

ICFHR2014 Competition on Handwritten Document Image Binarization (H-DIBCO 2014)

Konstantinos Ntirogiannis¹, Basilis Gatos¹ and Ioannis Pratikakis²

¹Computational Intelligence Laboratory,
Institute of Informatics and Telecommunications,
NCSR "Demokritos", Athens, Greece,
e-mail: {kntir,bgat}@iit.demokritos.gr

²Department of Electrical and Computer Engineering,
Democritus University of Thrace,
Xanthi, Greece
e-mail: ipratika@ee.duth.gr

Abstract— Document image binarization is an important step in the document image analysis and recognition pipeline. H-DIBCO 2014 is the International Document Image Binarization Competition which is dedicated to handwritten document images organized in conjunction with ICFHR 2014 conference. The objective of the contest is to identify current advances in handwritten document image binarization using meaningful evaluation performance measures. This paper reports on the contest details including the evaluation measures used as well as the performance of the 8 submitted methods along with a short description of each method.

Keywords- handwritten document image processing; binarization; performance evaluation

I. INTRODUCTION

Handwritten document image binarization contributes significantly in the success of the handwritten document image analysis and recognition challenging tasks. In this respect, it is imperative to create a framework for benchmarking purposes, i.e. a benchmarking dataset along with an objective evaluation methodology in order to capture the efficiency of current image binarization practices for handwritten document images. To this end, following the success of DIBCO series competitions (DIBCO 2009 [1] organised in conjunction with ICDAR'09, H-DIBCO 2010 [2] organized in conjunction with ICFHR 2010, DIBCO 2011 [3] organised in conjunction with ICDAR'11, H-DIBCO 2012 [4] organized in conjunction with ICFHR 2012 and DIBCO 2013 [5] organised in conjunction with ICDAR'13) the follow-up of this contest is organised in conjunction with ICFHR 2014. In H-DIBCO 2014 (Handwritten Document Image Binarization Contest), the general objective is to record recent advances in handwritten document image binarization using established evaluation performance measures. In this contest, we focused on the evaluation of document image binarization methods using a variety of scanned handwritten documents for which the corresponding binary ground truth image has been created. The authors of submitted methods had initially registered in the competition and downloaded representative document image 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 (10 handwritten images with the associated ground truth) along with the evaluation software has been released as publicly available in the following link:

<http://users.iit.demokritos.gr/~bgat/HDIBCO2014/benchmark>

The remainder of the paper is structured as follows: Each of the methods submitted to the competition is briefly described in Section II. The evaluation measures are detailed in Section III. Experimental results are shown in Section IV while in Section V conclusions are drawn.

II. METHODS AND PARTICIPANTS

Seven (7) distinct research groups have participated in the competition with eight (8) different algorithms since one participating research group has submitted two algorithms. Brief descriptions of the methods are given in the following (the order of appearance reflects the chronological order of expressing an interest to participate in the competition).

1) Golestan University, Gorgan, Iran (Majid Ziaratban and Fatemeh Bagheri): Since usually the texts include high frequencies in an image, low frequencies are reduced to cancel soft textures of backgrounds. By applying a primary thresholding method, an initial binary image is obtained. In this binary image, average size of characters is estimated. Some morphological operations are designed and used to estimate the overall slant of the text. Based on the estimated character size and overall slant, suitable directional filters are applied to emphasis the strokes with the similar direction to the overall slant. Furthermore, statistical distributions of the text pixels in images are used to decrease the effects of non-text objects. The threshold value is computed by analyzing the intensity histogram of the enhanced gray image. After applying the second thresholding process, some text parts may be ignored in the binarized image. To solve this problem, an iterative process is implemented, in which a neighborhood is considered for each detected foreground component. If some conditions satisfied in the neighborhood area, the threshold value is modified and recomputed for this area. Based on the modified threshold, some new foreground pixels are added to the binarized image. The iterations are continued until the conditions are satisfied.

2) Smith College, Department of Computer Science, Northampton (MA), USA (Nicholas R. Howe): This submission is based primarily upon the algorithm described in [6]. This algorithm finds the binarization that minimizes a global energy function based upon the image Laplacian, with a smoothness term that incorporates knowledge of the Canny edge locations. Several free parameters in the base binarization method are set using a stability criterion: the best parameter settings usually fall in regions where small changes to the parameters result in minimal change to the resulting binarization. The submitted entry differs from the algorithm described in [6] in that it uses an improved technique for choosing the high Canny threshold: rather than computing the full binarization for a range of values and looking for the most stable region, the new method simply looks for stability in the edge image itself. This speeds up the parameter choice time while also giving better results on the test images from prior contests.

3) Dept. of CSE, University of Kalyani and CVPR Unit, Indian Statistical Institute, India (Chandranath Adak and Bidyut B. Chaudhuri): The team has proposed an unsupervised method for binarization with two variations that are presented as following.

(a) At first the method converts the color image into gray, and all the peaks are found from the gray-level histogram. Then *Fuzzy C-Means* (FCM) clustering is used to segment the gray image. Here the number of clusters (C-value) is the number of peaks (\pm low threshold T1) of that histogram. Then a global threshold T2 is chosen (on the basis of the number of histogram peaks) and applied on the segmented image to produce a binary image.

(b) The second variation of the aforesaid method depends on the choice of the threshold T2. It is actually the threshold selected by Otsu [7]. But the T2 is selected from the segmented image (after segmentation of the gray image using FCM clustering).

4) Image Processing and Multimedia Laboratory, Dept. of Electrical and Computer Engineering, Democritus University of Thrace, Greece (Nikolaos Mitianoudis and Nikolaos Papamarkos): The method is divided into three steps: a) Background Removal, b) LCM clustering, c) Post-processing. In the Background Removal stage, the image is iteratively median filtered with increasing window sizes until the standard deviation of every possible 3×3 image patch is smaller than a threshold value of $S_T = 6$. We then form the absolute distance between the estimated background and the document image and construct a normalised histogram (pdf) of this distance. This pdf is generally a decreasing function. The pixels whose probability is greater than a portion q of the maximum probability are considered background pixels and are set to white. The next step is LCM clustering. The remaining pixels that are considered text pixels are transformed into the following LCM representation. For each of these pixels,

we use a 3×3 neighborhood around each pixel to construct a family of 8 3D vectors. These 8 vectors are constructed using the intensity value of the central pixel, the intensity value of the neighboring pixels and a modified contrast value for the neighborhood (the combination of the central pixel with itself is not included, giving us 8 vectors instead of 9). If we perform a 3D histogram of the vector population, we can see two concentrations: one corresponding to lower intensities and greater contrast that belong to ink entries and one corresponding to higher intensities and lower contrast mainly corresponding to noise. Mixtures of Gaussians clustering is used to segment the two clusters and perform binarization using the rule: "if any LCM data point in each 3×3 neighborhood is classified to the character cluster, then the corresponding central pixel is set to black. The remaining pixels are set to white". The final stage of post-processing identifies connected components and removes those who contain less than 20 pixels to remove possible noise stains left by the previous stage.

5) Synchromedia Lab, Icole de technologie supirieure, University of Quebec, Montreal, Quebec, Canada (Hossein Ziaei Nafchi and Reza Farrahi Moghaddam and Mohamed Cheriet): This method is a modified version of the binarization method proposed in [8]. It uses phase-derived features of images to model background and foreground. These features are: i) Denoised image with phase preserved, ii) Maximum moment of phase congruency covariance and iii) Locally weighted mean phase angle. Also, adaptive median and gaussian filters are applied for further enhancement.

6) Centro de Informática, Universidade Federal de Pernambuco, Brazil (Rafael Galvão Mesquita, Carlos Alexandre Barros de Mello, Ricardo Martins de Abreu Silva, Péricles Barbosa Cunha de Miranda): Our approach first simulate the way the human visual system perceives distant objects by applying the first step of the algorithm proposed in [9]. Afterwards, we apply Howe's binarization method [6]. The reason for combining these two techniques is that Howe's method may fail in some situations in which its parametric choice is optimal for only a part of the image, as in cases of degradation generated by smudges or uneven illumination. As the algorithm presented in [9] showed good results in these situations, we expect that this combination can achieve promising results. We have used DIBCO'11 dataset to tune the parameters (minimum angle of resolution and the radiuses of the structuring elements) of the algorithm in [9] using the I/F-Race [10] approach.

7) Dept. of CSE, University of Kalyani and CVPR Unit, Indian Statistical Institute, India (Barun Biswas, Ujjwal Bhattacharya and Bidyut B. Chaudhuri): This method does not use the color information of the input image. If a color image is given, the method first converts it grayscale

equivalent and then apply a blurring filter on it. Canny edge detection is performed on the blurred image. The gray values of the two pixels of the grayscale image at the left and right of each Canny edge pixel is stored in a bin. The lowest valley in the histogram of these gray values is obtained and the same is used as a global threshold value. Each pixel of the grayscale image with gray value greater than the above global threshold is turned as background pixel. A small square window is considered around each of the remaining pixels and certain simple statistics are computed on the gray values of this small window based on which the said pixel is turned either background or foreground. Finally, the well-known shrink and swell filter is applied on the above binarized image to get rid of a few noise pixels.

III. EVALUATION MEASURES

For the evaluation, the measures used comprise an ensemble of measures that have been widely used for evaluation purposes. These measures consist of (a) F-Measure; (b) pseudo F-Measure; (c) PSNR; (d) Distance Reciprocal Distortion Metric.

A. F-Measure

$$FM = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (1)$$

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. pseudo F-Measure

This measure has been introduced in [11] and it uses pseudo-Recall R_{ps} and pseudo-Precision P_{ps} (following the same formula as F-Measure). The pseudo-Recall and pseudo-Precision metrics use distance weights with respect to the contour of the ground-truth (GT) characters. In the case of pseudo-Recall, the weights of the GT foreground are normalized according to the local stroke width. Generally, those weights are delimited between [0,1]. In the case of pseudo-Precision, the weights are constrained within an area that expands to the GT background taking into account the stroke width of the nearest GT component. Inside this area, the weights are greater than one (generally delimited between (1,2]) while outside this area they are equal to one. It is worth to mention that pseudo-FMeasure F_{ps} is different from the pseudo-FMeasure $p\text{-FM}$ used in H-DIBCO 2010 [2] and H-DIBCO 2012 [4].

C. PSNR

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

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 $M \times N$ images is. We consider that the difference between foreground and background equals to C .

D. Distance Reciprocal Distortion Metric (DRD)

The Distance Reciprocal Distortion Metric (DRD) has been used before to measure the visual distortion in binary document images [12]. It properly correlates with the human visual perception and it measures the distortion for all the S flipped pixels as follows:

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

where DRD_k is the distortion of the k -th flipped pixel and it is calculated using a 5×5 normalized weight matrix W_{Nm} as defined in [12]. DRD_k equals to the weighted sum of the pixels in the 5×5 block of the Ground Truth GT that differ from the centered k^{th} flipped pixel at (x,y) in the binarization result image 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) 8×8 blocks in the GT image.

IV. EXPERIMENTAL RESULTS

A. Dataset

Previous H-DIBCO datasets [2], [4] contained images that were taken from the Library of Congress [13], a well-known source of degraded handwritten document images. In the current competition, the document images originate from the collections of the project *transcriptorium* [14]. Therefore, the dataset is unknown and unexpected to the participants that may have "trained" their algorithms using the usual source [13].

The selection of the images in the current dataset was made so that should contain representative degradations which appear frequently in handwritten documents (e.g. variable background intensity, faint characters, smudge, low contrast, bleed-through). Overall, the H-DIBCO testing dataset consists of 10 handwritten document images for which the associated ground truth was built for the evaluation. A representative example of the dataset along with the associated ground truth image is shown in Fig. 1 (a)-(b).

B. Evaluation Results

The evaluation was based upon the four distinct measures presented in Section III. At Table I, for each encountered measure, the detailed performance of each algorithm along with the widely used binarization techniques of Otsu [7] and Sauvola et. al. [15], is given. The final ranking as shown in

Table I ('Score') was calculated by firstly, sorting the accumulated ranking value for all measures for each test image. The summation of all accumulated ranking values for all test images denote the final 'Score' which is shown in Table I. Let $R^i(j,m)$ be the i^{th} method concerning the j^{th} image when using the m^{th} measure. Then, for each binarization method i , the final Score S_i is given by Eq. (5):

$$S_i = \sum_{j=1}^K \sum_{m=1}^L R^i(j,m) \quad (5)$$

where K is the number of images used in the evaluation (i.e. $K=10$) and L is the number of the evaluation metrics (i.e. $L=4$).

To measure the processing time required by each method, we used only one image ("H01") from the testing dataset with an average size. The computer used was a laptop with i7 processor (4 real cores) at 2.20 GHz and 8 GB of Ram memory. It should be noted that the participants were not advised to optimize their algorithms in terms of execution time. Hence, the processing time is not considered as an evaluation criterion.

TABLE I. DETAILED EVALUATION RESULTS FOR ALL METHODS SUBMITTED TO H-DIBCO 2014.

Rank	Method	Score	FM (%)	F_{ps} (%)	PSNR	DRD	Time (sec)
1	6	66	96.88	97.65	22.66	0.902	17.43
2	2	71	96.63	97.46	22.40	1.001	7.23
3	5	149	93.35	96.05	19.45	2.194	16.20
4	1	182	89.24	90.71	18.49	4.502	21.19
5	4	193	89.77	90.98	18.46	4.227	14.84
6	3.b	231	86.24	91.46	17.43	5.011	4.65
7	7	257	79.15	85.72	15.92	7.278	0.30
8	3.a	291	79.54	88.25	15.87	6.639	4.61
-	Otsu	-	91.78	95.74	18.72	2.647	-
-	Sauvola	-	86.83	91.80	17.63	4.896	-

As it can be observed from Table I, the best performance is achieved by **Method 6** which has been submitted by Rafael Galvão Mesquita, Carlos Alexandre Barros de Mello, Ricardo Martins de Abreu Silva, and Péricles Barbosa Cunha de Miranda from Centro de Informática, Universidade Federal de Pernambuco, Brazil. Example binarization result of the winning method as well as of methods 2 and 5 are shown in Fig. 1. It is worth mentioning that the performance of the first two methods (No. 6 & No. 2) is quite similar. Indeed, after qualitative comparison, we observed that there are only small differences between the results of the two aforementioned algorithms (see Fig. 1c, 1d).

For the sake of clarity, the complete set of binarization results of each participating binarization method can be found in the following link:

<http://users.iit.demokritos.gr/~bgat/HDIBCO2014/results>

V. CONCLUSIONS

The H-DIBCO 2014 Handwritten Document Image Binarization Contest attracted 7 research groups that are currently active in document image analysis. The general objective of the contest is to identify current advances in handwritten document image binarization using meaningful evaluation performance measures. This objective is fulfilled by firstly, providing short descriptions with the main idea of each submitted algorithm, thus, enabling the interested researchers to be aware of the highly performing algorithms and be able to push forward the state of the art by a new more advanced approach. Secondly, the public availability of the testing dataset and the evaluation software permits further benchmarking and comparison with H-DIBCO 2014 results.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 600707 - tranScriptorium.

REFERENCES

- [1] B. Gatos, K. Ntirogiannis and I. Pratikakis, "ICDAR 2009 Document Image Binarization Contest (DIBCO2009)", 10th International Conference on Document Analysis and Recognition (ICDAR'09), Barcelona, Spain, pp. 1375-1382, 2009.
- [2] I. Pratikakis, B. Gatos and K. Ntirogiannis, "H-DIBCO 2010 – Handwritten Document Image Binarization Competition", 12th International Conference on Frontiers in Handwriting Recognition (ICFHR'10), Kolkata, India, pp. 727-732, 2010.
- [3] I. Pratikakis, B. Gatos and K. Ntirogiannis, "ICDAR 2011 Document Image Binarization Contest (DIBCO 2011)", 11th International Conference on Document Analysis and Recognition (ICDAR 2011), pp. 1506 - 1510, Beijing, China, 2011.
- [4] I. Pratikakis, B. Gatos and K. Ntirogiannis, "ICFHR 2012 Competition on Handwritten Document Image Binarization (H-DIBCO 2012)", 13th International Conference on Frontiers in Handwriting Recognition (ICFHR'12), pp. 813-818, Bari, Italy, 2012.
- [5] I. Pratikakis, B. Gatos and K. Ntirogiannis, "ICDAR 2013 Document Image Binarization Contest (DIBCO 2013)", 12th International Conference on Document Analysis and Recognition (ICDAR 2013), pp. 1471 - 476, Washington, DC, USA, 2013.
- [6] N. Howe, "Document Binarization with Automatic Parameter Tuning", International Journal on Document Analysis and Recognition (IJRAR), Volume 16, Issue 3, pp 247-258, September 2013.
- [7] N. Otsu, "A threshold selection method from graylevel histograms", IEEE Trans. Systems Man cybernet, Vol. 9 (1), pp. 62–66, 1979.
- [8] H. Ziaei Nafchi, R. Farrahi Moghaddam, and M. Cheriet, "Historical document binarization based on phase information of images", Lecture Notes in Computer Science: Asian Conference on Computer Vision (ACCV'12 Workshops), Springer Berlin / Heidelberg, vol. 7729, pp. 1–12, 2013.
- [9] R.G. Mesquita, C.A.B. Mello and L.H.E.V. Almeida, "A new thresholding algorithm for document images based on the perception of objects by distance", Integrated Computer-Aided Engineering, 21 , pp. 133-146, 2014.
- [10] M. Birattari, Z. Yuan, P. Balaprakash and T. Stutzle, "F-Race and Iterated F-Race: An Overview", Experimental Methods for the Analysis of Optimization Algorithms, pp. 311-336, June 2010.

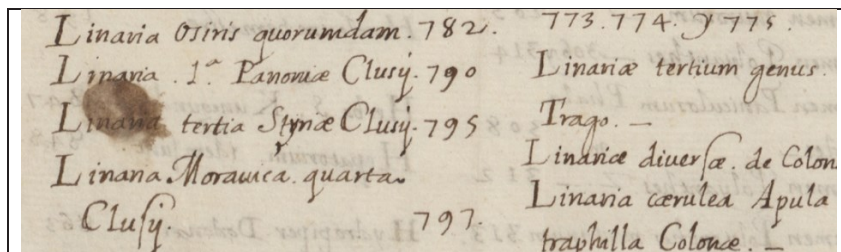
[11] K. Ntirogiannis, B. Gatos and I. Pratikakis, "Performance Evaluation Methodology for Historical Document Image Binarization", IEEE Transactions on Image Processing, vol. 22, no.2, pp. 595-609, Feb. 2013.

[12] 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.

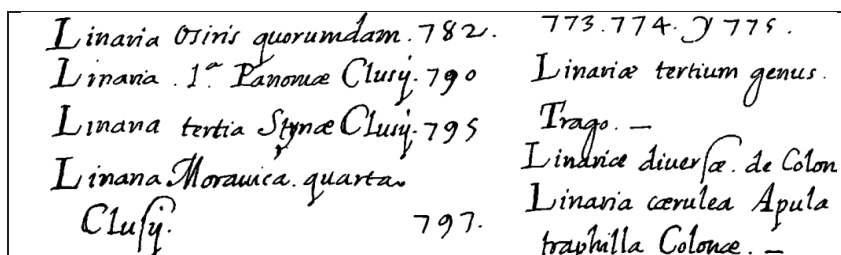
[13] <http://memory.loc.gov/ammem>

[14] <http://transcriptorium.eu>

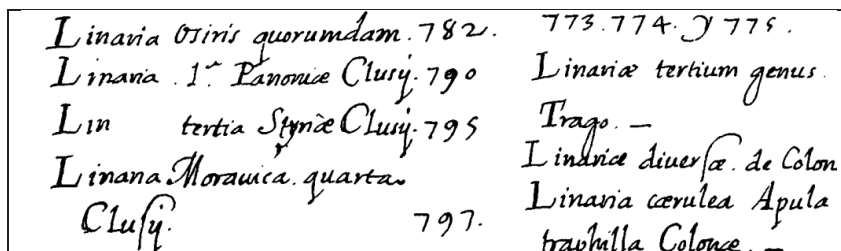
[15] J. Sauvola and M. Pietikainen. "Adaptive Document Image Binarization". Pattern Recognition. vol. 33. no. 2. pp. 225-236, 2000.



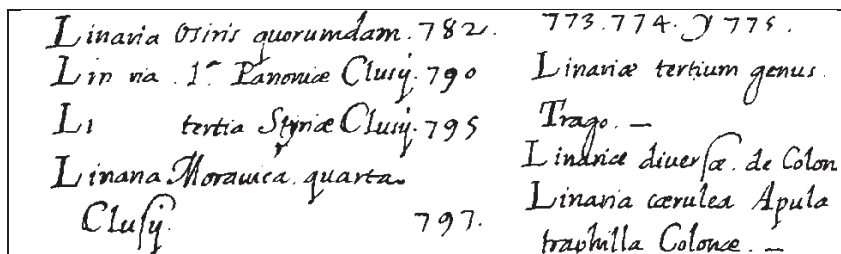
(a)



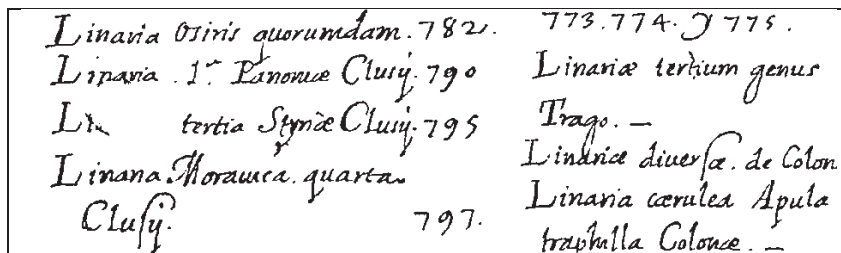
(b)



(c)



(d)



(e)

Figure 1. (a) Original image "H07"; (b) Ground truth image; (c) Binarization result from the H-DIBCO 2014 winning Algorithm 6; (d) Binarization result from the H-DIBCO 2014 Algorithm 2; (e) Binarization result from the H-DIBCO 2014 Algorithm 5.