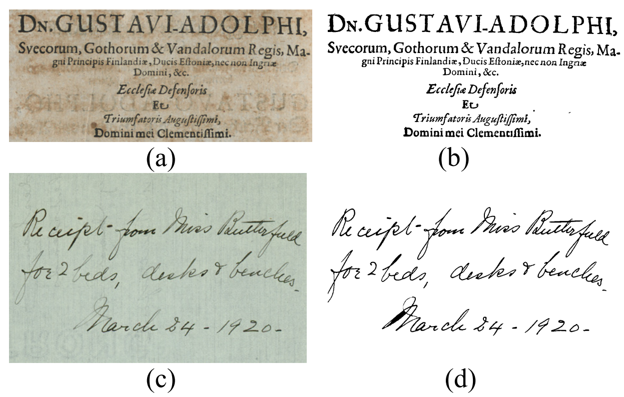

Figure 1. Representative samples and corresponding binarization results from the winner algorithm of DIBCO 2011 (a) Original printed image; (b) Binarized machine printed image; (c) Original handwritten image; (d) Binarized handwritten image.

The evaluation was based upon the four distinct measures presented in Section III. The final ranking is shown in Table I. It was calculated after first, sorting the accumulated ranking value for all measures for each test image. Thereafter, the summation of all accumulated ranking values for all test images denote the final score which is shown in Table I. At Table II, the detailed performance for the top 3 algorithms is also given. In this Table, HW1-8 denote the handwritten test images while PR1-8 denote the machine-printed test images. Overall, the best performance is achieved by ***Algorithm 10*** which has been submitted by **T. Lelore** and **F. Bouchara** of the **South University of Toulon-Var** in France. Example binarization results of this algorithm is shown in Fig. 1(b),(d).

REFERENCES

[1] B. Gatos, K. Ntirogiannis and I. Pratikakis, DIBCO 2009 - Document Image Binarization Contest, *IJDAR*, vol. 14, pp. 35-44, 2011

[2] I. Pratikakis, B. Gatos and K. Ntirogiannis, "H-DIBCO 2010 – Handwritten Document Image Binarization Competition", ICFHR'10, Kolkata, India, pp. 727-732, 2010

[3] K. Ntirogiannis, B. Gatos and I. Pratikakis, "An Objective Evaluation Methodology for Document Image Binarization Techniques", DAS 2008, Nara, Japan, pp. 217-224, September 2008.

[4] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Trans. Systems Man Cybernet.* Vol. 9 (1), pp. 62–66, 1979.

[5] T. Obafemi-Ajayi and G. Agam. "Statistical multi-resolution schemes for historical document binarization". In *Document Recognition and Retrieval XVIII, Proc. SPIE*, 2011.

[6] I. Ben Messaoud, H. Amiri, H. El Abed, and V. Märgner, "New binarization approach based on text block extraction", in *International Conference on Document Analysis and Recognition (ICDAR)*, 2011, in press.

[7] A. Hanbury and J. Serra. "A 3D-Polar Coordinate Colour Representation suitable for Image Analysis", Technical Report PRIP-TR-077, TU Wien, 2002.

[8] F. Kleber, M. Diem and R. Sablatnig. "Scale Space Binarization Using Edge Information Weighted by a Foreground Estimation", ICDAR 2011, in press.

[9] R.F., Moghaddam, M. Cheriet, "A multi-scale framework for adaptive binarization of degraded document images", *Pattern Recognition*, 42, pp. 2186-2198, 2010.

[10] MA. Ramírez-Ortegón, and E. Tapia, and R. Rojas, and E. Cuevas, "Transition Pixel: A Concept for Binarization Based on Edge Detection and Gray-Intensity Histograms", *Pattern Recognition*, Vol. 43, Issue 4, pp. 1233-1243, April 2010.

[11] MA. Ramírez-Ortegón, and E. Tapia, and R. Rojas, and E. Cuevas, "Transition thresholds and transition operators for binarization and edge detection", *Pattern Recognition*, Vol. 43, issue 10, pp 3243-3254, October 2010.

[12] MA. Ramírez-Ortegón, E. Dueñez-Guzmán, R. Rojas, and E. Cuevas, "Unsupervised Evaluation Measures for Binarization", *Pattern Recognition*, Vol. 44, issue 3, pp 491-502, 2011.

[13] H. Lu, A. C. Kot and Y.Q. Shi, "Distance-Reciprocal Distortion Measure for Binary Document Images", *IEEE Signal Processing Letters*, vol. 11, No. 2, pp. 228-231, 2004.

TABLE I. EVALUATION RESULTS WITH RESPECT TO THE MEASURES USED FOR ALL METHODS SUBMITTED TO DIBCO 2011

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Method	**10**	**8**	**11**	**6**	**4**	**5**	**7**	**9**	**1a**	**2**	**16**	**3**	**12**	**17**	**13**	**1b**	**14**	**15**
Score	309	346	429	470	489	515	532	600	610	620	630	649	676	682	715	792	835	1045

TABLE II. EVALUATION RESULTS FOR EACH TEST IMAGE WITH RESPECT TO THE MEASURES USED FOR THE TOP 3 METHODS SUBMITTED TO DIBCO 2011

	Method	**HW1**	**HW2**	**HW3**	**HW4**	**HW5**	**HW6**	**HW7**	**HW8**	**PR1**	**PR2**	**PR3**	**PR4**	**PR5**	**PR6**	**PR7**	**PR8**
F-Measure	**10**	88.2	95.1	92.8	89.5	95.2	92.2	92.0	94.0	94.9	77.2	94.8	95.0	92.3	9.9	4.6	86.1
	8	80.2	93.7	92.1	87.9	95.1	76.4	91.1	93.4	92.9	82.0	93.8	92.0	92.7	92.6	21.1	86.2
	11	79.1	94.4	93.2	89.1	90.6	87.3	88.5	94.6	94.2	70.3	96.5	94.8	94.8	84.9	79.1	88.5
PSNR	**10**	15.1	23.4	19.8	17.3	19.7	19.5	22.0	22.6	17.8	11.9	17.3	19.6	16.7	0.6	0.2	14.6
	8	12.3	22.6	19.5	16.8	19.6	15.3	21.6	22.3	16.4	13.2	16.5	17.7	17.1	21.4	7.6	14.6
	11	11.8	22.9	20.0	17.1	16.4	17.4	20.2	23.0	17.2	10.2	18.9	19.5	18.5	17.9	19.2	15.3
DRD	**10**	6.6	1.4	1.8	2.5	1.6	2.0	1.7	1.3	2.5	12.8	1.8	2.0	2.4	575.0	1052.7	3.6
	8	13.8	1.7	2.0	3.0	1.8	6.3	2.0	1.5	3.5	9.0	2.3	3.5	2.0	3.1	191.3	3.8
	11	15.3	1.7	1.8	2.8	4.6	3.9	3.4	1.3	3.0	19.6	1.3	2.0	1.5	9.3	11.0	3.2
MPM	**10**	14.0	0.1	0.2	0.7	1.1	0.1	0.1	0.0	1.2	34.9	0.4	0.1	1.0	478.5	498.0	0.4
	8	41.1	0.1	0.1	0.7	1.0	0.7	0.0	0.1	2.4	26.0	0.9	0.1	0.2	0.1	71.1	0.5
	11	48.0	0.8	0.6	3.1	12.0	2.3	2.0	0.1	3.4	53.1	0.5	0.1	0.3	5.9	4.1	2.6