An MLP for binarizing images of old manuscripts

Toufik Sari, LabGED Laboratory,
Computer Science Department, Badji Mokhtar University,
BP-12
Annaba, Algeria
e-mail: sari@labged.net

Abderrahmane Kefali, LabGED Laboratory,
Computer Science Department, Badji Mokhtar University,
BP-12
Annaba, Algeria kefali@labged.net

Halima Bahi, LabGED Laboratory,
Computer Science Department, Badji Mokhtar University,
BP-12
Annaba, Algeria
bahi@labged.net

***Abstract*–Ancient Arabic manuscripts' processing and analysis are very difficult tasks and are likely to remain open problems for many years to come. In this paper we tackle the problem of foreground/background separation in old documents. Our approach uses a back-propagation neural network to directly classify image pixels according to their neighborhood. We tried several multilayer Perceptron topologies and found experimentally the optimal one. Experiments were run on synthetic data obtained by image fusion techniques. The results are very promising compared to state-of-the-art techniques.**

I. INTRODUCTION

In the last two decades, the scientific community renewed with the reclamation necessity of ancient documents and especially of old manuscripts. Several projects were launched and several research papers and PhD thesis have, consequently, focused in such documents. The pattern recognition and image processing societies are in the center of such moving. In most cases, we should first deal with foreground/background separation. For document images, the binarization algorithm separates the text, "the foreground", from the page, "the background" using generally a cut-off value. In fact, this step is critical, bad value may cause the loss of relevant information or in contrast may add some noise to the image. This task is more difficult for historical documents with various types of degradations from the digitization process, aging effects, humidity, marks, fungi, dirt, etc. making the automatic processing of such documents difficult at many levels. The literature is rich of methods for document image binarization based on thresholding [1][2][3]. Therefore, we propose a new binarization method that not relies on thresholding, but on classification using an artificial neural network. The remainder of this paper is devoted to the presentation and evaluation of the proposed method by detailing the various sections.

II. RELATED WORKS

Artificial Neural Networks (ANN) have contributed since their introduction in the 1950s to solve more complex problems in many areas including prediction, approximation, pattern recognition, etc ... ANN are capable to generalize, from a set of learned data, their behavior on other data they have not met before.

An image binarization algorithm tend to classify image pixels as "blacks" and "whites" and indeed ANNs can excel in such a classification task. In spite of the ANN success in various image processing applications, their contribution in binarization is still limited and had not generated much research [3][4]. M. Egmont-Petersen et al. in [4] reported a review of more than 200 works related to ANNs for image processing but, however, only three of them have focused on image binarization and thresholding. N. Papamarkos et *al.* in [5] proposed a multithresholding approach to image binarization by means of Principal Component Analyzer (PCA) and a Kohonen Self-Organized Feature Map (SOFM) neural network. The input layer codes the 255 elements of the gray-level histogram and the output layer is 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 set as input in conjunction with the median filter value and Rank-Ordered Absolute

978-0-7695-4774-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICFHR.2012.176

Differences (ROAD) of the same pixel. A same approach was proposed earlier in [7].

III. PROPOSED METHOD

In this section we will describe a new method for binarizing images of ancient Arabic documents. The classification, or the assignment, of image pixels to "black" or "white" is performed using a Multilayer Perceptron (MLP) with back-propagation. In our approach we do not compute any threshold but rather we perform an exhaustive or direct classification.

The idea is as follows: for each pixel *p* of the image, we introduce to the MLP the grey values of the pixel *p* with those of its neighbors from an *N***N* window centered on *p*. The MLP should then output 0 for black or 1 for white. Figure 1

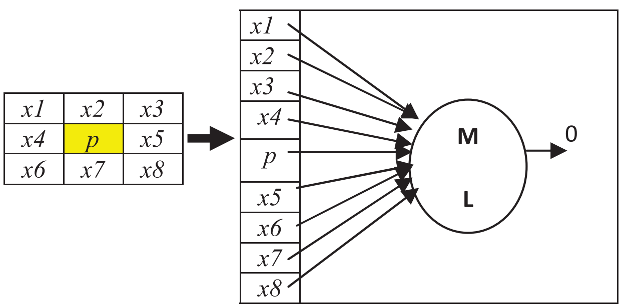


Figure 1. a MLP for classifying a pixel *p* according to its current value and of those of its neighbors.

The first thing we need to do is set the “adequate” structure of the MLP (number of hidden layers and number of neurons per layer), then we should prepare qualitative and quantitative data for training and testing. The next step is to train the MLP and finally come the validation step.

A. The MLP structure

As we said, we preferred to work with Multilayer Perceptrons 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. According to G. V. Cybenco [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 [9]. The activation function chosen is the sigmoid *function* defined by:

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

This is justified by the fact that the result of this function is a positive value in the interval [0, 1] in compliance with the response that the MLP should have.

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.). 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 calculated for each configuration the miss-classified ratio of pixels in the test set. The obtained results show that the configuration that minimizes the rate of miss-classifications in the test set is:

- Window size = 3*3 and therefore the number of inputs is 9.
- One hidden layer of 9 neurons and one neuron as output.

B. Training and validation data

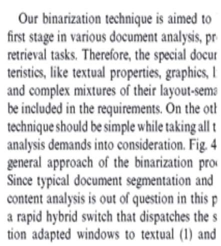
To ensure high accuracy, the MLP should be trained on huge examples in order to specify its free parameters (links weights). We used back propagation for training. In practice, using a training set only can result a phenomenon called *over-learning* where the ANN learns the training data but is not able to generalize to other date not seen in the training phase. To overcome this problem, we use a different set of data called ‘validation set’; and during the learning, at every epoch we check, in addition to the learning error, the validation error and compare it to the previous value to determine the deterioration time.

The training and validation sets are composed of several input vectors with the corresponding desired output. In our case, each vector represents the gray values of a pixel with that of those neighborhoods and the corresponding expected output is the binary value of the related pixel in the ground truth image (a binary 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.

C. Images of training and validation

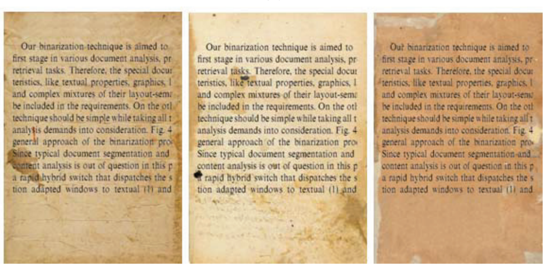
We created a set of images using the *image fusion* (mosaicing) technique proposed by P. Stathis et *al.,* in [10]. The idea is as follow, we start with some images of documents in black white which represent the ground truth and some backgrounds extracted from old documents we apply a fusion procedure to get as many different images of old documents. Note here, that the used backgrounds include different types of noise (transparency effects, holes, ink stains, etc.) to allow the MLP to learn the different types of noise (Figure 2).

a.



Our binarization technique is aimed to
first stage in various document analysis, pr
retrieval tasks. Therefore, the special docu
teristics, like textual properties, graphics, l
and complex mixtures of their layout-sema
be included in the requirements. On the otl
technique should be simple while taking all t
analysis demands into consideration. Fig. 4
general approach of the binarization pro
Since typical document segmentation and
content analysis is out of question in this p
a rapid hybrid switch that dispatches the s
tion adapted windows to textual (1) and

b.



c.

Figure 2. Images of documents obtained by fusing binary images and backgrounds. a) Some backgrounds from old documents, b) a ground truth binary image and c) the resulting synthetic image.

IV. EXPERIMENTATION AND RESULTS

Our application was developed in Java with the neural networks framework *Neuroph*. We constructed a set of 240 images of documents from 20 different backgrounds and 12 binary images.

As a first attempt, we used 70 images for training and 30 images 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 55,593,000 vectors. Then, we considered selecting a portion of vectors to use in training and validation. Thus, we selected randomly 12,000 vectors for the training and 6,000 others for the validation.

During the training we calculate at each time (After the MLP processes each example of the training set) training and validation errors. As known theoretically, 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 need to stop learning process. This happens at the epoch #2738 (Figure 3).

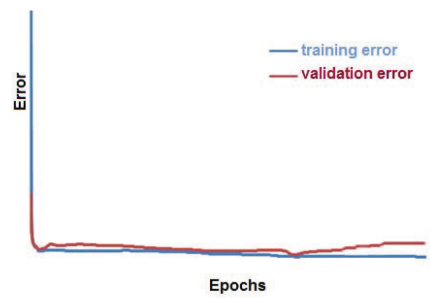


Figure 3. Errors of training and validation

To evaluate the generalization ability of our neural network, we tested it on the collection of 100 images of new old documents, and compared the resulting binary images with the corresponding ground truth ones. We used the rate of mill-classified pixels (MCR) as the evaluation metric.

$$MCR = \frac{Number\ of\ miss-classified\ pixels}{Number\ of\ image\ pixels} \quad (2)$$

We also compared our method to the well known state-of-the-art methods [11]. The results are summarized in Table 1. Figure 4 shows the mean rates of miss-classifications for various methods.

The histogram of figure 5 shows that our method is ranked third after Sauvola’s method with an average MCR=0.596%. This result is in our opinion very satisfactory considering the difficulty of processing old Arabic documents.

TABLE 1. SOME OBTAINED RESULTS FROM DIFFERENT BINARIZATION METHODS.

	Globale									Hybride		Locale					
	Abso.	Adap.	Otsu	ISODATA	Pun	Kapur	Li, Lee	Cheng, Chen	Global Nick	Otsu	ISODATA	Bernsen	Niblack	Sauvola	Wolf	Nick	RNA
IM_1	0,23	0,24	0,21	0,21	13,77	0,51	0,27	30,08	4,76	19,36	17,29	26,48	20,11	0,42	0,28	0,21	0,34
IM_2	13,41	13,67	12,89	12,89	13,41	4,02	12,37	20,22	14,29	14,29	13,97	24,17	13,08	0,74	0,20	0,37	2,24
IM_3	0,06	0,29	0,37	0,37	13,36	1,22	0,11	36,58	7,76	29,66	29,66	28,04	24,33	0,16	0,58	0,28	0,31
IM_4	0,02	0,05	0,14	0,14	15,13	0,81	0,05	25,77	0,95	15,13	0,17	25,82	18,94	0,12	0,44	0,17	0,25
IM_5	0,08	0,14	0,11	0,11	12,64	0,87	0,15	45,74	3,82	55,63	52,39	25,70	21,04	0,29	0,42	0,21	0,30
IM_6	0,02	0,09	0,19	0,19	12,98	0,95	0,01	36,88	2,00	24,68	0,21	20,78	24,06	0,11	0,58	0,26	0,24
IM_7	11,57	20,18	14,70	14,70	12,59	1,31	14,16	29,00	14,16	19,34	19,34	29,34	22,28	0,63	0,39	0,44	0,51
IM_8	0,34	0,34	0,55	0,55	15,64	1,16	0,34	40,65	5,30	8,90	8,48	25,19	17,91	0,26	0,66	0,38	0,30
IM_9	2,14	10,67	5,04	4,48	12,69	1,44	1,87	35,24	9,01	19,68	18,38	28,12	23,34	0,42	0,80	0,46	0,36
IM_10	0,09	0,27	0,38	0,38	14,20	0,83	0,13	28,18	7,25	18,80	17,90	18,99	13,55	0,14	0,75	0,41	0,24
IM_11	11,26	37,35	27,66	16,96	12,57	1,32	1,10	40,57	10,05	37,35	37,35	27,42	20,45	0,87	0,29	0,43	0,56
IM_12	0,33	0,64	2,31	2,19	13,76	2,08	0,41	28,22	9,80	16,33	16,33	22,45	19,63	0,22	0,66	0,34	0,35
IM_13	2,72	3,03	2,84	2,84	13,96	5,55	2,80	30,06	3,92	3,08	3,07	26,82	19,00	1,05	1,26	1,08	3,03
IM_14	0,10	0,29	0,53	0,53	16,54	1,08	0,16	22,31	6,62	10,99	10,57	21,15	10,19	0,10	0,52	0,24	0,30
IM_15	0,68	0,87	2,26	2,26	12,61	2,81	1,19	46,90	8,26	69,93	69,93	29,31	22,23	0,21	1,15	0,69	0,36
IM_16	4,70	13,52	0,87	0,92	12,41	1,07	0,59	35,70	7,99	31,31	31,31	23,37	11,37	0,79	0,17	0,33	0,45
IM_17	0,89	0,49	0,58	0,58	12,98	1,19	0,49	40,63	2,08	63,25	60,29	24,18	12,96	0,70	0,15	0,31	0,42
IM_18	0,32	0,35	0,31	0,31	14,31	0,63	0,36	32,76	1,93	43,15	0,92	30,82	25,21	0,61	0,28	0,36	0,39
IM_19	0,24	0,74	1,34	1,28	14,73	2,03	0,56	32,32	6,57	10,36	9,91	19,74	23,30	0,31	1,11	0,64	0,28
IM_20	0,02	0,18	0,25	0,25	12,38	0,84	0,06	37,90	5,61	28,04	0,26	27,02	26,37	0,13	0,52	0,24	0,23

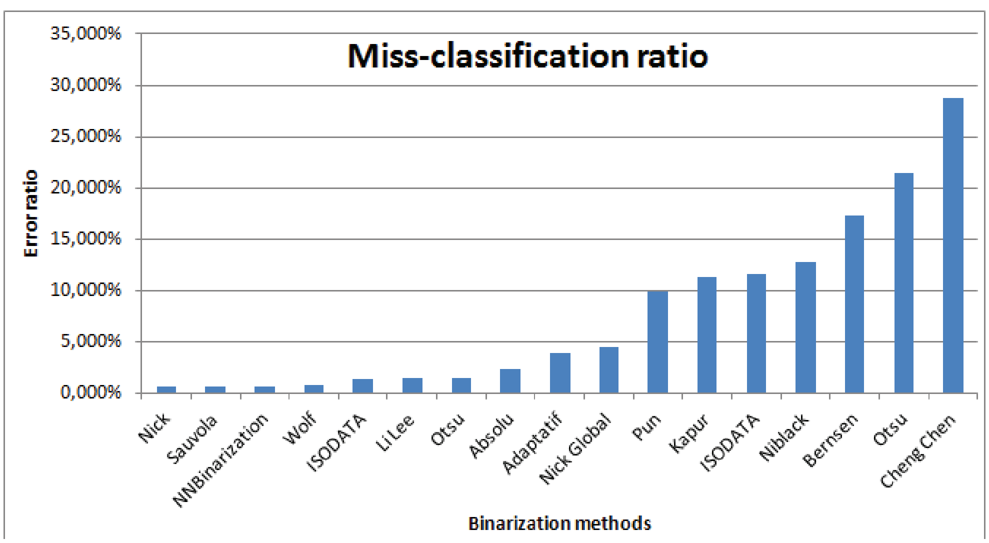


Figure 4. Average rates of miss-classifications.

V. CONCLUSION

In this paper, we presented our method for binarizing old documents based on neural networks. The binarization is an important task in all systems of image processing. 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 tasks. Indeed, as reported in 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 real failures and try to fix them.

REFERENCES

[1] I. Pratikakis, B. Gatos, K. Ntirogiannis: H-DIBCO 2010 - Handwritten Document Image Binarization Competition. ICFHR 2010: 727-732.

[2] B. Gatos, K. Ntirogiannis, I. Pratikakis: DIBCO 2009: document image binarization contest. IJDAR 14(1): 35-44 (2011)

[3] M. Sezgin and B. Sankur, “Survey over image thresholding techniques and quantitative performance evaluation”, Journal of Electronic Imaging, 13(1), 146–165, (2004).

[4] M. Egmont-Petersen, D. de Ridder and H. Handels, “Image processing with neural networks—a review”, Pattern Recognition, 35, 2279–2301, (2002).

[5] N. Papamarkos,C. Strouthopoulos and I. Andreadis, “Multithresholding of color and gray-level images through a neural network technique”, Image and Vision Computing, 18, 213–222, (2000)

[6] G. Kaliraj and S. Baskar, “An efficient approach for the removal of impulse noise from the corrupted image using neural network based impulse detector”, Image and Vision Computing, 28, 458–466, (2010).

[7] H. Kong and L. Guan, “A noise-exclusive adaptive filtering framework for removing impulse noise in digital images”, IEEE Trans. Circ. Syst. —I: Analog Digital Signal Process. 45(3), (1998).

[8] G. V. Cybenco, (1989), “Approximation by superposition of a sigmoidal function” Mathematics of control, Signals, and Systems, 2:303–314.

[9] G. Dreyfus & all, "Neural Networks Methodology and Application", Group Eyrolles, 2004.

[10] P. Stathis, E. Kavallieratou and N. Papamarkos, “An Evaluation Survey of Binarization Algorithms on Historical Documents”, 19th ICPR'08, vol. III, pp. 742-745, 2008

[11] A. Kefali, T. Sari and M. Sellami, "Evaluation of several binarization techniques for old Arabic documents images", First International Symposium on Modeling and Implementing Complex Systems (MISC), Constantine, Algeria, pp: 88-99, 2010.