Abstract—The first competition on the binarization of historical Persian documents and manuscripts (PHIBC 2012) has been organized in conjunction with the first Iranian conference on pattern recognition and image analysis (PRIA 2013). The main objective of PHIBC 2012 is to evaluate performance of the binarization methodologies, when applied on the Persian heritage images. This paper provides a report on the methodology and performance of the three submitted algorithms based on evaluation measures has been used.

Keywords—Document image processing, Historical document binarization, Persian heritage manuscripts, Binarization contest

I. INTRODUCTION

There are many old manuscripts and documents in the libraries and museums of Iran. Many of them include historically important data which needs automatic processing and reading. However, less attention has been made to preserve these valuable types of documents. The PHIBC 2012 is a primitive attempt toward evaluation of binarization methods, when applied on the Persian manuscripts.

Persian heritage documents and manuscripts following their similarity to Arabic documents [1] are in the form of handwritten document images. As usual, for handwritten documents, preservation of strokes and sub-strokes is of great interest. Previously, five Latin datasets of historical manuscripts has been publicly available for DIAR researchers [2]–[6]. PHIBC 2012 introduces the first dataset that developed for binarization of Persian heritage documents. The dataset used for PHIBC 2012 consisted of ten historical documents used with permission from "Documents and old manuscripts treasury of Mirza Mohammad Kazemaini (affiliated with Hazrate Emamzadeh Jafar), Yazd, Iran". The images in PHIBC 2012 suffered from various types of degradation, including bleed-through, deterioration of cellulose structure, faded ink and alien ink, among others [1]. For each image in the dataset, a ground truth is generated with a semi-automatic approach. At first, document is processed with phase congruency features used in [8], [9] to produce a rough binarized image. Then, the final ground truth is generated manually by human expert from the rough binary image produced in the first step. We will report the ground truth generation methodology in a dedicated report. Figure 1 shows sample images and corresponding ground truth images used in the PHIBC 2012.

II. DESCRIPTION OF METHODOLOGIES

In the Persian heritage image binarization competition (PHIBC 2012), three groups submitted three algorithms. The description of each methodology is provided by these groups and is as follows.

1- Su Bolan†, Tian Shangxuan‡, Lu Shijian‡ and Tan Chew Lim† (†School of Computing, National University of Singapore, and ‡Department of Computer Vision and Image Understanding Institute for Infocomm Research, Singapore).

There are four main steps in our proposed method. First, local image contrast which is evaluated by local maximum and minimum and local image gradient are combined using an exponential function with an adaptive factor. Second, The local image contrast is combined with the edge map to extract an accurate text character edge image. Third, the document image is binarized by a local threshold which is decided based on the constructed edge map and estimated stroke width. At last, some post-processing work is applied to produce better results.

2- Syed Ahsen Raza Ali Hamdani (National University of Sciences and Technology (NUST), Islamabad, Pakistan).

The algorithm is based on three processing steps: preprocessing, thresholding and postprocessing. In preprocessing, conditional noise removal and edge based processing is performed. Thresholding step involves a computation of final threshold for background and text segmentation based on an average value computed through multiple thresholds (based on 4 different Niblack inspired...
thresholding formulas). In final step of post processing, again conditional noise removal and constrained morphological operations are performed to get the final binarised image.

Proposed method is mainly based on fuzzy measures introduced in [11]. It finds a global threshold and apply it for whole image. The main advantage of this method against Otsu’s method, is its better classification of bleed-through in the case of documents that include large amount of this type of degradation. Afterward, an approach similar to Niblack and some post-processing processes are applied to improve the final binarization result.

III. EVALUATION MEASURES

For an objective comparison between submitted algorithms, six evaluation measures used [10]. These are F-Measure [12], pseudo F-Measure [13], peak signal-to-noise ratio (PSNR), distance reciprocal distortion metric (DRD) [14], misclassification penalty metric (MPM) [15] and negative rate metric (NRM) [2], [3].

Let’s $bwout$ and $GT$ denotes the binarized image and ground truth image, respectively. PSNR for binary images can be defined as:

$$MSE = \frac{\sum_{x=1}^{N} \sum_{y=1}^{M} [GT(x,y) - bwout(x,y)]}{N \times M} \quad (1)$$

$$PSNR = 10 \times \log\left(\frac{1}{MSE}\right) \quad (2)$$

F-Measure can be considered as an intelligent alternative for PSNR because it takes into account the number of foreground and background pixels. Let’s $TP$, $FP$, $FN$ and $TN$ denote the true positive, false positive, false negative and true negative, respectively. Recall, precision and F-Measure can be defined as:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$F - Measure = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (5)$$

Also, pseudo F-Measure is computed like F-Measure except that recall value is taken from skeletonized ground truth:

$$pseudo \ F - Measure = \frac{2 \times Recall_{skel} \times Precision}{Recall_{skel} + Precision} \quad (6)$$

Furthermore, NRM can be computed as:

$$NRM = \frac{NR_{FN} + NR_{FP}}{2} \quad (7)$$

where:

$$NR_{FN} = \frac{FN}{FN + FP}, \quad NR_{FP} = \frac{FP}{FP + TN} \quad (8)$$

DRD measures the distortion for all the $S$ flipped pixels as follows:
\[ DRD = \frac{\sum_{k=1}^{8} DRD_k}{NUBN}. \]  

(9)

where, \( DRD_k \) is the distortion of the \( k \)-th flipped pixel and it is calculated using a 5 \( \times \) 5 normalized weight matrix \( W_{Nm} \) \([14]\). \( DRD_k \) equals to the weighted sum of the pixels in the 5 \( \times \) 5 block of the ground truth \( GT \) that differ from the centered \( k \)-th flipped pixel at \( (x, y) \) in the binarization result image \( B \).

\[ DRD_k = \sum_{i=-2}^{2} \sum_{j=-2}^{2} [GT_k(i,j) - B_k(x, y)] \times W_{Nm}(i,j) \]  

(10)

NUBN is the number of the non-uniform (not all black not all white pixels) 8 \( \times \) 8 blocks in the \( GT \) image. MPM is a measure of how well the resulting image representing the contour of ground truth image and defined as:

\[ MPM = \frac{1}{2D} \left( \sum_{i=1}^{FN} d_{i,PN} + \sum_{j=1}^{FP} d_{j,FP} \right). \]  

(11)

where, \( d_{i,PN} \) and \( d_{j,FP} \) denote the distance of the \( i \)-th false positive and the \( j \)-th false negative pixel from the contour of the text in the ground truth image. The factor \( D \) is the sum of all the pixel-to-contour distances of the ground truth object.

A higher value for F-Measure, pseudo F-Measure and PSNR measures, indicate to better classification, while a lower value for DRD, MPM and NRM measures, shows better performance.

IV. EXPERIMENTAL RESULTS

In this section, characteristics of test images used in the PHIBC 2012 are described. Images in the PHIBD 2012 suffered from various types of degradation, include uneven illumination changes, various types of bleed-through, etc. Table I has summarized the degradation types of each document image used in the PHIBC 2012.

For each image in the dataset, best value of each measure between all of the methods is considered. A method with best value for a measure takes a score of 1, and other methods takes a fraction of 1 by a comparison with the best value. Since there are six evaluation measures and ten test images, we can compute the score of a method as:

\[ S_{k=1}^{10} = \sum_{i=1}^{6} \sum_{j=1}^{10} \left( \frac{\text{Best}_{i,j}}{\text{value}_{k,i,j}}, \frac{\text{value}_{k,i,j}}{\text{Best}_{i,j}} \right). \]  

(12)

where, \( k \) denotes the number of participators and \( \text{value} \) is the measure value obtained by a method. First fraction is used for those measures in which a lower value indicates to better score, and the second one is used for measures with inverse behavior. Finally, methods with higher scores take higher rank. Table II provides detailed experimental results of the binarization algorithms participated in PHIBC 2012 and some state-of-the-art binarization methods. Between three participated algorithms, the algorithm submitted by Su Bolan, Tian Shangxuan Lu Shijian and Tan Chew Lim (\( \dagger \)School of Computing, National University of Singapore, and \( \dagger \)Department of Computer Vision and Image Understanding Institute for Infocomm Research, Singapore) achieved the best performance. Figure 2 shows binarization results of the winner of PHIBC 2012.

V. CONCLUSION

This paper provides a report on the first Persian heritage image binarization competition (PHIBC 2012) which has been organized in conjunction with the first Iranian conference on pattern recognition and image analysis (PRIA 2013). The main objective of this competition is to evaluate the performance of the binarization methods, when applied on the historical Persian document images. The images used in PHIBC 2012 include wide range of degradation types and their associated ground truth are publicly available. Six evaluation measures has been used for comparison between submitted algorithms to PHIBC 2012. Based on the performance of the groups participated in the competition and state-of-the-art binarization

Fig. 2. Sample binarization results from the winner of PHIBC 2012.
methodologies, there is a lot of room for development of higher performance binarization algorithms.

ACKNOWLEDGMENT

The organizers of PHIBC 2012 would like to thank "Documents and old manuscripts treasury of Mirza Mohammad Kazemaini (affiliated with Hazrate Emamzadeh Jafar), Yazd, Iran" for providing us the images used in the PHIBC 2012.

REFERENCES

[1] R. Farrahi Moghaddam, M. Cheriet, M. M. Adankan, K. Filonenko, and R. Wiscovsky, “IBN SINA: A database for research on processing and understanding of arabic manuscripts images,” in Document Analysis Systems, 2010, pp. 11–18.

[2] B. Gatos, K. Nitrogiannis, and I. Pratikakis, “ICDAR 2009 document image binarization contest (DIBCO 2009),” in International Conference on Document Analysis and Recognition, 2009, pp. 1375–1382.

[3] I. Pratikakis, B. Gatos, and K. Nitrogiannis, “H-DIBCO 2010 handwritten document image binarization competition,” International Conference on Frontiers in Handwriting Recognition, pp. 727–732, 2010.

[4] ——, “ICDAR 2011 document image binarization contest (DIBCO 2011),” in International Conference on Document Analysis and Recognition, 2011, pp. 1506–1510.

[5] ——, “ICFHR 2012 competition on handwritten document image binarization competition,” in International Conference on Frontiers in Handwriting Recognition, 2012, pp. 813–818.

[6] R. Rowley-Brook, F. Pite, and A. Kokaram, “A ground truth bleed-through document image database,” in International Conference on Theory and Practice of Digital Libraries, 2012, pp. 185–196.

[7] “Special issue on recent advances in applications to visual cultural heritage,” IEEE Signal Processing Magazine, vol. 12, no. 1, pp. 234–778, 2008.

[8] H. Ziaei Nafchi, R. Farrahi Moghaddam, and M. Cheriet, “Historical document binarization based on phase information of images,” in Lecture Notes in Computer Science: Asian Conference on Computer Vision (ACCV’12 Workshops). Springer Berlin / Heidelberg, 2013, vol. 7729, pp. 1–12.

[9] H. Ziaei Nafchi and H. Rashidy Kanan, “A phase congruency based document binarization,” in IAPR International Conference on Image and Signal Processing, 2012, pp. 113–121.

[10] R. Farrahi Moghaddam and H. Ziaei Nafchi, “Objective evaluation of binarization methods,” MATLAB Central File Exchange, http://www.mathworks.com/matlabcentral/fileexchange/276522013.

[11] N. Vieira Lopes, P. A. Mogadouro do Couto, and H. Bustinse, “Automatic histogram threshold using fuzzy measures,” IEEE Transactions on Image Processing, vol. 19, no. 1, pp. 199–204, 2010.

[12] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” Information Processing and Management, vol. 45, pp. 427–437, 2009.

[13] K. Nitrogiannis, B. Gatos, and I. Pratikakis, “An objective evaluation methodology for handwritten image document binarization techniques,” in Document Analysis Systems, 2008, pp. 217–224.

[14] 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.

[15] D. P. Young and J. M. Ferriyman, “Pets metrics: On-line performance evaluation service,” in Proceedings of the 14th International Conference on Computer Communications and Networks, 2005, pp. 317–324.

[16] N. Otsu, “A threshold selection method from gray-level histograms,” IEEE Trans. Systems, Man, and Cybernetics, vol. 9, no. 2, pp. 62–66, 1979.

[17] R. Farrahi Moghaddam and M. Cheriet, “A multi-scale framework for adaptive binarization of degraded document images,” Pattern Recognition, vol. 43, no. 6, pp. 2186–2198, 2010.

[18] ——, “AdOtsu: An adaptive and parameterless generalization of otsu’s method for document image binarization,” Pattern Recognition, vol. 45, pp. 2419–2431, 2012.

[19] B. Su, S. Lu, and C. Tan, “Binarization of historical document images using the local maximum and minimum,” in Document Analysis Systems, 2010, pp. 159–166.

[20] N. Howe, “A laplacian energy for document binarization,” in International Conference on Document Analysis and Recognition, 2011, pp. 6–10.