Optimal combination of document binarization techniques using a self-organizing map neural network

E. Badekas and N. Papamarkos

Image Processing and Multimedia Laboratory
Department of Electrical & Computer Engineering
Democritus University of Thrace,
67100 Xanthi, Greece
papamark@ee.duth.gr
http://ipml.ee.duth.gr/~papamark/

Mailing Address:
Professor Nikos Papamarkos
Democritus University of Thrace
Department of Electrical & Computer Engineering
Image Processing and Multimedia Laboratory
67100 Xanthi, Greece
Telephone: +30-25410-79585
FAX: +30-25410-79569
Email: papamark@ee.duth.gr

Abstract

This paper proposes an integrated system for the binarization of normal and degraded printed documents for the purpose of visualization and recognition of text characters. In degraded documents, where considerable background noise or variation in contrast and illumination exists, there are many pixels that cannot be easily classified as foreground or background pixels. For this reason, it is necessary to perform document binarization by combining and taking into account the results of a set of binarization techniques, especially for document pixels that have high vagueness. The proposed binarization technique takes advantage of the benefits of a set of selected binarization algorithms by combining their results using a Kohonen Self-Organizing Map neural network. Specifically, in the first stage the best parameter values for each independent binarization technique are estimated. In the second stage and in order to take advantage of the binarization information given by the independent techniques, the neural network is fed by the binarization results obtained by those techniques using their estimated best parameter values. This procedure is adaptive because the estimation of the best parameter values depends on the content of images. The proposed binarization technique is extensively tested with a variety of degraded document images. Several experimental and comparative results, exhibiting the performance of the proposed technique, are presented.

Keywords: *Binarization, Thresholding, Document Processing, Segmentation, Page Layout Analysis, Evaluation, Detection.*

1. Introduction

In general, digital documents include text, line-drawings and graphics regions and can be considered as mixed type documents. In many practical applications there is need to recognize or improve mainly the text content of the documents. In order to achieve this, a powerful document binarization technique is usually applied. Documents in the binary form can be recognized, stored, retrieved and transmitted more efficiently than the original gray-scale ones. For many years, the binarization of gray-scale documents was based on the standard bilevel techniques that are also called global thresholding algorithms (Otsu, 1978, Kittler and Illngworth, 1986, Reddi et al., 1984, Kapur et al., 1985, Papamarkos and Gatos, 1994, Chi et al., 1996). These techniques, which can be considered as clustering approaches, are suitable for converting any gray-scale image into a binary form but are inappropriate for complex documents, and even more, for degraded documents. However, the binarization of degraded document images is not an easy procedure because these documents contain noise, shadows and other types of degradations. In these special cases, it is important to take into account the natural form and the spatial structure of the document images. For these reasons, specialized binarization techniques have been developed for degraded document images.

In one category, local thresholding techniques have been proposed for document binarization. These techniques estimate a different threshold for each pixel according to the gray-scale information of the neighboring pixels. In this category belong the techniques of Bernsen (1986), Chow and Kaneko (1972), Eikvil (1991), Mardia and Hainsworth (1988), Niblack (1986), Taxt et al. (1989), Yanowitz and Bruckstein (1989), Sauvola et al. (1997), and Sauvola and Pietikainen (2000). In another category belong the hybrid techniques, which combine information of global and local thresholds. The best known techniques in this category are the methods of Gorman (1994), and Liu and Li (1997).

For document binarization, probably, the most powerful techniques are those that take into account not only the image gray-scale values, but also the structural characteristics of the characters. In this category belong binarization techniques that are based on stroke analysis, such as the stroke width (SW) and characters geometry properties. The most powerful techniques in this category are the Logical Level Technique (LLT) (Kamel and Zhao, 1993) and its improved Adaptive Logical Level Technique (ALLT) (Yang and Yan, 2000), and the Integrated Function Algorithm Technique (IFA) (White and Rohrer, 1983) and its advanced "Improvement of Integrated Function Algorithm" (IIFA) (Trier and Taxt, 1995b). For those two techniques further improvements were proposed by Badekas and Papamarkos (2003). Recently, Papamarkos (2003) proposed a new neuro-fuzzy technique for binarization and gray-level (or color) reduction of

mixed-type documents. According to this technique a neuro-fuzzy classifier is fed by not only the image pixels' values but also with additional spatial information extracted in the neighborhood of the pixels.

Despite the existence of all these binarization techniques, it is proved, by evaluations that have been carried out (Trier and Taxt (1995a), Trier and Jain (1995), Leedham et al. (2003), Sezgin and Sankur (2004)), that there is not any technique that can be applied effectively to all types of digital documents. Each one of them has its own advantages and disadvantages.

The proposed binarization system, takes advantages of binarization results obtained by a set of the most powerful binarization techniques from all categories. These techniques are incorporated into one system and considered as its components. We have included document binarization techniques that gave the highest scores in evaluation tests that have been made so far. Trier and Taxt (1995a) found that Niblack's and Bernsen's techniques are the fastest of the best performing binarization techniques. Furthermore, Trier and Jain's evaluation tests (1995) identify Niblack's technique, with a post-processing step, as the best. Kamel and Zhao (1993) use six evaluation aspects (subjective evaluation, memory, speed, stroke-width restriction, number of parameters of each technique) to evaluate and analyze seven character/graphics extraction based binarization techniques. The best method, in this test, is the Logical Level Technique. Recently, Sezgin and Sankur (2004), compare 40 binarization techniques and they conclude that the local based technique of Sauvola and Pietikainen (2000) as well as the technique of White and Rohrer (IIFA) (1995b) are the best performing document binarization techniques. Apart from the above techniques, we include in the proposed system the powerful global thresholding technique of Otsu (1978) and the Fuzzy C-Mean (FCM) technique (Chi et al., 1996).

The main subject of the proposed document binarization technique is to build a system that takes advantages of the benefits of a set of selected binarization techniques by combining their results, using the Kohonen Self-Organizing Map (KSOM) neural network. This is important, especially for the fuzzy pixels of the documents, i.e. for the pixels that cannot be easily classified. The techniques that are incorporated in the proposed system are the following: Otsu (1978), FCM (Chi et al. 1996), Bernsen (1986), Niblack (1986), Sauvola and Pietikainen (2000), and improvement versions of ALLT (Yang and Yan, 2000 and Badekas and Papamarkos, 2003), and IIFA (Trier and Taxt, 1995b and Badekas and Papamarkos, 2003). Most of these techniques, especially those coming from the category of local thresholding algorithms, have parameter values that must be defined before their application to a document image. It is obvious that different values of the parameter set (PS) of a technique lead to different binarization results which means that there is not a set of best PS values for all types of document images. Therefore, for every

technique, in order to achieve the best binarization results the best PS values must be initially estimated.

Specifically, in the first stage of the proposed binarization technique, a Parameter Estimation Algorithm (PEA), is used to detect the best PS values of every document binarization technique. The estimation is based on the analysis of the correspondence between the different document binarization results obtained by the application of a specific binarization technique to a document image, using different PS values. The proposed method is based on the work of Yitzhaky and Peli (2003) which has been proposed for edge detection evaluation. In their approach, a specific range and a specific step for each one of the parameters is initially defined. The best values for the PS are then estimated by comparing the results obtained by all possible combinations of the PS values. The best PS values are estimated using a Receiver Operating Characteristics (ROC) analysis and a Chi-square test. In order to improve this algorithm, a wide initial range for every parameter is used and in order to estimate the best parameter value an adaptive convergence procedure is applied. Specifically, in each iteration of the adaptive procedure, the parameters' ranges are redefined according to the estimation of the best binarization result obtained. The adaptive procedure terminates when the ranges of the parameters values cannot be further reduced and the best PS values are those obtained from the last iteration.

In order to combine the best binarization results obtained by the independent binarization techniques (IBT) using their best PS values, the Kohonen Self-Organizing Map (KSOM) neural network (Strouthopoulos et al, 2002 and Haykin, 1994) is used as the final stage of the proposed method. Specifically, the neural network classifier is fed using the binarization results obtained from the application of the IBT and a corresponding weighting value that is calculated for each one of them. After the training stage, the output neurons specify the classes obtained, and using a mapping procedure, these classes are categorized as classes of the foreground and background pixels.

The proposed binarization technique was extensively tested by using a variety of document images most of which come from the old Greek Parliamentary Proceedings and from the Mediateam Oulu Document Database (Sauvola and Kauniskangas, 1999). Characteristic examples and comparative results are presented to confirm the effectiveness of the proposed method. The entire system has been implemented in a visual environment.

The main stages of the proposed binarization system are analyzed in Section 2. Section 3, describes analytically the method that is used to detect the best binarization result obtained by the application of the IBT. Also, the same section illustrates the algorithm used for the estimation of the weighting values for each one of the IBT. Section 4, analyses the method used for the

detection of the proper parameter values. Section 5, gives a recent description of the IBT included in the proposed binarization system. Finally, Section 6 presents the experimental results and Section 7 the conclusions.

2. Description of the proposed binarization system

The proposed binarization system performs document binarization by combining the best results of a set of IBT, most of which are developed for document binarization. That is, a number of powerful IBT are included in the proposed binarization system. Specifically, the IBT that were implemented and included in the proposed evaluation system are:

- Otsu (1978),
- FCM (Chi et al. 1996),
- Bernsen (1986),
- Niblack (1986),
- Sauvola and Pietikainen (2000),
- An improved version of ALLT (Yang and Yan, 2000 and Badekas and Papamarkos, 2003) and
- An improved version of IIFA (Trier and Taxt, 1995b and Badekas and Papamarkos, 2003).

The structure of the KSOM used in the proposed technique is depicted in Figure 1. The number of the input neurons is equal to the number of the IBT included in the binarization system. The number of the output neurons is usually taken equal to four. Usually, the KSOM for its convergence needs 300 epochs. The initial neighboring parameter d is taken equal to the number of the output neurons and the initial learning rate α is equal to 0.05. The KSOM is competitively trained according to the following weight update function:

$$\Delta w_{ij} = \begin{cases} a(y_i - w_{ij}), & \text{if } |c - j| < d \\ 0, & \text{otherwise} \end{cases} \qquad (1)$$

where y_i are the input values, c the winner neuron and i, j are the numbers of input and output neurons. The values of parameters a and d are reduced during the learning process.

A step by step description of the proposed document binarization technique follows:

Stage 1   Initially choose the N IBT that will participate in the binarization system.

Stage 2   For those of the selected IBT that have parameters, estimate the best PS values of them for the specific image. Section 4, analyzes the Parameter Estimation Algorithm which is

a converge procedure used to estimate the best PS values for each one of the selected IBT.

Stage 3 Apply the IBT, using their best PS values, to obtain $I_n(x, y)$, $n = 1, ..., N$ binary images.

Stage 4 Compare the binarization results obtained by the IBT and calculate a weighting value $w_n, n = 1, .., N$ for each binary image according to the comparison results obtained using the chi-square test. Analytical description of how the weighting values are calculated is given in Section 3.

Stage 5 Define the S_p set of pixels, the binary values of which will be used to feed the KSOM. If N_p is the number of pixels obtained, and N IBT are included in the system, then we have $N_T = N \cdot N_p$ training samples. The S_p set of pixels must include mainly the "fuzzy" pixels, i.e. the pixels that cannot be easily classified as background or foreground pixels. For this reason, the S_p set of pixels is usually obtained by using the FCM classifier (Chi et al. 1996). Thus, these pixels can be defined as the pixels with high vagueness, as they come up from the application of the FCM method. To achieve this, the image is firstly binarized using the FCM method and then the fuzzy pixels are defined as those pixels having membership function (MF) values close to 0.5. That is, the pixels with MF values close to 0.5 are the vague ones and their degree of vagueness depend on how close to 0.5 are these values.

According to above analysis, the proper S_p set is obtained by the following procedure:

(a) Apply the FCM globally to the original image and for two classes only. After this, each pixel has two MF values: MF_1 and MF_2.

(b) Scale the MF values of each pixel (x, y) to the range $[0, 255]$ and produce a new gray-scale image I_v, using the relation:

$$I_V(x, y) = round\left(255 \frac{(\max(MF_1, MF_2) - \min(MF_1, MF_2)}{1 - \min(MF_1, MF_2)} \right) \qquad (2)$$

(c) Apply the global binarization technique of Otsu to image I_v and obtain a new binary image I_p. The set of pixels S_p is now defined as the pixels of I_p that have value equal to zero.

Instead of using the FCM the S_p set of pixels can be alternatively obtained as:

➢ Edge pixels extracted from the original image by the application of an edge extraction mask. This is a good choice because the majority of fuzzy pixels are close to the edge pixels.

➢ Random pixels sampled from the entire image. This option is used in order to adjust the number of training pixels as a percentage of the total number of image pixels.

Stage 6 The training set S_T is defined as the set of the binary values of the N binary images that correspond to the positions of the S_p set pixels. In this stage the weighting values $w_n, n = 1,..,N$, can be used to adjust the influence of each binary image to the final binarization result via the KSOM. Using the weighting values the training set S_T is composed of the binary values of the N binary images multiplied by the corresponding weighting values.

Stage 7 Define the number of the output neurons of the KSOM and feed the neural network with the values of the S_T set. From our experiments, it has been found that a proper number for the output neurons is $K = 4$. After the training, the centers of the output classes obtained correspond to $O_k, k = 1,...,K$ vectors.

Stage 8 Classify each output class (neuron) as background or foreground by examining the Euclidean distances of their centers from the $[0,...,0]^T$ and $[1,...,1]^T$ vectors, which represent the background and foreground positions in the feature space, respectively. That is

$$O_k = \begin{cases} backgound\ class, & if\ \sqrt{\sum_{i=1}^{N} O_k^2(i)} < \sqrt{\sum_{i=1}^{N} \left[O_k(i) - 1 \right]^2} \\ \\ foreground\ class,\ otherwise \end{cases} \qquad (3)$$

Stage 9 This stage is the mapping stage. Each pixel corresponds to a vector of N elements with values the binary results obtained by the N IBT multiplied by their weighting values. During the mapping process, each pixel is classified by the KSOM to one of the obtained classes, and consequently to background or foreground pixel.

Stage 10 A post-processing step can be applied to improve and emphasize the final binarization results. This step usually includes a size filtering or the application of the MDMF filters (Ping and Lihui, 2001).

3. Obtaining the best Binarization Result

It is obvious that when a document is binarized, it is not known initially the optimum result that must be obtained. This is a major problem in comparative evaluation tests. In order to have comparative results, it is important to estimate a ground truth image. By estimating the ground truth image we can compare the different binarization results obtained, and therefore, we can choose the best. This ground truth image, known as Estimated Ground Truth (EGT) image, can be selected from a list of Potential Ground Truth (PGT) images, as it is proposed by Yitzhaky and Peli (2003) for edge detection evaluation.

Consider N document binary images $D_j(k,l), j=1,...,N$ of size $K \times L$, obtained by the application of a document binarization technique using N different PS values or by N IBT. In order to estimate the best binary image it is necessary to obtain the EGT image. Then, the independent binarization results are compared with the EGT image using the ROC analysis or a Chi-square test.

Let $pgt_i(k,l), i=1,..,N$ be the N PGT images and $egt(k,l)$ the EGT image. In the entire procedure is described below "0" and "1" are considered the background and foreground pixels, respectively.

Stage 1 For every pixel, it is calculated how many binary images consider this as foreground pixel. The results are stored in a matrix $C(k,l)$, $k=0,..,K-1$ and $l=0,..,L-1$. It is obvious that the values of this matrix will be between 0 and N.

Stage 2 N $pgt_i(k,l), i=1,..,N$ binary images are produced using the matrix $C(k,l)$. Every $pgt_i(k,l)$ image is considered as the image that has as foreground pixels all the pixels with $C(k,l) \geq i$.

Stage 3 For each $pgt_i(k,l)$ image, four cases are defined:

- A pixel is a foreground pixel in both $pgt_i(k,l)$ and $D_j(k,l)$ images:

$$TP_{pgt_i,D_j} = \frac{1}{K \cdot L} \sum_{k=1}^{K} \sum_{l=1}^{L} pgt_{i_1}(k,l) \cap D_{j_1}(k,l) \qquad (4)$$

- A pixel is a foreground pixel in $pgt_i(k,l)$ image and background pixel in $D_j(k,l)$ image:

$$FP_{pgt_i,D_j} = \frac{1}{K \cdot L} \sum_{k=1}^{K} \sum_{l=1}^{L} pgt_{i_1}(k,l) \cap D_{j_0}(k,l) \qquad (5)$$

- A pixel is a background pixel in both $pgt_i(k,l)$ and $D_j(k,l)$ images:

$$TN_{pgt_i,D_j} = \frac{1}{K \cdot L} \sum_{k=1}^{K} \sum_{l=1}^{L} pgt_{i_0}(k,l) \cap D_{j_0}(k,l) \qquad (6)$$

- A pixel is a background pixel in $pgt_i(k,l)$ image and foreground pixel in $D_j(k,l)$ image:

$$FN_{pgt_i,D_j} = \frac{1}{K \cdot L} \sum_{k=1}^{K} \sum_{l=1}^{L} pgt_{i_0}(k,l) \cap D_{j_1}(k,l) \qquad (7)$$

$pgt_{i_0}(k,l)$ and $pgt_{i_1}(k,l)$ represent the background and foreground pixels in $pgt_i(k,l)$ image, while $D_{j_0}(k,l)$ and $D_{j_1}(k,l)$ represent the background and foreground pixels in $D_j(k,l)$ image. According to the above definitions, for each $pgt_i(k,l)$ the average value of the four cases resulting from its match with each of the individual binarization results $D_j(k,l)$, is calculated:

$$TP_{pgt_i} = \frac{1}{N} \sum_{j=1}^{N} TP_{pgt_i,D_j} \qquad (8)$$

$$FP_{pgt_i} = \frac{1}{N} \sum_{j=1}^{N} FP_{pgt_i,D_j} \qquad (9)$$

$$TN_{pgt_i} = \frac{1}{N} \sum_{j=1}^{N} TN_{pgt_i,D_j} \qquad (10)$$

$$FN_{pgt_i} = \frac{1}{N} \sum_{j=1}^{N} FN_{pgt_i,D_j} \qquad (11)$$

Stage 4 In this stage, the sensitivity TPR_{pgt_i} and specificity $(1 - FPR_{pgt_i})$ values are calculated according to the relations:

$$TPR_{pgt_i} = \frac{TP_{pgt_i}}{P} \qquad (12)$$

$$FPR_{pgt_i} = \frac{FP_{pgt_i}}{1-P} \qquad (13)$$

where $P = TP_{pgt_i} + FN_{pgt_i}$, $\forall i$

Stage 5 This stage is used to obtain the $egt(k,l)$ image, which is selected to be one of the $pgt_i(k,l)$ images. There are two methods that can be used:

The ROC analysis

The ROC analysis is a graphical method based on a diagram constituted by two curves (CT-ROC diagram). The first curve (the ROC curve) constituted of N points with coordinates (sensitivity, 1-specificity) or $(TPR_{pgt_i}, FPR_{pgt_i})$ and each one of the points is assigned to a $pgt_i(k,l)$ image. The points of this curve are the correspondence levels of the diagram. A second line, which is considered as diagnosis line, is used to detect the Correspondence Threshold (CT). This line has two points with coordinates $(0,1)$ and (P,P). The pgt_i point of the ROC curve which is closest to the intersection point of the two curves is the CT level and defines which $pgt_i(k,l)$ image will be considered as the $egt(k,l)$ image. An example of a CT ROC diagram is shown in Figure 2 for $N=9$. The detected CT level in this example is the fifth.

The Chi-square test

For each pgt_i, the $X^2_{pgt_i}$ value is calculated, according to the relation:

$$X^2_{pgt_i} = \frac{(sensitivity - Q_{pgt_i}) \cdot (specificity - (1 - Q_{pgt_i}))}{(1 - Q_{pgt_i}) \cdot Q_{pgt_i}} \tag{14}$$

where $Q_{pgt_i} = TP_{pgt_i} + FP_{pgt_i}$. A histogram from the values of $X^2_{pgt_i}$ is constructed (CT-Chi-square histogram). The best CT will be the value of i that maximizes $X^2_{pgt_i}$. The $pgt_i(k,l)$ image in this CT level will be then considered as the $egt(k,l)$ image. An example of a CT Chi-square histogram is shown in Figure 3 for $N=9$. The detected CT level in this example is the fifth.

Stage 6 For each image $D_j(k,l)$, four cases are defined:

- A pixel is a foreground pixel in both $D_j(k,l)$ and $egt(k,l)$ images:

$$TP_{D_j,egt} = \frac{1}{K \cdot L} \sum_{k=1}^{K} \sum_{l=1}^{L} D_{j_1}(k,l) \cap egt_1(k,l) \tag{15}$$

- A pixel is a foreground pixel in $D_j(k,l)$ image and background pixel in $egt(k,l)$ image:

$$FP_{D_j,egt} = \frac{1}{K \cdot L} \sum_{k=1}^{K} \sum_{l=1}^{L} D_{j_1}(k,l) \cap egt_0(k,l) \qquad (16)$$

- A pixel is a background pixel in both $D_j(k,l)$ and $egt(k,l)$ images:

$$TN_{D_j,egt} = \frac{1}{K \cdot L} \sum_{k=1}^{K} \sum_{l=1}^{L} D_{j_0}(k,l) \cap egt_0(k,l) \qquad (17)$$

- A pixel is a background pixel in $D_j(k,l)$ image and foreground pixel in $egt(k,l)$ image:

$$FN_{D_j,egt} = \frac{1}{K \cdot L} \sum_{k=1}^{K} \sum_{l=1}^{L} D_{j_0}(k,l) \cap egt_1(k,l) \qquad (18)$$

Stage 7 Stages 4 and 5 are repeated to compare each binary image $D_j(k,l)$ with the $egt(k,l)$ image, using the relations (15)-(18) rather than the relations (8)-(11) which are used in Stage 3. According to the Chi-square test, the maximum value of $X^2_{D_j,egt}$ indicates the $D_j(k,l)$ image which is the estimated best document binarization result. Sorting the values of the Chi-square histogram, the binarization results are sorted according to their quality.

The values of the chi-square test obtained in Stage 7 are considered as the weighting values that are used in Stage 6 of the proposed binarization technique, described in Section 2. The relation that gives the weighting value for each one of the IBT is:

$$w_j = X^2_{D_j,egt} \qquad (19)$$

4. Parameter Estimation Algorithm

In the first stage of the proposed binarization technique it is necessary to estimate the best PS values for each one of the IBT. This estimation is based on the method of Yitzhaky and Peli (2003) proposed for edge detection evaluation. However, in order to increase the accuracy of the estimated best PS values we improve this algorithm by using a wide initial range for every parameter and an adaptive convergence procedure. That is, the ranges of the parameters are redefined according to the estimation of the best binarization result obtained in each iteration of the adaptive procedure. This procedure terminates when the ranges of the parameters values cannot be further reduced and the best PS values are those obtained from the last iteration. It is important to notice that this is an adaptive procedure because it is applied to every processing document image.

The main stages of the proposed PEA, for two parameters (P_1, P_2), are as follows:

Stage 1   Define the initial range of the PS values. Consider as $[s_1, e_1]$ the range for the first parameter and $[s_2, e_2]$ the range for the second one.

Stage 2   Define the number of steps that will be used in each iteration. For the two parameters case, let St_1 and St_2 be the numbers of steps for the ranges $[s_1, e_1]$ and $[s_2, e_2]$, respectively. In most cases $St_1 = St_2 = 3$.

Stage 3   Calculate the lengths L_1 and L_2 of each step, according to the following relations:

$$L_1 = \frac{e_1 - s_1}{St_1 - 1}, \qquad L_2 = \frac{e_2 - s_2}{St_2 - 1} \tag{20}$$

Stage 4   In each step, the values of parameters P_1, P_2 are updated according to the relations:

$$P_1(i) = s_1 + i \cdot L_1, \quad i = 0, ..., St_1 - 1 \tag{21}$$

$$P_2(i) = s_2 + i \cdot L_2, \quad i = 0, ..., St_2 - 1 \tag{22}$$

Stage 5   Apply the binarization technique to the processing document image using all the possible combinations of (P_1, P_2). Thus, N binary images $D_j, j = 1, ..., N$ are produced, where N is equal to $N = St_1 \cdot St_2$.

Stage 6   Examine the N binary document results, using the algorithm described in Section 3 and the chi-square histogram. The best value of a parameter, for example parameter P_1, is equal to the value (from the set of all possible values) that give the maximum sum of level values corresponding to this specific parameter value. For example, in the chi-square histogram of Figure 4, we have nine levels produced by the combination of two parameters having each one three values. As it can be observed, each parameter value appears in three levels. In order to determine the best value of a parameter, we calculate the sums of all level values that correspond to each parameter value and the maximum sum indicates the best value of the specific parameter. In our example, the levels (1, 2, 3) are summed and they compared with the sum of the levels (4, 5, 6) and levels (7, 8, 9). The maximum value defines that the best value of the W parameter is equal to five. In order to estimate the best value of the second parameter k, we compare the sums of the levels (1, 4, 7), (2, 5, 8) and (3, 6, 9) and conclude that the best value is equal to 0.1. Let $(P_{1B}, P_{2B})^T = (5, 0.1)^T$ be the best parameters' values and $(StNo_{1B}, StNo_{2B})^T = (2, 1)^T$

13

the specific steps of the parameters that give these best values. The values of $StNo_{1B}$ and $StNo_{2B}$ must be between $[1, St_1]$ and $[1, St_2]$, respectively.

Stage 7 Redefine the lengths (L_1', L_2') of the steps for the two parameters that will be used during the next iteration of the method:

$$L_1' = \frac{L_1}{2} \text{ and } L_2' = \frac{L_2}{2} \tag{23}$$

Stage 8 Redefine the ranges for the two parameters as $[s_1', e_1']$ and $[s_2', e_2']$ that will be used during the next iteration of the method, according to the relations:

$$s_1' = P_{1B} - (StNo_{1B} - 1) \cdot L_1' \text{ and } e_1' = P_{1B} + (St_1 - StNo_{1B}) \cdot L_1' \tag{24}$$

$$s_2' = P_{2B} - (StNo_{2B} - 1) \cdot L_2' \text{ and } e_2' = P_{2B} + (St_2 - StNo_{2B}) \cdot L_2' \tag{25}$$

Stage 9 Redefine the steps St_1', St_2' for the ranges that will be used in the next iteration according to the relations:

$$St_1' = \begin{cases} \text{if } St_1 > e_1' \text{-} s_1' + 1 \text{ then } St_1' = e_1' \text{-} s_1' + 1 \\ \text{else } St_1' = St_1 \end{cases} \tag{26}$$

$$St_2' = \begin{cases} \text{if } St_2 > e_2' \text{-} s_2' + 1 \text{ then } St_2' = e_2' \text{-} s_2' + 1 \\ \text{else } St_2' = St_2 \end{cases} \tag{27}$$

Stage 10 If $St_1' \cdot St_2' > 3$ go to Stage 4 and repeat all the stages. The iterations are terminated when the calculated numbers of the new steps have a product less or equal to 3 ($St_1' \cdot St_2' \leq 3$). The best PS values are those estimated during the Stage 6 of the last iteration.

5. The binarization techniques included in the proposed system

The seven binarization techniques, which are implemented and included in the proposed binarization system, are already referred in Section 2. The Otsu's technique (1978) is a global binarization method while FCM (Chi et al., 1996) performs global binarization by using fuzzy logic. We use FCM with a value of fuzzyfier m equal to 1.5.

In the category of local binarization techniques belong the techniques of Bernsen (1986), Niblack (1986) and Sauvola and Pietikainen (2000). Each one of these techniques uses two parameters to calculate a local threshold value for each pixel. Then, this threshold value is used locally in order to decide if a pixel is considered as foreground or background pixel. The relations that give the local threshold values $T(x, y)$ for each one of these techniques are:

1. Bernsen's technique

$$T(x, y) = \begin{cases} \dfrac{P_{low} + P_{high}}{2}, & \text{if } P_{high} - P_{low} \geq L \\ \\ GT, & \text{if } P_{high} - P_{low} < L \end{cases} \tag{28}$$

where P_{low} and P_{high} are the lowest and the highest gray-level value in a $N \times N$ window centered in the pixel (x, y), respectively and GT a global threshold value (for example a threshold value that is calculated from the application of the method of Otsu to the entire image). The window size N and the parameter L are the two independent parameters of this technique.

2. Niblack's technique

$$T(x, y) = m(x, y) + k \cdot s(x, y) \tag{29}$$

where $m(x, y)$ and $s(x, y)$ are the local mean and standard deviation values in a N×N window centered on the pixel (x, y), respectively. The window size N and the constant k are the two independent parameters of this technique.

3. Sauvola and Pietikainen's technique

$$T(x, y) = m(x, y)\left[1 + k\left(1 - \frac{s(x, y)}{R}\right)\right] \tag{30}$$

where $m(x, y)$ and $s(x, y)$ are the same as in the previous technique and R is equal to 128 in most cases. The window size N and the constant k are the two independent parameters of this technique.

The ALLT (Kamel and Zhao, 1993, Yang and Yan, 2000) and the IIFA (White and Rohrer, 1983, Trier and Taxt, 1995b) belong to the category of document binarization techniques that are based on structural text characteristics such as the characters' stroke width. In order to improve further the document binarization results, obtained by these techniques, significant improvements are proposed by Badekas and Papamarkos (2003) and these versions of techniques are included in the proposed evaluations system. Our experiments show that these techniques are very stable when they are applied to different types of document images. However, ALLT is more sensitive in the definition of its parameter a compared to the definition of the IIFA's parameter T_p.

6. Experimental Results

The proposed system was tested with a variety of document images most of which come from the old Greek Parliamentary Proceedings and from the Mediateam Oulu Document Database (Sauvola and Kauniskangas, 1999). In this section, four characteristic experiments that include comparative results are presented.

Experiment 1

In this experiment the proposed binarization technique is used to binarize the degraded document image shown in Figure 5, which is coming from the old Greek Parliamentary Proceedings. As it is mentioned in Section 5, the techniques of Niblack, Sauvola and Pietikainen, Bernsen and the ALLT and IIFA have parameters the best values of which must be defined before their application. For the first three binarization techniques, which have two parameters, we use the proposed iterative procedure (3 iterations), while for the rest two techniques which have just one parameter to define, the best value is detected in one iteration, using a wide initial range and 16 steps of their parameter values. The best PS values obtained using the PEA and the initial ranges for the parameters of the techniques are given in Table 1. Table 2 shows all the PS values obtained during the iterations and also the best PS values estimated in each iteration. As it is described in Section 4, the best PS value in every iteration defines the parameters ranges for the next iteration. The best PS values obtained, for the IBT, correspond to the binary document images shown in Figures 6-10. Furthermore, the document binary images obtained by the application of the Otsu and FCM techniques are shown in Figures 11-12. The comparison of the best document binarization results obtained in the previous stage (Figures 6-12) is shown in the chi-square histogram of Figure 13.

Feeding the KSOM with the above binarization results obtain the image shown in Figure 14. In order to achieve the best possible binarization result, the values of the chi-square histogram of Figure 13 are used as weighting values during the feeding of the KSOM. In this way, the final binarization result is obtained and it is shown in Figure 15. The number of the output neurons of the KSOM is taken equal to four and the number of samples in the training is taken equal to 2000. There is not any post-processing technique applied in this experiment.

The binarization results obtained by the proposed technique can be evaluated, comparing them with the binarization results obtained by the seven IBT. The nine binary images (Figures 6-12 and 14-15) can be compared using the method described in Section 3. As it is shown in the Chi-square histogram of Figure 16, the binarization result produced by the KSOM (Figure 14) is better than any of the results of the IBT and the binarization result obtained by the KSOM using the weighting values (Figure 15) is the best one.

Experiment 2

The proposed technique has been extensively tested with document images from the Mediateam Oulu Document Database (Sauvola and Kauniskangas, 1999). Specific examples of the application of the proposed technique to grayscale images obtained from the above database are discussed in this experiment. We use the same procedure, as in the previous experiment, in order to binarize three document images, each one belonging to a different category: Advertisement (Figure 17(a)), Form image (Figure 18(a)) and Article (Figure 19(a)). It should be noticed that for the binarization of the Form image presented in Figure 18(a), is used as a post-processing step, a size filter that removes from the final binary image objects consisting of less than five pixels. The binarization results obtained are shown in Figures 17(b), 18(b) and 19(b), respectively. The histograms of Figure 20 are shown the comparative results obtained by the chi-square test for the three images used in this experiment. From these histograms it is proved that the proposed binarization technique, using the weighting values, gives the best binarization result.

It is important to notice that due to the combination of the binarization results obtained by the IBT, the proposed technique leads to appropriate binarization of the images included in the processing documents. This is obvious from the binarization results of this experiment, where it can be observed that the proposed binarization technique leads to satisfactory results not only in the text regions but also in the image regions included in the documents.

Experiment 3

This is a psycho-visual experiment that is performed asking a group of people to compare the results of the proposed binarization technique with the results obtained by IBT. This test is based on ten document images coming from the Mediateam Oulu Document Database. For these documents we obtain and print the binary images produced by the proposed technique and the seven IBT. The printed documents were given to 50 persons asking them to detect visually the best binary image of each document. All persons conclude that the binary documents obtained by the application of the proposed technique to all testing images are always between the best three binarization results. A percentage of 91.8% (459/500) of the answers, state that the proposed technique, using the weighting values, result to the best binary document.

Experiment 4

This experiment demonstrates the application of the proposed document binarization technique to a large number of document images. The same procedure, as it is described in Experiment 1, was applied to 50 degraded document images obtained from the old Greek Parliamentary Proceedings. In order to obtain the proper value of the output neurons of the

KSOM, the proposed technique is applied to each document image, with 3, 4 or 5 output neurons. For each number of output neurons the KSOM is tested using or not the weighting values. In this way, 6 binary images are produced by the proposed binarization technique and these images are compared with the seven binarization results obtained by the IBT. The comparison of the 13 binary images is made by using, as in the previous experiments, the procedure described in Section 3. In order to have a statistical result for the 13 binarization techniques, the mean value of the chi-square histograms constructed for each processing document image is calculated and the histogram shown in Figure 21 is constructed using these values. It is obvious that the maximum value of this histogram is assigned to the binarization technique which has the best performance. According to the evaluation results obtained, it is concluded that the proposed binarization technique, using the KSOM with four output neurons and weighting values, gives in most cases, the best document binarization results.

7. Conclusions

This paper proposes a new document binarization technique suitable for normal and degraded digital documents. The main idea of the proposed binarization technique is to build a system that takes advantages of the benefits of a set of selected binarization techniques by combining their results using a Kohonen self-organizing neural network. In order to improve further the binarization results, the best parameter values of each document binarization technique are first estimated. Using these values each independent binarization technique leads to best binarization results which are then used to feed the Kohonen Self-Organizing Map neural network. After this, the final binary document image obtained combines the binary information produced by the independent techniques. The proposed technique is suitable to classify pixels that have high vagueness. That is, pixels which belong to edges, shadow areas and generally pixels that cannot be easily classified as foreground or background pixels. The entire system was extensively tested with a variety of degraded document images. Many of them came from standard databases such as the Mediateam Oulu Document Database and the old Greek Parliamentary Proceedings. Several experimental results are presented that confirm the effectiveness of the proposed binarization technique. It is important to notice that the proposed binarization technique lead to satisfactory binarization results not only in text areas but also in image areas included in the documents.

References

Badekas, E., Papamarkos, N., 2003. A system for document binarization, 3rd International Symposium on Image and Signal Processing and Analysis (ISPA), Vol. 2, pp. 909 - 914, Rome.

Bernsen, J., 1986. Dynamic thresholding of grey-level images, Proc. Eighth International Conference on Pattern Recognition, Paris, 1251-1255.

Chi, Z., Yan, H., Pham, T., 1996. Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition, World Scientific Publishing, 225 pages.

Chow, C.K., Kaneko, T., 1972. Automatic detection of the left ventricle from cineangiograms, Computers and Biomedical Research 5, 388-410.

Eikvil, L., Taxt, T., Moen, K., 1991. A fast adaptive method for Binarization of document images, Proceedings of International Conference on Document Analysis and Recognition (ICDAR), France, 435-443.

Gorman, L.O., 1994. Binarization and multithresholding of document images using connectivity, Graphical Models Image Processing (CVGIP), 56 (6), 494-506.

Haykin, S., 1994. Neural Networks: A Comprehensive Foundation, Macmillan College Publishing Company, New York.

Kamel, M., Zhao, A., 1993. Extraction of binary character / graphics images from gray-scale document images, Graphical Models Image Processing (CVGIP), 55 (3), 203-217.

Kapur, J.N, Sahoo, P.K., Wong, A.K., 1985. A new method for gray-level picture thresholding using the entropy of the histogram, Computer Vision Graphics and Image Processing 29, 273-285.

Kittler, J., Illingworth, J., 1986. Minimum error thresholding, Pattern Recognition, 19 (1), 41-47.

Leedham, G., Yan, C., Takru, K., Mian, J.H., 2003. Comparison of Some Thresholding Algorithms for Text/Background Segmentation in Difficult Document Images, Proceedings of 7th International Conference on Document Analysis and Recognition (ICDAR) (2), Scotland, 859 –865.

Liu, Y., Srihari, S.N., 1997. Document image binarization based on texture features, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19 (5), 540-544.

Mardia, K.V., Hainsworth, T.J., 1988. A spatial thresholding method for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 10 (8), 919-927.

Niblack, W., 1986. An Introduction to Digital Image Processing, Englewood Cliffs, N.J. Prentice Hall, 115-116.

Ostu, N., 1978. A thresholding selection method from gray-level histogram, IEEE Trans. Systems Man Cybernet. SMC-8, 62-66.

Papamarkos, N., 2003. A neuro-fuzzy technique for document binarization, Neural Computing & Applications, 12 (3-4), 190-199.

Papamarkos, N., Gatos, B., 1994. A new approach for multithreshold selection, Computer Vision Graphics and Image Processing 56 (5), 357-370.

Ping, Z., Lihui, C., 2001. Document Filters using morphological and geometrical features of characters, Image and Vision Computing 19, 847-855.

Reddi, S.S., Rudin, S.F., Keshavan, H.R., 1984. An optimal multiple threshold scheme for image segmentation, IEEE Transactions on System Man and Cybernetics, 14 (4), 661-665.

Sauvola, J., Kauniskangas, H., 1999. MediaTeam Document Database II, a CD-ROM collection of document images, University of Oulu Finland.

Sauvola, J., Pietikainen, M., 2000. Adaptive Document Image Binarization, Pattern Recognition 33, 225–236.

Sauvola, J., Seppanen, T., Haapakoski, S., Pietikainen, M., 1997. Adaptive Document Binarization, Proceedings of International Conference on Document Analysis and Recognition (ICDAR), Ulm Germany, 147-152.

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

Strouthopoulos, C., Papamarkos, N., Atsalakis, A., 2002. Text extraction in complex color documents, Pattern Recognition, 35(8), 1743-1758.

Taxt, T., Flynn, P.J., Jain, A.K., 1989. Segmentation of document images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (12), 1322-1329.

Trier, O.D., Jain, A.K., 1995. Goal-Directed Evaluation of Binarization Methods, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17 (12), 1191-1201.

Trier, O.D., Taxt, T., 1995a. Evaluation of binarization methods for document images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17 (3), 312-315.

Trier, O.D., Taxt, T., 1995b. Improvement of 'Integrated Function Algorithm' for binarization of document images, Pattern Recognition Letters 16, 277-283.

White, J.M., Rohrer, G.D., 1983. Image segmentation for optical character recognition and other applications requiring character image extraction, IBM J. Res. Dev., 27 (4), 400-411.

Yang, Y., Yan, H., 2000. An adaptive logical method for binarization of degraded document images, Pattern Recognition 33, 787-807.

Yanowitz, S.D., Bruckstein, A.M., 1989. A new method for image segmentation, Computer Vision, Graphics and Image Processing, 46 (1), 82-95.

Yitzhaky, Y., Peli, E., 2003. A Method for Objective Edge Detection Evaluation and Detector Parameter Selection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25 (8), 1027-1033.

Table 1. Initial ranges and the estimated best *PS* values for Experiment 1

	Technique	Initial ranges	Best *PS* values
1.	Niblack	$W \in [1,9]$, $k \in [0.1,1]$	$W = 5$ and $k = 0.32$
2.	Sauvola	$W \in [1,9]$, $k \in [0.1,0.4]$	$W = 5$ and $k = 0.17$
3.	Bernsen	$W \in [1,9]$, $L \in [10,100]$	$W = 5$ and $L = 89$
4.	ALLT	$a \in [0.1,0.25]$	$a = 0.10$
5.	IIFA	$T_p \in [10,40]$	$T_p = 24$

Table 2. The detection of the best *PS* values for each binarization technique

Iterations	Niblack	Sauvola	Bernsen	ALLT	IIFA
First	1. W=1, k=0,1 2. W=1, k=0,55 3. W=1, k=1 4. W=5, k=0,1 (Best) 5. W=5, k=0,55 6. W=5, k=1 7. W=9, k=0,1 8. W=9, k=0,55 9. W=9, k=1	1. W=1, k=0,1 2. W=1, k=0,25 3. W=1, k=0,4 4. W=5, k=0,1 (Best) 5. W=5, k=0,25 6. W=5, k=0,4 7. W=9, k=0,1 8. W=9, k=0,25 9. W=9, k=0,4	1. W=1, L=10 2. W=1, L=55 3. W=1, L=100 4. W=5, L=10 5. W=5, L=55 6. W=5, L=100 (Best) 7. W=9, L=10 8. W=9, L=55 9. W=9, L=100	1. a=10 (Max) 2. a=11 3. a=12 4. a=13 5. a=14 6. a=15 7. a=16 8. a=17 9. a=18	1. Tp=10 2. Tp=12 3. Tp=14 4. Tp=16 5. Tp=18 6. Tp=20 7. Tp=22 8. Tp=24 (Max) 9. Tp=26
Second	1. W=3, k=0,1 2. W=3, k=0,32 3. W=3, k=0,54 4. W=5, k=0,1 5. W=5, k=0,32 (Best) 6. W=5, k=0,54 7. W=7, k=0,1 8. W=7, k=0,32 9. W=7, k=0,54	1. W=3, k=0,1 2. W=3, k=0,17 3. W=3, k=0,24 4. W=5, k=0,1 5. W=5, k=0,17 (Best) 6. W=5, k=0,24 7. W=7, k=0,1 8. W=7, k=0,17 9. W=7, k=0,24	1. W=3, L=56 2. W=3, L=78 3. W=3, L=100 4. W=5, L=56 5. W=5, L=78 6. W=5, L=100 (Best) 7. W=7, L=56 8. W=7, L=78 9. W=7, L=100	10. a=19 11. a=20 12. a=21 13. a=22 14. a=23 15. a=24 16. a=25	10. Tp=28 11. Tp=30 12. Tp=32 13. Tp=34 14. Tp=36 15. Tp=38 16. Tp=40
Third	1. W=4, k=0,21 2. W=4, k=0,32 3. W=4, k=0,43 4. W=5, k=0,21 5. W=5, k=0,32 (Best) 6. W=5, k=0,43 7. W=6, k=0,21 8. W=6, k=0,32 9. W=6, k=0,43	1. W=4, k=0,14 2. W=4, k=0,17 3. W=4, k=0,2 4. W=5, k=0,14 5. W=5, k=0,17 (Best) 6. W=5, k=0,2 7. W=6, k=0,14 8. W=6, k=0,17 9. W=6, k=0,2	1. W=4, L=78 2. W=4, L=89 3. W=4, L=100 4. W=5, L=78 5. W=5, L=89 (Best) 6. W=5, L=100 7. W=6, L=78 8. W=6, L=89 9. W=6, L=100		
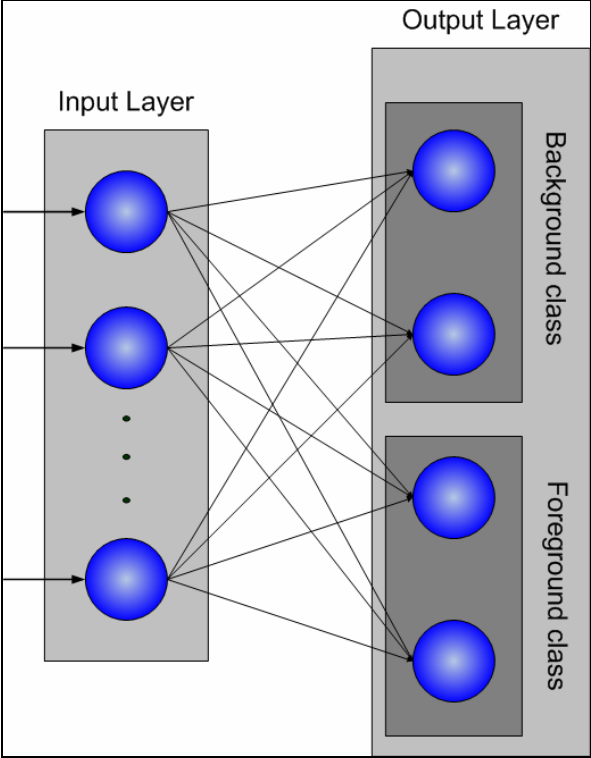
Figure 1. The structure of the Kohonen Self-Organizing Map neural network
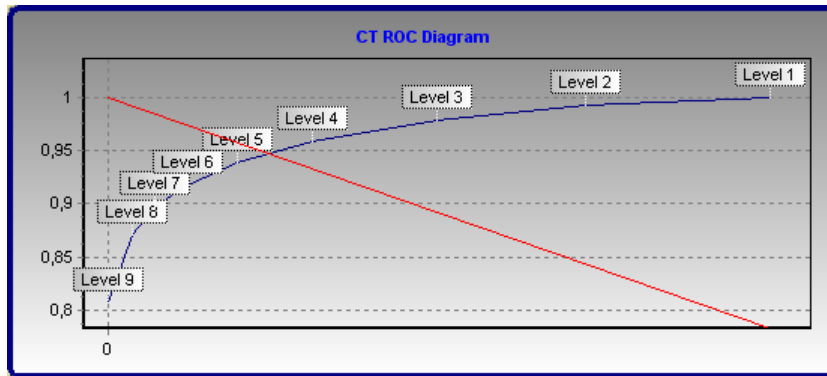
Figure 2. An example of a CT ROC diagram. The 5th level is the CT level
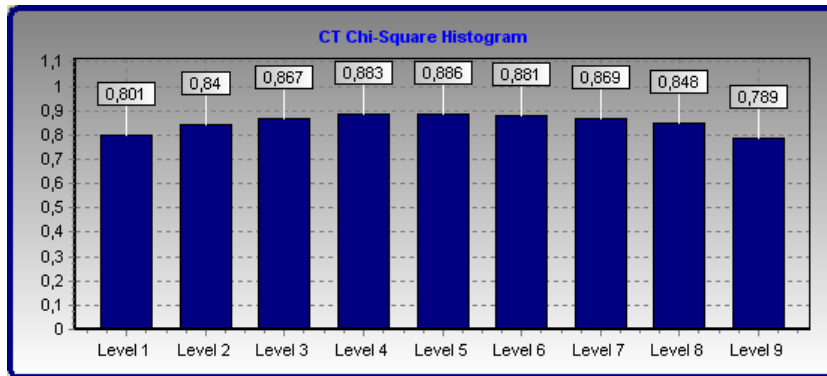

Figure 3. An example of a CT Chi-square histogram. The 5th level is the CT level
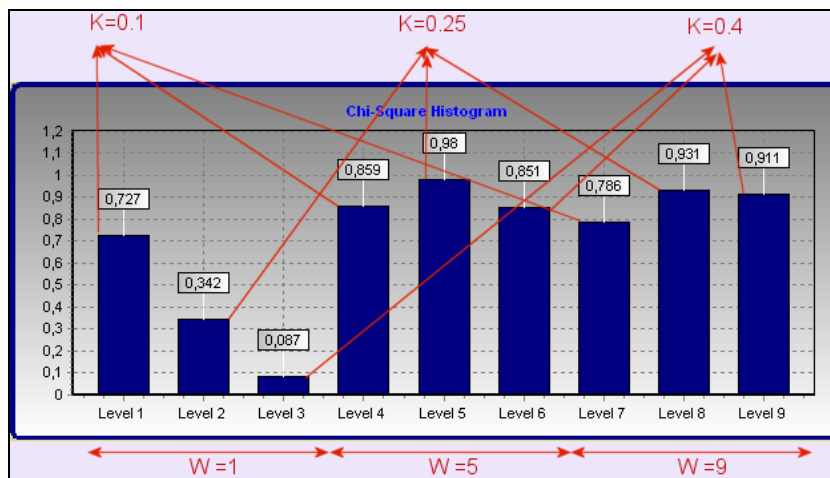

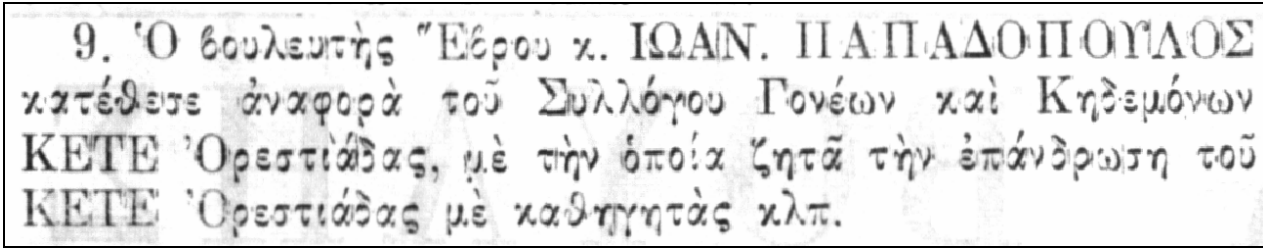**Figure 4**. A 9-level Chi-square histogram. The detected best values are $W = 5, k = 0.25$ in the fifth level

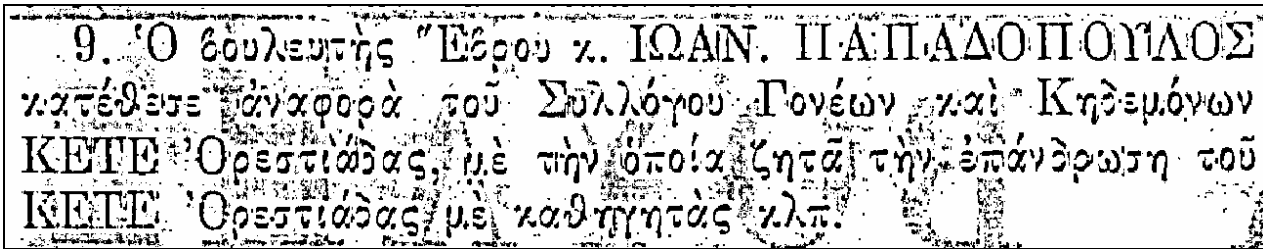**Figure 5**. Initial gray-scale document image for Experiment 1

Figure 6. Binarization result of Niblack's technique with $W = 5$ and $k = 0.32$
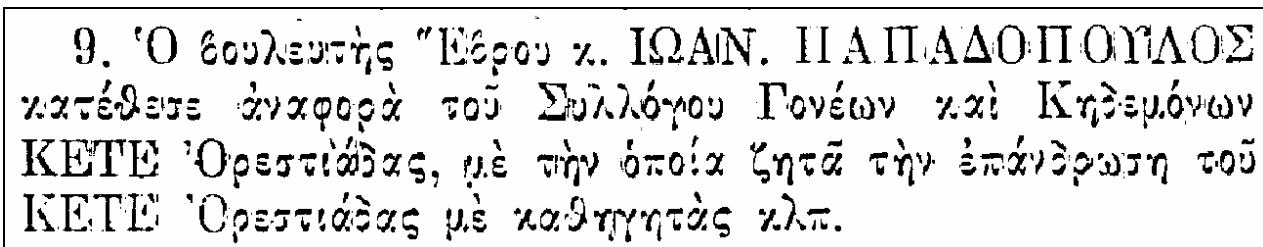
Figure 7. Binarization result of Sauvola's technique with $W = 5$ and $k = 0.17$
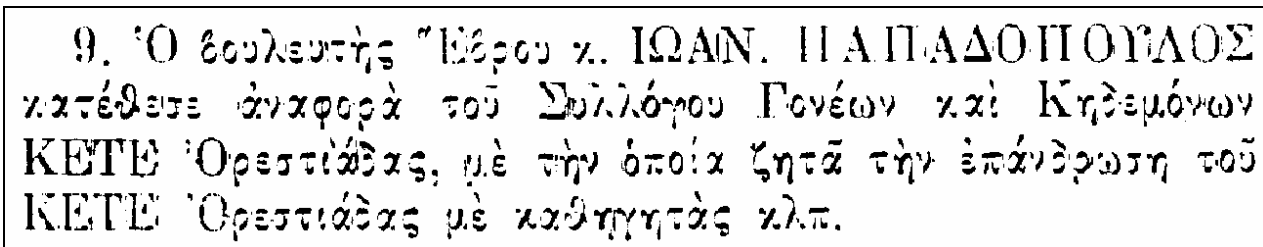
Figure 8. Binarization result of Bernsen's technique with $W = 5$ and $L = 89$
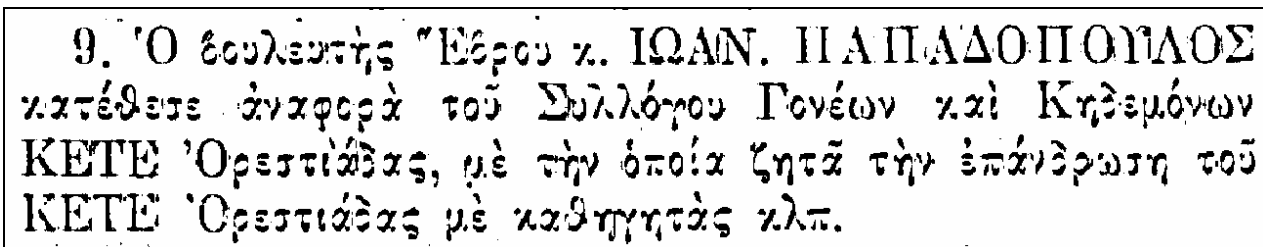
Figure 9. Binarization result of ALLT with $a = 0.1$

Figure 10. Binarization result of IIFA with $T_p = 24$

Figure 11. Binarization result of Otsu's technique

Figure 12. Binarization result of FCM
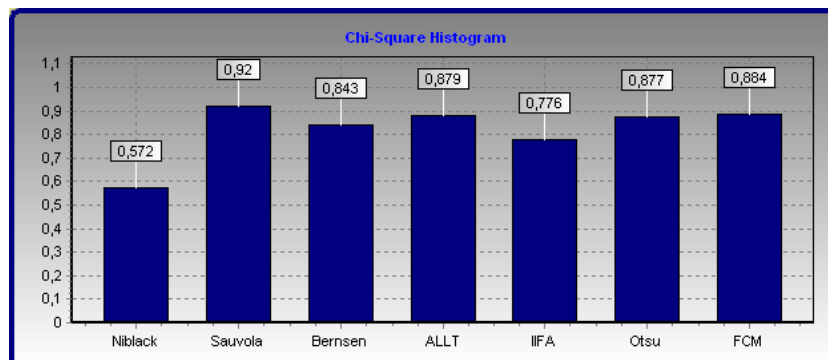
Figure 13. Comparison of the binarization results obtained by the IBT

Figure 14. The binarization result obtained feeding the KSOM by the IBT

Figure 15. The final binarization result using the KSOM and the weight values obtained by the comparison of the IBT
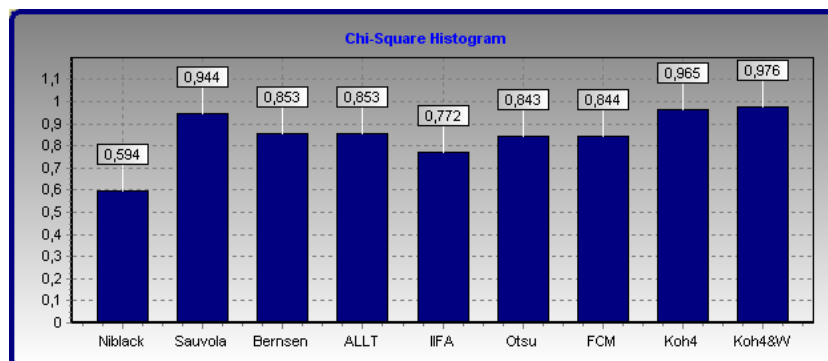

Figure 16. Comparison of the binarization results obtained by the IBT and the proposed technique

(a)

(b)

Figure 17. (a) The initial gray-scale advertisement image (b) The binarization result obtained by the proposed technique

TIME

SUBSCRIPTION SERVICE FORM

☐ Moving? Check this box. Add your magazine label and name and address below. Airmail to: TIME Magazine, 5 Ottho Heldringstraat, Amsterdam 1018 Netherlands.

☐ Subscribing? Check this box. Also check ☐ new or ☐ renewal. Give name and address below. Mail coupon with payment according to list below

Please attach magazine label here, list new address below, and mail six to eight weeks before you move. If you are receiving duplicate copies of TIME, please send both labels. Or if you have a question about your subscription, attach label here and clip this form to your letter.

Name   (BLOCK LETTERS, please)   (first)   (last)

Address

City   Country   Postcode

VAT 1

(a)

TIME

SUBSCRIPTION SERVICE FORM

☐ Moving? Check this box. Add your magazine label and name and address below. Airmail to TIME Magazine, 5 Ottho Heldringstraat, Amsterdam 1018 Netherlands

☐ Subscribing? Check this box. Also check ☐ new or ☐ renewal. Give name and address below. Mail coupon with payment according to list below

Please attach magazine label here, list new address below, and mail six to eight weeks before you move. If you are receiving duplicate copies of TIME, please send both labels. Or if you have a question about your subscription, attach label here and clip this form to your letter.

Name   (BLOCK LETTERS, please)   (first)   (last)

Address

City   Country   Postcode

VAT 1

(b)

Figure 18. (a) The initial gray-scale form image (b) The binarization result obtained by the proposed technique
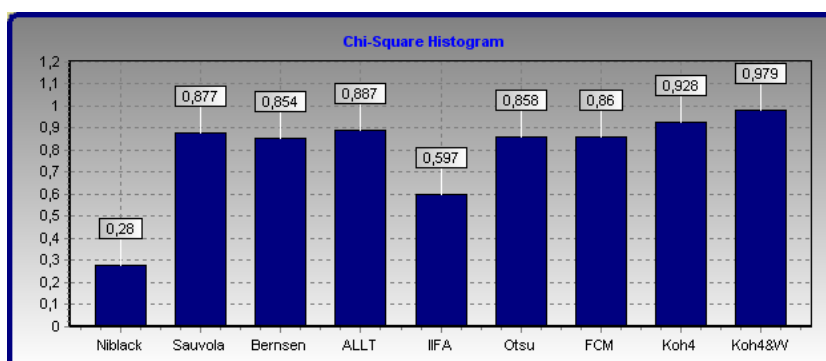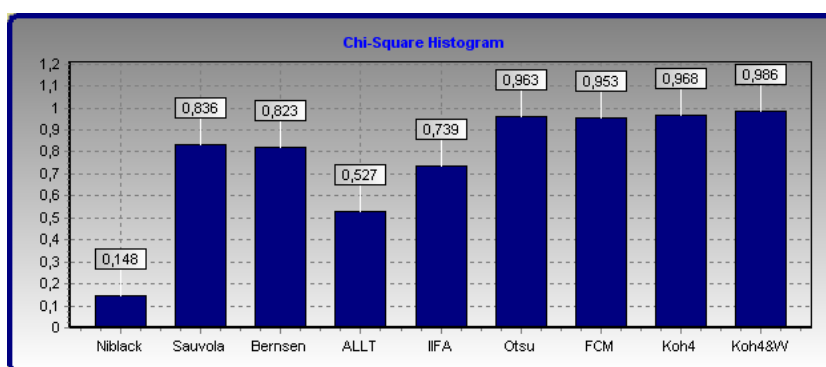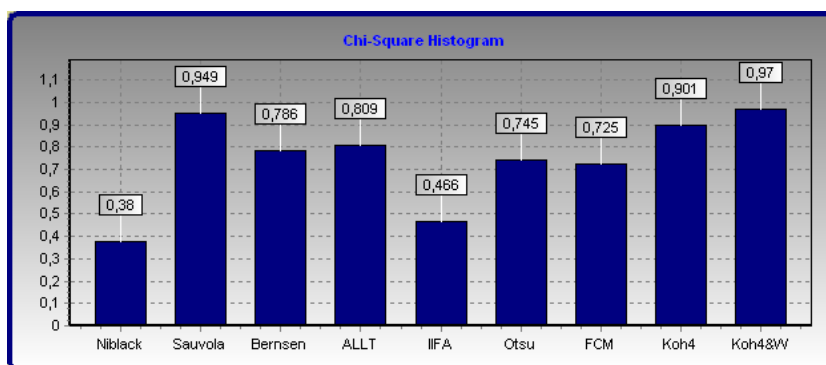

(a)

(b)

Figure 19. (a) The initial gray-scale article image (b) The binarization result obtained by the proposed technique

(a)

(b)

(c)

Figure 20. Comparison of the binarization results obtained for the document images of (a) Figure 17 (b) Figure 18 and (c) Figure 19.
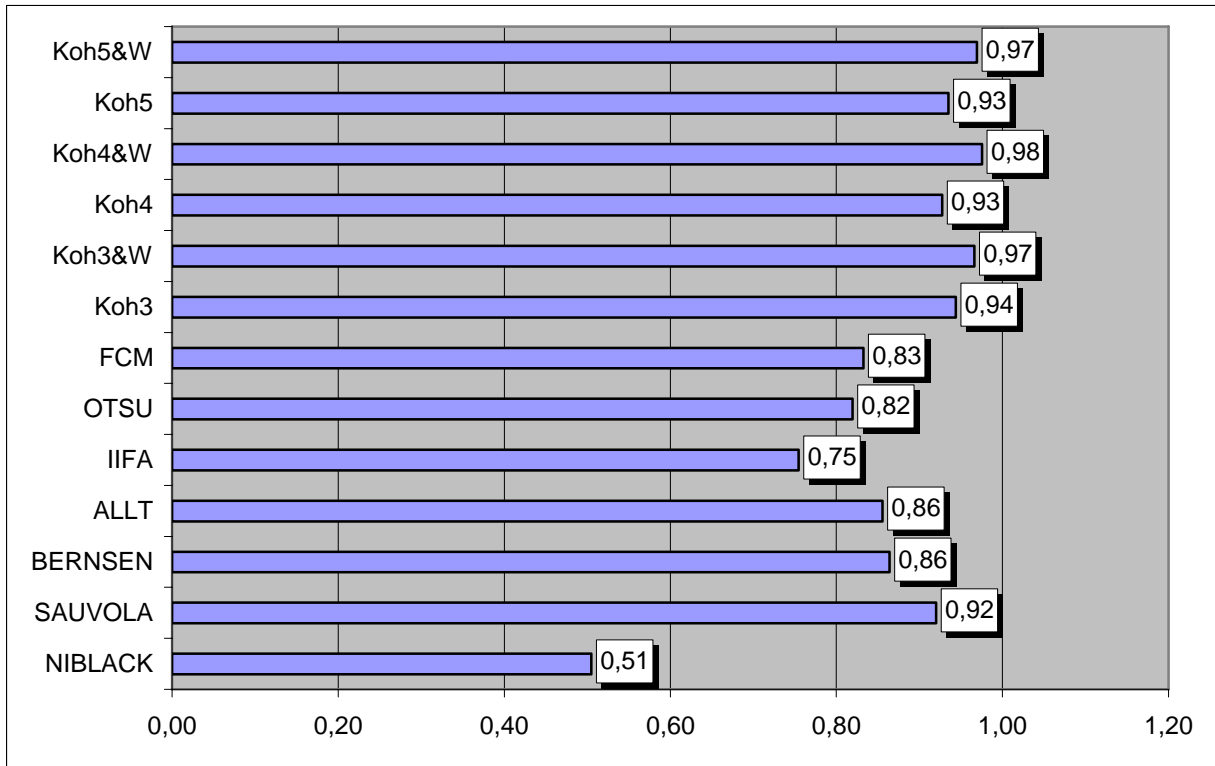
Figure 21. The histogram constructed by the mean values estimated for each binarization technique by the chi-square tests in Experiment 4.