Foreground-Background Separation by Feed-forward Neural Networks in Old Manuscripts

Abderrahmane Kefali, Toufik Sari and Halima Bahi
LabGED Laboratory
Computer Science Department, Badji Mokhtar University, BP-12
Annaba, 23000, Algeria
E-mail: {kefali, sari, bahi}@labged.net

Keywords: foreground-background separation, binarization, ANN, MLP, image fusion, PSNR

Received: February 16, 2014

Artificial Neural Networks (ANNs) are widely used techniques in image processing and pattern recognition. Despite of their power in classification tasks, for pattern recognition, they show limited applicability in the earlier stages such as the foreground-background separation (FBS). In this paper a novel FBS technique based on ANN is applied on old documents with a variety of degradations. The idea is to train the ANN on a set of pairs of original images and their respective ideal black and white ones relying on global and local information. We ran several experiments on benchmark and synthetic data and we obtained better results than state-of-the art methods.

Povzetek: V tem prispevku je predlagan nova binarizacija tehnika, ki temelji na umetni nevronski mreži za stare dokumente z različnimi nivoji poslabšanja.

1 Introduction

Important documentary collections exist currently in the libraries, museums and other institutions in pedagogic or sociopolitical matters. Historical documents of old civilizations and public archives are typical examples of such abundance which represent the patrimony, and nation's history. In the last two decades, the scientific community renewed the interest in the restoration of old documents. Several R&D projects were initiated and several research papers and PhD thesis have, consequently, focused on ancient documents processing and analysis. The pattern recognition and image processing researchers are in the core of this interest group.

In the most cases, document image processing should deal with foreground-background separation[1] (FBS). The goal of a FBS algorithm is to extract the relevant information (text, figures, tables, etc.), i.e. "the foreground", from the page, i.e. "the background". Actually, image binarization is critical in the sense that bad separation will cause the loss of pertinent information and/or add useless information. FBS is more difficult for old documents images which have various types of degradations from the digitization process itself, aging effects, humidity, marks, fungus, dirt, etc.

The FBS is generally done using a cut-off value "the threshold" [3][11]. The literature is rich in methods for document image binarization based on thresholding [1][2][3][11]. These methods use different ways to calculate the threshold and are grouped into two main classes: global methods when a single threshold is used for the whole image and local methods when one threshold per image area is computed. Mixtures can be done combining global and local information to find optimal threshold values.

ANNs have contributed since their introduction in the fifties to solve more complex problems in many areas including classification, prediction, approximation, pattern recognition, etc. ANNs are able to generalize, after a learning stage, their behavior on new data they had not "seen" before. In spite of the ANNs' success in various image processing applications, their contribution in image binarization is still limited and have not generated much research [3][4].

In this paper, we exploit the generalization facility of ANNs to devise a novel binarization approach that relies not on thresholding, but on classification using a Multilayer Perceptron (MLP). This assumption can be justified by definition that the binarization is in fact a classification process where the set of image pixels should be divided into two classes "blacks" and "whites" and, for this, ANNs can excel. The benefit of using a supervised neural network for FBS is the reduction of the complexity of the conventional thresholding methods and the availability of large databases for training and validation.

The remainder of this paper is organized as follow. We first give a brief description of kinds of thresholding techniques and focus on ANNs based methods. Next, we describe and detail our proposed approach. After that, we

[1] FBS is also called « binarization » meaning producing black and white (BW) images from grayscale or color ones.

present the evaluation results with several well-known methods, to finish with a conclusion.

2 Related works

Numerous surveys and competitions on document image binarization have been completed not only for comparison purposes but even for categorization, evaluation, and of course for scholars too. According to [3], the binarization techniques can be divided into six groups:

- Histogram-based methods: the methods of this class perform a thresholding based on the form of the histogram. [20][21];
- Clustering-based methods: These methods assign the image pixels to one of the two clusters: object and background. [22][23];
- Entropy-based methods: These algorithms use the information theory to obtain the threshold. [24][25];
- Object attribute-based methods: Find a threshold value based on some similarity measurements between original and binary images. [26][27];
- Spatial binarization methods: find the optimal threshold value taking into account spatial measures. [28];
- Locally adaptive methods: These methods are designed to give a new threshold for every pixel. Several kinds of adaptive methods exist. We find methods based on local gray range [29], local variation [30], etc.

Recently, ANNs have been used for binarization. As stated in [16], most of these works dealt with noisy handwritten document images and non-document images but, however, did not inspect the special case of historical documents.

M. Egmont-Petersen et al., [4] reviewed more than 200 works on image processing with ANNs from which only three were on image binarization. N. Papamarkos et al. in [5] proposed a multithresholding approach using the Principal Component Analysis (PCA) and a Kohonen Self-Organized Feature Map (SOFM) neural network. The input layer of SOFM coded the 255 elements of the gray-level histogram and the output layer is a 1D map in which only one winner neuron is activated for each input vector. In [6], the authors used a feed forward neural network to find the “clean” pixel corresponding to the noisy one as input in conjunction with the median filter value and Rank-Ordered Absolute Differences (ROAD) of the same pixel. We found that the same approach was proposed earlier in [7]. In [31], a PNN network is used but the histogram is compacted. On the other hand, in [32] every pixel is used to obtain the thresholded image.

A combination of a thresholding technique and an artificial neural network for leaf veins extraction is carried out in [12]. Padekas and Papamarkos [13] combined several binarization techniques using a Self-Organizing Map, for the binarization of normal and degraded printed documents for character visualization and recognition.

Yang et al. [14], used text image segmentation method with a neural network for document image processing. In [15], a multi-layer perceptron is used and trained using noisy document images to produce enhanced document images.

In [16], authors combined a global thresholding method, namely the Mass-Difference (MD) thresholding and supervised neural network for selecting a global optimum threshold value. This last is used then to binarize the degraded document image. MD thresholding is first applied to calculate several local optimum threshold values. Afterwards a supervised neural network is trained using these optimum threshold values as its inputs and a single global optimum threshold value as its output. The neural network consists of an input layer having 256 neurons receiving the local threshold values, one hidden layer containing 22 neurons and only one neuron in the output layer which represents the global threshold value.

Làzaro and al. [17] proposed to obtain an optimum threshold for each image using a semantic description of the histogram and a general regression neural network, for high precision OCR algorithms over a limited number of document types. The histogram of the input image is first smoothed in order to eliminate false minima and maxima and its discrete derivate is found. Using a polygonal version of the derivate and the smoothed histogram, a new description of the histogram is calculated. Once this last is generated, a semantic description is inferred. The obtained semantic description is used by a general regression neural network to obtain the optimal threshold value.

In [18], authors used multi-layer perceptron (MLP) as a threshold transfer to select the visually satisfied threshold by a modified back propagation algorithm. The proposed technique starts with the histogram of the contrasted image. The histogram is scaled to [0, 1] and divided into 128. The three-layer MLP has an input layer of 128 neurons, two hidden layers of 256 and 512 neurons respectively and an output layer with 128 neurons.

[19] Introduced three neural based binarization techniques. The three approaches start with a Self Organizing Map (SOM) trained over a part of the image in order to extract its most representative grey levels or colors. The SOM inputs are the pixels of the image. Then, the classification is done differently in the three approaches. In the first approach, the neurons of the SOM are segmented into N regions (for N classes) using the Kmeans algorithm [9]. Once the pixels are clustered, the SOM is used to classify the pixels of the entire image. Each pixel is given as input to the SOM in order to be classified. The second approach is applied on color images. In this approach an MLP is used where the inputs are the SOM neurons, and its outputs are the classes of the same neurons. After training, the MLP is applied on the whole image to classify each pixel. In the third approach, Sauvola or Niblack thresholds are applied on the neurons of the trained SOM, and a global threshold is extracted. This global threshold is applied on the entire image.

Although the binarization record is long and will continue to lengthen (see ICDAR 2013 competitions), few comparison studies were found. A large discussion on performances of the works mentioned above is not possible at this stage. A good comparison should be done, in our opinion, but on standard and benchmark databases of document images and using well renowned performance metrics.

3 Proposed method

In this section we will describe a new method for binarizing images of old documents. The proposed method may be placed among the hybrid class. In this method, the separation of image pixels to "blacks" or "whites" is performed by a Multilayer Perceptron (MLP) trained with back-propagation (see [49] for more details). In this approach, the MLP does not compute or learn any threshold but runs a direct binarization by classifying the image pixels into two classes. As we described above the binarization is a kind of two-class classification problem.

The MLP is an ANN with supervised learning dedicated especially for classification purposes. It has the ability of separating non-linearly separable classes of patterns. We have chosen to use MLPs as gray-levels of pixels may overlap in some cases. Even if this is not always true, MLPs go surely further than many other classifiers [19]. The motivation of using learning-based approach for FBS is that, as detailed in Figure 1, the pages of one same volume have similar defects as they are written in the same type of paper, in the same period of time and are conserved in the same conditions. As a result, one binarization algorithm will equally succeed on all the pages of the same manuscript. Therefore, we can train the algorithm on a limited number of pages, or some specific areas from different pages, and it can perform well on the other pages. For other volumes, we can re-train the classifier to adapt to the new data. The same scheme is also possible for one single page; we can train the algorithm on only the data from some areas of the page and it generalizes its behavior on the remainder of the page.

Figure 1: (a) a collection of documents in the same conservation conditions. (b) & (c) different pages of the same volume with similar degradations.

For each pixel, global and local information from the neighborhood are used to train the MLP. Local information is the grey values of the pixel *p* with those of its neighbors from an $N \times N$ window centered on *p*. In addition, we introduce global information, namely the global mean (*M*) and global standard deviation (*S*) of the whole image. Indeed, as noted in [18], these two last and other statistical parameters of the image are widely used by the most binarization methods, [22] [23] for instance, to compute the thresholds.

The MLP should then output 0 for black or 1 for white. See Figure 2.

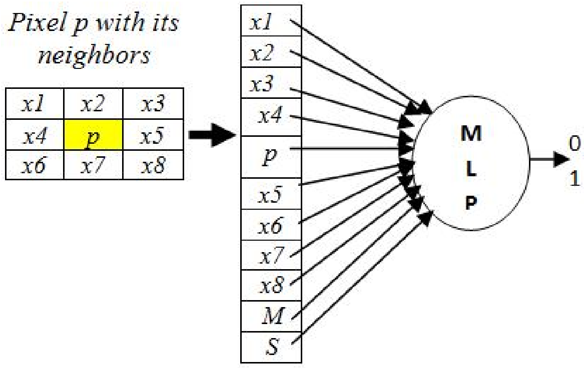


Figure 2: MLP for classifying a pixel p according to its current value and of those of its neighbors in a 3×3 window.

The proposed approach consists of four steps: *image preprocessing*, *MLP definition*, *MLP training*, and finally *Binarization* step. The block diagram of the approach is shown in Figure 3.

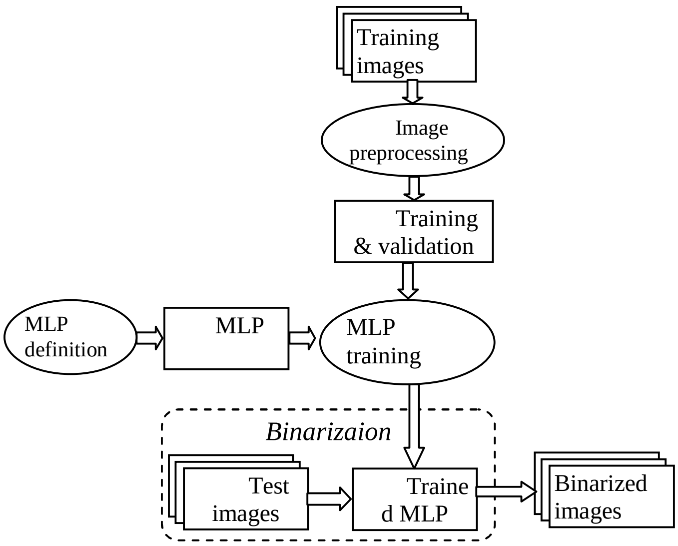


Figure 3: Block diagram of the proposed method.

3.1 MLP definition

The first thing we need to do is to set the “adequate” structure of the neural network (number of hidden layers and number of neurons per layer).

As we said, we preferred to work with Multilayer Perceptron which showed proven abilities in classification problems. Despite the importance of the optimal topology of an MLP for a given problem, it is not always easy to devise and almost not necessary.

Kolmogorov [50] showed that all continuous functions of *n* variables have an exact representation in terms of finite superpositions and compositions of a small number of functions of one variable [8]. In terms of networks this means that every continuous function of many variables can be a network with two hidden layers [52]. In addition, according to G.V. Cybenco in [8] *"a MLP network that has only one hidden layer is able to approximate almost any type of non linear mapping"*.

So, we used the trial-and-error procedure to find the ideal topology of our MLP. We devised an MLP with a single output layer corresponding to the binary value of the pixel (0 or 1). The number of inputs is the number of pixels in the window (9 for a 3×3 window, 25 for 5×5, etc.), in addition to the Mean (*M*) and the standard deviation (*S*) of the whole image. The activation function chosen is the sigmoid *function* defined by:

$$f(x) = \frac{1}{1+e^{-ax}} \quad . \tag{1}$$

However, one question remains: which window size will give better results? To answer this question, we tested several window sizes, with different neural networks of varying number of layers, and neurons in each layer; and trained all these configurations and evaluate for each configuration its performance on the test set, using several evaluation measures. The obtained results show that the optimal configuration is:

- Window size of 3×3 and therefore the number of inputs is 11,
- One hidden layer of 11 neurons,
- One neuron as output.

3.2 Image preprocessing

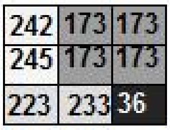
This phase has as goal the preparation and the extraction of training (and validation) data from a sample of training images and it runs offline.

3.2.1. Training and validation data

To guarantee a high accuracy, the MLP should be trained on a huge set of patterns in order to determine with high precision its free parameters (the links weights).

In practice, when using only a training set produces, in the most cases, the phenomenon of *over-learning*; where the ANN learns the training data but is not able to generalize to other data not seen during the training phase. To overcome this problem, we use another set of data called ‘validation set’. During the training phase, at every epoch we check in addition to the learning error the validation error and compare it to previous values to determine the moment of performance drop.

The training and validation sets are composed of several input vectors with the corresponding desired output. The size of the input vector is the number of input neurons of the MLP. In our case, each vector represents the gray values of a pixel with that of those neighborhoods, in addition to the global mean (*M*) and global standard deviation (*S*) of the whole image (Figure 4). The corresponding expected output is the binary value of the related pixel in the ground truth image (the black and white image). As the pixel values are in [0, 255], the data should be normalized or scaled to the interval [0, 1] (0 for black and 1 for white) to ensure the proper functioning of the neuron in sensitive regions of the activation function.



<242,173,173,245,**173**,173,223,233,36, 185.66, 64.58>

Figure 4: Input vector for one pixel in a 3×3 window.

3.2.2. Images of training and validation

We used two sets of images for creating the training and validation data. The first is a public set composed of document images from the four collections proposed within the context of the competitions DIBCO 2009[2], H-DIBCO 2010[3], DIBCO 2011[4] and H-DIBCO 2012[5].

These four collections contain a total of 50 real documents images (37 handwritten and 13 printed) coming from the collections of several libraries, with the associated ground truth images. All the images contain representative degradations which appear frequently (e.g. variable background intensity, shadows, smear, smudge, low contrast, bleed-through). Figure 5 show some images from these collections.

Figure 5: Images taken from the collections of: (a) DIBCO 2009, (b) H-DIBCO 2010, (c) DIBCO 2011, (d) H-DIBCO 2012.

The second set of images is a synthetic collection prepared in order to include most of the degradations in old documents. It is created by applying the *image mosaicing by superimposing technique for blending* [33]. The idea is as follow, we start with some images of documents in black & white, which represent the ground truth, and with some backgrounds extracted from old documents and we apply a *fusion procedure* to get as many different images of old documents. However, P. Stathis et *al.*, in [10] proposed two different techniques for the blending: *maximum intensity* and *image averaging*. We adopt to use the *image averaging technique* in order to have a more natural result.

[2] http://users.iit.demokritos.gr/~bgat/DIBCO2009/benchmark/
[3] http://www.iit.demokritos.gr/~bgat/H-DIBCO2010/benchmark
[4] http://utopia.duth.gr/~ipratika/DIBCO2011/benchmark
[5] http://utopia.duth.gr/~ipratika/HDIBCO2012/benchmark

The Mosaicing process used can be summarized by the following pseudo-code:

```
Input:  GT: the ground truth image
        BG: the background
Output: R:  the resulting image
For Each pixel (i,j)
   If BG(i,j) is More_Darker_Than GT(i,j) Then
     R(i,j) = BG (i,j)
   Else R(i,j) = (GT(i,j)+BG(i,j))/2
End For
```

We note here that we used a large amount of backgrounds with different varieties of degradations (transparency effects, holes, stains, etc.) to allow the MLP to learn from a vast diversity of possible cases (Figure 6).

(a) *(b)* *(c)*

Figure 6: Images of documents obtained by fusing binary images and backgrounds. (a) background from old documents, (b) ground truth binary image, (c) the resulting synthetic images.

3.3 MLP training

In this phase, the MLP is trained using the training and validation data prepared before. For that, we used back propagation algorithm. It is a supervised learning algorithm, where the system is provided with samples of inputs and the corresponding expected, or desired, output values.

The training process is done as following: we introduce the training vectors to the MLP, and we calculate the error between the outputs of the MLP (estimated output) and the real data (desired outputs). Next, we update the weights of all neurons. After that, we provide the validation vectors to the MLP, and we calculate the validation error (as before), without updating the neurons weights. The process is repeated in order to minimize both the learning and the validation error. The training is interrupted when the validation error begins to increase.

After training the neural network learns (usually) to provide the correct output value when it is presented with the input value only.

3.4 Binarization

Once learning is achieved, the last phase of the proposed method is the use of the trained MLP to binarize different document images. It is done by running the trained neural network with one forward pass using the final training weights. For each pixel of the processed image, we provide the MLP the feature vector as in the training phase and it should output a binary value for the pixel.

4 Experimentation and results

Our application was developed in Java with the framework *Neuroph*[6].

As we said before, we used two sets of documents for the training and validation. The first set is a public collection of a total of 50 real documents images (37 handwritten and 13 printed) coming from the collections of several libraries, with the associated ground truth. The second set is composed of 240 synthetic images of documents constructed by our fusion algorithm from 20 different backgrounds and 12 binary images.

As a first attempt, we used 15 images of the first set and 70 images of the second set for training, 10 images of the first set and 30 images of the second set for validation, and the remainders for testing. We also used all the pixels in all images of training and validation sets. This was not practicable because the images are larger (average 783,000 pixels) resulting in a massive training set of about 66,555,000 vectors. Then, we considered selecting a portion of vectors to use in training and validation. Thus, we selected randomly 500 vectors for training and validation image, resulting 42,500 vectors for training and 20,000 others vectors for validation.

During the training we calculate at each time the training and validation errors. As common, the training error decreases continuously and gradually but the validation error decreases at the beginning and then starts to deteriorate which mean that over-learning had occurred and so we should stop learning process. For our case, this happened at the epoch #25,382 (Figure 7).

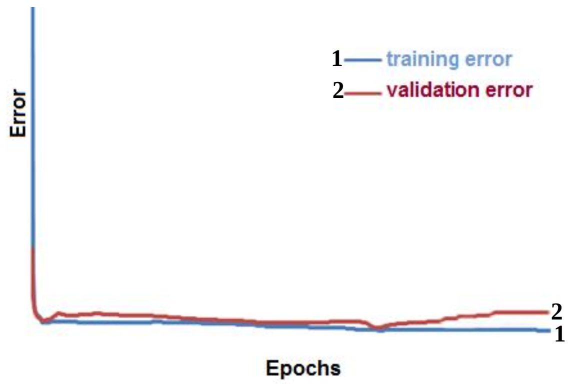


Figure 6: Errors of training and validation.

To assess the generalization ability of our neural network, we tested it on the testing collection containing 165 images of new old documents that were not presented to the MLP (25 of the first set and 140 of the second set), and compared the resulting binary images with the corresponding ground truth ones. The comparison is done quantitatively by using standard measures that have been widely used for evaluation purposes, especially in DIBCO 2009 [2], H-DIBCO 2010 [1], DIBCO 2011 [53], H-DIBCO 2012 [54] competitions. These measures consist of: FMeasure, PSNR, NRM, MPM, and DRD.

[6] http://neuroph.sourceforge.net

Noting *TP*, *TN*, *FP*, *FN* the True positive, True Negative, False positive and False negative values, respectively.

i) FMeasure

F-Measure was introduced first by Chinchor in [55].

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

Where:

$$Recall = \frac{TP}{TP + FN} \quad \text{and} \quad Precision = \frac{TP}{TP + FP}$$

ii) PSNR (Peak Signal to Noise Ratio)

PSNR is a similarity measure between two images. However, the higher the value of PSNR, the higher the similarity of the two images [2][34].

$$PSNR = 10.\log\left(\frac{C^2}{MSE}\right). \quad (3)$$

Where: $MSE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y) - I_2(x,y))^2}{MN}$

I and I_2 represents the two images matched. *M* and *N* there height and width respectively. *C* the difference between foreground and background (here 255).

iii) NRM (Negative Rate Metric)

NRM is based on the pixel-wise mismatches between the Ground Truth and the binarized image [51]. It combines the false negative rate NR_{FN} and the false positive rate NR_{FP}. It is denoted as follows:

$$NRM = \frac{NR_{FN} + NR_{FP}}{2}. \quad (4)$$

With : $NR_{FN} = \frac{FN}{FN + TP}$ and $NR_{FP} = \frac{FP}{FP + TN}$

The better binarization quality is obtained for lower NRM.

iv) MPM (Misclassification Penalty Metric)

The Misclassification penalty metric MPM evaluates the binarization result against the Ground Truth on an object-by-object basis [51].

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

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

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

v) DRD (Distance Reciprocal Distortion Metric)

DRD is an objective distortion measure for binary document images, and it was proposed by Lu et al. in [34]. This measure properly correlates with the human visual perception and it measures the distortion for all the *S* flipped pixels as follows:

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN}. \quad (6)$$

NUBN is the number of the non-uniform 8×8 blocks in the GT image.

DRD_k is the distortion of the k^{th} flipped pixel of coordinate *(x, y)* and it is calculated using a 5×5 normalized weight matrix W_{Nm}. This last is defined in [34] as follow:

$$W_{Nm}(i,j) = \frac{W_m(i,j)}{\sum_{i=1}^{m}\sum_{j=1}^{m} W_m(i,j)}$$

Such as:

$$W_m(i,j) = \begin{cases} 0, & \text{for } i = i_C \text{ and } j = j_C \\ \frac{1}{\sqrt{(i - i_C)^2 + (j - j_C)^2}}, & \text{otherwise.} \end{cases}$$

With $m = 5$, and $i_C = j_C = (1 + m) / 2$.

DRD_k is given as follow (eq.7):

$$DRD_k = \sum_{i=-2}^{2}\sum_{j=-2}^{2} \left| I_{GT}(x+i, y+j) - I_B(x,y) \right| \times W_{Nm}(i+2, j+2)$$

We also compared our method with several well known state-of-the-art methods of document binarization [3][10][11]. For the local methods, we reimplemented, all used values of parameters are those indicated in the original papers. The average results obtained with all compared methods over each test set are summarized in Table.1. The average results between the two sets are shown in Table.2. The final ranking of the compared methods is shown in Table.3, which also summarizes the partial ranks of each method according to each evaluation measure and the sum of ranks. Afterwards, We provide in Fig.8 graphs representing the average performance of the compared methods in terms of FMeasure and PSNR.

From the Tables and Fig.8, our proposed technique is ranked second after Sauvola and Pietikainen method for both test sets, and it has good performances in terms of FMeasure and PSNR. This result is due to the generalization ability of the MLP despite of the characteristic of degradation and the variation of noise types present in the documents.

	First test set					Second test set				
Method	FM	PSNR	NRM	MPM	DRD	FM	PSNR	NRM	MPM	DRD
Otsu [22]	0.8571	15.886	0.0628	0.1915	1.755	0.6245	15.101	0.1590	2.8783	4.6768
ISODATA [37]	0.8544	15.734	0.0632	0.1938	1.833	0.6153	15.059	0.1666	2.9310	4.7320
Kittler and Illingworth [23]	0.4673	6.2878	0.1630	9.7698	26.13	0.4239	8.3621	0.2163	17.727	28.682
Kapur and al.[24]	0.8432	14.989	0.0509	0.2673	2.161	0.3618	11.493	0.3096	3.3474	7.6726
Mello and Lins [35]	0.6496	12.708	0.2055	0.3465	3.051	0.3574	13.713	0.3314	0.7512	4.3837
Tsallis Entropy Based Algorithms [36]	0.5203	13.092	0.2712	0.0901	3.092	0.1387	11.532	0.4547	0.3561	3.9472
Iterative global thresholding (IGT) [38]	0.7622	13.683	0.1162	1.7347	4.215	0.6049	13.637	0.1711	2.3889	4.0455
Niblack [30]	0.5321	7.7542	0.1259	8.3005	17.85	0.3996	7.5398	0.1937	9.4157	17.059
Sauvola and Pietikainen. [39]	0.8360	15.537	0.0938	0.2486	1.634	0.6718	16.130	0.1835	0.7630	2.2288
Nick [40]	0.7764	13.882	0.0977	1.5542	4.116	0.6778	15.477	0.1656	0.9653	2.5658
Feng and Tan. [41]	0.7798	13.826	0.0820	2.2608	5.016	0.6744	15.148	0.1576	1.0867	2.7634
Bernsen [29]	0.5150	8.1056	0.1807	10.936	19.62	0.4081	8.1610	0.2069	8.7728	14.517
Meen-Gradient (Leedham and al.) [42]	0.7914	13.872	0.1009	0.2631	2.834	0.6023	12.832	0.1858	1.2630	3.7058
Cavalcanti and al. [43]	0.4926	11.928	0.3117	0.0807	4.255	0.3443	12.815	0.3454	0.8949	4.0059
Tabatabei and Bohlool. [44]	0.8002	14.744	0.0671	2.4013	5.309	0.6377	15.660	0.1572	2.2360	3.6973
Gangamma and al.[45]	0.6714	12.879	0.1765	4.1804	6.549	0.6037	14.317	0.2179	1.0963	2.4942
Improved IGT [46]	0.7530	13.695	0.1364	1.3784	3.751	0.5980	13.486	0.1921	1.8568	3.7286
Using Local Max and Min [47]	0.7891	14.342	0.0971	2.2913	5.439	0.5713	11.622	0.1835	1.3603	4.5757
Sari and al.[48]	0.8362	15.628	0.0719	0.7248	1.970	0.4979	9.4425	0.1764	4.3272	10.974
Proposed method	0.8436	15.555	0.0599	0.7468	2.597	0.6608	15.396	0.1456	1.806	2.9860

Table 1: Average results from different binarization methods on each test set.

Method	FM	PSNR	NRM	MPM	DRD
Otsu [22]	0.7408	15.4939	0.1109	1.5349	3.2159
ISODATA [37]	0.7348	15.3966	0.1149	1.5624	3.2825
Kittler and Illingworth [23]	0.4456	7.3250	0.1897	13.7486	27.406
Kapur and al.[24]	0.6025	13.2417	0.1803	1.8073	4.9168
Mello and Lins [35]	0.5035	13.2112	0.2684	0.5488	3.71735
Tsallis Entropy Based Algorithms [36]	0.3295	12.3124	0.3629	0.2231	3.5196
Iterative global thresholding (IGT) [38]	0.6835	13.6609	0.1437	2.0618	4.13025
Niblack [30]	0.4658	7.6470	0.1598	8.8581	17.4545
Sauvola and Pietikainen. [39]	0.7539	15.8343	0.1386	0.5058	1.9314
Nick [40]	0.7271	14.6801	0.1316	1.2597	3.3409
Feng and Tan. [41]	0.7271	14.4877	0.1198	1.6737	3.8897
Bernsen [29]	0.4615	8.1333	0.1938	9.8547	17.0685
Meen-Gradient (Leedham and al.) [42]	0.6969	13.3522	0.1433	0.7630	3.2699
Cavalcanti and al. [43]	0.4185	12.3722	0.3286	0.4878	4.13045
Tabatabei and Bohlool. [44]	0.7190	15.2025	0.1122	2.3186	4.50315
Gangamma and al.[45]	0.6376	13.5984	0.1972	2.6384	4.5216
Improved IGT [46]	0.6755	13.5911	0.1643	1.6176	3.7398
Using Local Max and Min [47]	0.6802	12.9824	0.1403	1.8258	5.00735
Sari and al.[48]	0.6670	12.5353	0.1241	2.5260	6.472
Proposed method	0.7522	15.4759	0.1028	1.2764	2.7915

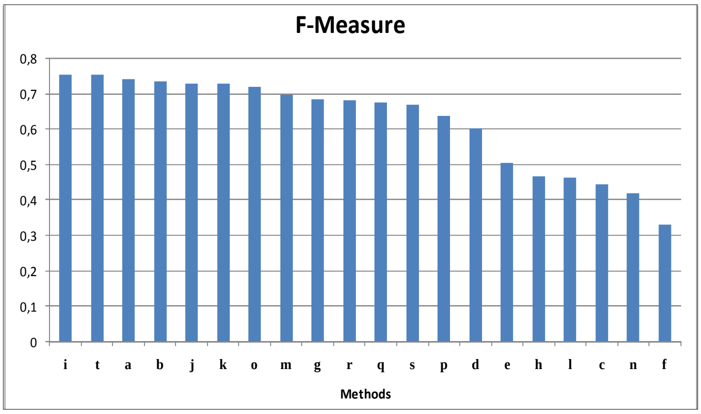
Table 2: Average results from different binarization methods between the two test sets.

5 Conclusion

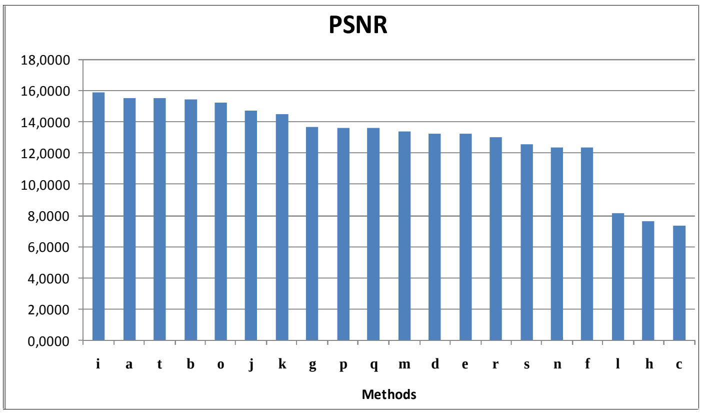
Document binarization is an important and crucial step in all systems of document analysis and recognition. This task becomes more difficult for historical documents containing various types of noises and degradations. In this paper, we proposed a new ANN based method for old document binarization. The purpose of using ANN, and especially Multilayer Perceptrons, for image binarization is to fill the lack of employing the techniques of soft computing and machine learning in such problem, and to take advantages of the generalization abilities of the MLP. Indeed, as confirmed by the experimentation results, the MLP presented a reliable behavior for the complex task of foreground/background separation from significantly degraded document images. Many extensions are possible and we will continue to enhance the proposed method. Other experimentations are needed in order to identify the applicability of the proposed solution in mobile devises of real life utilization.

Rank	Method	FM	PSNR	NRM	MPM	DRD	Sum of ranks
1	Sauvola and Pietikainen. [39]	1	1	8	3	1	14
2	Proposed method	2	3	1	7	2	15
3	Otsu [22]	3	2	2	8	3	18
4	ISODATA [37]	4	4	4	9	5	26
5	Nick [40]	5	6	7	6	6	30
6	Meen-Gradient (Leedham and al.) [42]	8	11	10	5	4	38
7	Feng and Tan. [41]	6	7	5	11	10	39
8	Tabatabei and Bohlool. [44]	7	5	4	15	13	44
9	Improved IGT [46]	11	10	13	10	9	53
10	Iterative global thresholding (IGT) [38]	9	8	11	14	12	54
11	Mello and Lins [35]	15	13	18	4	8	58
12	Using Local Max and Min [47]	10	14	9	13	16	62
13	Tsallis Entropy Based Algorithms [36]	20	17	20	1	7	65
14	Sari and al.[48]	12	15	6	16	17	66
15	Kapur and al.[24]	14	12	14	12	15	67
16	Cavalcanti and al. [43]	19	16	19	2	11	67
17	Gangamma and al.[45]	13	9	17	17	14	70
18	Niblack [30]	16	19	12	18	19	84
19	Bernsen [29]	17	18	16	19	18	88
20	Kittler and Illingworth [23]	18	20	15	20	20	93

Table 3: Final ranking of the compared methods on the two test sets.



(a)



(b)

A	Otsu
b	ISODATA
c	Kittler and Illingworth
d	Kapur and al.
e	Mello and Lins
f	Tsallis Entropy Based Algorithms
g	Iterative global thresholding (IGT)
h	Niblack
i	Sauvola and Pietikainen.
j	Nick
k	Feng and Tan.
l	Bernsen
m	Meen-Gradient (Leedham and al.)
n	Cavalcanti and al.
o	Tabatabei and Bohlool.
p	Gangamma and al.
q	Improved IGT
r	Using Local Max and Min
s	Sari and al.
t	Proposed method

Figure 7: Graphs showing the performance of the compared binarization methods in terms of (a) FMeasure and (b) PSNR.

References

[1] I. Pratikakis, B. Gatos, K. Ntirogiannis (2010). H-DIBCO 2010 - Handwritten Document Image Binarization Competition. *In proceedings of the International Conference on Frontiers in Handwriting Recognition ICFHR*, pp. 727-732.

[2] B. Gatos, K. Ntirogiannis, I. Pratikakis (2009). DIBCO 2009: document image binarization contest.

International Journal of Document Analysis and Recognition IJDAR, Vol.14, No. 1, pp. 35-44.
[3] M. Sezgin, B. Sankur (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, Vol. 13, No. 1, pp. 146-165.
[4] M. Egmont-Petersen, D. de Ridder, H. Handels (2002). Image processing with neural networks - a review. *Pattern Recognition*, Vol. 35, No. 10, pp. 2279-2301.
[5] N. Papamarkos, C. Strouthopoulos, I. Andreadis (2000). Multithresholding of color and gray-level images through a neural network technique. *Image and Vision Computing*, Vol. 18, No. 3, pp. 213-222.
[6] G. Kaliraj, S. Baskar (2010). An efficient approach for the removal of impulse noise from the corrupted image using neural network based impulse detector. *Image and Vision Computing*, Vol. 28, No. 3, pp. 458-466.
[7] H. Kong, L. Guan (1998). A noise-exclusive adaptive filtering framework for removing impulse noise in digital images. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 45, No. 3, pp. 422-428.
[8] G. Cybenco (1989). Approximation by superposition of a sigmoidal function. *Mathematics of control, Signals, and Systems*, Vol. 2, No. 4, pp. 303-314.
[9] J. B. MacQueen (1967). Some methods for classification and analysis of multivariate observations. *In Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.
[10] P. Stathis, E. Kavallieratou, N. Papamarkos (2008). An Evaluation Survey of Binarization Algorithms on Historical Documents. *In proceedings of the 19th International Conference on Pattern Recognition ICPR*, pp. 742-745.
[11] A. Kefali, T. Sari, M. Sellami (2010). Evaluation of several binarization techniques for old Arabic documents images. *In proceedings of the First International Symposium on Modeling and Implementing Complex Systems MISC.* Constantine, Algeria, pp. 88-99.
[12] H. Fu, Z. Chi (2006). Combined Thresholding and Neural Network Approach for Vein Pattern Extraction from Leaf Images. *IEE Proceedings-Vision, Image and Signal Processing*, Vol. 153, No. 6, pp. 881-892.
[13] E. Badekas, N. Papamarkos (2007). Optimal Combination of Document Binarization Techniques Using a Self Organizing Map Neural Network. *Engineering Applications of Artificial Intelligence*, Vol. 20, No. 1, pp. 11-24.
[14] Y. Yang, K. Summers, M. Turner (2004). A Text Image Enhancement System Based on Segmentation and Classification Method. *In Proceedings of the First ACM Workshop on Hardcopy Document Processing HDP*, pp. 33-40.
[15] J.L. Hidalgo, S. Espana, M.J. Castro. J.A. Perez (2005). Enhancement and Cleaning of Handwritten Data by Using Neural Networks. *Lecture Notes in Computer Science*, Vol.3522. Springer-Verlag, Berlin Heidelberg New York, pp. 376-383.
[16] A. Khashman, B. Sekeroglu (2007). Global Binarization of Document Images Using a Neural Network. *Third International IEEE Conference on Signal-Image Technologies and Internet-Based System SITIS*, Shanghai, pp. 665-672.
[17] J. Lazaro, J.L Martin, J. Arias, A. Astarloa, C. Cuadrado (2010). Neuro Semantic Thresholding for High Precision OCR Applications. *Image and Vision Computing*, Vol. 28, No. 4, pp. 571-578.
[18] T. Chen, M. Takagi (1993). Image Binarization by Back Propagation Algorithm. *International Archives of Photogrammetry and Remote Sensing*, Vol.29, pp. 345-345.
[19] H. Hamza, E. Smigiel, A. Belaid (2005). Neural based binarization techniques. *In Proceedings of the 8th International Conference on Document Analysis and Recognition ICDAR*, pp. 317-321.
[20] M.I. Sezan (1990). A peak detection algorithm and its application to histogram-based image data reduction. *Computer Vision, Graphics, and Image Processing*, Vol. 49, No. 1, pp. 36-51.
[21] A. Rosenfeld, P. de la Torre (1983). Histogram concavity analysis as an aid in threshold selection. *IEEE Transactions on System, Man, and Cybernetics*, Vol. 13, No. 2, pp. 231-235.
[22] N. Otsu (1979). A thresholding selection method from grayscale histogram. *IEEE Transactions on System, Man, and Cybernetics*, Vol. 9, pp. 62-66.
[23] J. Kittler, J. Illingworth (1986). Minimum error thresholding. *Pattern Recognition*, Vol. 19, No. 1, pp. 41-47.
[24] J.N. Kapur, P.K. Sahoo, AK.C. Wong (1985). A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, Vol. 29, No. 3, pp. 273-285.
[25] H.D. Cheng, J.R. Chen, J. Li (1998). Threshold selection based on fuzzy c-partition entropy approach. *Pattern Recognition*, Vol. 31, No. 7, pp. 857-870.
[26] L. Hertz, R.W. Schafer (1988). Multilevel thresholding using edge matching. *Computer Vision, Graphics, and Image Processing*, Vol. 44, No. 3, pp. 279-295.
[27] L.K. Huang, M.J.J. Wang (1995). Image thresholding by minimizing the measures of fuzziness. *Pattern Recognition*, Vol. 28, No. 1, pp. 41-51.
[28] A.S. Abutableb (1989). Automatic thresholding of gray-level pictures using two-dimensional entropy. *Computer Vision, Graphics, and Image Processing*, Vol. 47, No. 1, pp. 22-32.
[29] J. Bernsen (1986). Dynamic thresholding for gray-level images. *In Proceedings of the 8th International Conference on Pattern Recognition ICPR*, Paris, French, pp. 1251-1255.

[30] W. Niblack (1985). *An introduction to Digital Image Processing*. Strandberg Publishing Company, Birkeroed, Denmark.

[31] K.L. Chung, W.Y. Chen (2003). Fast Adaptive PNN-Based Thresholding Algorithms. *Pattern Recognition*, Vol. 36, No. 12, pp. 2793-2804.

[32] C.Y. Chang, P.C. Chung (2001). Medical Image Segmentation Using a Contextual-Constraint-Based Hopfield Neural Cube. *Image and Vision Computing*, Vol.19, No. 9-10, pp. 669-678.

[33] L.G. Brown (1992). A survey of Image Registration Techniques. *ACM Computing Surveys*, Vol. 24, No. 4, pp. 325-376.

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

[35] C.A.B. Mello, R.D. Lins (2000). Image segmentation of historical documents. *Visual 2000*, Mexico City, Mexico, Vol. 30.

[36] C.A.B. Mello, A.L.I.D. Oliveira, A. Sanchez (2008). Historical Document Image Binarization. *In proceedings of the International Conference on Computer Vision Theory and Applications*, Funchal, Portugal, Vol. 1, pp. 108-113.

[37] F.R.D. Velasco (1980). Thresholding using the isodata clustering algorithm. *IEEE Transaction on Systems, Man and Cybernetics*, Vol. 10, No. 11, pp. 771-774.

[38] E. Kavallieratou (2005). A Binarization Algorithm Specialized on Document images and Photos. *In Proceedings of the 8th International Conference on Document Analysis and Recognition ICDAR*, pp. 463-467.

[39] J. Sauvola, M. Pietikainen (2000). Adaptive document image binarization. *Pattern Recognition*, Vol. 33, No. 2, pp. 225-236.

[40] K. Khurshid, I. Siddiqi, C. Faure, N. Vincent (2009). Comparison of Niblack inspired Binarization methods for ancient documents. *In Proceedings of the 16th Document Recognition and Retrieval DRR*, USA.

[41] M.L. Feng, Y.P. Tan (2004). Adaptive binarization method for document image analysis. *In Proceedings of the IEEE International Conference on Multimedia and Expo ICME*, pp. 339-342.

[42] G. Leedham, C. Yan, K. Takru, J.H.N. Tan, L. Mian (2003). Comparison of Some Thresholding Algorithms for Text/Background Segmentation in Difficult Document Images. *In Proceedings of the 7th International Conference on Document Analysis and Recognition ICDAR*, Vol.2, pp. 859-864.

[43] G.D.C Cavalcanti, E.F.A. Silva, C. Zanchettin, B.L.D. Bezerra, R.C. Doria, J.C.B. Rabelo (2006). A Heuristic Binarization Algorithm for Documents with Complex Background. *In Proceedings of the International Conference on Image Processing ICIP*, pp. 389-392.

[44] S.A. Tabatabaei, M. Bohlool (2010). Novel method for binarization of badly illuminated document images. *In Proceedings of the IEEE 17th International Conference on Image Processing ICIP*, Hong Kong, pp. 3573-3576.

[45] B. Gangamma, M.K. Srikanta (2011). Enhancement of Degraded Historical Kannada Documents. *International Journal of Computer Applications*, Vol.29, No.11, pp. 1-6.

[46] E. Kavallieratou, S. Stathis (2006). Adaptive Binarization of Historical Document Images. *In Proceedings of the 18th Conference on pattern recognition ICPR*, pp. 742-745.

[47] B. Su, S. Lu, C.L. Tan (2010). Binarization of Historical Document Images Using the Local Maximum and Minimum. *In Proceedings of the 9th IAPR International Workshop on Document Analysis Systems DAS*, Boston, MA, USA, pp. 159-166.

[48] T. Sari, A. Kefali, H. Bahi (2012). An MLP for binarizing images of old manuscripts. *In Proceedings of the International Conference on Frontiers in Handwriting Recognition ICFHR*, pp. 247-251.

[49] S. B. Kotsiantis (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, Vol.31, pp. 249-268.

[50] A.N. Kalmogorov (1957). On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Dokl. Acad. Nauk SSR*, 114, pp. 953-956

[51] J. Aguilera, H. Wildenauer, M. Kampel, M. Borg, D. Thirde, J. Ferryman (2005). Evaluation of motion segmentation quality for aircraft activity surveillance. *In Proceedings of the Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance VS-PETS*, pp. 317-324.

[52] F. Girosi, T. Poggio (1989). Representation properties of networks: Kolmogorov's theorem is irrelevant. *Neural Computation*, Vol 2, No. 4, pp. 465-469.

[53] I. Pratikakis, B Gatos, K. Ntirogiannis (2011). ICDAR 2011 Document Image Binarization Contest (DIBCO 2011). *In Proceedings of the 2011 International Conference on Document Analysis and Recognition ICDAR*, Beijing, China, pp. 1506-1510.

[54] I. Pratikakis, B. Gatos, K. Ntirogiannis (2012). ICFHR 2012 competition on handwritten document image binarization (H-DIBCO 2012), *In Proceedings of the International Conference on Frontiers in Handwriting Recognition ICFHR*, pp. 817-822.

[55] N. Chinchor (1992). MUC-4 Evaluation Metrics. *In Proceedings of the Fourth Message Understanding Conference*, pp. 22–29.