

## ICFHR 2016 Handwritten Document Image Binarization Contest (H-DIBCO 2016)

Ioannis Pratikakis<sup>1</sup>, Konstantinos Zagoris<sup>1</sup>, George Barlas<sup>1</sup> and Basilis Gatos<sup>2</sup>

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

<sup>2</sup>Computational Intelligence Laboratory,  
Institute of Informatics and Telecommunications,  
NCSR “Demokritos”,  
15310 Athens, Greece,  
e-mail: bgat@iit.demokritos.gr

**Abstract** — H-DIBCO 2016 is the international Handwritten Document Image Binarization Contest organized in the context of ICFHR 2016 conference. The general objective of the contest is to identify current advances in document image binarization of handwritten document images using performance evaluation measures that are motivated by document image analysis and recognition requirements. This paper describes the contest details including the evaluation measures used as well as the performance of the 12 submitted methods along with a brief description of each method.

**Keywords** - handwritten document image, binarization, performance evaluation

### I. INTRODUCTION

Handwritten 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, following the success of DIBCO series competitions dedicated to handwritten document images, i.e. H-DIBCO 2010 [1], H-DIBCO 2012 [2], H-DIBCO 2014 [3] organized in conjunction with ICFHR 2010, 2012 and 2014, respectively, the follow-up of these contests, namely H-DIBCO 2016 is organized in the framework of ICFHR 2016. In this contest, we focused on the evaluation of handwritten document image binarization methods using a variety of scanned handwritten documents for which we created the binary image ground truth. The authors of submitted methods registered in the competition and downloaded representative samples along with the corresponding ground truth from previous DIBCO contests available in the competition’s site (<http://vc.ee.duth.gr/h-dibco2016/>). 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 handwritten images, the associated ground truth as well as the evaluation

software are publicly available at the following link: <http://vc.ee.duth.gr/h-dibco2016/benchmark>.

### II. METHODS AND PARTICIPANTS

Nine (9) research groups have participated in the competition with twelve (12) distinct algorithms (Participant 3 submitted three algorithms while Participant 7 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 approach employs a Fully Convolutional Network (FCN) [4] that takes a color image as input and outputs the probability that each pixel in the image is part of the foreground. An FCN is a Convolutional Neural Network (CNN) composed only of many convolution layers (no fully connected layers). As such, the FCN can take as input any sized image and return an appropriately sized output. In this work, the FCN uses no down-sampling and zero-padded convolution, so the output image is the same size as the input image.

Binarization can be viewed as a pixel-wise classification problem with two classes: foreground and background. The FCN is trained using input/output image pairs. The input is a 3-channel 256x256 RGB image. A single channel grayscale image can be converted to this format by copying the gray channel into each of the 3 RGB channels. The output is a single channel 256x256 image encoding per-pixel probabilities of foreground. The target output image is a binary image where foreground pixels have value 0 and background pixels have value 1. The loss function is the sum of the per-pixel cross-entropy between the predicted and target distribution. The architecture is composed of 6 convolution layers. Each convolution layer, except the last, has 16 learned kernels and is followed by Batch Normalization [5] and element-wise ReLU activation. The middle 4 layers also have residual connections [6]. The last convolution layer has a single kernel and uses an element-wise sigmoid operation to transform output values into proper probabilities. All kernels are square and their sizes (first to last) are 15, 11, 7, 7, 7, 1.

It takes a large number of images to train an FCN, far more than the 50 images provided by the competition. It is common for CNNs (and FCNs) to be first trained on an auxiliary task (e.g. ImageNet), and then fine-tuned on the task of interest. We pre-train our FCN using English and German handwritten parish records from the 1800s. The corresponding "ground truth" binarization is automatically generated using the binarization method of Wolf *et al.* [7]. While these target output binarizations do contain noise, they provide a good starting point for the FCN weights.

Fine-tuning is applied using the curated ground truth image pairs used in all previous DIBCO and HDIBCO competitions. This helps the FCN not make the same errors that the Wolf *et al.* binarization method does. For inference, the binarization is accomplished by a single forward pass through the network and thresholding the probabilities (at 50%). For memory efficiency, the image is fragmented into subimages, which are independently fed into the network and then reassembled into the output image.

## **2) Technion – Israel institute of Technology, Israel (Nati Kligler and Ayellet Tal)**

The proposed method is composed of three stages, where the novelty is a new pre-processing step:

### **(a) Pre-processing – Creating the Visibility Score Map:**

The image is considered as a 3D point set (X-Y-intensity). This set is linearly transformed from the 3D Euclidean space onto a spherical surface. When applying our specific transformation, it can be shown that *concavities on the sphere's surface correspond to text in the original image*. Then, in order to detect these *concavities*, we use of the *Hidden Point Removal (HPR)* operator [8]. This operator, which was originally aimed at detecting the visibility of point sets, is used here for the first time in image processing. It detects the concavities mentioned earlier, and hence detects the text. Intuitively, this is so since the HPR operator is proved to be less likely to find points at concavities as visible. The challenge is to define the viewpoint utilized in the HPR operator. The output of this stage is a *visibility score map*, which assigns each point (pixel) the probability that it resides in a concavity, i.e., whether it is a foreground (text) pixel. This map is utilized later instead of the given intensity map.

(b) **Binarization:** The preprocessing above is independent of the type of image to be binarized. For the task defined in the H-DIBCO contest, the best results were achieved using the method presented by [9], applied to the map of Stage (a).

(c) **Post-processing:** We use standard denoising on the result of Stage (b).

## **3) University of Bordeaux, France and Qatar University, Qatar (Yazid Hassaine, Abdelaali Hassaine and Somaya Al Maadeed)**

**Method 1:** This method was adapted from a technique for restoration of optical soundtracks of old movies [10] in which the text part is considered as the opaque region of the optical soundtrack.

**Method 2:** This method classifies the edges of the Otsu binarization method as true edges or wrong edges using the geometric features introduced in [11]. Regions are eliminated if the majority of their edges are classified as wrong edges.

**Method 3:** This method combines the two above methods and train them on all DIBCO databases as well as the QUWI handwriting database [12].

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

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.

## **5) Universidade Federal de Pernambuco (UFPE), Brasil (Leandro Henrique Espindola V. De Almeida and Carlos Alexandre Barros de Mello)**

The binarization algorithm submitted can be divided into two stages: stroke width detection and pixel classification. In the first stage, a rough background is calculated and used to highlight the textual components presented in the image. All textual components are evaluated separately and, for each pixel of the component, it is defined a particular stroke width value. After all foreground pixels get their stroke width evaluated, the calculation is extended for the background pixels, defining the stroke width matrix. In the second stage, the stroke width matrix is used to refine the previously calculated background, and to calculate the combined structural contrast image, that will be used as a marker for the final classification round, using a neighborhood window with size defined by the stroke width associated with the actual pixel.

## **6) Universidade Federal de Pernambuco (UFPE), Brasil (Edward Roe and Carlos Alexandre Barros de Mello)**

The binarization method makes use of a local image equalization process and an extension to the standard difference of Gaussians, (called XDoG). The binarization is achieved after three main steps: the first step removes undesirable degradation artifacts and enhances edges using the local image equalization and Otsu binarization algorithm. The second step uses global image equalization and XDoG edge detection operator to binarize the text. The final step combines the two previous steps, performing a clean up to remove remaining degradations artifacts and to fix possible missing text, to produce the final result.

## **7) University of Quebec, Canada (Hossein Ziaei Nafchi and Rachid Hedjam and Reza Farrahi Moghaddam and Mohamed Cheriet)**

**Method 1:** This method is a modified version of the binarization method proposed in [13]. 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.

*Method 2:* This method is a modified version of the binarization method proposed in [13]. Method of [13] is applied on a visually enhanced image rather than the original image. The enhancement method which is based on an iterative smoothing procedure is used to remove stain, shadow, or other similar degradations.

**8) Aliah University, Kolkata, India** (*Tauseef Khan, Ayatullah F. Mollah*)

In the proposed method, the input image is pre-processed to remove noises and to improve the quality of images. After that a variant model of Sauvola's Text Binarization Method is applied to binarize the pre-processed images. Finally, connected component based post-processing is applied to eliminate noisy elements such as small components that are isolated from the surroundings.

**9) Badji Mokhtar University, LabGED laboratory, Algeria** (*Abderrahmane Kefali, Toufik Sari, and Halima Bahi-Abidet*)

This method is an extension of our previous technique described in [14].

The proposed method is a hybrid thresholding-based technique. It uses two thresholds T1 and T2 and it runs in two passes. In the first pass, a global thresholding is performed in order to class the most of pixels of the image. All pixels having a gray-level higher than T2 are removed (becomes white) because they represent the background pixels. All pixels having a gray-level lower than T1 are considered as foreground pixels and therefore they are kept and colored in black. These two thresholds are estimated from the gray-levels histogram of the original image and they represent the average intensity of the foreground and background respectively. To obtain these two thresholds, we first compute a global threshold T using a global thresholding algorithm (Otsu's algorithm [15]). T separates the gray-levels histogram of the image into two classes: foreground and background. T1 and T2 are then estimated from T.

The remaining pixels are left to the second pass in which they are locally binarized by combining the results of several local thresholding methods to select the most probable binary value. The local methods included are: Niblack's [16], Sauvola and Pietikainen's [17], Nick [18], Binarization using local Maximum and Minimum [19] methods, and our neural-based thresholding method proposed in [20].

### 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 ( $FM$ ), (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  $F_{ps}$  is introduced in [21] and it uses pseudo-Recall  $R_{ps}$  and pseudo-Precision  $P_{ps}$  (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 [22]. 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  $NUBN$  is the number of the non-uniform (not all black or white pixels)  $8 \times 8$  blocks in the GT image, and  $DRD_k$  is the distortion of the  $k$ -th flipped pixel that is calculated using a  $5 \times 5$  normalized weight matrix  $W_{Nm}$  as defined in [22].  $DRD_k$  equals to the weighted sum of the pixels in the  $5 \times 5$  block of the  $GT$  that differ from the centered  $k^{\text{th}}$  flipped pixel at  $(x,y)$  in the binarization result image  $B$  (Eq. 4).

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

#### IV. EXPERIMENTAL RESULTS

The H-DIBCO 2016 testing dataset consists of 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 document images of this dataset originate from collections that belong to READ project [23] contributed by the Archive Bistum Passau (ABP) and by Staatsarchiv Marburg (StAM) which concerns the Grimm Collection. The ABP collection contains sacramental register and index pages like baptism, marriage and death entries containing around 18000 document images. The StAM – Grimm collection contains around 36000 document images from the Grimm brothers comprising mainly letters, postcards, greeting cards, etc. The used images are shown in Figure 1(a).

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 [15] and Sauvola [17] are also presented. Overall, the best performance is achieved by *Method 2* which has been submitted by *Nati Kligler and Ayellet Tal* affiliated to *Technion – Israel institute of Technology, Israel*. The binarization results of this algorithm for each image of the testing dataset is shown in Fig. 1(b).

TABLE I. DETAILED EVALUATION RESULTS FOR ALL METHODS SUBMITTED TO H-DIBCO 2016.

| Rank | Method  | Score | FM          | $F_{ps}$    | PSNR       | DRD        |
|------|---------|-------|-------------|-------------|------------|------------|
| 1    | 2       | 166   | 87.61±6.99  | 91.28±8.36  | 18.11±4.27 | 5.21±5.28  |
| 2    | 3-3     | 174   | 88.72±4.68  | 91.84±4.24  | 18.45±3.41 | 3.86±1.57  |
| 3    | 3-2     | 187   | 88.47±4.45  | 91.71±4.38  | 18.29±3.35 | 3.93±1.37  |
| 4    | 6       | 188   | 87.97±5.17  | 91.57±6.82  | 18.00±3.68 | 4.49±2.65  |
| 5    | 3-1     | 219   | 88.22±4.80  | 91.42±4.53  | 18.22±3.41 | 4.01±1.49  |
| 6    | 7-2     | 237   | 88.11±4.63  | 91.17±6.42  | 18.00±3.41 | 4.38±1.65  |
| 7    | 7-1     | 239   | 87.60±4.85  | 90.87±6.70  | 17.86±3.51 | 4.51±1.62  |
| 8    | 1       | 270   | 85.57±6.75  | 91.05±6.18  | 17.50±3.43 | 5.00±2.60  |
| 9    | 5       | 272   | 86.24±5.79  | 90.84±5.53  | 17.52±3.42 | 5.25±2.88  |
| 10   | 8       | 329   | 84.32±6.81  | 85.64±6.15  | 16.59±2.99 | 6.94±3.33  |
| 11   | 4       | 417   | 76.28±9.71  | 77.99±10.57 | 14.21±2.21 | 15.14±9.42 |
| 12   | 9       | 422   | 76.10±13.81 | 79.60±12.87 | 15.35±3.19 | 9.16±4.87  |
| -    | Otsu    | -     | 86.61±7.26  | 88.67±7.99  | 17.80±4.51 | 5.56±4.44  |
| -    | Sauvola | -     | 82.52±9.65  | 86.85±8.56  | 16.42±2.87 | 7.49±3.97  |

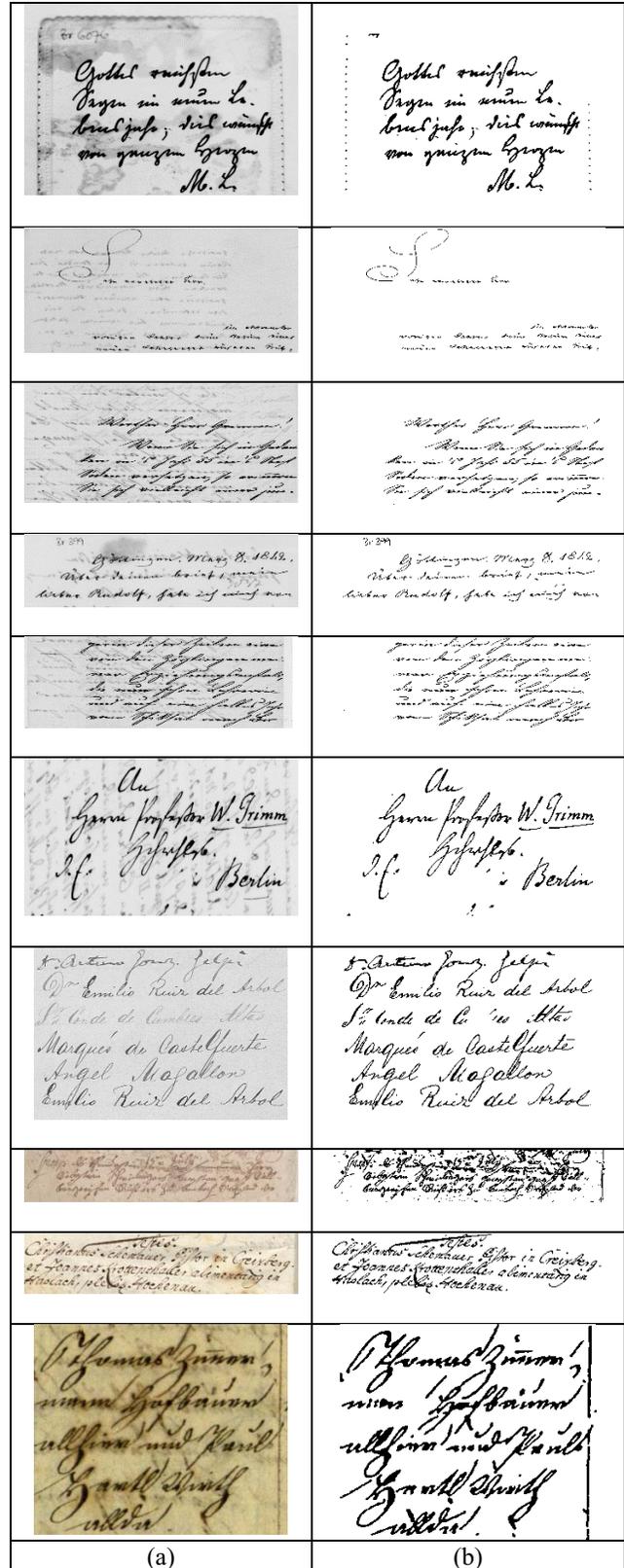


Figure 1. (a) The H-DIBCO 2016 testing dataset (b) Binarization results from the winner algorithm of H-DIBCO 2016.

## V. CONCLUSIONS

Taking into account the final evaluation, several conclusions are drawn that could provide a fruitful feedback for the research community working on improving handwritten document image binarization. It is worth noting that the winner method relies upon an already published method [9] that has already participated in previous years DIBCO challenges, which was enriched by a novel preprocessing as well as postprocessing stage. At this point, it should be noted that the proposed preprocessing stage is inspired by another context i.e. computational geometry rather than the document image analysis. The same holds for the second ranked method which is adapted from a technique for restoration of optical soundtracks of old movies [10]. Another useful observation is that still, standard approaches like the global Otsu algorithm [15] and the locally adaptive Sauvola algorithm [17] are fully involved in newly proposed approaches which in most of the cases result in increasing binarization performance while in particular examples they compare favorably with the overall best method. Last but not least, it is worth mentioning that the performance achieved by the use of pre-processing and postprocessing stages is proving that those stages have a major impact on the success of the binarization process.

## ACKNOWLEDGMENT

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

## REFERENCES

- [1] I. Pratikakis, B. Gatos and K. Ntirogiannis, "H-DIBCO 2010 – Handwritten Document Image Binarization Competition", 12<sup>th</sup> International Conference on Frontiers in Handwriting Recognition (ICFHR'10), Kolkata, India, pp. 727-732, 2010.
- [2] I. Pratikakis, B. Gatos and K. Ntirogiannis, "ICFHR 2012 Competition on Handwritten Document Image Binarization (HDIBCO 2012)", 13<sup>th</sup> International Conference on Frontiers in Handwriting Recognition (ICFHR'12), pp. 813-818, Bari, Italy, 2012.
- [3] K. Ntirogiannis, B. Gatos and I. Pratikakis, "ICFHR 2014 Competition on Handwritten Document Image Binarization (H-DIBCO 2014)", 14<sup>th</sup> International Conference on Frontiers in Handwriting Recognition (ICFHR'14), Crete island, Greece, pp. 809-813, IEEE Computer Society Press, ISBN-978-1-4799-4335-7, 2014.
- [4] Long, Jonathan, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
- [5] S. Ioffe, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167, 2015.
- [6] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," arXiv preprint arXiv:1512.03385 (2015).
- [7] C. Wolf, J-M. Jolion and F. Chassaing, "Text Localization, Enhancement and Binarization in Multimedia Documents," Proceedings of the International Conference on Pattern Recognition (ICPR), volume 4, pages 1037-1040, IEEE Computer Society. August 11th-15th, 2002, Quebec City, Canada.
- [8] S. Katz and A. Tal, "Direct Visibility of Point Sets," SIGGRAPH, vol. 26, no. 3, 2007.
- [9] N. Howe, "Document Binarization with Automatic Parameter Tuning," International Journal on Document Analysis and Recognition, vol. 16, no. 3, pp. 247-258, 2013.
- [10] A. Hassaïne, E. Decencièrre, B Besserer, "Efficient restoration of variable area soundtracks", Image Analysis & Stereology 28 (2), 113-119, 2009.
- [11] A. Hassaïne, S. Al-Maadeed and A. Bouridane, "A set of geometrical features for writer identification," Neural Information Processing. Springer Berlin Heidelberg, 2012.
- [12] S. Al-Maadeed, W. Ayouby, A. Hassaïne and J. Aljaam, "QUWI: An Arabic and English Handwriting Dataset for Offline Writer Identification." Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on. IEEE, 2012.
- [13] 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.
- [14] T. Sari, A. Kefali, H. Bahi, "Text Extraction from Historical Document Images by the Combination of Several Thresholding Techniques", Advances in Multimedia, vol. 2014, Article ID 934656, 10 pages, 2014.
- [15] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", IEEE transactions on Systems, Man and Cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.
- [16] W. Niblack, "An introduction to Digital Image Processing", Strandberg Publishing Company, Birkerød, Denmark, 1985.
- [17] J. Sauvola, M. Pietikainen, "Adaptive document image binarization", Pattern Recognition, Vol. 33, No. 2, pp. 225-236, 2000.
- [18] K. Khurshid, I. Siddiqi, C. Faure, N. Vincent, "Comparison of Niblack inspired Binarization methods for ancient documents", In Proceedings of the 16<sup>th</sup> Document Recognition and Retrieval DRR, USA, 2009
- [19] B. Su, S. Lu, C.L. Tan, "Binarization of Historical Document Images Using the Local Maximum and Minimum", In Proceedings of the 9<sup>th</sup> IAPR International Workshop on Document Analysis Systems DAS, Boston, MA, USA, pp. 159-166, 2010.
- [20] A. Kefali, T. Sari, H. Bahi, "Foreground-Background Separation by Feed-forward Neural Networks in Old Manuscripts", Informatica, vol. 38, No. 4, pp. 329-338, 2014.
- [21] 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.
- [22] 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.
- [23] READ project (<http://read.transkribus.eu/>)