ICDAR2017 Competition on Document Image Binarization (DIBCO 2017)

Ioannis Pratikakis[1], Konstantinos Zagoris[1], George Barlas[1] and Basilis Gatos[2]

[1]Visual Computing Group,
Department of Electrical and Computer Engineering,
Democritus University of Thrace,
67100 Xanthi, Greece
e-mail: {ipratika, kzagoris, gbarlas}@ee.duth.gr

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

***Abstract* — DIBCO 2017 is the international Competition on Document Image Binarization organized in conjunction with the ICDAR 2017 conference. The general objective of the contest is to identify current advances in document image binarization of machine-printed and handwritten document images using performance evaluation measures that are motivated by document image analysis and recognition requirements. This paper describes the competition details including the evaluation measures used as well as the performance of the 26 submitted methods along with a brief description of each method.**

Keywords – machine-printed, handwritten document image, binarization, performance evaluation

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. 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, the DIBCO series competitions is active since 2009 [1] which is dedicated to benchmarking binarization algorithms of not only handwritten document images, (e.g. the recent H-DIBCO 2016 [2] organized in conjunction with ICFHR 2016) but also both machine-printed and handwritten document images (e.g. the DIBCO 2013 [3] organized in conjunction with ICDAR 2013.

In this paper, we present the results of DIBCO 2017, organized in conjunction with ICDAR 2017, focused on the evaluation of machine-printed and handwritten document image binarization methods using document images with various complexity for which we created the binary image ground truth. The authors of submitted methods registered in the competition and downloaded representative document images along with the corresponding ground truth from previous DIBCO contests available in the competition's site (https://vc.ee.duth.gr/dibco2017/). In the sequel, all registered participants were required to submit their binarization executable. After the evaluation of all candidate methods, the testing dataset which comprises 10 machine-printed and 10 handwritten images, the associated ground truth as well as the evaluation software are publicly available at: http://vc.ee.duth.gr/dibco2017/benchmark.

II. METHODS AND PARTICIPANTS

Eighteen (18) research groups have participated in the competition with twenty six (26) distinct algorithms (Participant 5 submitted three algorithms while Participants 1, 3, 11, 15, 17, 18 submitted two algorithms). Brief descriptions of the methods are given in the following (the order of appearance is the chronological order of the algorithm's submission).

1) Brigham Young University, UT, USA (*Christopher Tensmeyer*)

This method performs binarization using an ensemble of 5 deep Fully Convolutional Networks (FCNs) that operate over multiple image scales, including full resolution. The networks take in 3 input components: (i) raw grayscale input image, (ii) The Howe binarization of the image [4], and (iii) Relative Darkness (RD) features [5] densely computed over the image. RD features are simply a count how of many pixels in a neighborhood are darker, lighter, or similar than the central pixel, where darker, lighter, and similar are determined by a threshold. The RD features were trained by using a 5x5 window with a threshold of 10 (i.e. darker pixels are those with intensity at least 10 below that of the central pixel). As a side note, FCNs do quite well when trained on just the raw grayscale images, so these additional features should not be viewed as critical to success of this method, but do empirically improve quality of results ~1%. Raw predicted probabilities are averaged per-pixel across the 5 FCNs in the ensemble. The resulting average probabilities are post-processed using a Densely Connected Conditional Random Field (CRF) with Gaussian edge potentials [6] (Method 1a). Inference is done using the mean-field approximation for 5 iterations. A variation of the algorithm without post-processing has also been tested (Method 1b).

2) School of ICT, Griffith University, Australia; CVPR Unit, Indian Statistical Institute, India; School of Software, University of Technology Sydney,

2379-2140/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDAR.2017.228

Australia (*Chandranath Adak; Bidyut B. Chaudhuri; Michael Blumenstein*)

This method is mainly motivated by [7] and [8] for combining local and global adaptive binarization. The foreground (object ink pixels) is separated from the background through an inpainting-based background estimation, followed by image normalization. Here we used a hybrid sparse representation-based inpainting method. The inpainting mask is the dilated Niblack form as used in [7]. The global binarization is performed on the normalized image as like [7]. The local binarization is performed as used in [8] by an adaptive image contrast followed by ink-edge pixel detection and local threshold estimation. The local and global binarized output is combined and post-processed by small component removal.

3) **Hubei University of Technology, Wuhan, P.R. China** (*XIONG Wei, XU Jing-jing, ZHAO Shi-yun, DONG Cong-xiang, XIAO Zi-yu, LI Min*)

Method a: This method comprises three main steps. First, morphological bottom-hat transform is carried out to enhance the document image contrast, and a disk-shaped mask is used to apply morphological operation whose size is determined by the stroke width transform (SWT) [9]. Second, Howe's binarization method [4] based upon graph cut the Laplacian energy minimization with Canny edge detection is then performed on the enhanced document images. Finally, image post-processing is adopted to produce better results.

Method b: This method consists of four main steps. First, a bilateral filter is a non-linear, edge- preserving and noise-reducing smoothing filter applied to the input image. Second, the local image contrast and local image gradient are combined [10] to indicate text stroke edge pixels. Third, the document image is binarized by a locally adaptive thresholding method, and the neighborhood window size is determined by the SWT. Finally, image post-processing is adopted to produce better results.

4) **Jadavpur University, India** (*Showmik Bhowmik, Bishwadeep Das, Ram Sarkar*)

This is a game theory-inspired binarization technique which has two major steps namely background separation and binarization. This method initially takes a grayscale image as input and tries to estimate as well as eliminate the background from the input image. For the purpose of background separation, an 'inpainting' method is applied following the work described in [7]. This process, which also helps eliminating the noise present in the input image, produces a background separated image for which the binarization is performed. In the next step, a two-player game, inspired by game theory is implemented at the pixel level. For the implementation of the game, an 3×3 overlapping window in scanned over the separated image. For each window, the central pixel is considered as the first player and rest of the eight pixels together considered as the second player. From this two-player game the Nash equilibrium is computed. The purpose of conducting this game is to extract feature for the central pixel present in each overlapping window under consideration. From this game, payoff for the central pixel, at Nash equilibrium state, is considered as a feature. Along with that, the central pixel value adjusted with the background contrast and the minimum value of the 3x3 overlapping window (excluding the central pixel) are considered as other two features. Based on these three features, all the pixels are grouped into two clusters, i.e. foreground and background. For grouping the pixels, K-means clustering algorithm is applied. As K-means algorithm suffers from initial point selection, in this work, initial cluster centers are fixed dynamically for each input image.

5) **Document Image and Pattern Analysis (DIPA) Center, Islamabad, Pakistan** (*Syed Ahsen Raza*)

Method a: The initial and modified phase of this algorithm is based on some basic steps like: preprocessing, thresholding and postprocessing. In preprocessing, conditional noise removal is done using a cascade of filtering operations followed by edge-based processing. The thresholding step involves the computation of final threshold for background and text segmentation based on an average value computed through multiple thresholds (based on 2 different Niblack inspired thresholding formulas). Computation of final threshold is an iterative process. In final step of post processing again conditional noise removal and constrained morphological operations are performed to get the final binarized image.

Method b: The proposed method for handwritten documents binarization is based on three main steps. First, conditional noise removal is performed based on the aspect ratio of the noise in the image. In the next step, actual binarization is performed using the modified version of Niblack thresholding algorithm. At third and final step again conditional noise removal procedure is performed using a mix of noise removal filters. This step is carried out to preserve the information of interest and discard unwanted artifacts.

Method c: This procedure is based on a mixture of image processing operation. In first step, image is prepared for actual binarization by noise removal using a customized noise removal filter. In the binarization step, a customized version of local adaptive binarization is used to compute a threshold in an iterative process. In third and final step, conditional and adaptive noise removal along with some morphological operations is performed to generate the final image.

6) **Northwest MinZu University, China** (*Zhenjiang Li*)

The whole algorithm is divided into three steps: In the first step, character stroke edges are extracted as appeared in [8] for which the final edge map is produced after filtering out small areas. In the second step, for each point, the Sauvola method is used to compute the local threshold, in which K is adaptively acquired, and the value is the mean of the contrast of the edge points in the neighborhood of the current pixel. In the final step, there are 2 conditions to judge a point as the foreground spot. First, the pixel value is smaller than the local threshold. Secondly, the number of edge points in the neighborhood of the current point is larger than the diameter of the neighborhood.

7) **University of Alicante, Spain,** (*Jorge Calvo-Zaragoza and Antonio Javier Gallego*)

This method views the image binarization problem as a two-class classification task at pixel level. It relies upon the use of a convolutional auto-encoder devoted to learning an end-to-end map from an input image of a fixed size to its selectional output, in which activations indicate whether the pixel must be classified as foreground or background. Once trained, documents can therefore be binarized by parsing them through the model and applying a threshold. For training, the dataset provided by previous DIBCO editions is considered. The full configuration consists of 3 encoding layers of 3x3 convolutions with 120 filters and Rectifier Linear Unit activations, followed by 2x2 max-pooling operations. The decoding function replicates the encoding stage but replacing pooling by up-sampling. Batch normalization and dropout units of 20 % are included in each convolutional block. Our model accepts fixed-size images of 256x256 but document images can be larger, and also variable in size. These cases can be easily processed by dividing the input images into equal pieces of those dimensions, and then combining the independent outputs provided. To boost the performance a pre-processing of the input images is considered as follows. The image is inverted so that higher greyscale values indicate a higher probability of being classified as foreground. After this, the mean value of the training images is computed and subtracted to each pixel, keeping 0 when negative. Finally, a min-max filter is applied so as to make the pixel values be closer to either the maximum or the minimum. More details about the operation of this method can be found in [10].

8) **Institute of Automation, Chinese Academy of Sciences, Beijing, P.R, China** (*FuxiJia, Cunzhao Shi, Kun He, Chunheng Wang and Baihua Xiao*)

This is an effective approach for the local threshold binarization of degraded document images. We utilize the structural symmetric pixels (SSPs) to calculate the local threshold in neighborhood and the voting result of multiple thresholds will determine whether one pixel belongs to the foreground or not. The SSPs are defined as the pixels around strokes whose gradient magnitudes are large enough and orientations are symmetric opposite. The compensated gradient map is used to extract the SSP so as to weaken the influence of document degradations. To extract SSP candidates with large magnitudes and distinguish the faint characters and bleed-through background, we propose an adaptive global threshold selection algorithm. To further extract pixels with opposite orientations, an iterative stroke width estimation algorithm is applied to ensure the proper size of neighborhood used in orientation judgement. At last, we present a multiple threshold vote based framework to deal with some inaccurate detections of SSP.

9) **École nationale supérieure d'informatique (ESI), Algiers, Algeria** (*Omar Boudraa and Walid Khaled*)

The proposed algorithm is based on three steps, namely, preprocessing, hybrid binarization and post-processing. In preprocessing, weak contrast is enhanced using CLAHE (Contrast Limited Adaptive Histogram Equalization) method [11]. The binarization step involves a hybridation between three well-known thresholding algorithms (OTSU [12], Multilevel OTSU [13] and Niblack Method [14]) while local average contrast value represents the decision criterion. In the final step, post processing comprises noise removal by connected components analysis and morphological operations which aim to improve the final binarized image [15].

10) **Smart Engines Ltd, Moscow, Russia** (Dmitrii Ilin, *Pavel Bezmaternykh, Dmitry Nikolaev*)

For this binarization contest we used U-Net convolutional network architecture for accurate pixel classification. Such a network can be trained from very few images and outperforms in many domains. Strong usage of data augmentation results in effective utilisation of provided training samples. Due to strong restriction of contest images model small amount of augmentation strategies were applied. Instead of sliding window tiling coverage can be used for effective local threshold application.

11) **Institute of Mathematics and Statistics (IME), University of São Paulo (USP), Brasil** (*Igor dos Santos Montagner, Mateus Espadoto, Nina S. T. Hirata, Nury Yuleny Aorsquipa and Roberto Hirata Jr.*)

<u>Method 1</u>: This algorithm comprises the following stages: (i) Document image conversion to grayscale intensities, (ii) Noise removal from the background. In particular, in order to get uniform background brightness, a normalization per percentile is applied. It is made by removing pixel values in the 0.2 percentage range at the beginning and at the end of the histogram, after that, the intensity profile is normalized. That improves the contrast between the letters and the background. Furthermore, a closing operation process is executed to remove small dark objects with only minor changes to bigger ones, in that procedure we use a disk of radius 5 as the structuring element. Finally, we divide, pixel by pixel, the original image by the background image obtained, (iii) Stroke width detection which is obtained by a variation of [9], (iv) false characters removal and true positive ones filling. In particular, small holes in the foreground (text) are filled using closing morphological operator and reduce the boundary noise in the characters of the text. Then opening operator is applied in order to remove small objects from the background. Finally, (v) an area

opening operator is used to remove objects smaller than 80 pixels, which are considered as noise.

Method 2: The proposed method is applied on image patches and consists of three steps: (i) applying morphological TopHat with structuring elements of varying sizes followed by Otsu's and Yen's thresholding separately, to determine which combination of structuring element and thresholding algorithm minimizes the difference between the noisy patch and the ground truth; (ii) training a classifier to identify the best transformation combination for each patch, according to the data obtained in the previous step; (iii) classifying patches of new images to apply the transformation estimated as the best one.

12) **Chonnam National University, Gwangju, South Korea** (*Quang Nhat Vo, Gueesang Lee, Soo-Hyung Kim, Hyung-Jeong Yang*)

The binarization method is mainly based on the Deep Supervised Network (DSN) [16][17]. We have developed a DSN model with a multiscale structure to learn text-like features from document images itself to classify text and background from degraded document images. Specifically, we consider three properly designed DSN architectures: DSN_C3, DSN_C4, and DSN_C5 that contain three, four, and five groups of convolutional layers, respectively. Each DSN structure is trained independently using document image patches as input and binary maps as ground truth. The target of our design is to predict the foreground maps at three different feature levels. We observe that predicted maps produced at end layers have fewer noises in the background due to the construction of high-level features. However, the detail of the text is lost after pooling layers. On the other hand, predicted maps generated at first layers have clearer text strokes but contain more background noises. Therefore, a better result is achieved by integrating the output of three DSNs. Three foreground maps are predicted for each image patch using our proposed DSN architecture.

13) **Blekinge Institute of Technology, Karlskrona, Sweden** (*Florian Westphal, Håkan Grahn, Niklas Lavesson*)

This approach follows the general idea of Afzal et al. [18] to train recurrent neural networks (RNNs) for image binarization. Similar to their approach, we divide each image into smaller patches for faster processing and process each patch starting from each of its four corners, converting it into a sequence of input pixels. In contrast to [18], we use Grid Long Short-Term Memory (Grid LSTM) cells [19] as RNN cells, which allow us to take more image context into consideration when processing a particular input. By using 4-dimensional Grid LSTM cells, our approach processes per time step the input pixels to be binarized, the input pixels of the row above the currently processed one, as well as the surrounding pixels presented in a scaled down version of the currently processed patch. The configured 5 Grid LSTM cells per dimension yield for the 4 different processing directions 20 different views on the processed patch in the output dimension. This dimension is configured as priority dimension, so that its output at the current time step is directly influenced by the outputs of the other dimensions at that time step.

These 20 different views are then further processed by a bidirectional LSTM layer consisting of 2-dimensional Grid LSTM cells, producing two binarized images. Those two images are then combined into one final output image by a weighted sum. The described neural network architecture is trained using weighted cross entropy loss using only the images from DIBCO 2009, H-DIBCO 2010, DIBCO 2011, H-DIBCO 2012, H-DIBCO 2014 and H-DIBCO 2016. The images from DIBCO 2013 are used as validation dataset to determine when training should be stopped.

14) **Jinling Institute of Technology, Jiangsu Province, P.R. China** (*CHEN Sheng-guo, HU Yong*)

This method is an adaptive modification to Sauvola's method [20]. The idea of Sauvola' s method is to vary the threshold over the image, based on the local mean and local standard deviation computed in a small neighborhood of each pixel, and the dynamic range of standard deviation, R. For some images, Sauvola's method is sensitive to the value of parameter k. Our method calculates the parameter k adaptively based on the normalized k-values of a gray image which often approximate a normal distribution. Our method directly gets the binary image from the k-values with a threshold which is calculated by considering the mean value, and the standard deviation of the normalized k-values.

15) **Larbi Tebessi University, Tebessa, Algeria and Ecole nationale Superieure d'Informatique (ESI), Algiers, Algeria** (*Abdeljalil Gattal, Abdenour Sehad, Youcef Chibani*)

Method a: This method is based on the texture features extracted from the Basic Image Features (BIFs)[21][22]. Every location in the image is categorized into one of the seven local symmetry classes according to local symmetry type, which can be flat, slope, dark rotational, light rotational, dark line on light, light line on dark or saddle-like. In our case, the BIFs are generated using the scale parameter σ which is set to 0.5 and different values of the parameter $\varepsilon \in \{0.13, 0.14, 0.15, ..., 0.20\}$ from the handwritten document image. In the next step, each BIF image with different parameter e is binarized by converting the six local symmetry type (slope, dark rotational, light rotational, dark line on light, light line on dark or saddle-like) to black pixels. These binarized images are combined to create the final binarized image after applying a post-processing step.

Method b: This work is mainly based on the well-known Sauvola's method. It consists of making the method independent of parameters such as k and the size of the sliding window. In Step 1, the input grayscale image is binarized with Sauvola's algorithm by setting up k=0.2 and selecting different size {11x11,21x21,31x31,..,101x101} of window pixels. In step 2, each binarized image with different size of window is compared to the binarized image with Niblack's method by computing a metrics as F-measure and then the best binarized image is selected.

16) Aliah University, Kolkata, India (*Tauseef Khan, Payel Sengupta, Ayatullah Faruk Mollah*)

The proposed method consists of the following steps: (i) Pre-processing technique (median filter) is applied to the given input image, to remove some unwanted noise from the image, (ii) After that a hybrid adaptive binarization technique is applied using fuzzy membership (iii) Finally, some post-processing methods (i.e. morphological operation, Connected Component Analysis) have been applied to eliminate small and isolated structures and try to remove the connections to the background to improve the overall result.

17) Institute of Automation, Chinese Academy of Sciences, Beijing, P.R. China, (*Yan-Ming Zhang, Jun-Yu Ye, Xu-Yao Zhang, Cheng-Lin Liu*)

We treat image binarization as a classification problem and use a deep fully convolutional neural network (FCN) to label each pixel as background/foreground. More specifically, our method consists of three steps: (i) We convert 3-channel color images into 1-channel gray images by the standard method: gray = 0.2989*r + 0.5870*g + 0.1140*b, and then apply a simple image normalization to scale the input to [-1, +1], (ii) FCN is trained from scratch. The training set comes from previous DIBCO competitions, and contains 86 images labeled at the pixel level. We have tried several data augmentation methods, but observed no noticeable improvement. Our two submissions are different in the model structure: The first model is like VGGNet which only involves convolutional Layers, while the second one contains two convolutional blocks which are composed of multi-scale convolutions, layer normalization and residual connections. To keep the resolution of feature maps, we make the strides of all operations in the networks equal to 1, (iii) We designed a simple procedure to filter out small isolated components.

18) Uppsala University, Uppsala, Sweden, (*Ekta Vats, Anders Hast and Prashant Singh*)

This team has proposed two variations of document image binarization methods: (i) a two band pass filtering approach (Method 18a), and (ii) an automatic adaptive two band pass filtering based on Bayesian optimization (Method 18b). Both document image binarization methods are primarily based upon a two band pass filtering approach for background removal developed by the team. The idea is to use a high frequency bandpass filter to separate the fine detailed text from the background. Since some noise is captured as well, a low frequency bandpass filter is then used as a mask in order to remove great parts of that noise. 'Method 18a' performs background removal on noisy input images using the two band pass filter and then grayscale image binarization using Otsu's method. Since there are certain parameter values such as mask size, text size, window size and threshold to be set dynamically, 'Method 18b' employs Bayesian optimization for automatic parameter selection. This method performs an adaptive band pass filtering within an optimum window size, mask size and threshold values that are selected automatically using Bayesian optimization.

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 image analysis and recognition. These measures consist of (i) F-Measure (*FM*), (ii) pseudo-FMeasure (F_{ps}), (iii) *PSNR* and (iv) Distance Reciprocal Distortion (*DRD*).

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-FMeasure

Pseudo-FMeasure $\boldsymbol{F_{ps}}$ is introduced in [23] and it uses pseudo-Recall $\boldsymbol{R_{ps}}$ and pseudo-Precision $\boldsymbol{P_{ps}}$ (following the same formula as F-Measure). The pseudo Recall/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.

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. The higher the value of *PSNR*, the higher the similarity of the two images. Note 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 to measure the visual distortion in binary document images [24]. It properly correlates with the human visual perception and it measures the distortion for all the $\boldsymbol{S}$ flipped pixels as follows:

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

where NUBN is the number of the non-uniform (not all black or white pixels) 8x8 blocks in the GT image, and DRD_k is the distortion of the k-th flipped pixel that is calculated using a 5x5 normalized weight matrix $\boldsymbol{W_{Nm}}$ as defined in [24]. DRD_k equals to the weighted sum of the pixels in the 5x5 block of the $\boldsymbol{GT}$ that differ from the centered k^{th} flipped pixel at (x,y) in the binarization result image $\boldsymbol{B}$ (Eq. 4).

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

IV. EXPERIMENTAL RESULTS

The DIBCO 2017 testing dataset consists of 10 machine-printed and 10 handwritten document images for which the associated ground truth was built manually for the evaluation. The selection of the images in the dataset was made so that representative degradations appear. The machine-printed documents of the dataset originate from collections that belong to the IMPACT project [25], while the handwritten document images originate from collections that belong to READ project [26]. Example testing images of handwritten and machine-printed images are shown in Fig. 1(a) and Fig. 2(a), respectively.

The evaluation was based upon the four distinct measures presented in Section III. The detailed evaluation results along with the final ranking are shown in Table I. The final Ranking was calculated after first, 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 at column "Score". Additionally, the evaluation results for the widely used binarization techniques of Otsu [12] and Sauvola [20] are also presented. At Table II, III, we provide the performance of each algorithm for only the handwritten and the machine-printed document images, respectively. Overall, the best performance is achieved by ***Method 10*** which has been submitted by ***Dmitrii Ilin, Pavel Bezmaternykh, Dmitry Nikolaev*** affiliated to **Smart Engines Ltd, Moscow, Russia**. The binarization results of this algorithm for each image of the testing dataset is shown in Fig. 1(b) and Fig. 2(b) in the case of handwritten and machine-printed document images, respectively.

V. CONCLUSIONS

Taking into account the participating methods and their performance, several conclusions are drawn that could provide a fruitful feedback for the research community working on improving both machine-printed and handwritten document image binarization. First of all, a dominant feature of DIBCO 2017 is the participation of an increased number of supervised approaches that mostly ranked at high positions. Needless to mention that the highest ranked method has been a supervised one. It is worth noting that the document images required for the training stage of the supervised approaches were gathered from the publicly available previous years' DIBCO datasets. As has also been observed in previous years' DIBCO challenges, standard approaches like the global Otsu algorithm [15] and the locally adaptive Niblack [27] and Sauvola algorithm [17] are fully involved in the newly proposed approaches. Furthermore, methods that appeared to be highly performant in previous years' DIBCO challenges, have been used from this year's submitted methods either as a whole or by means of a particular component of the algorithm. Typical examples are the methods of Howe [4] and Su et al. [8]. It is also worth mentioning that most of the methods use an explicit post-processing stage which in certain cases is coupled in the binarization pipeline with the use of a pre-processing stage, proving that those stages have a major impact on the success of the binarization process. Last but not least, it should be encountered a tendency of the submitted approaches towards selecting a strategy which used stages that dominate the research in document image binarization. Such stages concern the use of the stroke width transform [9] or background estimation approaches.

TABLE I. DETAILED EVALUATION RESULTS FOR ALL METHODS SUBMITTED TO DIBCO 2017.

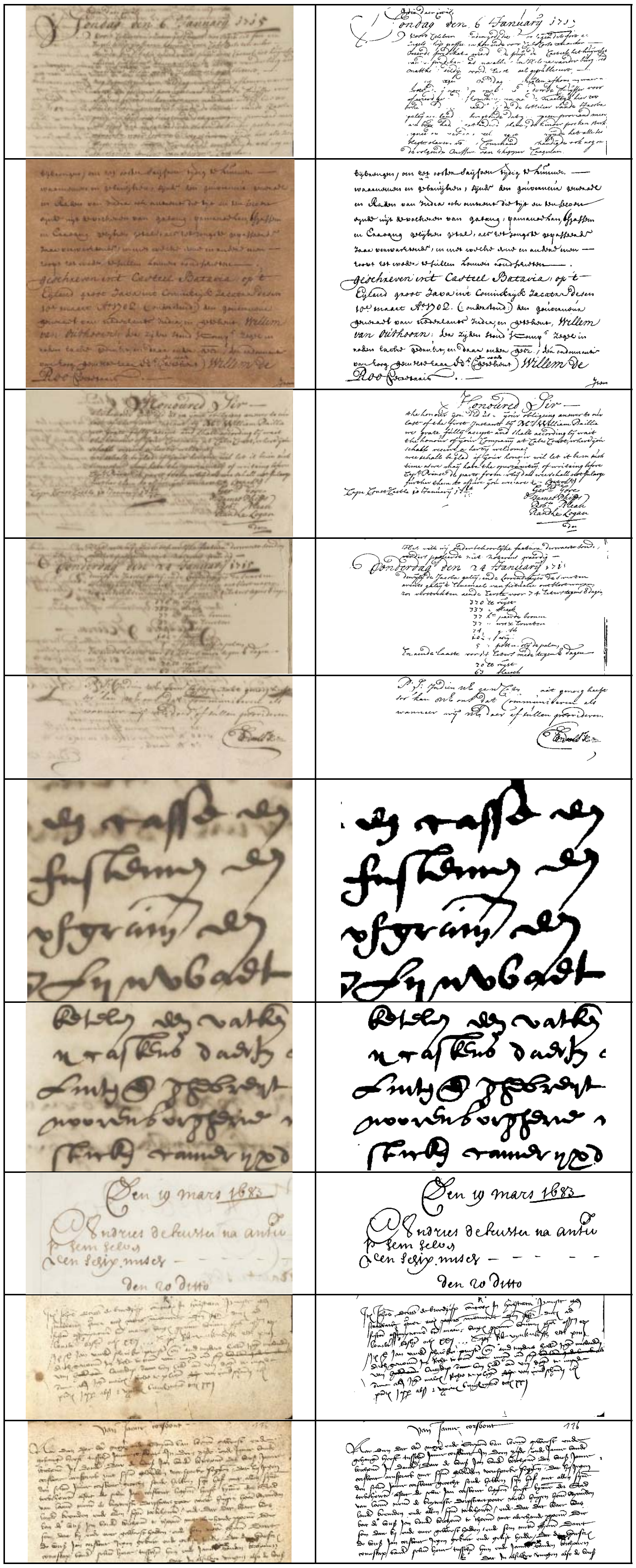
Rank	Method	Score	FM	F_{ps}	PSNR	DRD
1	**10**	**309**	**91.04**	**92.86**	**18.28**	**3.40**
2	17a	455	89.67	91.03	17.58	4.35
3	12	481	89.42	91.52	17.61	3.56
4	1b	529	86.05	90.25	17.53	4.52
5	1a	566	83.76	90.35	17.07	4.33
6	17b	608	88.37	89.59	17.10	4.94
7	3a	635	89.17	89.88	17.85	5.66
8	7	669	86.39	88.82	16.89	4.55
9	3b	806	86.49	87.02	16.33	6.57
10	13	831	86.33	89.64	16.50	4.75
11	8	895	85.34	86.06	16.25	8.18
12	2	904	84.39	87.41	15.74	7.54
13	9	1131	82.44	86.28	15.07	7.89
14	11a	1149	83.93	87.54	15.43	6.63
15	16	1183	79.62	84.32	15.09	6.77
16	4	1234	77.20	77.87	14.10	22.95
17	18a	1330	79.49	83.53	14.38	8.98
18	5a	1357	81.24	80.41	14.50	10.08
19	6	1423	78.60	83.79	14.24	9.68
20	18b	1452	78.72	79.87	14.14	10.32
21	14	1515	76.61	79.94	13.53	13.94
22	15b	1560	66.55	67.25	13.73	10.33
23	5b	1614	76.90	78.73	13.26	13.33
24	5c	1776	69.82	73.82	12.88	11.27
25	11b	1797	68.63	69.72	12.77	13.72
26	15a	1871	63.45	60.59	12.38	13.63
-	Otsu	-	77.73	77.89	13.85	15.54
-	Sauvola	-	77.11	84.1	14.25	8.85

TABLE II. DETAILED EVALUATION RESULTS FOR ALL METHODS SUBMITTED TO DIBCO 2017 FOR HANDWRITTEN DOCUMENT IMAGES.

Rank	Method	Score	FM	F_{ps}	PSNR	DRD
1	10	135	91.04	92.86	18.28	3.40
2	3a	181	89.17	89.88	17.85	5.66
3	17a	216	89.67	91.03	17.58	4.35
4	12	233	89.42	91.52	17.61	3.56
5	17b	287	88.37	89.59	17.10	4.94
6	1b	314	86.05	90.25	17.53	4.52
7	8	319	85.34	86.06	16.25	8.18
8	1a	357	83.76	90.35	17.07	4.33
9	3b	393	86.49	87.02	16.33	6.57
10	7	400	86.39	88.82	16.89	4.55
11	2	449	84.39	87.41	15.74	7.54
12	13	469	86.33	89.64	16.50	4.75
13	11a	510	83.93	87.54	15.43	6.63
14	16	577	79.62	84.32	15.09	6.77
15	9	632	82.44	86.28	15.07	7.89
16	4	651	77.20	77.87	14.10	22.95
17	5a	658	81.24	80.41	14.50	10.08
18	5b	685	76.90	78.73	13.26	13.33
19	6	743	78.60	83.79	14.24	9.68
20	18a	747	79.49	83.53	14.38	8.98
21	18b	783	78.72	79.87	14.14	10.32
22	15b	794	66.55	67.25	13.73	10.33
23	14	803	76.61	79.94	13.53	13.94
24	5c	845	69.82	73.82	12.88	11.27
25	15a	928	63.45	60.59	12.38	13.63
26	11b	931	68.63	69.72	12.77	13.72
-	Otsu	-	71.75	71.05	12.33	22.81
-	Sauvola	-	75.03	84.72	14.22	7.79

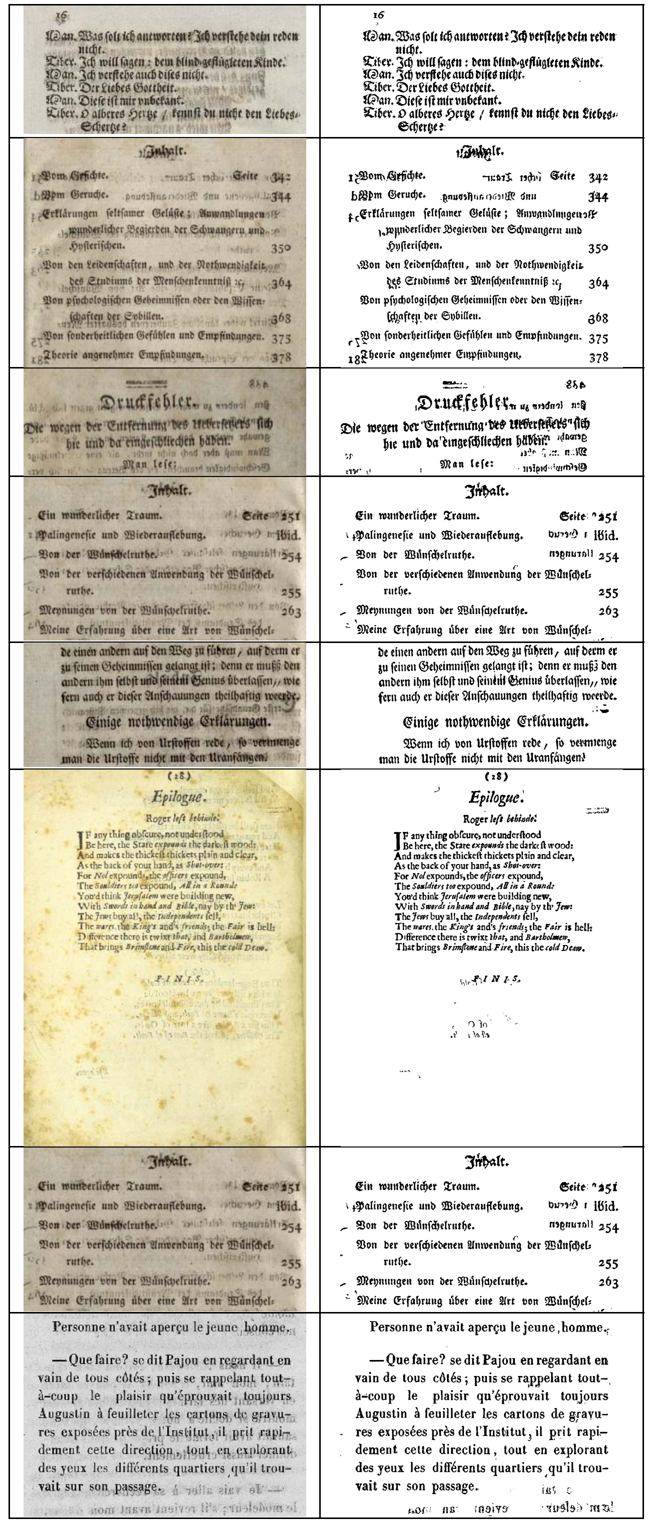
TABLE III. DETAILED EVALUATION RESULTS FOR ALL METHODS SUBMITTED TO DIBCO 2017 FOR MACHINE-PRINTED DOCUMENT IMAGES.

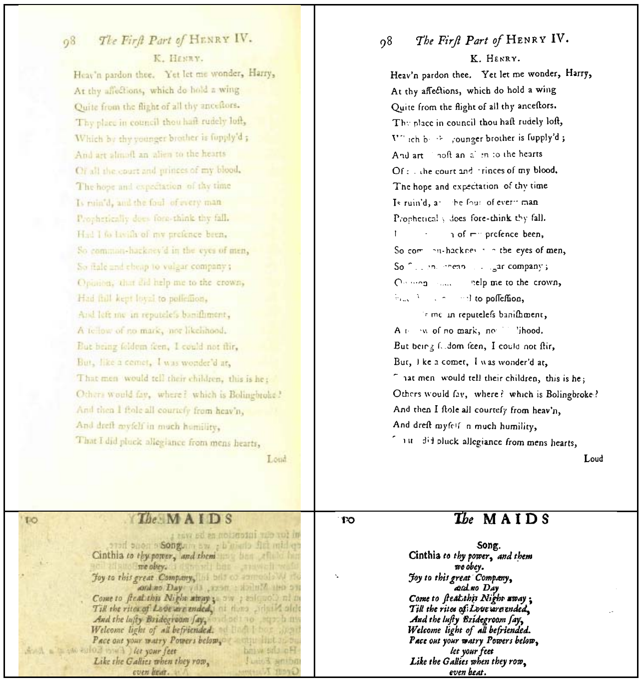
Rank	Method	Score	FM	F_{ps}	PSNR	DRD
1	10	174	91,04	92,86	18,28	3,40
2	1a	209	83,76	90,35	17,07	4,33
3	1b	215	86,05	90,25	17,53	4,52
4	17a	239	89,67	91,03	17,58	4,35
5	12	248	89,42	91,52	17,61	3,56
6	22	269	86,39	88,82	16,89	4,55
7	17b	321	88,37	89,59	17,10	4,94
8	13	362	86,33	89,64	16,50	4,75
9	3b	413	86,49	87,02	16,33	6,57
10	3a	454	89,17	89,88	17,85	5,66
11	2	455	84,39	87,41	15,74	7,54
12	9	499	82,44	86,28	15,07	7,89
13	8	576	85,34	86,06	16,25	8,18
14	18a	583	79,49	83,53	14,38	8,98
15	4	583	77,20	77,87	14,10	22,95
16	16	606	79,62	84,32	15,09	6,77
17	11a	639	83,93	87,54	15,43	6,63
18	18b	669	78,72	79,87	14,14	10,32
19	6	680	78,60	83,79	14,24	9,68
20	5a	699	81,24	80,41	14,50	10,08
21	14	712	76,61	79,94	13,53	13,94
22	15b	766	66,55	67,25	13,73	10,33
23	11b	866	68,63	69,72	12,77	13,72
24	5b	929	76,90	78,73	13,26	13,33
25	5c	931	69,82	73,82	12,88	11,27
26	15a	943	63,45	60,59	12,38	13,63
-	Otsu	-	83.7	84.73	15.36	8.26
-	Sauvola	-	79.19	83.47	14.29	9.91



(a) (b)

Fig. 1. (a) The DIBCO 2017 testing dataset of handwritten documents (b) Binarization results from the winner algorithm of DIBCO 2017.





(a) (b)

Fig. 2. (a) The DIBCO 2017 testing dataset of machine-printed documents (b) Binarization results from the winner algorithm of DIBCO 2017.

ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Union's H2020 Programme READ under grant agreement n° 674943.

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), July. 26-29, 2009, Barcelona, Spain, pp. 1375-1382.

[2] I. Pratikakis, B. Gatos and K. Ntirogiannis, "ICDAR 2013 Document Image Binarization Contest (DIBCO2013)", 12th International Conference on Document analysis and Recognition (ICDAR'13), August 25-29, 2013, Washington DC, USA

[3] I. Pratikakis, K. Zagoris, G. Barlas and B. Gatos, "ICFHR 2016 Handwritten Document Image Binarization Contest (H-DIBCO 2016)", 15th International Conference on Frontiers in Handwriting Recognition (ICFHR'16), October 23-26, 2016, Shenzhen, China, pp. 619-623, IEEE Computer Society Press.

[4] N. Howe, "Document Binarization with Automatic Parameter Tuning," International Journal on Document Analysis and Recognition, vol. 16, no. 3, pp. 247-258, 2013.

[5] Wu, Yue, et al., "Learning document image binarization from data", In IEEE Proceedings of International Conference on Image Processing (ICIP), 2016.

[6] Koltun, Vladlen. "Efficient inference in fully connected CRFs with Gaussian edge potentials" Adv. Neural Inf. Process. Syst 2.3 (2011): 4.

[7] K. Ntirogiannis, B. Gatos, I. Pratikakis, "A Combined Approach for the Binarization of Handwritten Document Images", Pattern Recognition Letters, vol. 35, pp. 3-15, 2014.

[8] B. Su, S. Lu, C.L. Tan, "Robust Document Image Binarization Technique for Degraded Document Images", IEEE Transactions on Image Processing, vol. 22, no. 4, pp. 1408-1417, 2013.

[9] B. Epshtein, E. Ofek, Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, 2010, pp. 2963-2970. doi: 10.1109/cvpr.2010.5540041

[10] J. Calvo-Zaragoza, and A.-J. Gallego, "A selectional auto-encoder approach for document image binarization", arXiv preprint arXiv:1706.10241 (2017).

[11] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization". In: P. Heckbert: Graphics Gems IV, Academic Press, 1994.

[12] N. Otsu, "A threshold selection method from gray-level histograms", IEEE Trans. Sys., Man., Cyber., vol. 9, 1979, p. 62–66

[13] D.Y. Huang, T.W. Lin, W.C. Hu, "Automatic Multilevel Thresholding Based on TwoStage Otsu's Method with Cluster Determination by Valley Estimation", Int. Journal of Innovative Computing, 2011, 7.

[14] K. Khurshid, I. Siddiq, C. Faure, and N. Vincent, "Comparison of niblack inspired binarization methods for ancient documents", In 16th Document Recognition and Retrieval Conference, DRR–2009, USA.

[15] S. Lu, B. Su, C.L. Tan, "Document image binarization using background estimation and stroke edges", In: International Journal on Document Analysis and Recognition (IJDAR), 13, 2010, pp. 303-314.

[16] C.Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," AISTATS, 2015.

[17] S. Xie and Z. Tu, "Holistically-Nested Edge Detection," ICCV, 2015.

[18] M.Z. Afzal, J. Pastor-Pellicer, F. Shafait, T. Breuel, A. Dengel, and M. Liwicki, "Document Image Binarization using LSTM: A Sequence Learning Approach", 3rd International Workshop on Historical Document Imaging and Processing, pp. 79-84, 2015.

[19] N. Kalchbrenner, I. Danihelka, and A. Graves, "Grid Long Short-Term Memory", Arxiv report, 2015-07-06 http://arxiv.org/abs/1507.01526

[20] J. Sauvola, M. Pietikainen, "Adaptive document image binarization", Pattern Recognition, Vol. 33, No. 2, pp. 225-236, 2000.

[21] L. Griffin, M. Lillholm, M. Crosier, and J. van Sande, "Basic image features (bifs) arising from approximate symmetry type", Scale Space and Variational Methods in Computer Vision, pages 343-355, 2009.

[22] Abdeljalil Gattal, Chawki Djeddi, Youcef Chibani, Imran Siddiqi, "Isolated Handwritten Digit Recognition Using oBIFs and Background Features", 12th IAPR International Workshop on Document Analysis Systems (DAS-2016), Santorini, Greece, 2016. pp. 305-310.

[23] 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, 2013.

[24] 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.

[25] IMPACT Project (http://www.impact-project.eu)

[26] READ project (http://read.transkribus.eu/)

[27] W. Niblack, "An introduction to Digital Image Processing", Strandberg Publishing Company, Birkeroed, Denmark, 1985.