NEAR EAST UNIVERSITY

GRADUATE SCHOOL OF APPLIED SCIENCES

NOVEL IMAGE BINARIZATION METHOD WITH APPLICATION TO DOCUMENT ENHANCEMENT

Boran Şekeroğlu

PhD Dissertation

Department of Computer Engineering

Nicosia – 2007

ABSTRACT

Thresholding is an efficient method for the binarization of the images where the relationship between pixel values in the images can provide an effective basis point for the separation of the background and foreground layers. Several image binarization methods have been developed and used for different types of applications, however, the efficiency of these methods can be impaired by the variation of gray levels in these different applications, thus causing over-thresholding, under-thresholding or noise addition. This dissertation presents a single-stage global thresholding method that enhance document images by clearly separating background and foreground layers within these images and investigates the use of the mean value in direct local thresholding of the images. The proposed method which is global, is named Mass-Difference (MD) thresholding. It finds an appropriate thresholding value for each image using the relationship between luminance value and mean intensity of the image without considering peak values in the gray level histogram. The investigated local method, named Pattern Averaging Threshold (PAT) determines the mean of the defined segments and uses this value as threshold point without any approximation. PAT is used to visualize the hidden information within the images and to prepare the inputs of an intelligent system to reduce the 'learning' time of the neural network. Experimental results of PAT suggest that, it can be used to visualize the hidden data which is important especially in security and the forensic sciences and it is also an effective data preprocessing task for the intelligent systems. The proposed MD and PAT methods are implemented using a database that was especially collected and constructed to have different types of challenging document images comprising 175 historical documents, specially created words and handwritten text. Both methods are compared with 12 benchmark and/or recently developed global and local thresholding methods. The evaluation of the thresholding methods aims at determining a superior thresholding method that can be efficiently applied to a variety of images such as scanned documents. Evaluation is performed using visual inspection and computed noise analysis; that uses three new PSNR-derived metric parameters. Experimental results suggest that the developed MD global method is superior in providing a fast and efficient text separation in document images.

Dedicated to my dear wife and daughter . . .

ACKNOWLEDGMENTS

I would like to thank everyone who provided help and advice during the preparation of this dissertation.

First, I would like to thank my supervisor Assoc. Prof. Dr. Adnan Khashman for his invaluable advice and belief in my work and myself over the course of this Ph.D. Research.

Second, I would like to express my gratitude to Near East University and Thesis Supervision Committee Members, Prof. Dr. Fahreddin M. Sadıkoğlu, Assoc. Prof. Dr. Rahib Abiyev and Assist. Prof. Dr. Hüseyin Sevay for their advice.

Third, I would like to thank my family for their constant encouragement, support and patience during the preparation of this dissertation.

Finally, I would also like to thank my wife Süsen D. Şekeroğlu and my daughter Dilara Naz Şekeroğlu for their existence.

CONTENTS

LIST OF ABBREVIATIONS

IN	Image Negatives
LT	Log Transformations
PLT	Power-Law Transformations
PLTF	Piecewise-Linear Transformation Functions
HE	Histogram Equalization
FT	Fourier Transform
DFT	Discrete Fourier Transform
ILPF	Ideal Low Pass Filters
BLPF	Butterworth Low Pass Filter
GLPF	Gaussian Low Pass Filter
IHPF	Ideal High Pass Filter
BHPF	Butterworth High Pass Filter
GHPF	Gaussian High Pass Filter
CT	Computed Tomography
MRI	Magnetic Resonance Image
FFT	Fast Fourier Transform
PDF	Probability Density Function
PAT	Pattern Averaging Thresholding
ALT	Adaptive Logical Thresholding
WFM	Water Flow Model
MD	Mass-Difference
PSNR	Peak Signal-to-Noise Ratio
APAR	Average PSNR Accuracy Rate
APD	Average PSNR Deviation
CPR	Combined Performance Rate
MSE	Mean-Squared Error
RW	Recognized Word
WP	White Paper
WBM	White Board Marker
YP	Yellow Envelope Paper
ICIS	Intelligent Coin Identification System

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Image Binarization (thresholding) is the low-level spatial domain image processing technique that is intended to enhance or segment the 'relevant data' or the 'region of interest' within the images. It is based on the assumption that objects ('region of interest') and background layers in the image can be distinguished by their gray level values. Binarization methods can be categorized into two groups as global thresholding and local (adaptive) thresholding methods. Global thresholding is a simple and efficient method where a defined or computed threshold value is used to separate foreground objects from background by considering whole image characteristics and Local (Adaptive) Thresholding is the assigning of a value to each pixel to determine whether it is a foreground or background pixel using local information from the image. Several thresholding methods that belong to these two groups have been developed. Both binarization groups carry some disadvantages beside their apparent advantages. Global methods have faster execution time that minimizes the computational cost and the noise in resultant images. However, local noise may affect the whole binarization process while change of partial characteristics of the image also changes whole characteristics that cause under or over thresholded images.

Local methods have variable execution time depends on the size of the defined segments – small sized segments have longer execution time and large sized segments have faster execution time– and the noise addition, variability of the segment sizes and the variable parameters are the main disadvantages of the local methods. Small segment sizes add additional noise into resultant images when the gathered information of a segment does not consist any information that belong to region of the interest. This yields the

visualization of the unnecessary information within the segments and causes additional noise within binarized images. Large segment sizes may decrease the noise addition, however they may also act as the global method and sometimes cause the loss of the relevant data within the segments. Although these disadvantages are the serious drawback of the local methods, the main advantage of them is the more clean and readable output of the relevant data when the segment size is small enough to enhance the region of interest and large enough to suppress the noise.

The main application areas of the image binarization are the fields that requires the enhanced or separated data for any system. However, document analysis is still the most popular area that uses image binarization for enhancing or separating the region of interest which is the text in document images. Digitized document analysis has recently become more significant with the advances in digital archiving and electronic libraries. Scanned document images, especially historical and handwritten documents, generally carry various levels of noise because of the age, paper, pen and pencil influences on the documents. Age factor adds irremovable noise and meaningless random shapes on the documents which prevent efficient separation and recognition of the layers. Paper properties such as patterned or colored papers; add different background layers to the scanned documents. In addition, the variety of pens and pencils produces different and various foreground layers for the documents. Therefore, efficient binarization of scanned paper-based documents is usually required prior to further processing. The efficiency of document image binarization depends on the efficient separation and classification of background and foreground layers and the efficiency of a binarization method can be defined as producing a background layer that does not contain any information belonging foreground (text) layer and the foreground layer that does not contain any noise from background layer.

With the existence of many global and local thresholding methods, deciding upon an optimum method for document image binarization is a challenging task; because the efficiency of existing thresholding methods is usually application-dependent where one methods performance appears superior when using a certain type of document, but fails on a different type of document. The solution to this problem would be in creating and using a comprehensive multi-applications document image database that accounts for

different types of documents, such as historical documents, degraded documents, artificially created words, and handwritten documents.

Several comparisons have been previously performed in order to evaluate existing thresholding methods and deciding upon an optimum thresholding method for document binarization in particular. The more comprehensive comparisons were performed by Trier and Taxt [1], Trier and Jain [2], Leedham et al. [3], Sezgin and Sankur [4] and He et al. [5].

These comparative studies have attempted to suggest an optimum thresholding method that can be efficiently used for document image binarization. However, results of these different evaluations suggested different methods as being superior; which is anticipated as the image databases differ from one evaluation to another; where one evaluation uses historical documents, others use created words, or artificially degraded document scans. Another problem is the insufficient number of document images used in some of these evaluations [1, 2, 3] which affects the significance of the evaluation outcome. In addition, using a large number of images that have similar noise and layer characteristics [5], does not provide an effective evaluation. Moreover, the use of visual inspection as in [1], without any computed analysis, as the only or main criteria for evaluation may not provide a robust evaluation outcome. On the other hand, the use of OCR module with some historical documents is not possible due to old different fonts that can not be recognized by the available OCR modules. Finally, there is a lack of clear categorization of thresholding methods into adaptive local methods and global methods when performing the evaluations. Such clear categorization would greatly aid in providing a more objective comparison and in suggesting an overall superior thresholding method or a category-based superior thresholding method.

This thesis presents a new global thresholding method named as Mass-Difference (MD) Thresholding. Additionally, Pattern Averaging Thresholding (PAT) which is based on the direct use of local mean values of images as threshold points, is investigated. Also a comprehensive comparative evaluation of MD, PAT and 12 benchmark and recent thresholding methods that can be used for document image binarization is provided. The objectives of the work presented in this thesis can be summarized as shown in next section.

1.1 Contribution

- Design and development of an efficient global thresholding method which is named as mass difference (MD) thresholding method for image binarization.
- Investigating the use of the mean value as a direct threshold value within the segments of local thresholding method which is named as pattern averaging thresholding (PAT) especially for the visualization of the hidden data within the images.
- Creating and using a comprehensive multi-applications document image database that includes historical documents, degraded documents, handwritten and artificially created words within bright and low-contrast and dark images with sufficient number of images.
- Implementing document image binarization using 14 thresholding methods, including the proposed and investigated methods, (seven global methods and seven local methods).
- Defining and implementing two evaluation and comparison criteria: visual inspection and computed noise analysis of binarized images.
- Comparing the performance of the 14 methods and determining a superior thresholding method for each group independently and for the overall groups.

1.2 Thesis Overview

The remaining chapters of this dissertation are organized as follows:

- *Chapter 2* briefly describes the fundamentals of basic spatial and frequency domain image enhancement methods.
- *Chapter 3* reviews the benchmark and recent global and local methods, and advantages and disadvantages of these methods.
- *Chapter 4* introduces the proposed global method, investigated local method, statistical and sufficiency experiments and comparisons.

- *Chapter 5* presents the multi-application document image database, the evaluation procedure (which includes three new evaluation parameters) and the performed comparative evaluation.
- Concluding remarks of this thesis and future work are given in *Chapter 6*.

CHAPTER 2

IMAGE ENHANCEMENT

2.1 Overview

Image enhancement is the process that intends to increase the visual appearance of digital images, graphics or photographs and, the enhancement methods are application-specific and are often developed empirically [6]. Thus, a method that is superior for enhancing X-ray images may not necessarily be appropriate for enhancing pictures of Mars transmitted by a space probe [7].

In this chapter, definitions of image enhancement, its techniques and application areas of these techniques will be described.

2.2 Image Enhancement Approaches

Image enhancement approaches can be divided into two categories: spatial domain methods and frequency domain methods. Spatial domain is the normal image space and frequency domain is the continuous signal of an image. The fundamental difference between these two approaches is the processing way of enhancement techniques. In the spatial domain approach, techniques are based on direct manipulation of pixels. In the frequency domain approach, techniques are based on the modification of the Fourier Transform [7].

2.2.1 Overview of Spatial Domain Image Enhancement Techniques

Spatial domain image enhancement techniques operate on pixels in image space and the processes are denoted as follows [7].

$$g(x, y) = T\left[f(x, y)\right] \tag{2.1}$$

where $f(x, y)$ is the input image, $g(x, y)$ is the processed image, and T is an operator on f, defined over some neighborhood of (x, y). So, *grayscale* (also called *intensity* or *mapping* [7]) *transformation function* can be obtained by determining neighborhood size T as 1×1. Consequently, in single pixel neighborhood, T becomes grayscale transformation function where g depends only on value of f at (x, y). This form can be rewritten as:

$$s = T(r) \tag{2.2}$$

where r and s are variables denoting, respectively the gray level of $f(x, y)$ and $g(x, y)$ at any point (x, y) [7].

Basic Gray Level Transformations in Spatial Domain

Several transformation functions and techniques had been developed by modifying the grayscale transformation function such as Image Negatives (IN), Log Transformations (LT), Power-Law Transformations (PLT) and Piecewise-Linear Transformation Functions (PLTF).

Image Negatives are used to obtain photographic negative of an image by applying the negative transformation which is given in *Equation 2.3*.

$$s = L - 1 - r \tag{2.3}$$

where L is the gray-level range of a given image defined as $[0, L - 1]$.

Logarithmic transformations are used to expand the spectrum of dark pixels while compressing the spectrum of higher value pixels in an image. The general form of the

logarithmic transformations is given in *Equation 2.4*.

$$s = c\log(1 + r) \tag{2.4}$$

where c is a constant. For specific applications, it is also possible to use the inverse logarithmic transformation to expand the spectrum of higher value pixels while compressing the spectrum of dark pixels.

The Power-Law transformation, given in *Equation 2.5*, provides a more flexible transformation curve than LT according to the value of c and γ. If $\gamma < 1$, PLT produces expanded spectrum of dark pixels while producing compressed spectrum of higher value pixels, and in other case, if $\gamma > 1$ it produces expanded spectrum of higher value pixels while produces compressed spectrum of dark pixels. Identity transformation is obtained if $\gamma = 1$ (Note that $c = 1$ for all cases).

$$s = cr^{\gamma} \tag{2.5}$$

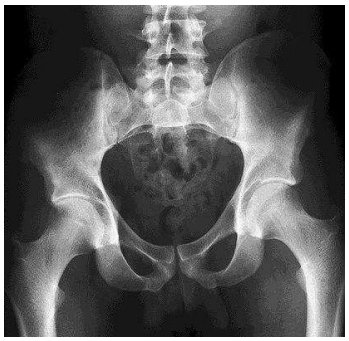
where c and γ are positive constants.

Piecewise-linear transformation consists of several functions such as contrast stretching, gray-level slicing and bit-plane slicing which are used for image enhancement.

Contrast stretching is one of the simplest and most important approaches for piecewise linear transformation. During image acquisition, images may become low-contrast because of poor illumination. The idea of contrast stretching is to increase the dynamic range of the gray levels in the image being processed [7], and the typical formula is given in *Equation 2.6* [3,4].

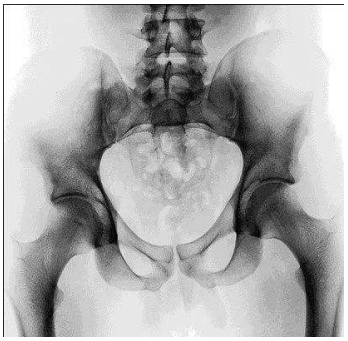
$$s = (r - c)\left(\frac{b - a}{d - c}\right) + a \tag{2.6}$$

where, s and r denotes output and input images respectively, a and b denotes lower and upper limits of image respectively (between 0 and 255 in 8 bit grayscale image) and c and d represent the lowest and highest pixel values in an image. *Figure 2.1* shows the implementation of IN, LT, and PLT. *Figure 2.2* shows contrast stretching.



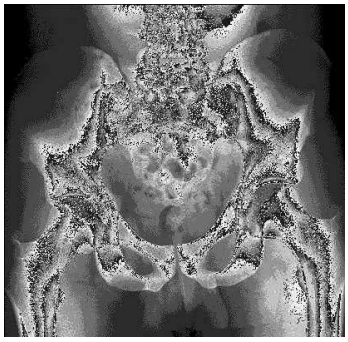
(a) Original image

(b) Image after log transformation



(c) Image after applying image negatives

(d) Image after power-law transformation with $\gamma = 0.8$



(e) Image after power-law transformation with $\gamma = 1.2, c = 1$

Figure 2.1: Implementation of various transformations on an X-ray image

(a) Original low-contrast image

(b) Enhanced image after contrast stretching

Figure 2.2: Contrast stretching on an X-ray image

Histogram Processing in Spatial Domain

In the spatial domain, histogram processing is an important approach for image enhancement and it is the basis for numerous processing techniques [7]. Histogram is the discrete function of digital image in the range k as $[0, L-1]$ and it is defined as :

$$h(r_k) = n_k \tag{2.7}$$

where r_k is the k^{th} gray level and n_k is the number of pixels in the image having gray level r_k. Thus, it is not complicated to say that probability of occurrence of gray level r_k ($p(r_k)$) is estimated by dividing its values by total number of pixels in the image, which is denoted as n in *Equation 2.8*. Also it is known as the normalization of a histogram.

$$p(r_k) = \frac{n_k}{n} \tag{2.8}$$

One of the basic applications of histograms is the determination of the contrast level (or image types [7]) of images such as dark image, bright image, low contrast image and high contrast image.

Dark image can be defined as the collection of image pixels in the range $[0, n]$, without having pixel values in the range $[n, L-1]$ where n is the gray level limit of image pixels and can be assumed as the central value of 8 bit gray level which is 128.

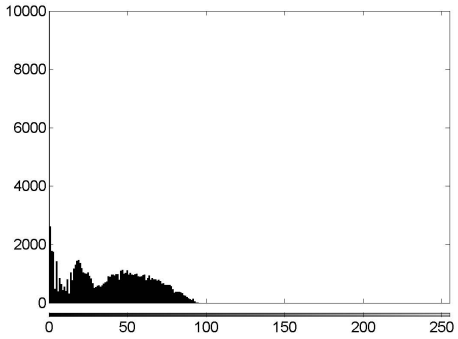
A bright image can be defined as the collection of image pixels in the range $[n, L-1]$, without having pixel values in the range $[0, n]$.

Low-contrast images have more complex relationship in the upper and lower limits of gray level values. An image can be classified as a low contrast image if the image pixels are collected in the range $[n-z, n+z]$ where z is a variable that determines the upper and lower limits of image pixels.

In ideal case, high-contrast image can be defined as the equal distribution of image pixels in the range $[0, L-1]$. Examples of dark, bright, low-contrast and high contrast image with their corresponding histograms are given in *Figure 2.3*.

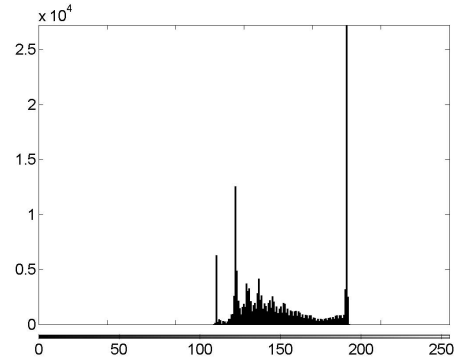
As mentioned above, probability of occurrence in an histogram can be computed us-

(a) Dark image

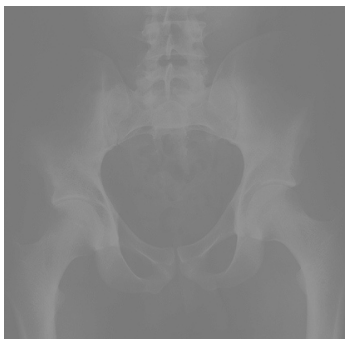


(b) Histogram of (a)

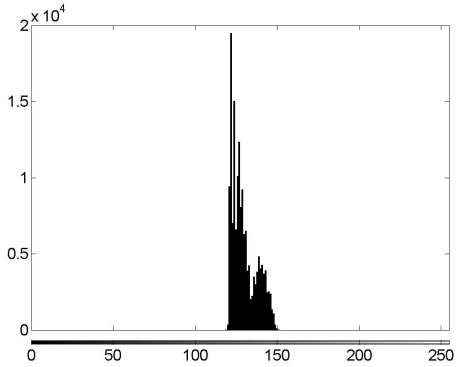
(c) Bright image



(d) Histogram of (c)

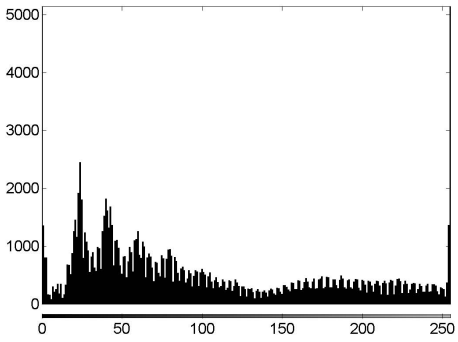


(e) Low-contrast image



(f) Histogram of (e)

(g) High-contrast image



(h) Histogram of (g)

Figure 2.3: The X-ray image at different levels of contrast, namely, dark, bright, low-contrast, and high-contrast, and the histogram corresponding to each contrast level

ing *Equation 2.8* and histogram equalization was defined as given in *Equation 2.9*:

$$s_k = T(r_k) = \sum_{j=0}^{k} p_r(r_j) \tag{2.9}$$

where T is the transformation function for histogram equalization, r_k is the k^{th} gray level, n_k is the number of pixels in the image having gray level r_k, s_k is the histogram equalized image, and $p(r_j)$ is the probability of the occurrence. By substituting *Equation 2.8* into the *Equation 2.9*, we can simplify histogram equalization as shown in *Equation 2.10* and histogram equalization applied to bright and low contrast images of *Figure 2.3* and their corresponding histograms can be seen in *Figure 2.4*.

$$\sum_{j=0}^{k} \frac{n_j}{n} \qquad \text{where } k = 0, 1, 2, \ldots, (L-1) \tag{2.10}$$

Spatial Filtering : Smoothing and Sharpening Filters

The methods and approaches that were presented in previous sections are explained as global methods; however, it is not complicated to apply these methods in local segments. For example, if transformation functions, such as Log and Power-Law transformations, or Histogram Equalization are applied in local segments which are mostly defined as square or rectangular in a whole image, they become local enhancement methods that each of the defined segments are independent from each other. *Figure 2.5* shows the segment operation on image with functions and coordinates.

In the spatial domain, the main use of the segments belongs to the filtering approaches which can be classified into two groups as smoothing filters and sharpening filters. Smoothing filters are used for blurring and for noise reduction [7]. Blurring is the removal of small details of image to provide more effective extraction of objects or other interests. Noise reduction is provided by applying some filters such as linear or non-linear. Linear filters are straight forward methods which are directly applied to the defined segments of image. They are generally replacing the center pixel of segment by the average of all pixels of segment. Because of this reason, sometimes they are called *averaging filters*, however, mostly they are know as *low-pass filters*. Typical formulae of lowpass filters can be written as shown in *Equation 2.11*.

(a) Bright image (b) Enhanced image of (a) after histogram equalization (c) Histogram of (b)

(d) Low-contrast image (e) Enhanced image of (d) after histogram equalization (f) Histogram of (e)

Figure 2.4: Implementation of histogram equalization for bright and low-contrast versions of the original X-ray image presented in *Figure 2.1(a)*

$$R = \frac{1}{m \times n} \sum_{i=1}^{m \times n} z_i \tag{2.11}$$

where R is the value to replace, m and n is segment dimensions, and z is the pixel value within segment neighborhood i.

Figure 2.6 shows the implementation of a typical low-pass filter to an x-ray image by using different segment sizes.

Non-linear filters which are generally called order statistics filters [7] in smoothing filters are based on the ranking of the pixels and replacing the center pixel with best ranking one. Most popular non-linear smoothing filter is median filter which is the best ranking was generally assumed the center pixel of sorted numbers which is 5^{th} in 3×3 segment and 13^{th} in 5×5 segment.

Figure 2.7 shows the implementation of a median filter to an x-ray image by using 3×3 segment size.

Another group of spatial domain filters is sharpening filters that are intended to enhance noisy details of images. These noise can be blurring effect or the noise which is

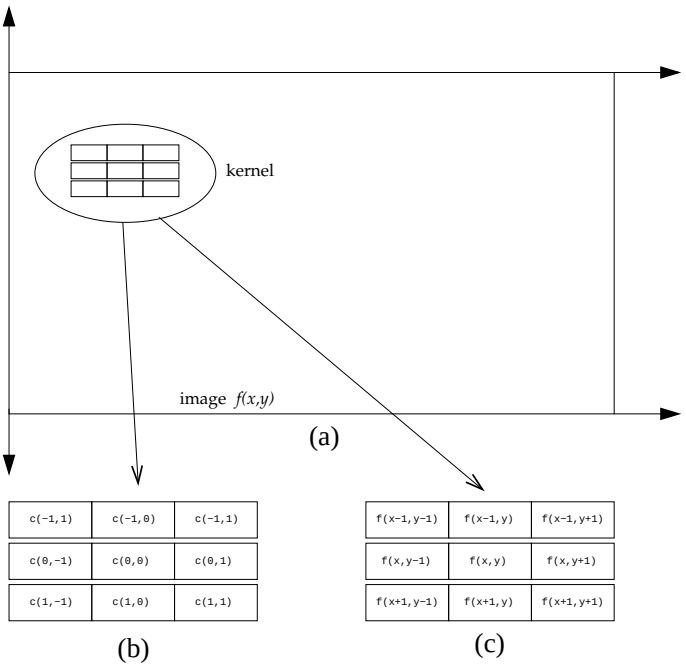


Figure 2.5: Kernel (segment) operation on image (a) 3×3 segment on image (b) represented coordinates of segment and (c) operations in segment. (original drawing courtesy of R.C Gonzalez and R.E. Woods [7]).

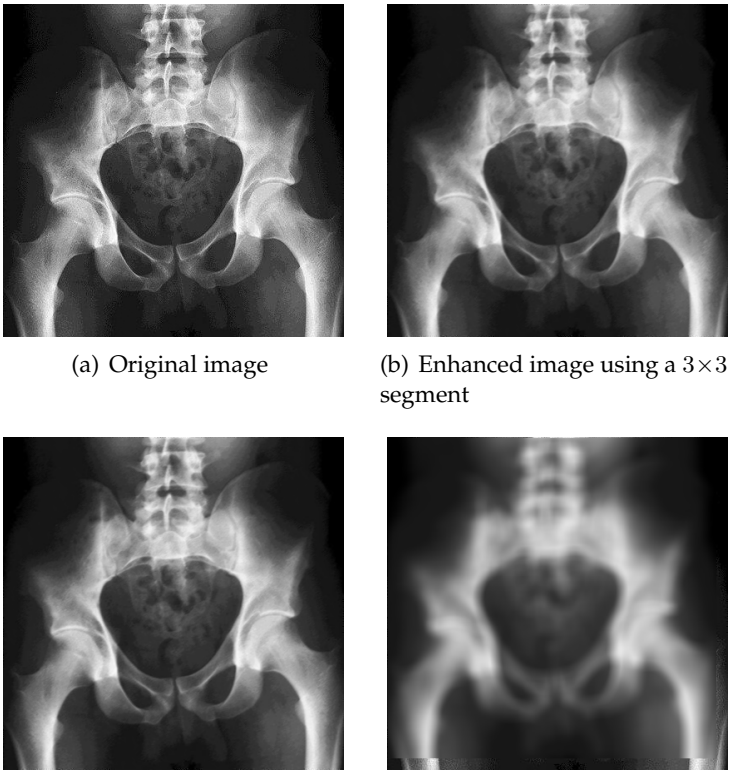
obtained during image acquisition. Sharpening filters are based on the first and second order derivatives of an image which can be formulated basically as shown in *Equation 2.12* and *Equation 2.13* respectively.

$$\nabla f = \frac{\partial f}{\partial x} + \frac{\partial f}{\mathrm{d}y} = f(x,y) - f(x,y) + f(x,y+1) - f(x,y) \tag{2.12}$$

$$\begin{aligned} \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\mathrm{d}y^2} &= f(x+1,y) - f(x-1,y) + \\ & 2f(x,y) + f(x,y+1) + f(x,y-1) + 2f(x,y) \end{aligned} \tag{2.13}$$

Implementation of second-order derivative of an image which is called Laplacian Filtering can be obtained by using a mask which is shown in *Figure 2.8*.

However, in image enhancement, the use of Laplacian Filtering has some additional features to obtain enhanced image. These additional features can be seen in *Equation 2.14* and the result of Laplacian Filtering can be seen in *Figure 2.8*.



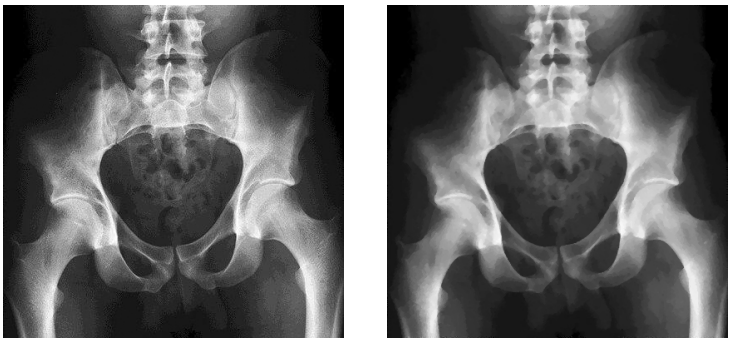
(a) Original image

(b) Enhanced image using a 3×3 segment

(c) Enhanced image using a 5×5 segment

(d) Enhanced image using a 15×15 segment

Figure 2.6: Implementation of low-pass filtering on the original X-ray image presented in (a) or *Figure 2.1(a)*



(a) Original image

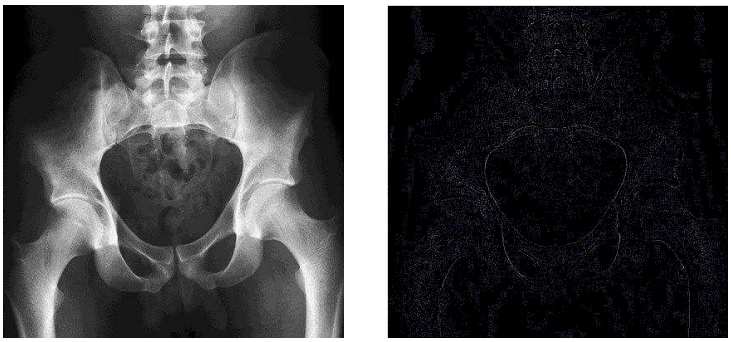
(b) Enhanced image using a 3×3 segment

Figure 2.7: Implementation of median filtering on the original X-ray image presented in (a) or *Figure 2.1(a)*

$$g(x,y) = \begin{cases} f(x,y) - \nabla^2 f(x,y) & \text{if the center coefficient of the Laplacian mask is positive} \\ f(x,y) + \nabla^2 f(x,y) & \text{if the center coefficient of the Laplacian mask is negative} \end{cases} \tag{2.14}$$

0	1	0
1	-4	1
0	1	0

Figure 2.8: Laplacian filtering mask



(a) Original image (b) Result of laplacian filtering

Figure 2.9: Laplacian filtering and enhancement of the example X-ray image

2.2.2 Overview of Frequency Domain Image Enhancement Techniques

In this section, basic definitions and the implementations of Discrete Fourier Transform (DFT) and the respected filters will be described.

In image processing, frequency domain always mentioned together with Discrete Fourier Transform (DFT) which is the discrete version of Fourier Transform (FT). The equations of single variable (one-dimensional) FT and DFT can be seen in *Equation 2.15* and *Equation 2.16* respectively.

$$F(u) = \int_{+\infty}^{-\infty} f(x)\, e^{-j2\pi ux} \qquad \text{where for } u = 0, 1, 2, 3, \ldots, M-1 \tag{2.15}$$

where $j = \sqrt{-1}$

$$F(u) = \frac{1}{M} \sum_{x=0}^{M-1} f(x)\, e^{-j2\pi ux/M} \qquad \text{for} \quad u = 0, 1, 2, 3, \ldots, M-1 \tag{2.16}$$

where $x = 0, 1, 2, 3, \ldots, M-1$.

Also, it is possible to obtain $f(x)$ by applying inverse Fourier Transformation which the continuous and discrete versions are given in *Equation 2.17* and *Equation 2.18* respectively.

$$f(x) = \int_{+\infty}^{-\infty} F(u)\, e^{-j2\pi ux} du \tag{2.17}$$

$$f(x) = \frac{1}{M} \sum_{x=0}^{M-1} F(u)\, e^{-j2\pi ux/M} \qquad \text{for} \quad x = 0, 1, 2, 3, \ldots, M-1 \tag{2.18}$$

Hence, we can express $F(u)$ in polar coordinates as shown in *Equation 2.19*.

$$F(u) = |F(u)|\, e^{-j\phi(u)} \tag{2.19}$$

where

$$|F(u)| = \left[R(u)^2 + I(u)^2 \right]^{\frac{1}{2}} \tag{2.20}$$

is called the *magnitude* or *spectrum* of the Fourier Transform and,

$$\phi(u) = \tan^{-1}\left[\frac{I(u)}{R(u)} \right] \tag{2.21}$$

is called the *phase angle* or *phase spectrum* and the *power spectrum* defined as the square of the Fourier Spectrum as shown in *Equation 2.22*.

$$P(u) = |F(u)|^2 = R(u)^2 + I(u)^2 \tag{2.22}$$

where $R(u)$ and $I(u)$ are the real and imaginary part of $F(u)$ respectively.

Also, we can express two-dimensional continuous and discrete FT and their respect-

ing inverse FT, phase angle and power spectrum as shown in Equations respectively.

$$F(u,v) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\, e^{-j2\pi(ux+vy)}\mathrm{d}x\,\mathrm{d}y \tag{2.23}$$

$$f(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(u,v)\, e^{-j2\pi(ux+vy)}\mathrm{d}u\,\mathrm{d}v \tag{2.24}$$

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\, e^{-j2\pi(ux/M+vy/N)} \tag{2.25}$$

$$f(x,y) = \frac{1}{MN}\sum_{x=0}^{M-1N-1}\sum_{y=0} F(u,v)\, e^{-j2\pi(ux/M+vy/N)} \tag{2.26}$$

$$|F(u,v)| = \left[I(u)^2 + R(u)^2\right] \tag{2.27}$$

$$\phi(u,v) = \tan^{-1}\left[\frac{I(u,v)}{R(u,v)}\right] \tag{2.28}$$

$$P(u,v) = |F(u,v)|^2 = I(u,v)^2 + R(u,v)^2 \tag{2.29}$$

Using Eulers formula as shown in *Equation 2.30*, we can express the *Equation 2.25* and *Equation 2.26* as shown in *Equation 2.31* and *Equation 2.32*.

$$e^{j\theta} = \cos\theta + j\sin\theta \tag{2.30}$$

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\left[\cos 2\pi(ux/M+vy/N) - j\sin 2\pi(ux/M+vy/N)\right] \tag{2.31}$$

$$f(x,y) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} F(u,v)\left[\cos 2\pi(ux/M+vy/N) - j\sin 2\pi(ux/M+vy/N)\right] \tag{2.32}$$

Application of the filtering process in frequency domain generally has same proce-

dure [7] which starts by the multiplication of input image by -1^{x+y} (after preprocessing if necessary) to center the transform and continues by computing $F(u, v)$ (DFT) of the image by using *Equation 2.25* or *Equation 2.31*. Any filtering function which is denoted as $H(u, v)$ can be applied at this time by the multiplication with $F(u, v)$. Then it is uncomplicated to apply inverse DFT and to obtain the real part of the results by using *Equation 2.26* or *Equation 2.32*. This is followed by the multiplication of these results by -1^{x+y} to normalize the centered transform. As a consequence, the application of any filtering function can be written as shown in *Equation 2.33*.

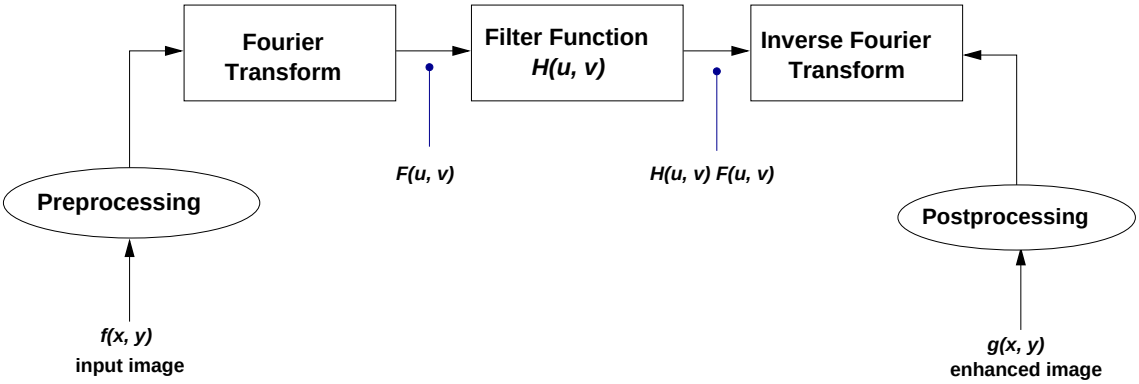


Figure 2.10: Filtering steps in the frequency domain

$$G(u, v) = H(u, v)\, F(u, v) \tag{2.33}$$

General block diagram of filtering process in frequency domain is given in *Figure 2.10*.

Similar to spatial domain filters, we can divide frequency domain filtering approaches into two groups such as smoothing and sharpening filters.

Smoothing Filters in Frequency Domain

Smoothing can be obtained by the attenuation of high frequency signals by using a specified range in the DFT of an image. As mentioned before, this attenuation can be achieved by applying filtering function which was defined in *Equation 2.33*.

Basic smoothing filters in frequency domain are Ideal Low Pass Filters (ILPF), Butterworth Low Pass Filter (BLPF) and Gaussian Low Pass Filter (GLPF).

One of the basic and simplest ILPFs is the 2D ILPF which is based on the defined distance D_0 from the centered DFT of an image. 2D ILPF cuts the higher frequency com-

ponents of image which distance $D(u,v)$ is greater than D_0. Transfer function of 2D ILPF is given in *Equation 2.34*.

$$H(u,v) = \begin{cases} 1 & \text{if } D(u,v) \leq 0 \\ 0 & \text{if } D(u,v) > 0 \end{cases} \tag{2.34}$$

Distance from any point (u,v) to the center of DFT can be expressed as:

$$D(u,v) = \left[(u - M/2)^2 + (v - N/2)^2\right]^{\frac{1}{2}} \tag{2.35}$$

Notice that, if the radius of a defined distance D_0 is relatively small, the image power will also be small and the resulting image will lose more information related to the loss of power. As a result, a more blurred image will be obtained because of the more "cutoff" of high frequency components. However, if the radius of D_0 is relatively large, power loss will be reduced and a more detailed image will be obtained. Example of 2D Ideal Low-pass Filter implementation of X-ray image with cutoff distance 10, 50 and 150 can be seen in *Figure 2.11*.

One of the most important and widely used low-pass filtering is Butterworth Low Pass Filtering (BLPF) which can be applied in *n*th order of image. Transfer function of BLPF is defined as shown in *Equation 2.36*.

$$H(u,v) = \frac{1}{1 + \left[\frac{D(u,v)}{D_0}\right]^{2n}} \tag{2.36}$$

Similar to ILPF, the effect of radius value D_0 is almost the same in BLPF. Example of Butterworth Low Pass Filter implementation of X-ray image in 2nd order with cutoff distance 10, 50 and 150 can be seen in *Figure 2.12*.

Another important Lowpass Filter in Frequency Domain is Gaussian Low Pass Filter (GLPF) which uses D_0 and $D(u,v)$ similar to other low-pass filters. The general formulae of Gaussian Low Pass Filter can be seen in *Equation 2.37*.

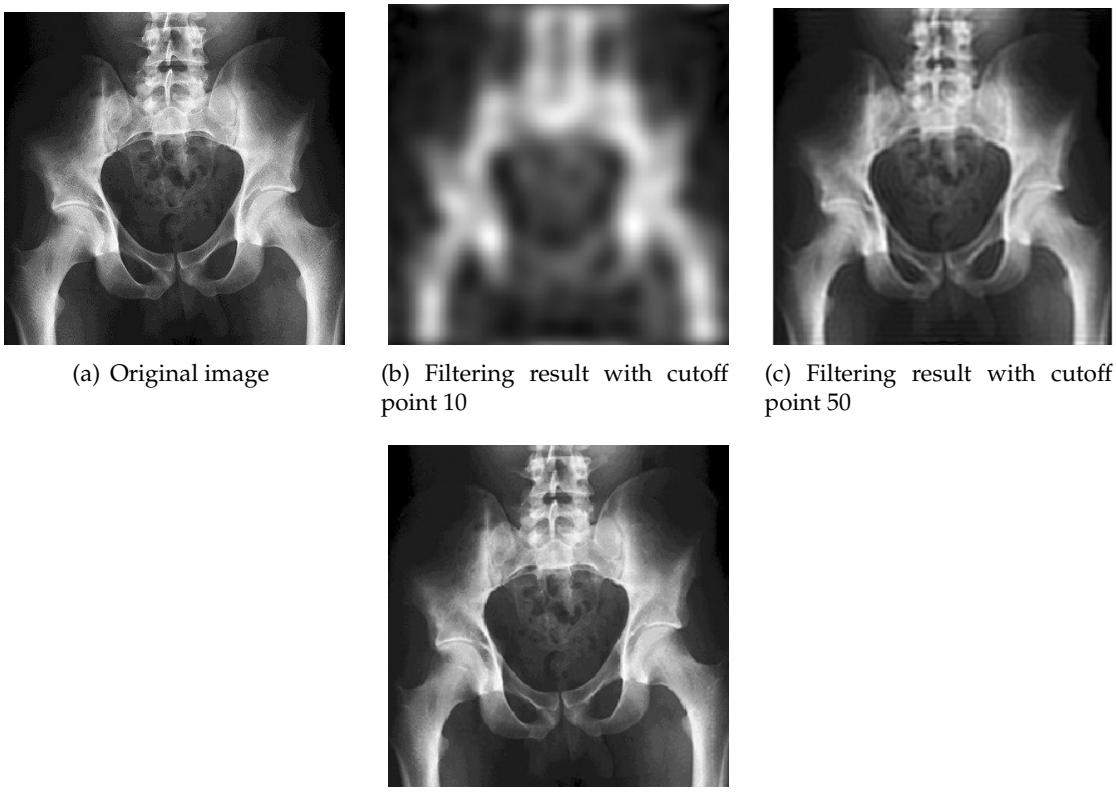
$$H(u,v) = e^{-D^2(u,v)/2\sigma^2} \tag{2.37}$$

where σ is the standard deviation of Gaussian Curve. However, it is possible to let $D_0 = \sigma$

and to express *Equation 2.37* as shown in *Equation 2.38*.

$$H(u,v) = e^{-D^2(u,v)/2D_0^2} \tag{2.38}$$

Example of Gaussian Low Pass Filter implementation of X-ray image with cutoff distance 10, 50 and 150 can be seen in *Figure 2.13*.



(a) Original image

(b) Filtering result with cutoff point 10

(c) Filtering result with cutoff point 50

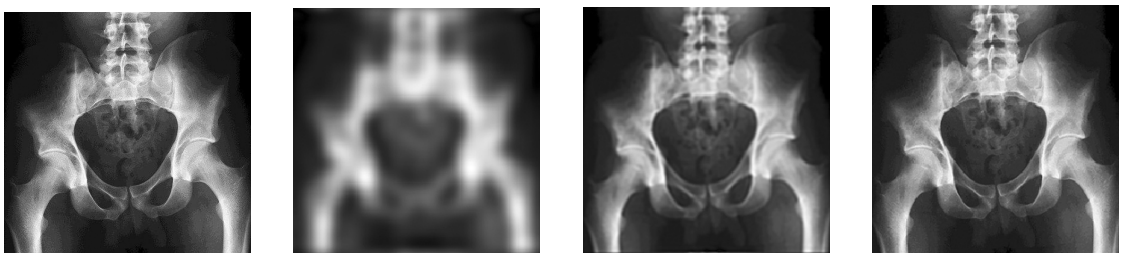
(d) Filtering result with cutoff point 150

Figure 2.11: 2D ILPF implementation of the original X-ray image. Note that the blurring effect in (b) with small size of cutoff point D_0.

Sharpening Filters in Frequency Domain

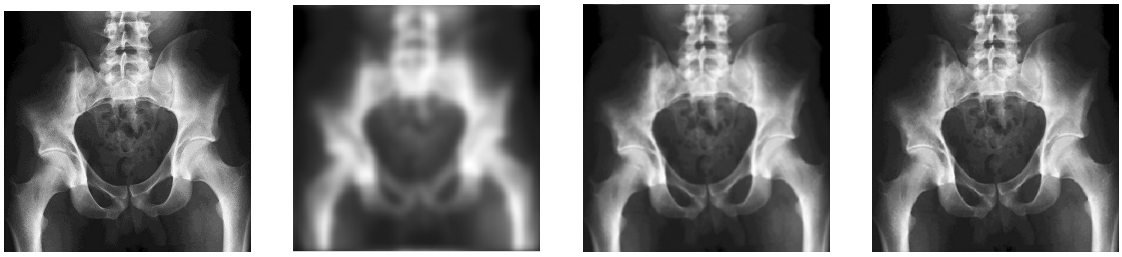
In the frequency domain, sharpening can be achieved using high-pass filters that attenuate the low frequency components without disturbing high frequency components [7]. Generally, high pass filtering is the reverse operation of low pass filtering and basically they can be described as given in *Equation 2.40*.

$$H_{HP}(u,v) = 1 - H_{LP}(u,v) \tag{2.39}$$



(a) Original image (b) Filtering result with cutoff point 10 (c) Filtering result with cutoff point 50 (d) Filtering result with cutoff point 150

Figure 2.12: The results of Butterworth low-pass filtering of the original X-ray image



(a) Original image (b) Filtering result with cutoff point 10 (c) Filtering result with cutoff point 50 (d) Filtering result with cutoff point 150

Figure 2.13: The results of Gaussian low-pass filtering of the original X-ray image

where H_{LP} the low-pass filtering transfer function.

Thus Ideal High Pass Filter, Butterworth High Pass Filter and Gaussian High Pass Filter can be expressed by using *Equation 2.39* as shown in *Equation 2.40*, *Equation 2.41* and *Equation 2.42* respectively.

$$H(u,v) = \begin{cases} 1 & \text{if } D(u,v) \leq 0 \\ 0 & \text{if } D(u,v) > 0 \end{cases} \tag{2.40}$$

$$H(u,v) = \frac{1}{1 + \left[\frac{D_0}{D(u,v)}\right]^{2n}} \tag{2.41}$$

$$H(u,v) = e^{-D^2(u,v)/2D_0^2} \tag{2.42}$$

Example of Ideal High Pass Filtering, Butterworth High Pass Filtering and Gaussian High Pass Filter implementation of X-ray image with cutoff distance 1,10 and 20 can be seen in *Figure 2.14*, *Figure 2.15* and *Figure 2.16* respectively.

(a) Original image (b) Filtering result with cutoff point 1 (c) Filtering result with cutoff point 10 (d) Filtering result with cutoff point 20

Figure 2.14: The results of ideal high-pass filtering of the original X-ray image

(a) Original image (b) Filtering result with cutoff point 1 (c) Filtering result with cutoff point 10 (d) Filtering result with cutoff point 20

Figure 2.15: The results of Butterworth high-pass filtering of the original X-ray image

(a) Original image (b) Filtering result with cutoff point 1 (c) Filtering result with cutoff point 10 (d) Filtering result with cutoff point 20

Figure 2.16: The results of Gaussian high-pass filtering of the original X-ray image

2.2.3 Main Application Areas of Image Enhancement

The use of image enhancement has increasing popularity in the fields that require increased visual appearance of images or objects. Most important application areas of image enhancement are medical imaging, military-security-forensic sciences, document analysis, and pattern preprocessing.

Enhancement in Medical Imaging

Medical Imaging consists of several areas where enhancement of images are required. Widely used medical imaging techniques are Digital X-Ray, Digital Mammography [8, 9, 10], CT Scans [11, 12], and MRI [13]. The aim of image enhancement in medical imaging is to improve visual appearance of images to provide faster diagnosis of diseases. For example, in an X-ray image, it is important to enhance images to see if there is any broken bones in the patient and in mammography, it is important to show all cells clearly to see if there are any cancer cells or tumors.

In the enhancement of medical images, either existing spatial domain approaches or frequency domain approaches can be used or new techniques can be developed based on these domains. For example, J.K.Kim et al. [8] developed a technique by using first derivatives and local statistics of images which belong to spatial domain approaches to improve the appearance of mammographic images, and a technique that was based on the Fast Fourier Transform (FFT) was presented by E.W. Abel et al. [14] to increase the visual appearance of cancerous bones of x-ray images.

Enhancement in Military, Security and Forensic Sciences

In military, security and forensic sciences, main application areas of image enhancement are the improvement of night-vision images [15], fingerprint images [16, 17], face components [18], and satellite images [19].

In night vision and satellite images, it is important to increase the visuality of each component of dark or noisy image, however in fingerprint and face images, it is more important to clear unnecessary data to extract features from the images.

Similar to all enhancement applications, any spatial or frequency domain approaches can be efficient to increase the visual appearance of images, however, it is not guaranteed

that a method should produce superior results for all night-vision, fingerprint, face or satellite images.

Enhancement in Document Analysis

In document analysis, the aims of the image enhancement methods can be listed as the extraction of the characters by providing effective reduction of the noise and the additional layers within the document images and to provide more clean document images for human readers or optical character recognition (OCR) modules.

Therefore, both aims of document analysis require different enhancement methods to achieve readable and separable documents. For example, the improvement of readability of the documents can be useful for fax documents to eliminate added noises which are obtained during the transmission [20], however separation can be useful for digitizing documents [21].

2.3 Summary

The visual appearance of images can be increased using several enhancement methods that belong to either spatial or frequency domain. In the spatial domain, methods are applied directly to the image. However, in the frequency domain, the methods or filters can be applied after obtaining the Discrete Fourier Transform of image.

For both domains, output images can be different or the same according to the applied techniques, applications and the characteristics of the images. So, it is almost impossible to determine which domain's techniques produce most successful results.

In the next chapter, image binarization, that is a low level image processing technique, will be described in details. In addition, benchmark and recently proposed twelve thresholding methods will be explained.

CHAPTER 3

IMAGE BINARIZATION METHODS

3.1 Overview

Image binarization (thresholding) is a low-level image processing method to separate and to enhance the region of interest to provide increased visual appearance of image. This enhancement and separation is provided by dividing image into two regions as background $(logical\ 1)$ and foreground $(logical\ 0)$. Ideally, separated image of foreground is expected to have a region of interest or object in image with a minimum loss of information and fuzziness. Consequently, it should not consist of any pixels belonging to the background and several techniques are developed to achieve this aim. In this chapter, basic definitions of image binarization, chronological development, detailed explanation about selected twelve methods and application areas will be presented.

3.2 Fundamentals of Image Binarization

Image Binarization is one of the basic spatial domain image processing techniques that is used to segment or enhance the region of interest within an image. It is based on the assumption that object and background can be distinguished by their gray level values [22] and the result of this assumption is the cause for the development of several thresholding methods which use various properties of images. General image binarization function can be expressed as given in *Equation 3.1*.

$$|g(x,y)| = T\Big[f(x,y)\Big] \tag{3.1}$$

where $f(x, y)$ is the input image, $g(x, y)$ is the processed image, and T is an operator on f, defined over some neighborhood of (x, y).

However, the main difference between the other spatial domain techniques which were described in Chapter 2, and image binarization, is the output image. In binarization, the output image consists only 0 $(binary\ 0)$ and 255 $(binary\ 1)$. Thus characteristic formulae of image binarization with threshold point Θ can be defined as shown in *Equation 3.2*.

$$g(x,y) = \begin{cases} 0 & \textit{if} \quad g(x,y) \leq T(f[x.y]) = \Theta \\ 255 & \textit{otherwise} \end{cases} \tag{3.2}$$

General properties of binarization methods are mostly common for all methods, especially for global ones. Gray level image histogram $h(g)$, probability density function (PDF) and its corresponding standard deviation (σ), mean (μ), priori probability $(p(T))$ and image entropy $(H(T))$ should be understood before implementing and analyzing any method.

Gray level image histogram which was defined in *Equation 2.7* is the distribution of the number of pixels that have same gray level value and was defined as follows [7]:

$$h(g) = n_g \tag{3.3}$$

where g is the gray level and n_g is the number of pixels in the image having gray level g. In image processing and binarization, probability density function is used to normalize the gray level histogram of images and it was defined as below:

$$PDF = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.4}$$

where σ and μ are the variance and the mean of the image and are given in *Equation 3.5* and *Equation 3.6* respectively:

$$\sigma^2(T) = \left[\sum_{g=a}^{b} (g - \mu(T))^2\ p_a(g)\right] \tag{3.5}$$

where g is the gray level, μ is the mean, $h(g)$ is the gray level histogram, $p_a(g)$ is the gray level distribution and a and b are the lowest and highest gray level value of the

distribution.

$$\mu(T) = \frac{\left[\sum_{g=a}^{b} h(g)g\right]}{P(T)} \tag{3.6}$$

Gray-level distribution is defined as follows:

$$p(T) = \frac{\sum_{g=a}^{b} h(g)}{N \times M} \tag{3.7}$$

where $h(g)$ is the gray level histogram, a and b are the lowest and highest gray level values of the distribution and N and M are the x and y dimension of the image or segment. A priori probability $P(T)$ was defined as follows:

$$P(T) = \sum_{g=a}^{b} p(g) \tag{3.8}$$

Image entropy is an other way to perform binarization methods. Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image and is defined as shown in *Equation 3.9*:

$$H(T) = \sum_{g=0}^{T} p(g) \log p(g) \tag{3.9}$$

In order to provide an efficient separation and enhancement of the region of interest within an image, several thresholding methods which can be classified into two groups such as global binarization methods and local binarization methods, were proposed.

Global thresholding methods consider the whole image and its global characteristics to determine a single threshold value, and the local thresholding methods divide the image into segments to determine individual threshold values for each segment. However, both groups carry out some disadvantages beside their advantages. Global methods have generally faster execution time and less noise in the resultant image than local methods, however, according to the characteristics of document images, for example, they can be over or under thresholded that cause some loss of relevant information. Local methods generally produce resultant images with less loss of relevant information than global

methods; however, the segment size, which is the main disadvantage of the local methods, brings some additional noise to these images in small sizes and they behave as global methods and can be over-thresholded in large sizes.

In literature, one of the first proposed thresholding methods is Riddler and Calvard [23] method which is based on the change of the foreground and background class means at iteration n. This method was followed by the Otsu [24] method which became one of the most popular global methods and uses variances within the image to determine the final threshold point (see *Section 3.3.1*). Nakagawa and Rosenfeld [25] proposed one of the first local thresholding methods which is known as Nakagawa and Rosenfeld implementation of Chow and Kaneko [26]. Then Pun [27], proposed the use of image entropy in threshold selection and at that time Yasuda et al. [28] proposed another local thresholding method.

White and Rohrer [29] proposed local thresholding which compares the gray level pixel values to the average of the gray level values in some neighborhood and if the pixel is significantly darker than the average, it is denoted as foreground; otherwise, it is classified as background. Rosenfeld et al. [30] proposed a histogram-based global thresholding method that is based on analyzing the concavities of the histogram $h(g)$ visand its convex hull. Kapur et al. [31] proposed an entropy based thresholding method that later become one of the most famous entropy-based methods (see *Section 3.3.5*). At that time, Lloyd [32] proposed another global method that divides the image histogram into two clusters and minimizes misclassification error between these clusters.

Then Kittler and Illingworth [22] proposed their Minimum Error Thresholding technique (see *Section 3.3.2*) which is based on clustering of image histogram similar to Lloyd method. Also, Niblack [33] and Bernsen [34] independently proposed their local thresholding methods, which are still the most popular and mostly compared and cited methods (see *Section 3.4.1* and *Section 3.4.5*). Palumbo et al. [35] proposed another local threshold method which consists in measuring the local contrast of five neighborhoods. Abutaleb [36] proposed a global thresholding method which was based on two-dimensional entropy of the image and Yanowitz and Bruckstein [37] proposed a local thresholding method that uses the discrete Laplacian of the surface, produced by using the combination of edge and gray level information.

Taxt et al. [38] proposed a local thresholding method for document image segmentation. Eikvil et al. [39] proposed a local thresholding method that is based on image clustering of a small window in a larger concentric window. At that time, Parker [40] proposed another local thresholding method that first detects the edges and the interior of objects is filled.

Li and Lee [41] proposed another entropy based method that minimizes the theoretic distance of information. Kamel and Zhao [42] proposed another local thresholding method that measures the difference of local mean and the local pixel and compare it with a predetermined value to determine the threshold point for each segment.

Yanni and Horne [43] proposed global thresholding method which uses the midpoint of the two assumed peaks of the gray level histogram of an image to determine the final threshold (see *Section 3.3.3*). Ramesh et al. [44] proposed global thresholding that uses a simple functional approximation to minimize the image histogram (see *Section 3.3.4*).

Then, Yen et al. [45], Pal [46] and Sahoo et al. [47] proposed another entropy based thresholding methods and recently Albuquerque et al. [48] proposed another entropy based method that uses Tsallis entropy (see *Section 3.3.6*). Oh and Lindquist [49] proposed a local method and this method was followed by the Sauvola et al. [50] method which recently became popular while improving the Niblack method (see *Section 3.4.2*). Solihin and Leedham [51] proposed a global thresholding method which is based on the integral ratio. Yibing and Yang [52] improved the Kamel and Zhao logical thresholding technique (see *Section 3.4.4*) to determine the required parameters automatically. Wold and Jolion [53] improved the Sauvola method to normalize contrast and the local mean of the image to decrease the amount of noise.

Leedham et al. [3] proposed the Mean-Gradient technique which is based on the local mean and the local mean gradient of an image (see *Section 3.4.3*) and at that time, Badekas and Papamarkos [54] improved the adaptive logical thresholding of Yibing and Yang. Sezgin and Sankur [55] proposed a global thresholding method that is based on sample moment function.

Recently, Park et al. [56] proposed a new method that uses 3D terrain of a grayscale image and simulates waterfall to binarize images (see *Section 3.4.6*), and Kavallieratou [57][58] proposed iterative global thresholding that calculates the difference of the mean

value and the current pixels and uses histogram equalization in each iteration to clean and binarize images. Leedham and Chen [59] proposed decompose algorithm which requires several processing steps that includes mean gradient method of Leedham et al. *Table 3.1* and *Table 3.2* shows chronological order of benchmark and recently proposed global and local thresholding methods respectively.

3.3 Global Binarization Methods

Global thresholding methods use a defined or a computed threshold value for the entire image and several techniques that intend to achieve appropriate thresholding point were proposed.

In the next subsections, benchmark and recently proposed six global methods will be described and in *Section 3.3.7* advantages and disadvantages of global binarization methods will be discussed.

Table 3.1: Chronological order of basic and recently proposed *global* thresholding methods

No	**Author**	**Features**
1	[23]	Iterative clustering
2	[24]	Class separability
3	[27]	Maximum Shannon's entropy
4	[30]	Histogram concavities and convex hull
5	[31]	Entropy
6	[32]	Clustering and minimizing error
7	[22]	Minimum error between clusters
8	[36]	High order entropy
9	[41]	Entropy and theoretic distance
10	[43]	Clustering and peak values
11	[44]	Functional approximation
12	[45]	Entropic correlation
13	[60]	Noise Attribute
14	[46]	Maximum entropy
15	[47]	Renyi entropy
16	[51]	Integral ratio
17	[55]	Sample Moment Function
18	[48]	Tsallis entropy
19	[57]	Iterative histogram equalization

These methods are: the Otsu Method [24], Kittler and Illingworth Minimum Error Technique [22], Yanni and Horne method [43], Ramesh et al. method [44], Kapur et al. Entropy Method [31] and Albuquerque et al. Entropy Method [48].

Table 3.2: Chronological order of basic and recently proposed *local* thresholding methods

No	**Author**	**Features**
1	[25]	Variable thresholding
2	[28]	Local intensity change
3	[29]	Based on local mean and neighbors
4	[33]	Local mean and deviation
5	[34]	Local based on neighbors
6	[35]	Local contrast
7	[37]	Threshold surface
8	[38]	Mixture of two Gaussian distribution
9	[39]	The pixels inside a small window are thresholded on the basis of clustering in larger window
10	[42]	Local contrast and logical level
11	[49]	Two-pass algorithm
12	[50]	Improvement of Niblack
13	[52]	Adaptive logical level
14	[53]	Improvement of Sauvola et. al.
15	[54]	Improvement of adaptive logical level
16	[3]	Local mean and gradient
17	[56]	Rainfall simulation
18	[59]	Decompose algorithm

3.3.1 Otsu Method

Otsu method [24] was proposed in 1979 as a selection method which was based on the image histogram. It uses discriminant analysis to divide the foreground and background by maximizing the discriminant measure. According to Ng and Lee [61], the threshold operation is regarded as the partitioning of the pixels of an image into two classes C_0 and C_1 (e.g., objects and background) at gray level t, i.e., $C_0 = 0, 1, \ldots t$ and $C_1 = t+1, t+2, \ldots l-1$. An optimal threshold point can be determined by minimizing one of the following equations using within-class variance, between-class variance, and the total variance, σ_b^2, σ_w^2, σ_T^2 respectively.

The operations of the Otsu method can be seen in *Figure 3.1*.

$$\begin{aligned} \lambda &= (\sigma_b^2/\sigma_w^2), \\ \eta &= (\sigma_b^2/\sigma_T^2), \\ k &= (\sigma_T^2/\sigma_w^2) \end{aligned} \tag{3.10}$$

Therefore, the optimal threshold value can be found using only the term:

$$\sigma_B^2(k) \cdot \sigma_B^2(k) = \frac{[\mu_T \omega(k) - \mu(k)]^2}{\omega(k)[1-\omega(k)]} \tag{3.11}$$

$$k^* = ArgMin(\eta) \tag{3.12}$$

3.3.2 Kittler and Illingworth method

The Kittler and Illingworth method [22], which is based on clustering the image, starts by choosing an arbitrary initial threshold T and compares both sides of T to determine error. Then, T is shifted and determined errors are compared to find a minimum error point which is assigned as a threshold point. The simplest formulae can be written as:

$$J(\tau) = \min_{T} J(T) \tag{3.13}$$

where $J(\tau)$ is the minimum error threshold and $J(T)$ is the criterion function. $J(T)$ can be written directly as:

$$\begin{aligned} J(T) = \quad & 1 + 2[P_1(T)\log\sigma_1(T) + P_2(T)\log\sigma_2(T)] \\ & -2[P_1(T)\log P_1(T) + P_2(T)\log P_2(T)] \end{aligned} \tag{3.14}$$

where P_1 and P_2 denote the priori probability and σ_1 and σ_2 denote standard deviations of left and right sides of T respectively. Operations of Kittler and Illingworth method can be seen in *Figure 3.2*.

3.3.3 Yanni and Horne Method

Yanni and Horne method [43] initializes the midpoint of two peaks of image histogram which is defined as:

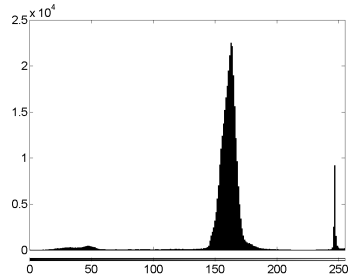
$$g_{mid} = \frac{(g_{max} + g_{min})}{2} \tag{3.15}$$

where g_{mid} is the midpoint of assumed peaks of image histogram and g_{max} and g_{min} are highest and lowest gray-levels respectively. The midpoint is updated using the mean of the two peaks on the right and left sides of the initial midpoint which can be written as:

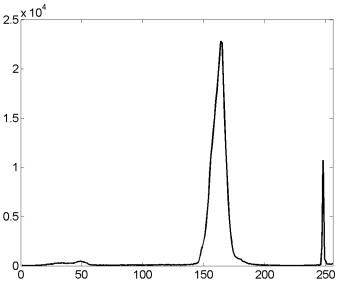
$$g^*_{mid} = \frac{(g_{peak1} + g_{peak2})}{2} \tag{3.16}$$

where g^*_{mid} is updated midpoint and g_{peak1} and g_{peak2} are the mean values of left and right

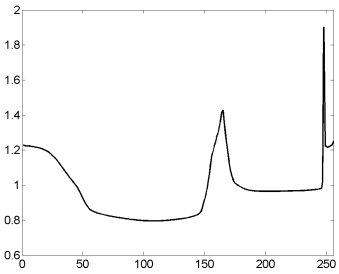
(a) Original image



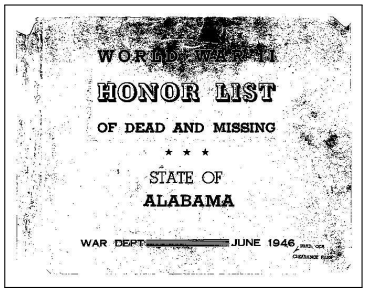
(b) Gray-level histogram of the original image



(c) Gaussian distribution of the gray level histogram of the original image

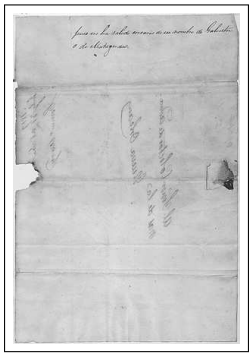


(d) Minimum arguments at $T = 154$

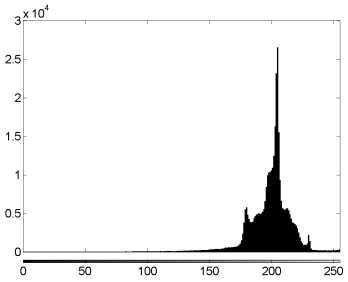


(e) Binarized image

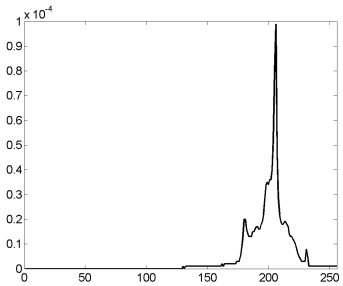
Figure 3.1: Otsu thresholding operations



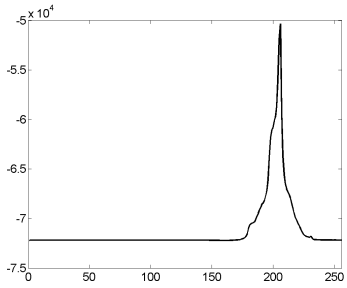
(a) Original image



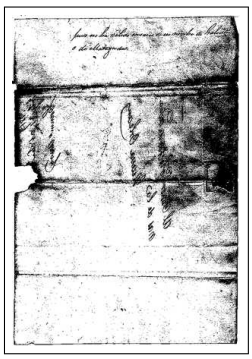
(b) Gray-level histogram of the original image



(c) Gaussian distribution of the histogram of the original image



(d) Error graph $J(T)$ with minimum error point $T = 195$



(e) Binarized image

Figure 3.2: Kittler and Illingworth thresholding operations

sides of initial midpoint respectively. Finally, threshold point is calculated as shown in *Equation 3.17*:

$$T_{top} = (g_{max} - g_{min}) \sum_{g=g_{min}}^{g^*_{mid}} \tag{3.17}$$

3.3.4 Ramesh et al. Method

Ramesh et al. method [44] is based on the approximation of the distributed gray level histogram of an image and it divides this distributed histogram into two parts T_0 and T_1, and finds the minimum argument of the summation of these parts, which is defined as:

$$T_{top} = ArgMin(T_0 + T_1) \tag{3.18}$$

where T_0 and T_1 are the left and right sides of histogram and can be defined as :

$$T_0 = \sum_{g=0}^{T} \left(\frac{\mu_0(T)}{P(T)} - g \right)^2 \tag{3.19}$$

$$T_1 = \sum_{g=T+1}^{L-1} \left(\left(\frac{\mu_1(T)}{1 - P(T)} \right) - g \right)^2 \tag{3.20}$$

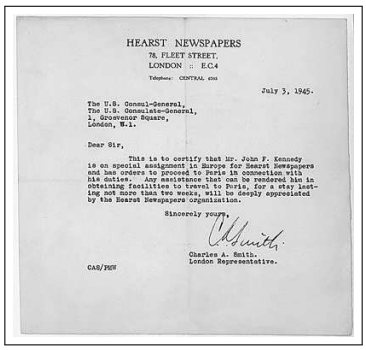
Operations of Ramesh method can be seen in *Figure 3.3*.

3.3.5 Kapur et al. Entropy Method

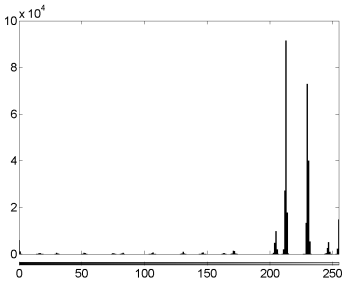
Kapur et al. method [31] divides an image into two classes such as background and foreground, and assumes these classes have different signal source. Maximum summation of these two classes entropies is considered as an exact threshold value, which is defined as:

$$T_{opt} = ArgMax\lfloor H_f(T) + H_b(T) \rfloor \tag{3.21}$$

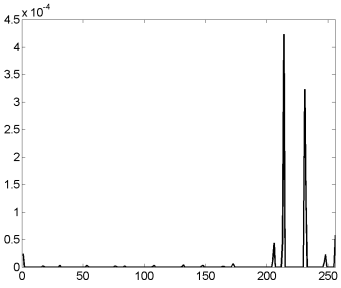
where $H_f(T)$ and $H_b(T)$ are the foreground and background entropies of image and are



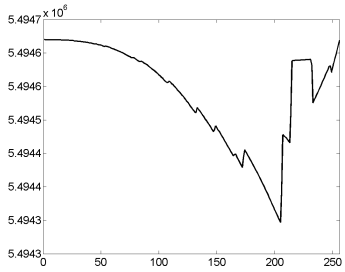
(a) Original image



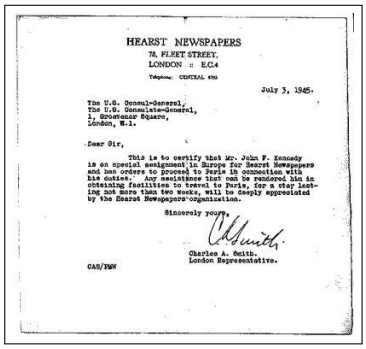
(b) Gray-level histogram of the original image



(c) Gaussian distribution of the histogram of the original image



(d) Argument graph with minimum summation point $T = 204$



(e) Binarized image

Figure 3.3: Ramesh et al. thresholding operations [44]

defined as:

$$H_f(T) = -\sum_{g=0}^{T} \left(\frac{p(g)}{P(T)} \cdot \frac{\log p(g)}{P(T)} \right) \tag{3.22}$$

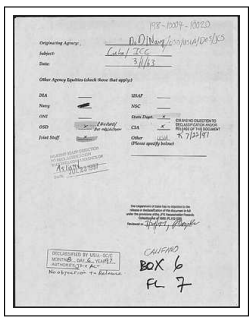
$$H_b(T) = -\sum_{g=T+1}^{G} \left(\frac{p(g)}{P(T)} \cdot \frac{\log p(g)}{P(T)} \right) \tag{3.23}$$

where $p(g)$ and $P(T)$ are probability mass function and area probability, respectively. Operations of the Kapur et al. method can be seen in *Figure 3.4*.

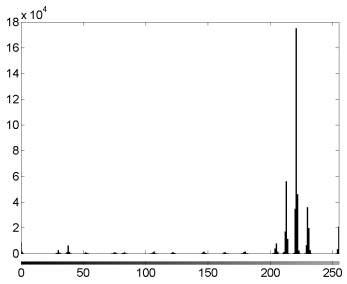
3.3.6 Albuquerque et al. Entropy Method

Albuquerque et al. Tsallis entropy thresholding [48] is based on Kapur et al. entropy method however, it uses Tsallis entropy form due to the presence of non-additive information in some classes of images.

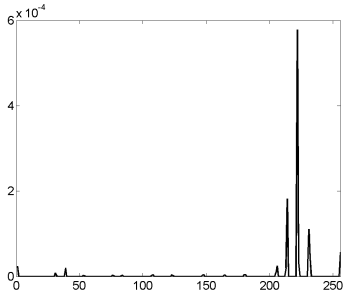
Similar to the Kapur et al. method, image is divided into two classes such as background and foreground, and maximum argument of calculated T is selected as the exact



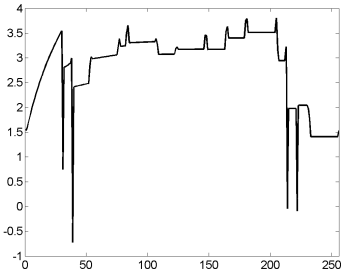
(a) Original image



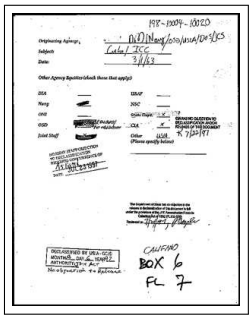
(b) Gray-level histogram of the original image



(c) Gaussian distribution of the histogram of the original image



(d) Summation graph of two classes with maximum argument point $T = 204$



(e) Binarized image

Figure 3.4: Kapur et al. thresholding operations [31]

threshold value. General formulae can be seen in *Equation 3.24*.

$$T_{opt} = ArgMax\left(S_q^A(t) + S_q^B(t) + (1-q)\cdot S_q^A(t)\cdot S_q^B(t)\right) \tag{3.24}$$

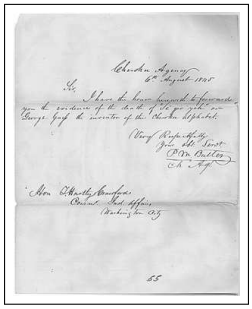
where q is an entropic index that characterizes the degree of non-extensivity, S_q^A and S_q^B are Tsallis entropy of image foreground and background which were defined as shown in *Equation 3.25* and *Equation 3.26*.

$$S_q^A(t) = 1 - \frac{\sum_{i=1}^{t}\left(\frac{p_i}{p^A}\right)^q}{(q-1)} \tag{3.25}$$

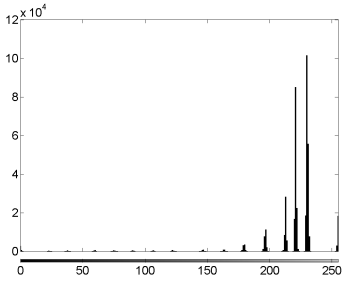
$$S_q^B(t) = 1 - \frac{\sum_{i=1}^{t}\left(\frac{p_i}{p^B}\right)^q}{(q-1)} \tag{3.26}$$

where p_i, p^A and p^B are probability distribution level, and probability distribution of foreground and background respectively. Operations of Albuquerque et al. method can

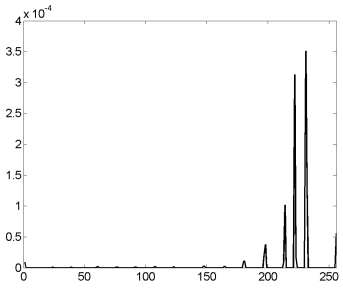
be seen in *Figure 3.5*.



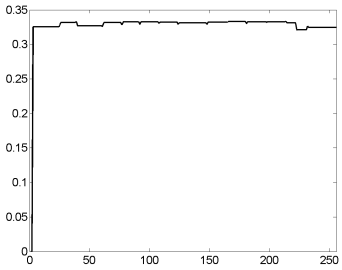
(a) Original image



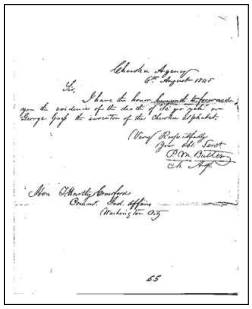
(b) Gray-level histogram of the original image



(c) Gaussian distribution of the histogram of the original image



(d) Summation graph of two classes with maximum argument point $T = 179$



(e) Binarized image

Figure 3.5: Albuquerque et al. thresholding operations [48]

These described global methods are included in the several comparisons that will be presented in *Chapter 4* and *Chapter 5* because of their popularity in document binarization. Almost every research in document binarization comprises the comparison of at least three of these methods. A recently proposed method Albuquerque et al. entropy method was proposed as the superior in entropy based methods, hence it was also included to these six methods.

3.3.7 Advantages and Disadvantages of Global Binarization Methods

Global binarization methods have some disadvantages besides their apparent advantages of binarizing images with various degrees of success depending on the type and the characteristics of the images.

The main advantages of global methods can be listed as faster execution time and less noise in resultant images. However, depending on the characteristics of the images, global methods can over or under threshold which causes some loss of relevant informa-

tion.

In less degraded images which generally comprise of uniform and low-level of noise, global methods produce more efficient binarization than local methods. However, non-uniform backgrounds, high-level of noise, and other irrelevant layers such as smears, shadows etc. , may cause some global methods to produce complete loss of information or similar to local methods, the detection of these layers and noise as objects. In *Figure 3.6*, correctly and incorrectly thresholded images by global methods are demonstrated to show the effects of irrelevant layers in binarization process.

3.4 Local Binarization Methods

Local thresholding methods use different threshold values for segments within the image and several methods had been proposed to determine these local values. The benchmark and recently proposed local methods are Niblack Method [33], Sauvola et al. Method [50], Mean-Gradient Method of Leedham et al. [3], Adaptive Logical Method [52], Bernsen Method [34], and Water Flow Model [56]. In next subsections, these methods will be explained in details and in *Section 3.4.7* , general advantages and disadvantages of local binarization methods will be listed.

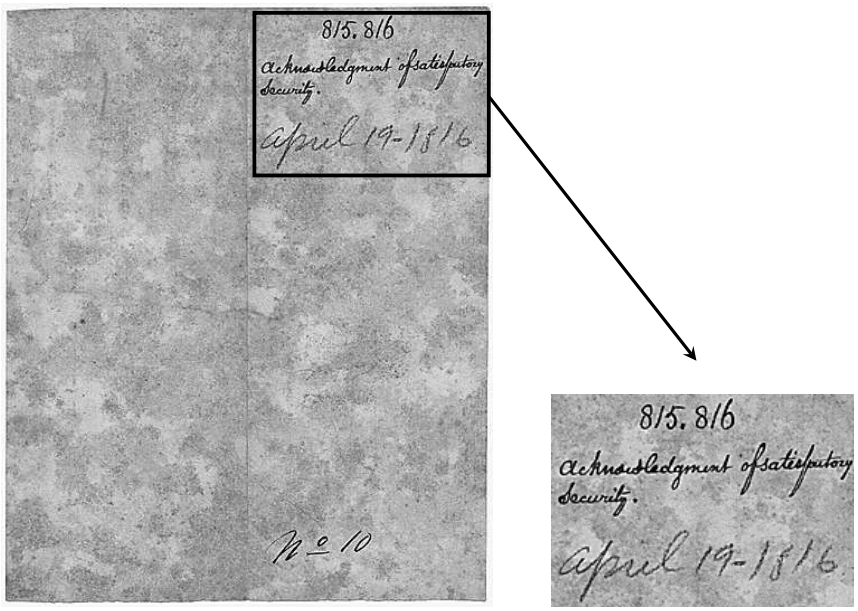
3.4.1 Niblack Thresholding Method

Niblack method [33] is based on varying the threshold value over the image, based on the local mean and local standard deviation. The threshold at pixel (x, y) is calculated as:

$$T(x, y) = \mu(x, y) + k \cdot s(x, y) \tag{3.27}$$

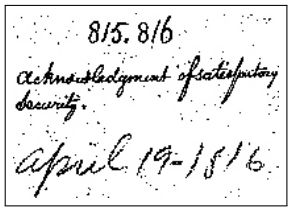
where $\mu(x, y)$ and $s(x, y)$ are the mean and standard deviation values, respectively, in a local neighborhood of (x, y) and k is the value to adjust object boundary.

Essentially, the Niblack Method uses the local mean and deviation that provides an approximation of the mean level by the amount of local deviation. However, increment or decrement of this mean value depends on a constant k, where if $k > 0$, mean value will be approximated to upper boundary, if $k < 0$, mean value will be approximated to lower boundary and if $k = 0$, mean value will become the threshold point. Operations of



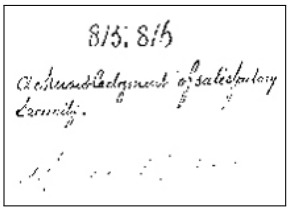
(a) Original image where the processed portion is indicated with a black box

(b) Portion of the original image processed (enlarged)

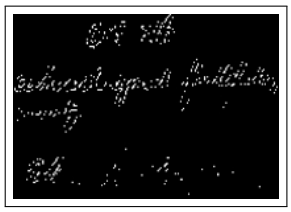


(c) Binarized image using the Kittler and Illingworth method

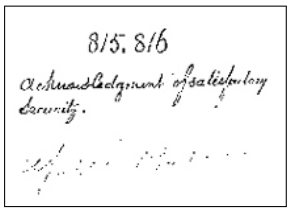
(d) Binarized image using the Otsu method



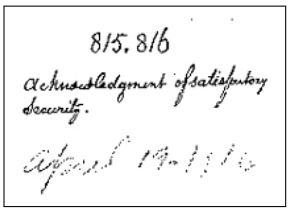
(e) Binarized image using the Yanni and Horne method



(f) Binarized image using the Ramesh et al. method



(g) Binarized image using the Albuquerque et al. method



(h) Binarized image using the Kapur al. method

Figure 3.6: The effects of irrelevant layers on global methods. The Kittler and Illingworth method produces some noise in the case of clear characters. The Otsu method detects all irrelevant data as objects. The Yanni and Horne method loses some information. The Ramesh et al method detects almost all pixels as objects. The Albuquerque et al. method's performance is similar to that of the Yanni and Horne method. That is, there is some loss of information. Finally, in the Kapur et al. method, there is little loss of information.

Niblack Method can be seen in *Figure 3.7*.

3.4.2 Sauvola et al. Thresholding Method

The Sauvola et al. method [50] is an improvement of Niblack thresholding method aimed at producing better results with degraded documents. It can be described as:

$$T(i,j) = \mu(i,j) + \left(1 + k \cdot \left[1 - \frac{\sigma(i,j)}{R}\right]\right) \tag{3.28}$$

where $\mu(i,j)$ and $\sigma(i,j)$ is the mean and variance value respectively in a local neighborhood of (i,j) with k and $R = 128$.

Sauvola et al. improved Niblack method to add more adaptive local deviation to the mean value. However, the effect of local deviation is mostly eliminated by R. Operations of Sauvola et al. method can be seen in *Figure 3.8*.

3.4.3 Mean-Gradient Thresholding Method

The mean-gradient method [3] was proposed by Leedham et al. and is the improved variant of Niblack's method and is based on the local mean and local mean-gradient values. The gradient and the mean gradient of intensity images were defined as shown in *Equation 3.29* and *Equation 3.30* respectively.

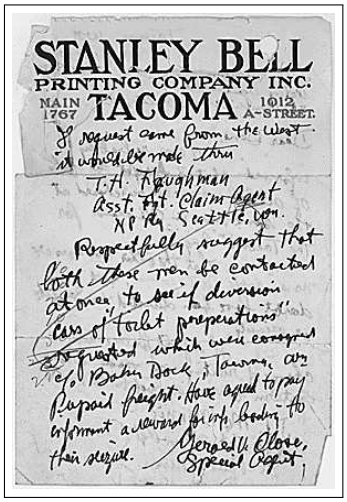
$$\nabla I(x,y) = \left[\frac{\partial I(x,y)}{\partial x}, \frac{\partial I(x,y)}{\partial y}\right] \tag{3.29}$$

$$G = \frac{\left(\sum_{x=0}^{i-1}\sum_{y=0}^{j-1}\left[\frac{\partial I(x,y)}{\partial x}, \frac{\partial I(x,y)}{\partial y}\right]\right)}{x \cdot y} \tag{3.30}$$

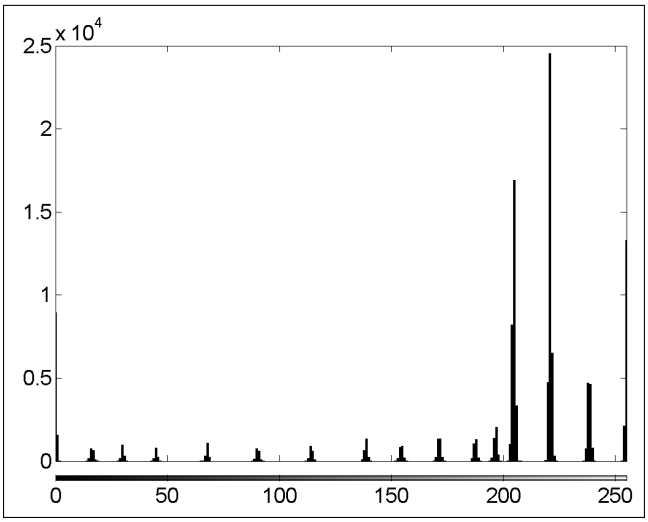
Then, the pre-condition was added to improve threshold selection:

$$\begin{aligned} &\textit{if } (c \leq R) \quad T(x,y) = \mu(x,y) + k \cdot G(x,y), \\ &\textit{else} \qquad\qquad T(x,y) = 0.5\,\mu(x,y) \end{aligned} \tag{3.31}$$

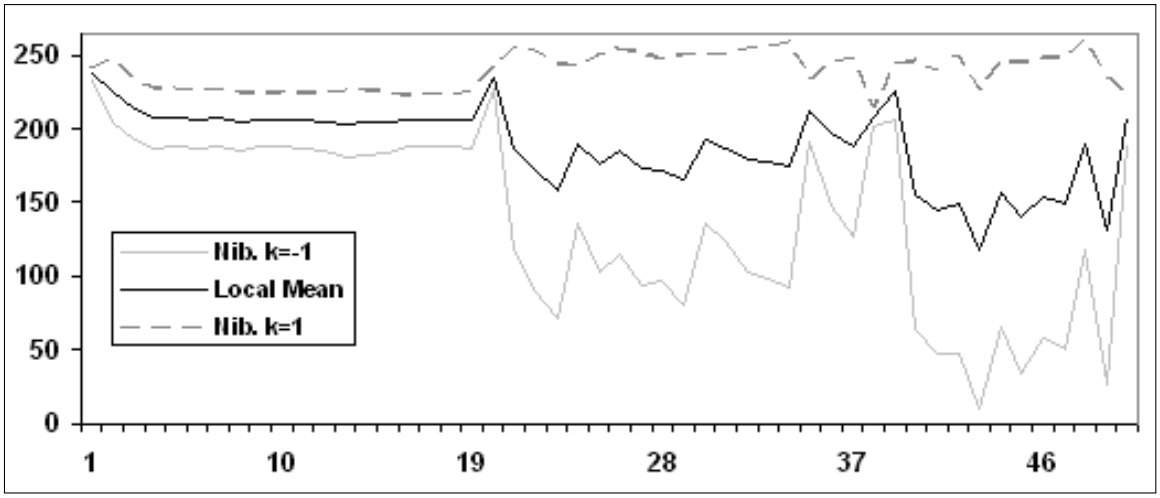
where $k = -1.5$, $R = 40$, $\mu(x,y)$ and $G(x,y)$ are the local mean and local-mean gradient of a segment respectively and local contrast c was defined as shown in *Equation 3.32*:



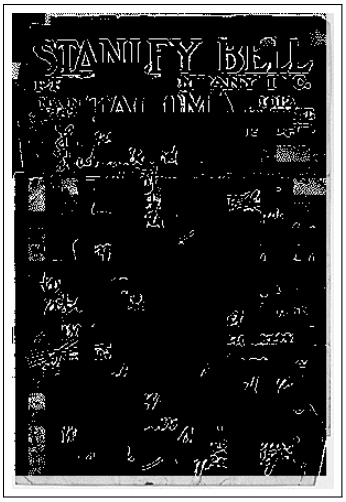
(a) Original image



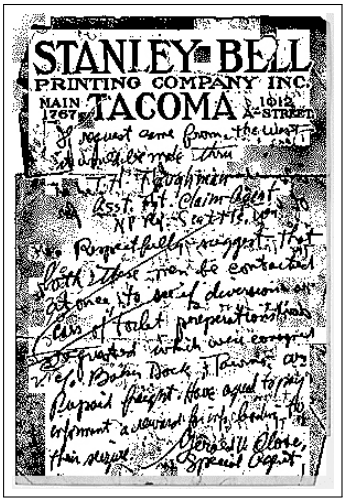
(b) Gray-level histogram of the original image



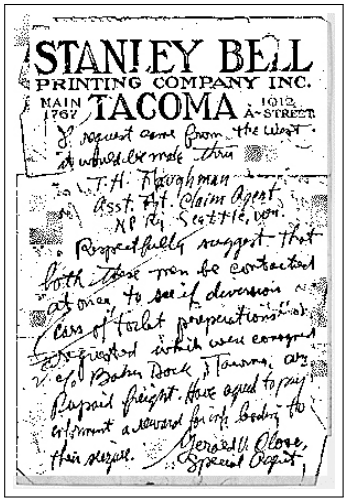
(c) Threshold values with different varieties of k



(d) The result using the Niblack method with $k = 1$

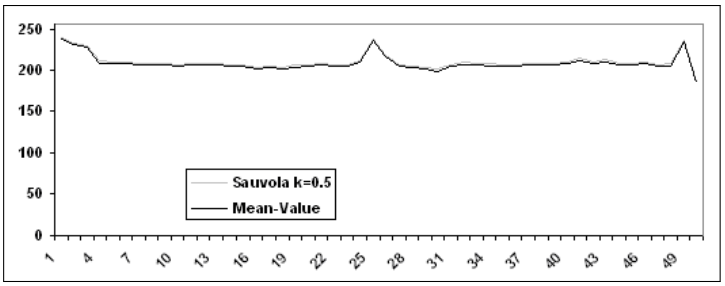


(e) The result using the Niblack method with $k = 0$

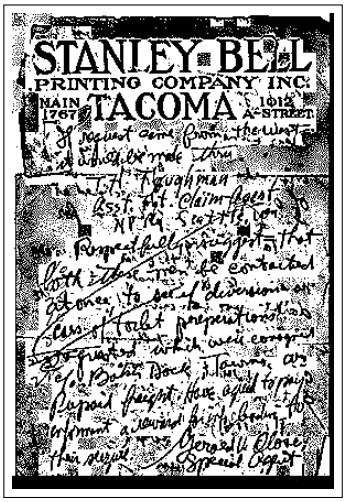


(f) The result using the Niblack method with $k = -1$

Figure 3.7: Niblack thresholding operations and examples of approximation of local mean values

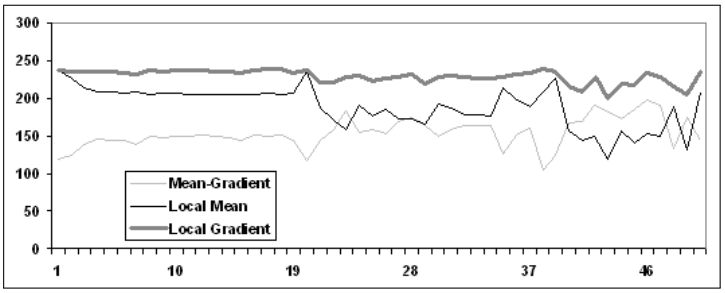


(a) Threshold value of the Sauvola et al. method with different $k = 0.5$ and mean values

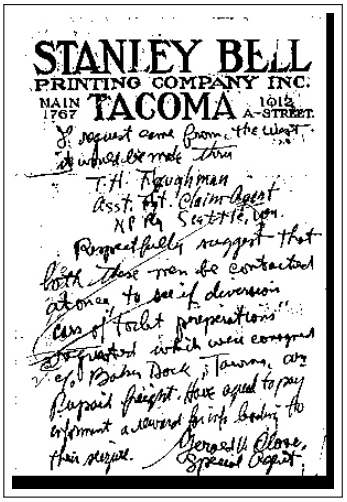


(b) Binarized image using the Sauvola et al. method, 12×12 segment, $k = 0.5$, $R = 128$

Figure 3.8: Sauvola et al. thresholding operations and examples of approximation of local mean values



(a) Mean-gradient threshold, local mean and local gradient values



(b) Binarized image: mean-gradient method, 15×15 segment, $k = -1.5$, $R = 40$

Figure 3.9: Mean-gradient thresholding operations

$$c = Z_{max} - Z_{min} \tag{3.32}$$

where Z_{max} and Z_{min} represent the maximum and minimum pixel values in a segment. Similar to the Bernsen method (see *Section 3.4.5*), local contrast c is used as an initial criteria for threshold selection which if predetermined value R is smaller than c, threshold value is determined as the half value of the mean of segment, otherwise the threshold value is determined by using Niblack's variant method which uses local gradient G instead of the local deviation $\sigma(i, j)$.

The operations of the mean-gradient method can be seen in *Figure 3.9*.

3.4.4 Adaptive Logical Thresholding (ALT)

ALT method [52] is the improvement of Kamel and Zhao logical level technique [42]. It is based on the idea of comparing the gray level of processed pixel or its smoothed gray level with some local averages in the neighborhoods about a few other neighboring pixels (*Equation 3.33*).

$$b(x,y) = \begin{cases} 1 & \textit{if} \quad \vee_{i=0}^{3} \lfloor L(P_i) \wedge L(P_i') \wedge L(P_{i+1}) \wedge L(P_{i+1}') \rfloor \\ 0 & \textit{otherwise} \end{cases} \tag{3.33}$$

where P_i, P_i', P_{i+1}, and P_{i+1}' are four points of the centered windows. P_i' and $L(P)$ were described as shown in *Equation 3.34*.

$$P_i' = P_{(i+4) \bmod 8} \qquad \textit{for} \quad i = 0, \ldots, 7,\ L(P) = \textit{ave}(P) - g(x,y) > T \tag{3.34}$$

where T is a predetermined parameter and $\textit{ave}(P)$ was defined as:

$$\textit{ave}(P) = \sum_{-sw \le i \le +sw} \sum_{-sw \le j \le +sw} \frac{f(P_x - i, P_y - j)}{(2SW+1)^2} \tag{3.35}$$

where P_x, P_y are the coordinates of P and $g(x, y) = f(x, y)$ and SW is a predetermined stroke width.

Yang and Yan improved this method to choose SW and T automatically which are stroke width and global parameter respectively.

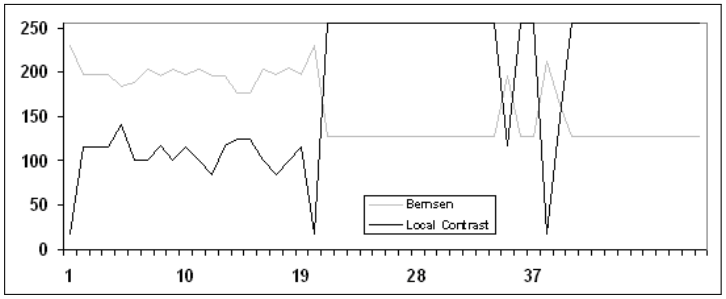
3.4.5 Bernsen Method

The Bernsen method [34] divides the image into defined $r \times r$ segments than finds the maximum Z_{max} and the minimum Z_{min} gray level values within a segment. After that, it measures the local contrast c, by using these values and compares it with a predefined value d to determine if the corresponding segment belongs to the foreground or the background layer. The formulae of Bernsen Method can be seen in *Equation 3.36* and *Equation 3.37*:

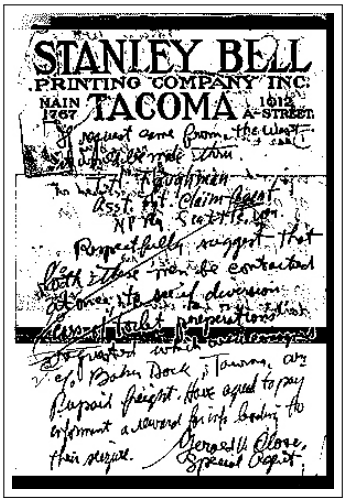
$$T(x,y) = \frac{Z_{low} + Z_{high}}{2} \tag{3.36}$$

$$T(x,y) = Z_{low} + Z_{high} < d \tag{3.37}$$

Therefore, if the local contrast is smaller than a predefined value d, segment determined as foreground or background directly. Otherwise *Equation 3.36* is used to determine the threshold point for the segment. Operations of Bernsen Method can be seen in *Figure 3.10*.



(a) Local contrast and calculated Bernsen threshold values for 15×15 segments



(b) Binarized image: Bernsen method, 15×15 segment, $d = 15$

Figure 3.10: Bernsen thresholding operations

3.4.6 Water Flow Model

Kim, Jung and Park [56] proposed the Water Flow Model (WFM) which is based on the property of water that always flows down to lower regions. In their method, it was assumed that lower gray levels denote characters and higher gray levels represent backgrounds. Rainfall fills the water in lower regions and it makes a lot of ponds in image terrain. After that, the final result is obtained by thresholding the amount of filled water using Otsu's thresholding method. The amount of water was defined as shown in *Equation 3.38*.

$$w_0 = \left\lceil \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left(f_{top} - f(x,y)\right)}{M \cdot N} \right\rceil \tag{3.38}$$

where M and N represent horizontal and vertical maximum distances of a terrain respectively, $f(x,y)$ denotes the height of the terrain and f_{top} denotes the maximum level of the terrain.

These described six local methods are also selected to perform comparisons beside six global methods with proposed methods in *Chapter 4* and *Chapter 5*, because of their popularity and effective binarization of document images. Also recently proposed methods Adaptive Logical Method and Water Flow Model are included into comparison since they were mentioned by authors as the superior of local methods.

3.4.7 Advantages and Disadvantages of Local Methods

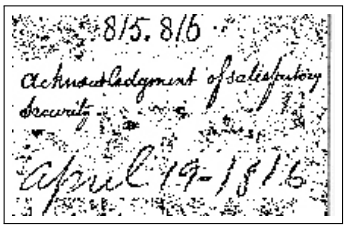
Local methods generally produce images with less loss of relevant information than global methods, however, the segment size can be considered the main disadvantages of local methods; as additional noise may be added to images using small segment sizes, whereas, large segment sizes behave as global methods and thus can over or under threshold the image.

Beside these, if segment does not consist of any interest object within, small changes in pixel values can affect the local methods to detect these pixels as objects. In *Figure 3.11*, correctly and incorrectly thresholded images by local methods are demonstrated to show

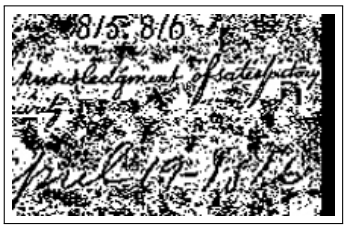
the effects of segments in binarization process.

3.5 Application Areas of Image Binarization

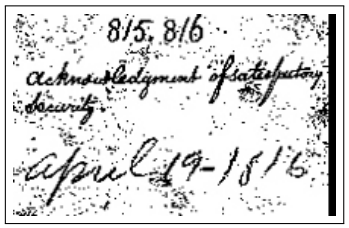
Binarization methods can either be used for low level process of any system to prepare preliminary data or lonely as a system to produce required results, and simplicity and efficiency of image binarization methods make them popular almost for every field of science and industry. However, the main application areas of image binarization are pattern recognition, biometrics, medical imaging and document analysis and understanding. In next subsections, the aims of binarization in these fields and performed researched will be presented briefly.



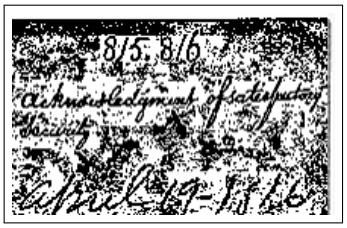
(a) Niblack method (15×15): most of the irrelevant layers detected as objects



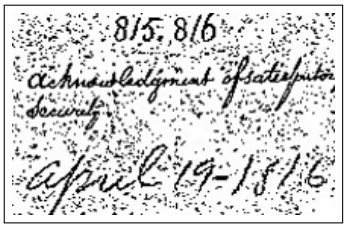
(b) Sauvola et al. (12×12): irrelevant layers and objects are mixed up together



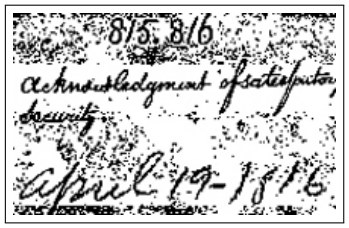
(c) Mean-gradient method (15×15): some of the irrelevant layers detected as objects



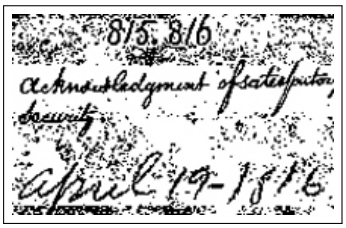
(d) PAT (15×15): similar to Sauvola method



(e) ALT (8×8) : correctly detection of characters with some noise



(f) Bernsen method (15×15): initial criteria affects some segments to the detect irrelevant layer as objects



(g) Water flow model (5×5 with $w = 17$): some loss of characters with noise

Figure 3.11: Binarization of the image in *Figure 3.6(b)* using local methods

3.5.1 Image Binarization in Pattern Recognition

Pattern Recognition applications mostly require low level processing of data to prepare effective input or training data for recognition systems and also they can be used to reduce the amount of unnecessary data. In oceanic applications, recognition of plankton images [62, 63, 64] and classification of sea-ices [65] use binarization as an initial process to prepare effective data for the recognition or classification systems. In industrial applications, such as recognition of coin patterns, thresholding is used to reduce the amount of data sent to a classifier that provides increment of recognition rate and decrement of time of training [66, 67, 68, 69, 70].

3.5.2 Image Binarization in Biometrics

In Biometrics, relevant data varies to applications. In fingerprint recognition, thresholding applications are mostly used to enhance and binarize the images to provide optimal effective and clean data for recognition systems. Therefore, several researches recently had been performed for this purpose [71, 72, 73, 74]. Also, thresholding applications had been used in face skin detection and retinal vessel recognition in personal authentication and security fields of biometrics [75, 76]. In spite of these researches, binarization did not become popular tool for biometrics because little loss of information may cause any biometrical system not to produce effective results for security or other fields that requires high level of accuracy.

3.5.3 Image Binarization in Medical Imaging

In medical imaging, the aim is to extract or enhance the region of interest within an image. Thus, binarization techniques are frequently used to achieve this aim. However, application dependency of proposed thresholding methods, mostly cause the use of different methods in different medical applications. Any method that produces most successful results for Magnetic Resonance Images may not produce effective results for X-Ray images. In spite of this disadvantage, thresholding methods are used almost in every field of medical imaging. Segmentation and enhancement of microscopic images [77, 78], radiotherapy imaging applications [79], enhancement of x-ray, MRI [80, 81, 82], computed tomography [83] and ultrasound images [84, 85], and detection of tumor cells

[86, 87, 88, 89, 90] are some of many applications of thresholding in medical imaging.

3.5.4 Image Binarization in Document Analysis and Understanding

Document analysis and understanding is the most popular application area of image binarization where effective separation of characters that belongs to foreground are required while discarding irrelevant information of the background layer. However, variety of distortions such as smears, smudges, shadows, acquisition noises, illumination effects, within documents make challenging task for binarization methods. Development of new techniques and comparison of conventional or recently proposed methods are still being performed by researchers and this also makes another challenging task which is the determination of superior method for document images. Therefore, similar to other application areas, application-dependency of methods makes some methods superior for some images, and another for other images.

3.6 Summary

Binarization is a low level image processing technique that is used for segmentation or enhancement of the region of interest within the images. Several methods that are classified into local and global groups had been proposed. However, each method brings some disadvantages beside their apparent advantages. In this chapter, basic definitions of image binarization, chronological development, detailed explanation about selected twelve methods and application areas were presented. In the next chapter, proposed global Mass-Difference (MD) thresholding and local Pattern Averaging thresholding (PAT) methods will be described. Also, performed experiments and the obtained results will be explained in details.

CHAPTER 4

THE PROPOSED THRESHOLDING METHOD

4.1 Overview

In this chapter, the proposed novel global mass difference (MD) thresholding method and the investigated local pattern averaging thresholding (PAT) method will be described. In addition, basic characteristics of document images and the performed test experiments will be explained in details.

4.2 Mass-Difference Thresholding Method

The proposed Mass-Difference (MD) method is a global single-stage thresholding method that finds a single threshold value using the global maxima (maximum luminance value) and the mass average (mean of the intensities) of an image. It was designed by considering the characteristics of document images which generally consist of darker foreground information than background.

4.2.1 The Hypothesis

MD thresholding method is designed according to the three hypothesis as listed below :

- Hypothesis I : The foreground layer of an image is always in the lower region of the mean value.
- Hypothesis II : The foreground layer of an image can be distinguished from its background by shifting this mean value by a variable amount.

- Hypothesis III : Symmetrically difference of the luminance value and the mean value provides exact separation and enhancement point.

In the next subsection, the mathematical description of the MD thresholding method will be described by using the hypothesis that are listed above.

4.2.2 Mathematical Description of the MD Thresholding Method

A previously developed basic global thresholding method that resembles the MD method is the background-symmetry algorithm, which assumes a distinct and dominant peak for the background that is symmetric about its maximum, and is defined as:

$$\theta = max\ p(p\% - max\ p) \tag{4.1}$$

where θ is the threshold point, $maxp$ is the maximum peak value in the gray level histogram of the image and $p\%$ is the non-object pixel side of that maximum. However, in background-symmetry algorithm, thresholding methods may need some adaptation if the brightness histogram has been changed.

MD thresholding method is different from the background-symmetry algorithm in that the maximum value is defined as the highest pixel value within the image, whereas in background-symmetry the maximum peak is found by determining the maximum value in the histogram.

Images that are separable by a single-stage threshold value, contain the 'region of interest' or 'relevant data' (which is the text in document images), in the lower region of the mean value of the image (see *Section 4.2.1*, Hypothesis I). Thus, the binarization process of these images can be achieved by using the following equation:

$$T\ =\ \mu\ -\ \theta \tag{4.2}$$

where μ represents the mean of the intensities (mass) within the image and defined as given in *Equation 4.3* and θ denotes the distance (difference) of the 'region of interest' from the mean point(μ).

$$\mu = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} I(x, y) \tag{4.3}$$

However the distance of the relevant data from the mean point is not constant for all type of images. Consequently, unique global characteristics of images should be considered to determine this distance.

Shifting the mean position to a point that separates background and foreground layers, could provide efficient separation of these layers (see *Section 4.2.1*, Hypothesis II). The luminance value of an image is the brightest pixel (maximum gray value) within the image and it defines the maximum gray level limit of the background layer in both uniform and non-uniform conditions.Luminance value L can be determined using the function that is given in *Equation 4.4*.

$$L = f_{max}(I) \tag{4.4}$$

where L denotes the luminance value, f_{max} is the function that finds the maximum pixel value within the original grayscale image I.

Thus, considering the deviation of the limit from the mean point of an image and symmetrically shifting this mean point by the amount of the deviation, provides exact separation of the 'region of interest' from the background layer (see *Section 4.2.1*, Hypothesis III)as shown in *Equation 4.5*.

$$\theta = L - \mu \tag{4.5}$$

where L is the luminance value of image and μ represents mean of the intensities within the image.

Therefore T can be rewritten as shown in *Equation 4.6*.

$$T = \mu - (L - \mu) \tag{4.6}$$

An MD enhanced image (MI) is obtained using equation *Equation 4.7*:

$$MI(x,y) = \begin{cases} 0 & \text{if } I(x,y) \leq T \\ 255 & \text{else} \end{cases} \tag{4.7}$$

Figure 4.1–Figure 4.4 show the basic operations of Mass-Difference thresholding method.

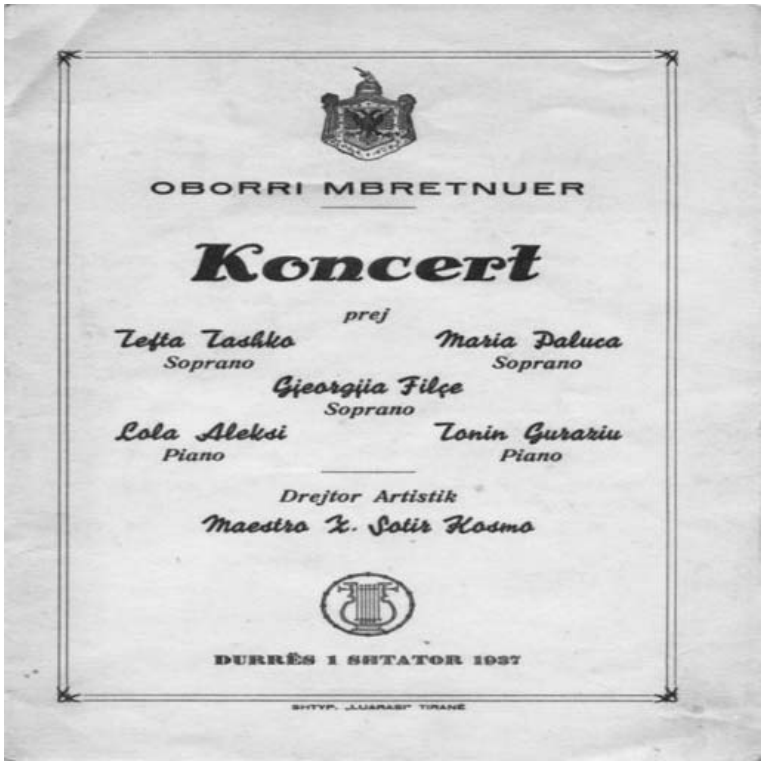


Figure 4.1: Example Image

Various experiments had been performed to test the behavior of the proposed method under different conditions.

Non-uniform Background Test

For any thresholding method, binarization of an image that has non-uniform background or bimodal gray level histogram, is a challenging task. Thus, the efficiency of a method

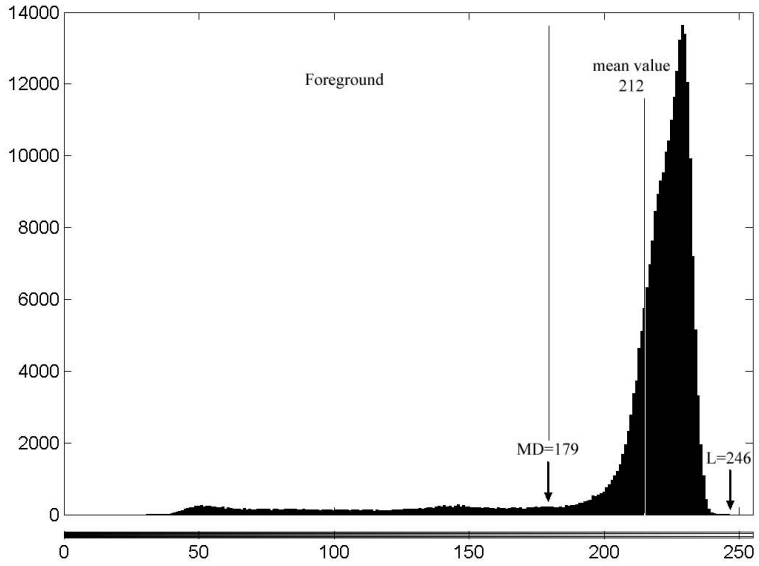


Figure 4.2: Corresponding Histogram and MD operations on Image *Figure 4.1*

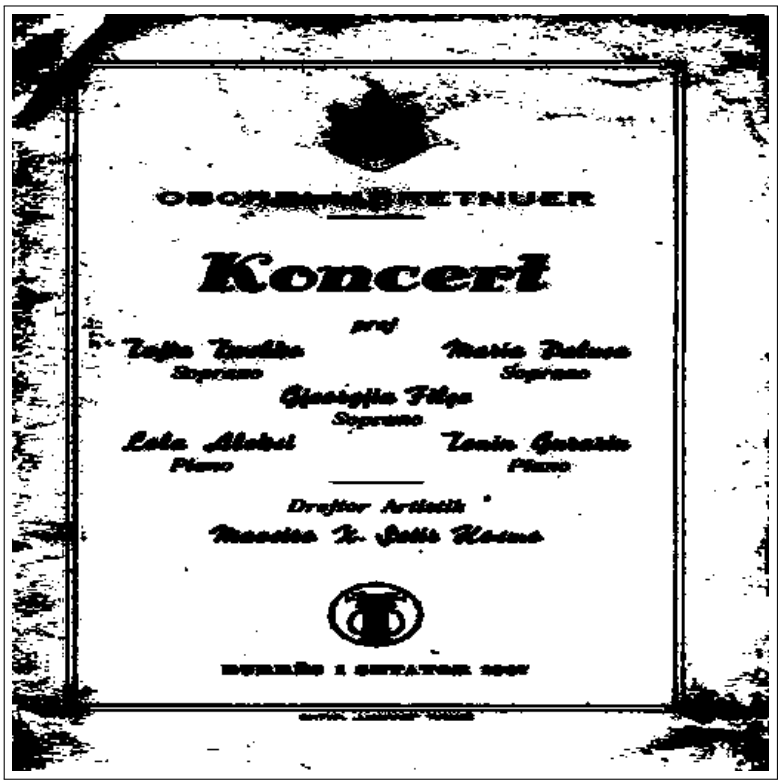


Figure 4.3: Binarization of Example Image Using Mass Value

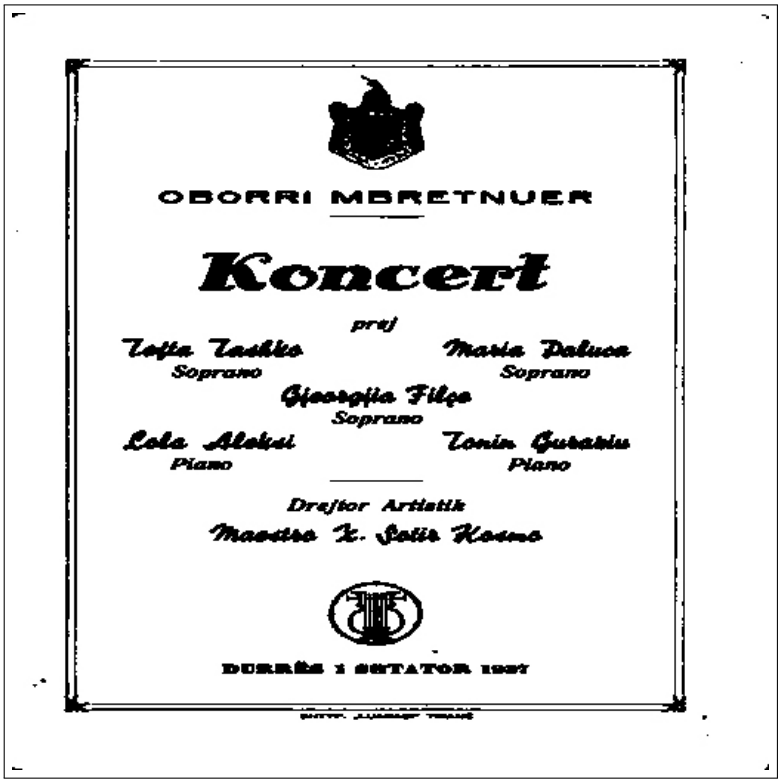
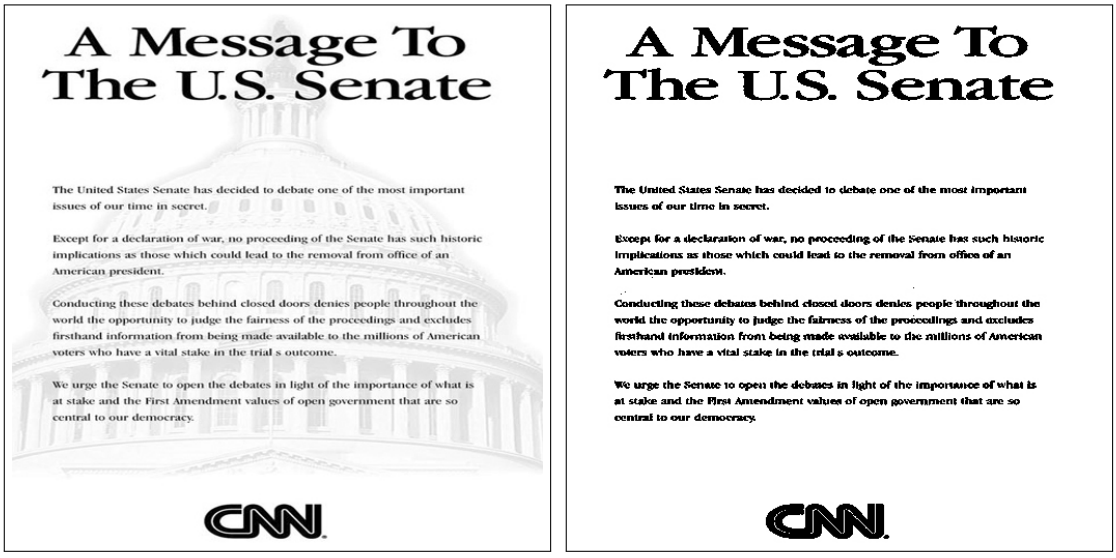


Figure 4.4: Binarization of Example Image Using MD Value

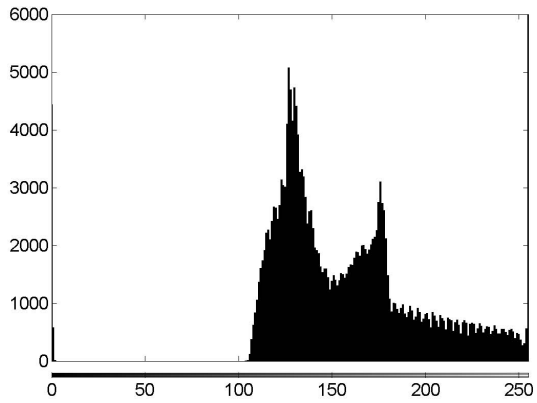


(a) Original non-uniform background image (b) Binarized image by MD thresholding

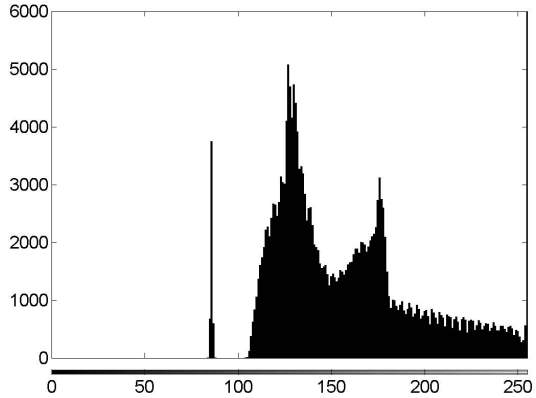
Figure 4.5: Binarization Example of MD thresholding method

(a) created non-uniform background and illumination with darker text color

(b) created non-uniform background and illumination with lighter text color



(c) Corresponding histogram of (a)



(d) Corresponding histogram of (b)

(e) binarization of 4.6(a) by proposed MD Method

(f) binarization of 4.6(b) by proposed MD Method

Figure 4.6: Testing of proposed MD method in bimodal images

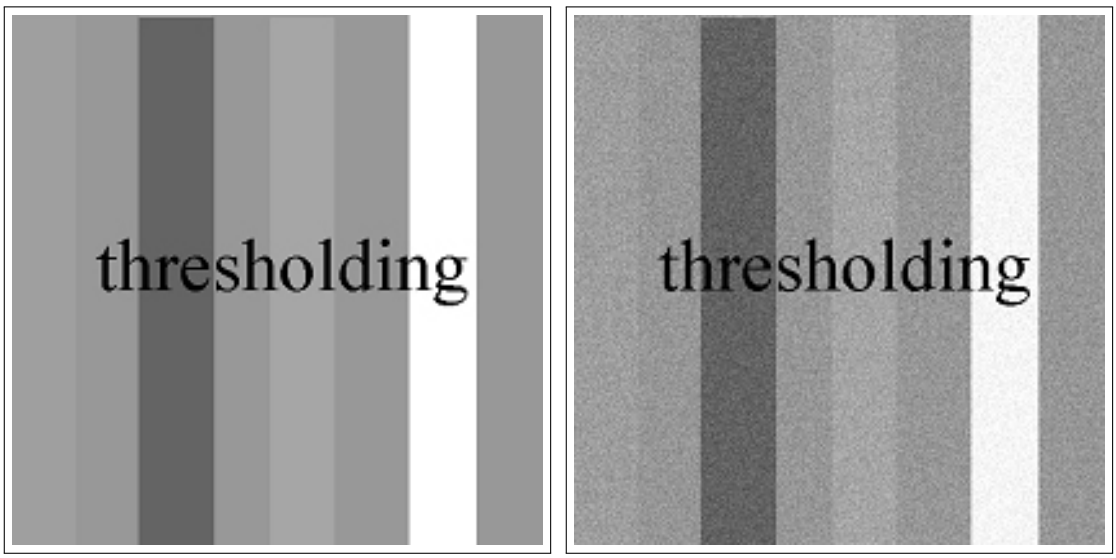
can also be demonstrated by testing its behavior under these conditions. Non-uniform background test is applied to the proposed MD thresholding method using both clean and noisy non-uniform background images to demonstrate its efficiency in the binarization of these kind of documents. The resultant images is expected to have only the region of interest which are all text in document images. *Figure 4.5* shows the result of non-uniform background test of MD thresholding method using a document image and *Figure 4.6* shows the binarization results of the MD thresholding method in artificially created text with bimodal histogram and different text color and illumination (which any of the other considered global and local methods could not produce effective separation for this image). *Figure 4.7* shows the binarization results of the MD thresholding method in artificially created text documents which consist non-uniform background and additional noise. *Figure 4.8* shows the binarized images of *Figure 4.6* (a) and (b) by using other global methods, and *Figure 4.9* shows the binarized images of *Figure 4.6* (a) and (b) by using local methods.

L-value Test

MD thresholding method determines the threshold point by using the luminance value L, thus if any noise occur within an image, which is represented only by a pixel with a value of maximum gray level 255, the amount of shifting will be changed. Consequently, it is required to test the behavior of the proposed MD thresholding method under this described condition. The results of this test, which was called $L-value$ test, are expected to remove the whole background without any noise addition and loss of information. *Figure 4.10*. shows the obtained results of the applied L-value test.

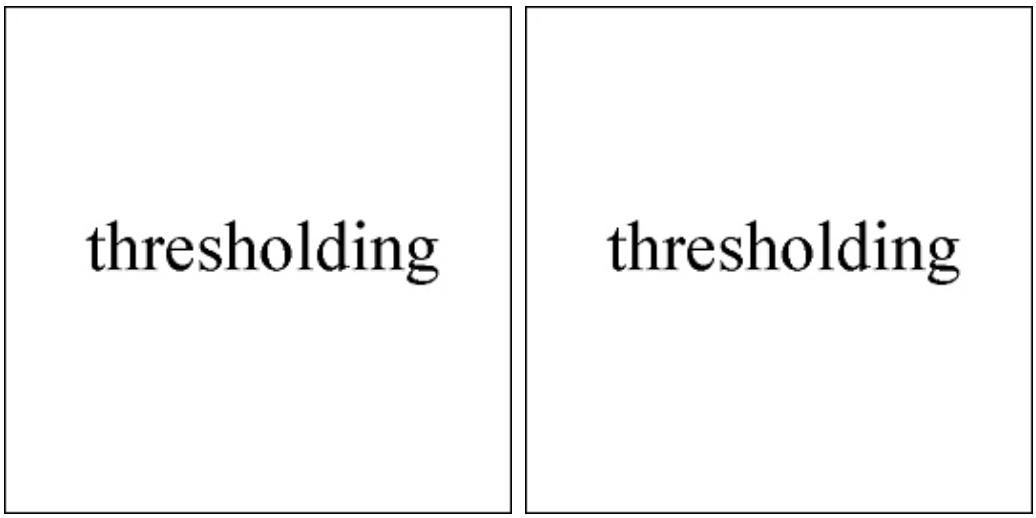
4.2.3 Statistical Experiments on the proposed MD Thresholding Method

Statistical experiments on the proposed MD thresholding method were performed to demonstrate that the selected separation point is the exact threshold value. Five sample images were selected to contain different illumination, background and noise, and were binarized by using all 8-bit gray level values which are $T = 0...255$. Then Mean Square Error (MSE) is used to find the minimum error point of 256 binarized images for each sample image . The results are compared to the error obtained using the MD



(a) created non-uniform background image

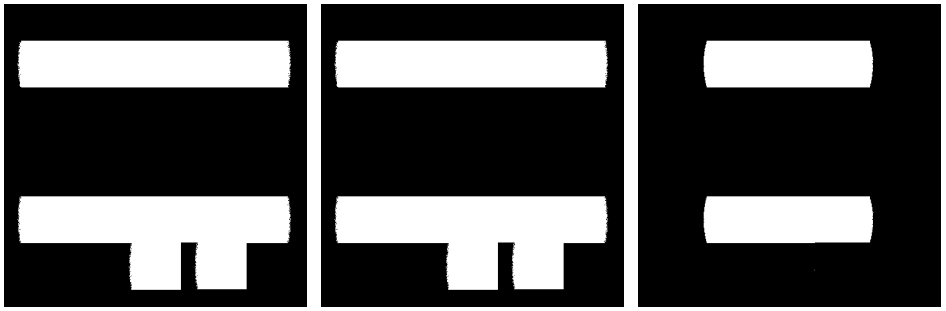
(b) created non-uniform background image with Gaussian noise (mean=0, variance=0.01)



(c) binarization of 4.7(a) by proposed MD method

(d) binarization of 4.7(b) by proposed MD method

Figure 4.7: Testing of proposed MD method under extreme conditions



(a) binarization of 4.6(a) by Otsu Method

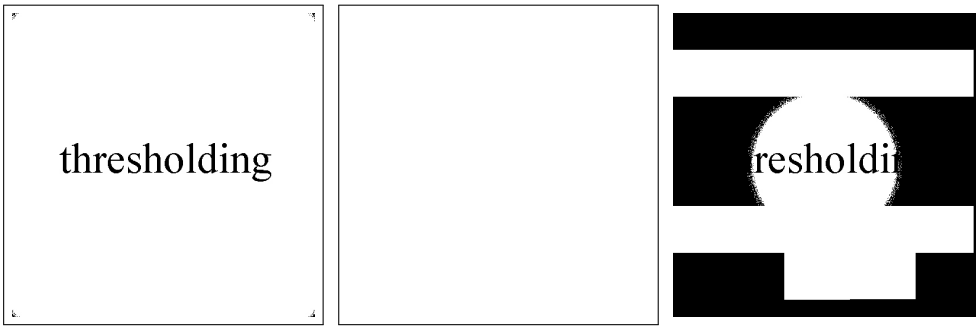
(b) binarization of 4.6(b) by Otsu Method

(c) binarization of 4.6(a) by Kittler and Illingworth Method

(d) binarization of 4.6(b) by Kittler and Illingworth Method

(e) binarization of 4.6(a) by Yanni and Horne Method

(f) binarization of 4.6(b) by Yanni and Horne Method



(g) binarization of 4.6(a) by Ramesh et al. Method

(h) binarization of 4.6(b) by Ramesh et al. Method

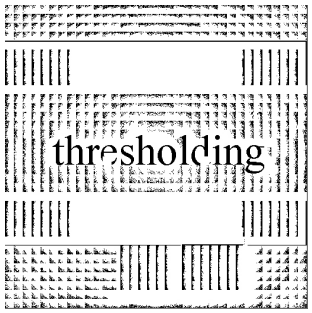
(i) binarization of 4.6(a) by Kapur et al. Method

(j) binarization of 4.6(b) by Kapur et al. Method

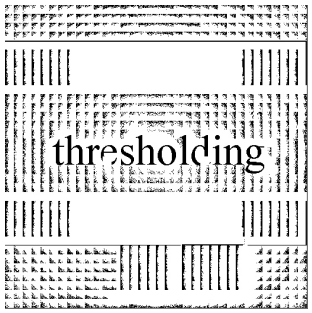
(k) binarization of 4.6(a) by Albuquerque et al. Method

(l) binarization of 4.6(b) by Albuquerque et al. Method

Figure 4.8: Binarization of *Figure 4.6* (a) and (b) images by global methods



(a) binarization of 4.6(a) by Niblack Method

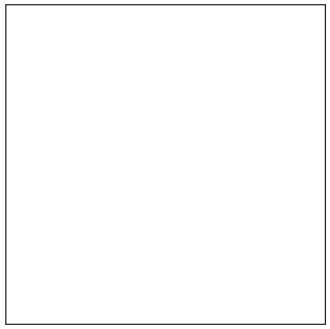


(b) binarization of 4.6(b) by Niblack Method

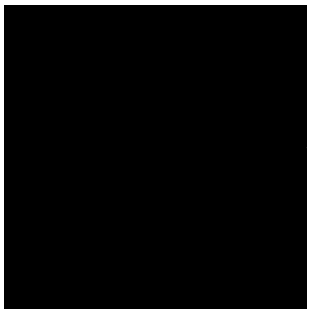
(c) binarization of 4.6(a) by Sauvola et al. Method

(d) binarization of 4.6(b) by Sauvola et al. Method

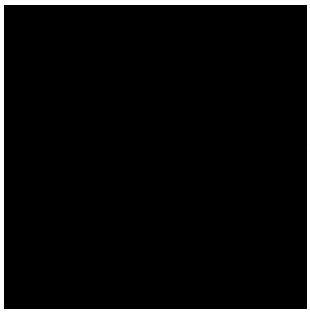
(e) binarization of 4.6(a) by Mean-Gradient Method



(f) binarization of 4.6(b) by Mean-Gradient Method



(g) binarization of 4.6(a) by Bernsen Method



(h) binarization of 4.6(b) by Bernsen Method

(i) binarization of 4.6(a) by ALT Method

(j) binarization of 4.6(b) by ALT Method

(k) binarization of 4.6(a) by WFM Method

(l) binarization of 4.6(b) by WFM Method

Figure 4.9: Binarization of *Figure 4.6* (a) and (b) images by local methods

(a) artificially created L-value test document with darker text : $\mu = 218, l = 255$

(b) artificially created L-value test document with lighter text : $\mu = 222, l = 255$

(c) binarization of (a) by proposed MD Method

(d) binarization of (b) by proposed MD Method

Figure 4.10: L-value test of proposed MD method under extreme conditions

threshold value, and a True Percentage Relative Error (ε_t) is determined. The equations of MSE and ε_t can be seen in *Equation 4.8* and *Equation 4.9* respectively:

$$MSE = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} (Y_{i,j} - \hat{X}_{i,j})^2 \tag{4.8}$$

where $(M \times N)$ is the size of the image, and $Y_{i,j}$ and $\hat{X}_{i,j}$ represent the pixel values at location (i, j) of the original and enhanced images, respectively.

$$\varepsilon_t = (True\ Error - True\ Value) \tag{4.9}$$

where $True\ Error$ is the calculated minimum MSE value and the $True\ Value$ is the MSE value of MD point.

Table 4.1 shows the results of the MSE analysis and *Figure 4.11* shows the MD threshold point on MSE graphs of the sample images.

After calculating the True Percentage Relative Error (ε_t), two tailed and two sampled equal variance t-Test is used to determine the significance of the threshold point obtained by the MD method.

T-test is the statistical significance level test of t distribution of hypothesis to be rejected or to be accepted [91, 92]. Particular significance d was defined as shown below [93]:

$$d = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \tag{4.10}$$

where $\bar{x}$ is the sample mean, n is the number of samples, μ_0 is the mean of the normally distributed samples and σ is the variance.

Confidence level was selected as 95% which is equal to 2.571 [94] in $n = 5$. Particular significance of MSE values that was determined by using MD method, is computed as 0.988 that yields the MD threshold point can be assigned as the exact threshold point.

Additionally, visual inspection of binarized sample images (as shown in *Figure 4.12*

and *Figure 4.13*) shows the effect of shifting the obtained value by MD threshold.

Table 4.1: True Percentage Relative Error (ε_t) Comparison

Image Name	Minimum Error Value	Minimum Error Point Range	MD Point	MD Point Error	ε_t value
S1	2.84	123-138	126	2.84	0%
S2	2.68	115-137	120	2.68	0%
S3	8.70	115-137	159	8.84	1.4%
S4	5.94	115-145	141	5.94	0%
S5	1.35	123-137	136	1.35	0%

4.2.4 Experiments on the MD Thresholding Method

Two experiments had been performed to test the success and efficiency of proposed method by comparing other methods. In next subsection, these two experiments will be explained briefly.

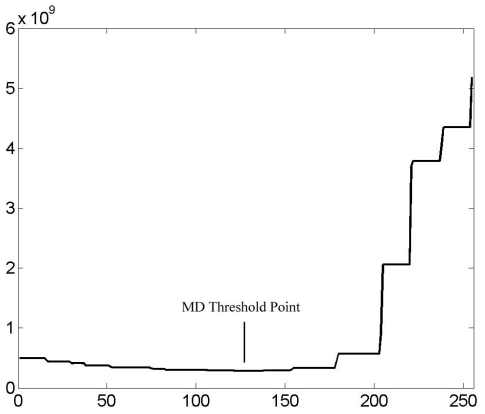
Experiment I

Experiment I [1] comprises the binarization of 30 document images by four global methods, namely Otsu method [24], Kittler and Illingworth method [22], Kapur et al. method [31] , Quadratic Integral Ratio (QIR) [51], besides MD, and a local method Parker [40].

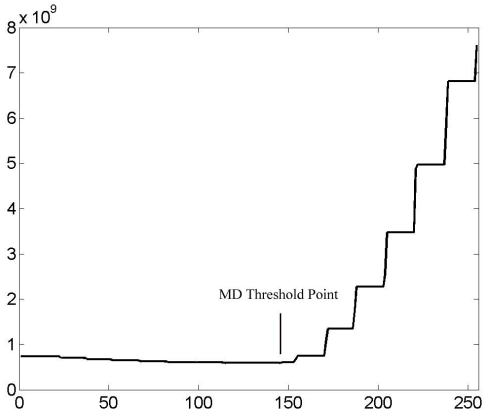
The efficiency of these methods was determined by the recognition rate of words in the thresholded or segmented document images. Visual inspection of the documents determined the number of recognized and readable words in a thresholded document. This was performed by 15 independent persons (see *Chapter 5* for the details of the visual inspection procedure). The total number of the words in the original document and the number of recognized and readable words after thresholding were used to determine the recognition rate.

Thirty images that comprise a total of 1205 words were divided into three groups as clean, degraded and highly degraded documents. The results of the experiment showed that MD was the superior method which was followed by Otsu and Kapur et al. Methods. *Table 4.2* shows the obtained results by implementation of 30 document images.

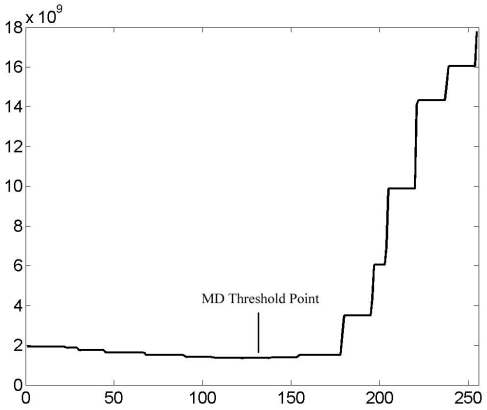
[1] Experiment I was published in IEEE International Conference on Industrial Technology, in December 2006 as "Novel Thresholding Method for Document Analysis".



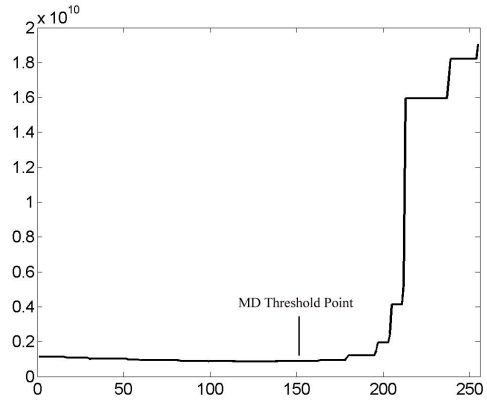
(a) MSE graph of sample image 1



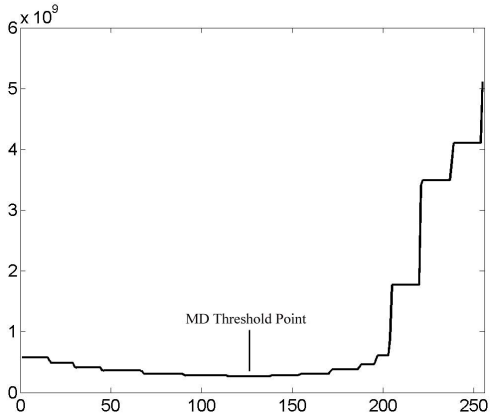
(b) MSE graph of sample image 2



(c) MSE graph of sample image 3



(d) MSE graph of sample image 4

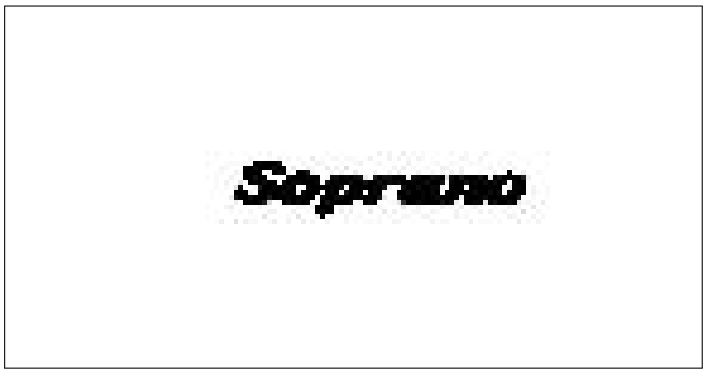


(e) MSE graph of sample image 5

Figure 4.11: MSE graphs of sample images

(a) MD Point : 126

(b) Loss of information : (at $T = 146$)

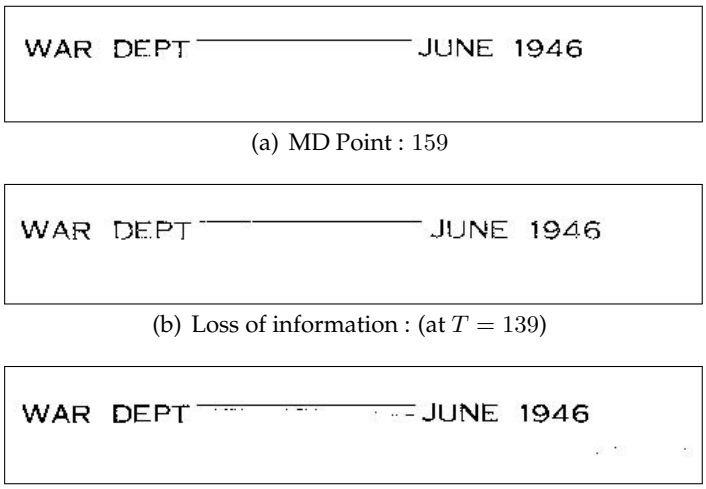


(c) Additional Noise : (at $T = 136$)

Figure 4.12: Threshold point effects in sample image 1

Table 4.2: Recognition Rates of Experiment I

Thresholding Method	Clean Words	Degraded Words	Highly Degraded Words	Total Words	Recognition Rate
Otsu	300/325	335/450	290/430	925/1205	77%
Kapur et al.	289/325	325/450	175/430	789/1205	65%
Parker	252/325	250/450	235/430	737/1205	61%
Kit. and Ill.	199/325	60/450	35/430	294/1205	24%
QIR	215/325	90/450	80/430	385/1205	32%
MD	308/325	350/450	300/430	958/1205	80%



(a) MD Point : 159

(b) Loss of information : (at $T = 139$)

(c) Additional Noise : (at $T = 179$)

Figure 4.13: Threshold point effects in sample image 3

Experiment II

Experiment II [2] contains the binarization of 55 images by using Otsu Method, Kittler and Illingworth method as global methods, and Niblack Method as local method. The implementations were carried out using two image sets of documents. The first set comprises 50 historical documents which contain a total of 2021 words with different contrasts and brightness. The second set contains 5 created words which contain a total of 45 characters with different backgrounds and different color to test the occurrence of the characters after thresholding.

Similar to Experiment I, the efficiency of these methods was determined by the recognition rate of words in the thresholded or segmented document images by 15 independent persons (see *Chapter 5* for the details of the visual inspection)for two sets separately. A general comparison is then performed by combining the results using the two sets of documents. General results are categorized as: $recognized$ or $unrecognized$ words in Set 1, and $clear$ or $unclear$ characters in Set 2.

Results showed that MD was the superior method which was followed by the Niblack method. *Table 4.3* and *Table 4.4* shows the obtained results by the implementation of Set 1 and Set 2 of Experiments II.

Also, commercial image processing packages such as Photoshop, ACDSee etc., are investigated to observe the efficiency of the methods within these packages, however,

[2] Experiment II was published in 11th Panhellenic Conference on Informatics, Patras, Greece, in May 2007, as "A Novel Thresholding Method for Text Separation and Document Enhancement".

Table 4.3: Recognition Rates of Characters in Set 1 of Experiment II

Thresholding Method	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Otsu	2021	1657	364	81.98 %
Kittler and Illingworth	2021	1045	976	51.70 %
Niblack	2021	1699	322	84.06 %
MD	2021	1730	291	85.60 %

Table 4.4: Recognition Rates of Characters in Set 2 of Experiment II

Thresholding Method	Total Characters	Clear Characters	Unclear Characters	Recognition Rate
Otsu	45	40	5	88.88%
Kit. and Illingworth	45	45	0	100%
Niblack	45	45	0	100%
MD	45	45	0	100%

none of the commercial packages contain the method that can be efficiently used for automatic image binarization.

4.3 Pattern Averaging Thresholding (PAT)

The usage of the mean value is popular in almost every field of science and engineering. Some examples of these usage are the pattern preparation phases in intelligent systems or the filtering approaches in image processing. However, direct usage of the mean value in image binarization has not been investigated in details.

4.3.1 The Hypothesis

The $mean\ value$ is the center or the location of distribution. Thus, for local thresholding methods, generally the mean value of the grayscale pixel values of the image segments were concerned as the initial thresholding line that requires the approximation or adjustment. Thus, mean-based local thresholding methods which were explained in *Chapter 3*, generally use local standard deviation or gradients to adjust these independent mean values.

- Hypothesis I : direct use of the mean value may produce similar sufficient results for particular type of images when applying as the threshold point.

4.3.2 Mathematical Description of the PAT Method

PAT [95] is based on averaging pixel values within segments of a pattern, thus yielding one average pixel value per segment. The output pattern would contain averaged segment values and is defined as in *Equation 4.11* and also can be rewritten by using Niblack's method formula as shown in *Equation 4.12*. Operations of PAT Method can be seen in *Figure 4.14*.

$$I(x,y) = \begin{cases} 255 & \text{if } \mu(x,y) \geq P(x,y) \\ 0 & \text{else} \end{cases} \tag{4.11}$$

where P and (x,y) denotes the original pixel value of an image and the PAT point of segment respectively; and I is the thresholded image.

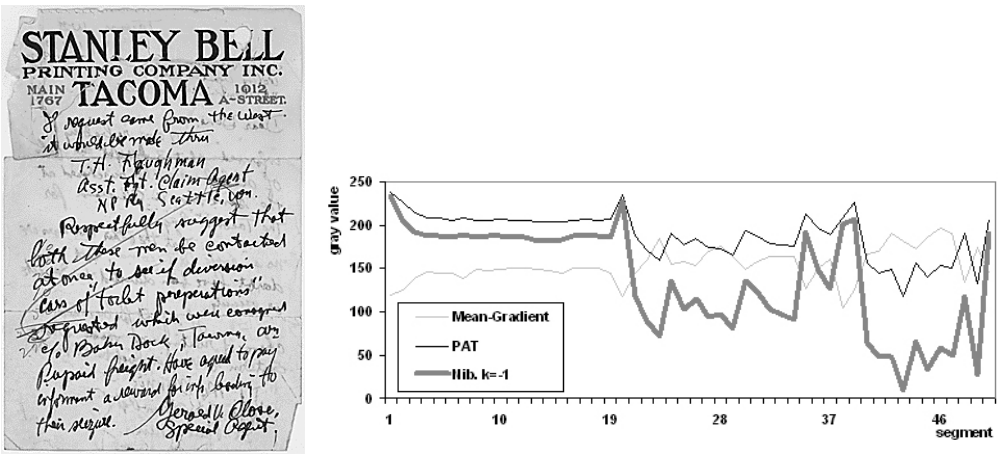
$$T[x,y] = \mu(x,y) + k.s(x,y) \tag{4.12}$$

where k is the adjustment value of the local mean and $s(x,y)$ is the local standard deviation. Hence, by using $k = 0$ we can simply obtain the PAT value as shown in *Equation 4.13*.

$$T[x,y] = \mu(x,y) \tag{4.13}$$

PAT is investigated to increase the visual appearance of hidden information such as watermarked text and figures within the images.

However, similar to the other mean-based local methods, the determination of segment size is a challenging task where small size of segments increase the noise addition into binarized images and large size of segments act as global methods and sometimes can produce over or under thresholded images that cause loss of relevant information.



(a) Original Image (b) Local Mean Graphs

Figure 4.14: PAT Operations

4.3.3 Experiments on PAT Method

In order to investigate the use of local mean values, two experiments were performed which are described in the next subsections.

Enhancement of Unclear Patterns using PAT

This section presents the experimental results of implementing PAT and the comparison of PAT with six other well-known thresholding algorithms for unclear image enhancement and visualization [3]. The comparison is drawn using six various images out of the 48 images in the collected database. Segment sizes of 8×8 pixels and 16×16 pixels were used for PAT to enhance the test images. Different thresholding values that represent the true pixel values of an image were obtained. *Figure 4.15 – Figure 4.17*. show the experimental results of implementing PAT (8×8 and 16×16 segmentation), Bernsen's method, Kapur's entropy, Kittler and Illingworth's method, Niblack's Method, Otsu's method and Parker's method, for the enhancement and visualization of three of the unclear patterns; namely official stamp, watermark and banknote.

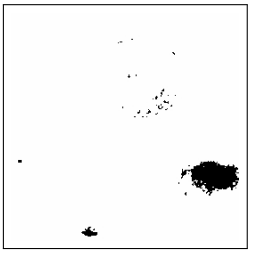
A close visual inspection of the enhanced images in figures, shows that the use of pattern average thresholding (PAT) is an effective method for visualizing and enhancing unclear patterns. In "official stamp" image (*Figure 4.15*), the stamp and fingerprints are al-

[3]This subsection was published in Proceedings of the 3^{rd} International Symposium on Electrical, Electronic and Computer Engineering, vol. 1, pp. 253-257, Nicosia, TRNC, in 2006 as "Enhancement of Unclear Patterns Using Pattern Average Thresholding".

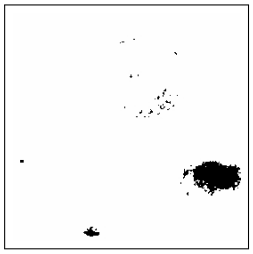
(a) Original Stamp and Fingerprint Image

(b) Binarization by PAT (8×8) method

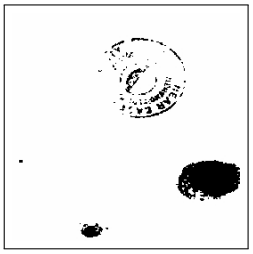
(c) Binarization by PAT (16×16) method



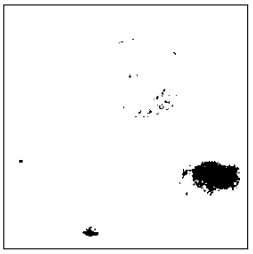
(d) Binarization by Bernsen method



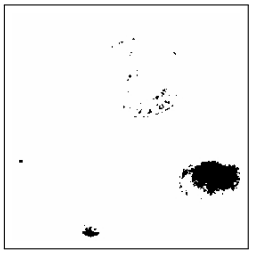
(e) Binarization by Kapur et al. method



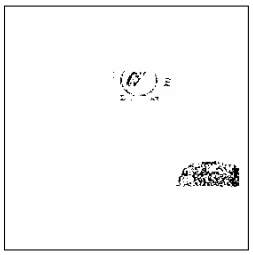
(f) Binarization by Kittler and Illingworth method



(g) Binarization by Niblack method



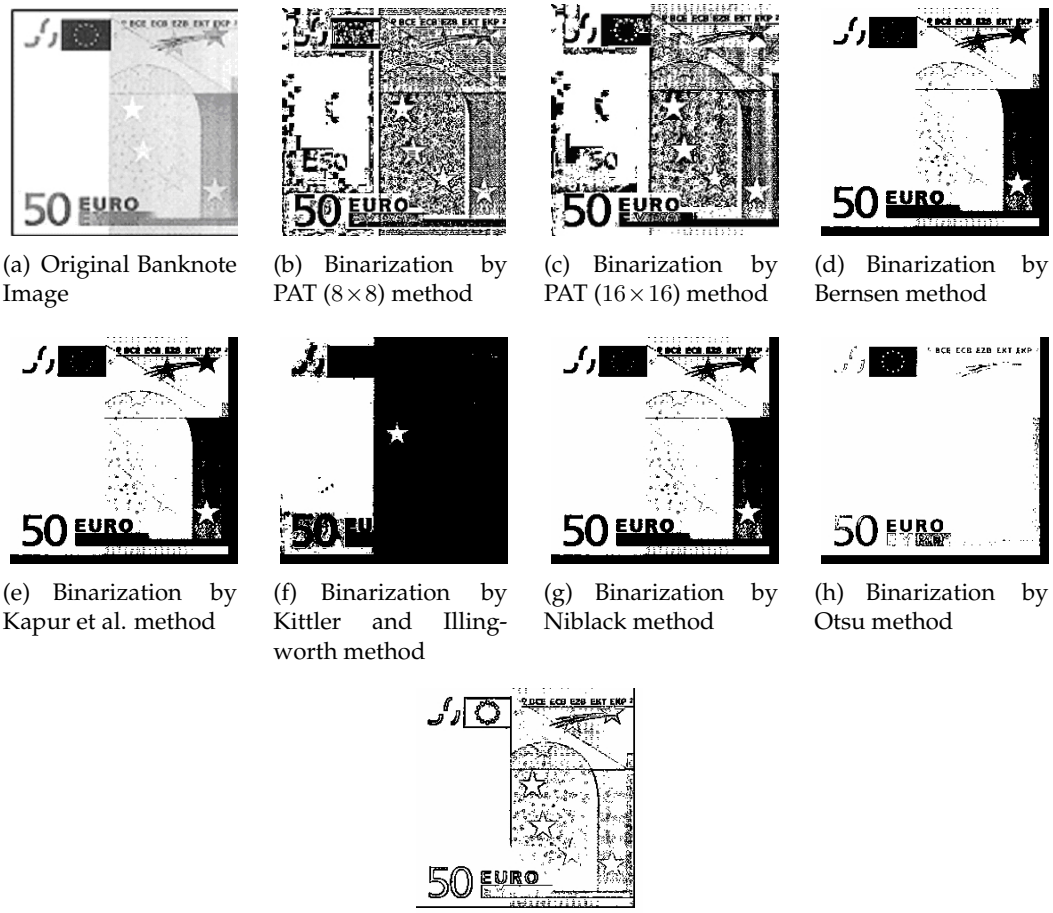
(h) Binarization by Otsu method



(i) Binarization by Parker method

Figure 4.15: Example Results of Fingerprint and Stamp Image

most invisible for the naked eye but are successfully enhanced and visualized using PAT, however, other thresholding techniques have some loss of information which causes data loss for forensic analyzers. In "hidden banknote pattern", expected result was the detection of watermark (E50) in a banknote which is used for the counterfeit banknote detection. While Kittler and Illingworth's, Otsu's, Niblack's, Bernsen's, Kapur's and Parker's methods could not enhance the (E50) watermark, PAT enhances this pattern successfully (*Figure 4.16*). In "watermark" image (*Figure 4.17*), while most of the thresholding techniques discarded watermark characters and figures while detecting dark patterns, PAT and Kittler and Illingworth's methods produced ideal results for detecting watermark patterns.



(a) Original Banknote Image

(b) Binarization by PAT (8×8) method

(c) Binarization by PAT (16×16) method

(d) Binarization by Bernsen method

(e) Binarization by Kapur et al. method

(f) Binarization by Kittler and Illingworth method

(g) Binarization by Niblack method

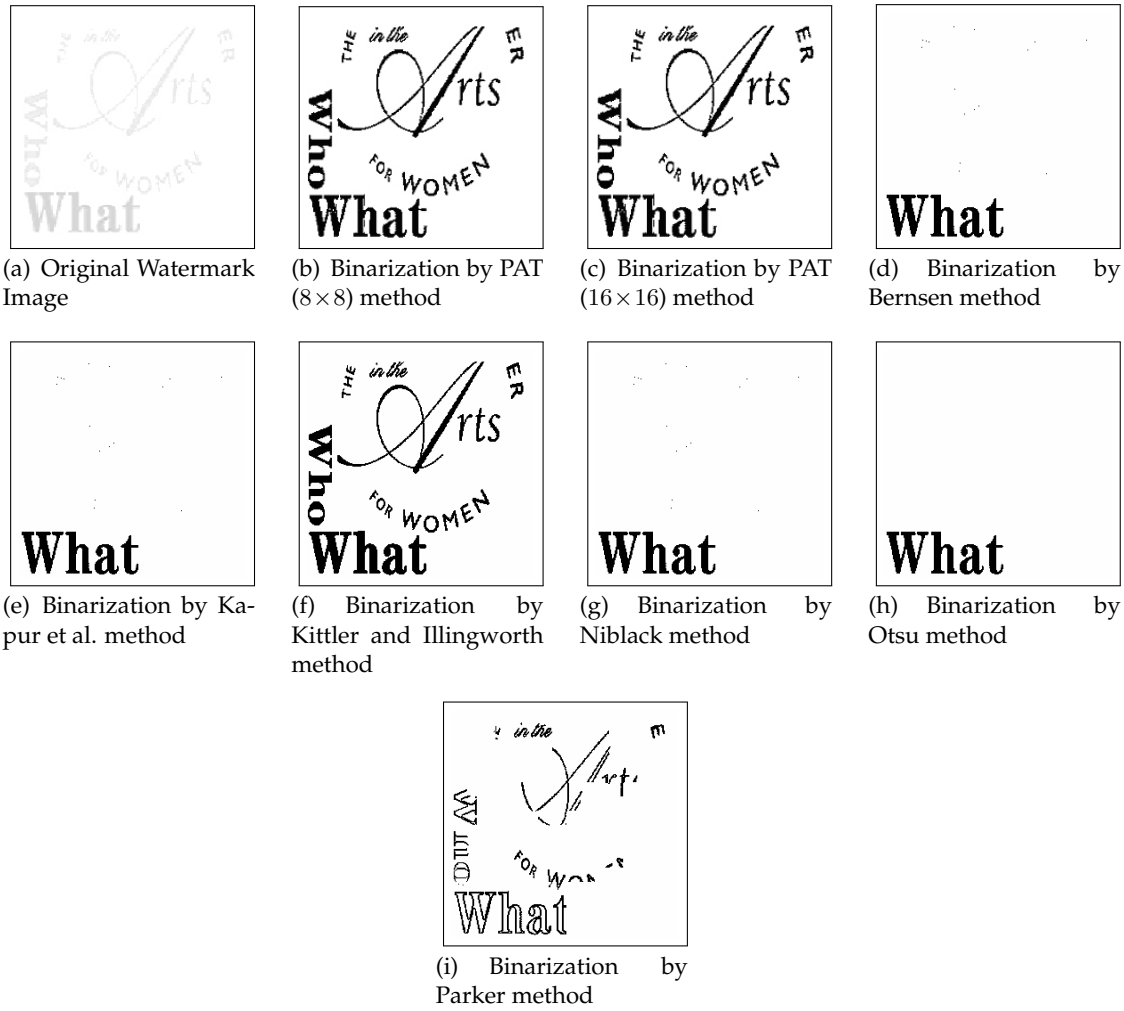
(h) Binarization by Otsu method

(i) Binarization by Parker method

Figure 4.16: Example Results of Banknote Image

The use of a smaller square segment size (8×8) for PAT method yields more details of the patterns but includes additional noise. To remove this additional noise, larger size

segments were used such as 16×16 segments, which yielded more clear patterns with some loss of information.



(a) Original Watermark Image

(b) Binarization by PAT (8×8) method

(c) Binarization by PAT (16×16) method

(d) Binarization by Bernsen method

(e) Binarization by Kapur et al. method

(f) Binarization by Kittler and Illingworth method

(g) Binarization by Niblack method

(h) Binarization by Otsu method

(i) Binarization by Parker method

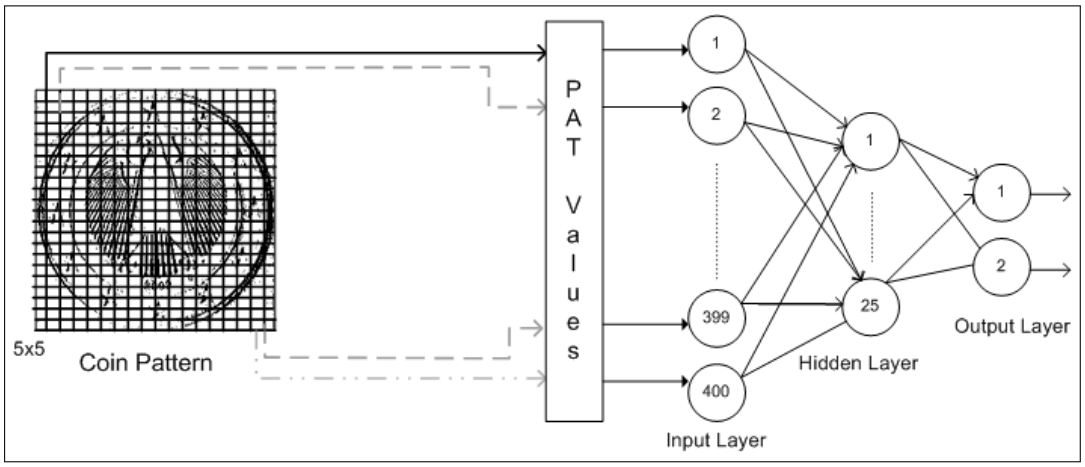
Figure 4.17: Example Results of Watermark Image

The use of PAT method for enhancing unclear patterns produces images with highly visible patterns that can be used in further image processing, or for other identification purposes.

Pattern Recognition Application using PAT

While the efficiency of the PAT in the visualization and the enhancement of unclear patterns is demonstrated, it can also be applied as an image preprocessing phase with a pattern recognition. For example, PAT has been used to enhance coin images as part of

a rotation invariant Intelligent Coin Identification System (ICIS)[4] [67],[68],[69]. The use of the PAT, reduces the computational expense and thus time cost while providing ICIS with sufficient data on for successful recognition. The use of PAT with ICIS is shown in *Figure 4.18*.



(a)

Figure 4.18: Pattern averaging threshold and neural network topology of intelligent system

4.4 Summary

This chapter proposed a novel global Mass-Difference (MD) thresholding method and investigated a local Pattern Averaging Threshold (PAT) method. In addition, performed tests and comparisons under different conditions, were presented.

The next chapter explains performed comparative evaluation of binarization methods which were considered in *Chapter 3* and proposed methods, for document image enhancement. Also document image database and evaluation criteria will be explained.

[4]Various experiments on this subsection were published in 1 - Lecture Notes in Control and Information Sciences, vol. 345, as "ICIS: A Novel Coin Identification System", 2006, 2- WSEAS Transactions on Signal Processing, issue 5, vol. 2, pp. 781 − 786, as "Intelligent Rotation-Invariant Coin Identification System", 2006, 3- Proceedings of the IEEE International Symposium on Intelligent Control (ISIC'06),as "Intelligent Coin Identification System",2006. and 4- Advances in Soft Computing –as a book chapter, Springer, vol.41, pp. 290 − 297, as "Rotated Coin Recognition Using Neural Network", 2007.

CHAPTER 5

COMPARATIVE EVALUATION OF THRESHOLDING METHODS FOR DOCUMENT IMAGE BINARIZATION

5.1 Overview

The efficiency of document image enhancement techniques depends on the efficient separation and classification of background and foreground layers. Thus, the initial purpose of these techniques is the effective preparation and separation of various layers in documents in order to provide sufficient and clear data for recognition systems and human readers. In ideal conditions, optimistic separations of relevant data which belong to foreground layer of texts and characters is required while discarding unnecessary information of background layer. However, document images generally carry various levels of noise which causes pessimistic separation of the layers. Such variety in noise levels creates a challenging task for thresholding methods and may prevent efficient separation and enhancement of document images. The determination of the most appropriate thresholding method for document image binarization is another challenging task because of the varieties of the considered databases and methods in performed comparisons. In this chapter, performed experiments on document image binarization, design of experiments, evaluation criteria and obtained results will be explained in details by performing comparative evaluation of considered methods [1].

[1] An article based on the content of this chapter was submitted to the Journal of Image and Vision Computing, Elsevier Science, in September 2007.

5.2 Recent Comparisons

With the existence of various global and local thresholding methods, deciding upon an superior method for document image binarization is a challenging task; because efficiencies of the existing thresholding methods are usually application-dependent where one method's performance appears superior when using a certain type of document, but the worst in a test on a different type of document. The solution to this problem would be in creating and using a comprehensive multi-applications document image database that accounts for different types of documents, such as historical documents, degraded documents, artificially created words, and handwritten documents.

Several comparisons have been previously performed [1, 2, 3, 4, 5] in order to evaluate existing thresholding methods and deciding upon an superior thresholding method for document binarization in particular.

Trier and Taxt [1] compared fifteen known thresholding methods which included eleven local thresholding methods and four global thresholding methods. The local methods were: Bernsen [34], Chow and Kaneko [26], Eikvil et al. [39], Mardia and Hainsworth [96], Niblack [33], Taxt et al. [38], Yanowitz and Bruckstein [37], Parker [40], White and Rohrer Dynamic thresholding [29], White and Rohrer Integrated Function Algorithm [29], and Trier and Taxt method [97]. The global methods were: Abutaleb [36], Kapur et al. [31], Kittler and Illingworth [22], and Otsu [24]. Yanowitz and Bruckstein [37], and White and Rohrer [29] used post-processing in their methods to remove ghost objects. Similar post-processing was also adopted by Bernsen [34], Eikvil et al. [39], Niblack [33], and Parker [40] in order to compare and evaluate modified versions of the two methods in [34,26]. The evaluation strategy in [1] was based on the visual inspection of binarized images of cable and hydro images, where an expert would visually inspect the images and evaluate the results according to the five criteria which were defined as: the broken line structures, broken symbols, text . . ., etc., blurring of lines, symbols and text, loss of complete objects, and noise in homogenous areas. In their evaluation, Trier and Taxt [1] concluded that thresholding methods with postprocessing produced better results than others, and the superior thresholding methods in local and global groups were modified Niblack method, and Otsu method, respectively. However, in their general evaluation and considering the overall performance of these methods, the modified

Niblack method was considered superior to others while Otsu method was ranked as 13^{th} in the general evaluation. They also concluded that Kittler and Illingworth method had the least successful overall result.

In a different work, Trier and Jain [2] performed a similar comparison using the same methods as in [1] but differing in that; their evaluation was based on the recognition of the characters within binarized images by an Optical Character Recognition (OCR) module. The results of this comparison and evaluation suggested similar ranking of the thresholding methods as found in [1]; where modified Niblack was proposed again as superior to the other methods in overall performance. However, Trier and Jain noticed that the pre-processed Niblack method has a disadvantage with the size of neighborhood, where the size should be small enough to preserve local details and large enough to suppress noise.

A more recent comparison of thresholding methods was performed by Leedham et al. [3]. They compared an improved version of Niblack method, Yanowitz and Bruckstein method, and Solihin and Leedham Quadratic Integral Ratio [51] method with their proposed locally adaptive thresholding methods: Mean-Gradient technique and Background subtraction. Their evaluation was based on recall, precision and accuracy, as suggested in [98], and was performed on 40 images which were: equally distributed historical handwritten images, cheque images, form images, and newspaper images. Consequently, Leedham et al. concluded that there is no single efficient algorithm for all types of images. Sezgin and Sankur [4] performed another comparison of 40 selected thresholding methods using two different image databases: degraded document images and nondestructive testing images. The degraded document images database consisted of 40 created document images with different fonts, sizes and typefaces. In order to obtain degradation of these documents, poor quality photocopied and faxed documents were used, and blur masks and speckle noise were added to these images. The evaluation procedure in [4] was based on five performance criteria which were: misclassification error, edge mismatch, relative foreground area error, modified Hausdroff distance and region non-uniformity. The ranks of these criteria were averaged for each image and used to measure their performance. Sezgin and Sankur noticed that, Kittler and Illingworth Minimum Error thresholding method was superior to all other global and local

thresholding methods in their evaluation using both image databases. Using document image database, Kittler and Illingworth method was followed by Sauvola and Pietikainen method; where the later method did not perform very well when using the nondestructive database.

Recently, He et al. [5] performed a comparison of thresholding methods using historical documents as their database. They used a single global threshold value of 165 for the database to simulate global effect and compared the results to Niblack, and Sauvola and Pietikainen methods. They also modified these methods to choose constant k and segment sizes sw automatically and called their modifications as Adaptive Niblack and Adaptive Sauvola methods. The evaluation was carried out on a set of 4435 images and was performed on only one typed word with dimensions 23×25 pixels on either grey or yellow background with black or red typing. The binarized images were then fed into OCR to determine the recognition rate of characters. He et al. concluded that, Niblack and Adaptive Niblack had slightly better performance than others.

The above described works on comparative evaluation of thresholding methods have attempted to suggest an superior thresholding method that can be efficiently used for document image binarisation. However, the results of these different evaluations suggested different methods as being superior; which is anticipated as the image databases differ from one evaluation to another; where one evaluation uses historical documents, others use created words, or artificially degraded document scans. Another problem is the insufficient number of document images used in some of these evaluations [1][2][4], which affects the significance of the evaluation outcome. In addition, using a large number of images that have similar noise and layer characteristics [5], does not provide an effective evaluation. Moreover, the use of visual inspection as in [1], without any computed analysis, as the only or main criteria for evaluation may not provide a robust evaluation outcome. On the other hand, the use of OCR module with some historical documents is not possible due to old different fonts that can not be recognized by the available OCR modules. Finally, there is a lack of clear categorisation of thresholding methods into adaptive local methods and global methods when performing the evaluations. Such clear categorisation would greatly aid in providing a more objective comparison and in suggesting an overall superior thresholding method or a category-based superior thresh-

olding method. The work presented within this chapter aims to solve these problems.

5.3 Experiment Design

In order to provide an efficient evaluation, large document image database, that includes historical documents, degraded documents, artificially created words, and handwritten within bright, low contrast images, is required. Clean document images can be used for evaluation, however, in the case of historical documents, extreme conditions such as shadows, non-uniform illumination, low contrast, signal dependent noise, smear and smudge may cause pessimistic foreground separation, thus document image database that used in experiments, include these degraded document images in order to perform a more effective evaluation.

The evaluation procedure should also be capable of performing accurate comparison of considered methods, thus, the use of both visual inspection and computed analysis based on two metrics which we derived using the Peak Signal-to-Noise Ratio (PSNR) of the enhanced images are proposed. The following subsections present the document image database and the evaluation procedure.

5.3.1 Document Image Database

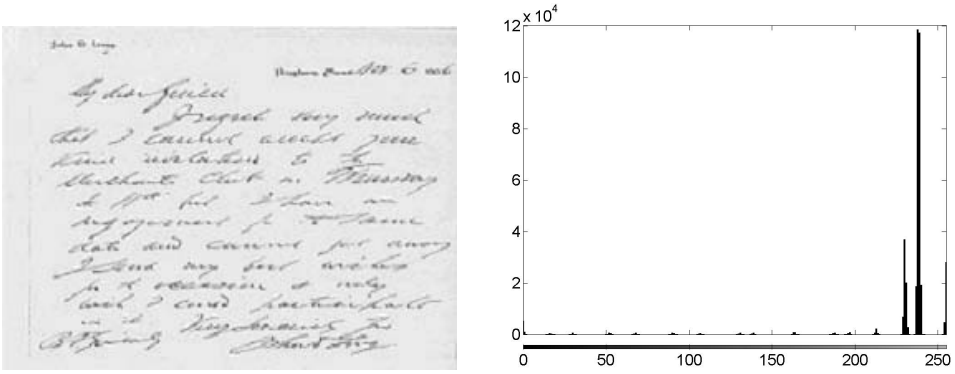
The database consists of 175 historical documents, handwritten documents, and specially created text; which are organized into three sets:

Image Set I

This set comprises 115 scanned historical documents which contain a total of 10291 words with different contrasts and brightness. This set has been classified, according to the images' corresponding grey level histograms [7]; into three groups, namely Bright, Low-Contrast and Dark images. Classification was performed by considering grey level histograms of the images.

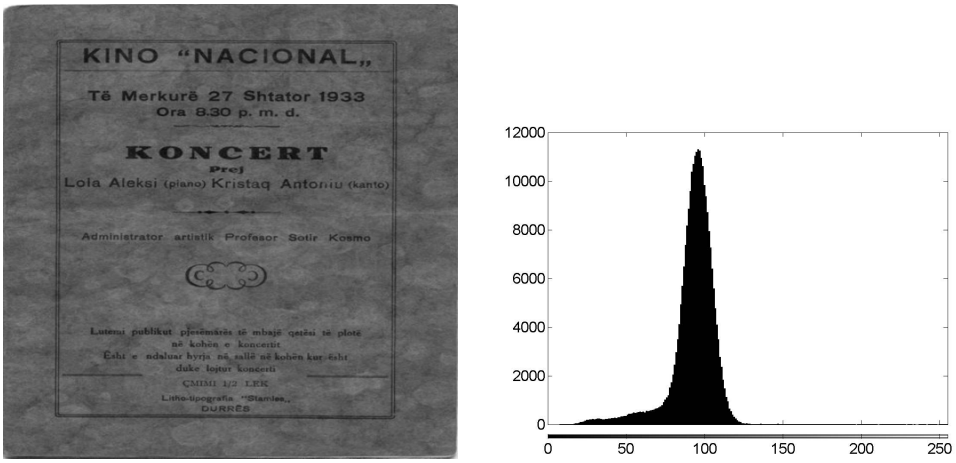
The bright image group consists of 77 degraded images which have different types of noise, shadows and smudge. The low-contrast image group consists of 33 marginally-degraded images which also have different types of noise, shadows and smudge. The

dark image group has five images which contain additional noise. Examples of these historical document images and their corresponding grey level histograms can be seen in *Figure 5.1–Figure 5.3* and other examples of bright and low contrast images of Set I can be seen in *Figure 5.4*.



(a) Example bright image (b) Corresponding gray level histogram

Figure 5.1: Example Bright Image of Set I

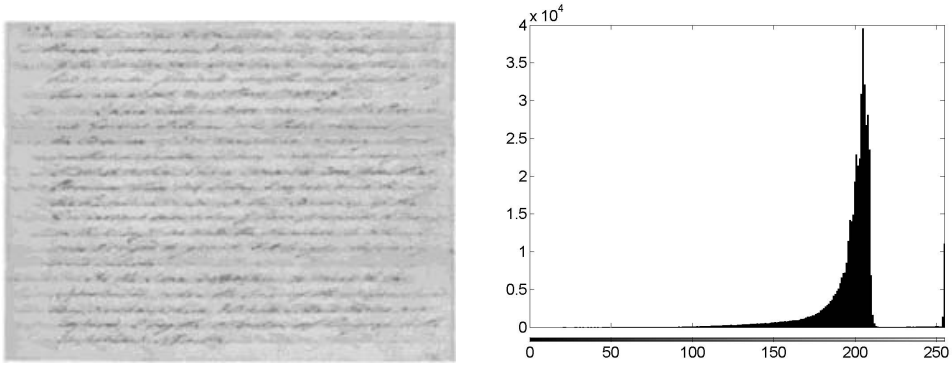


(a) Example dark image (b) Corresponding gray level histogram

Figure 5.2: Example Dark Image of Set I

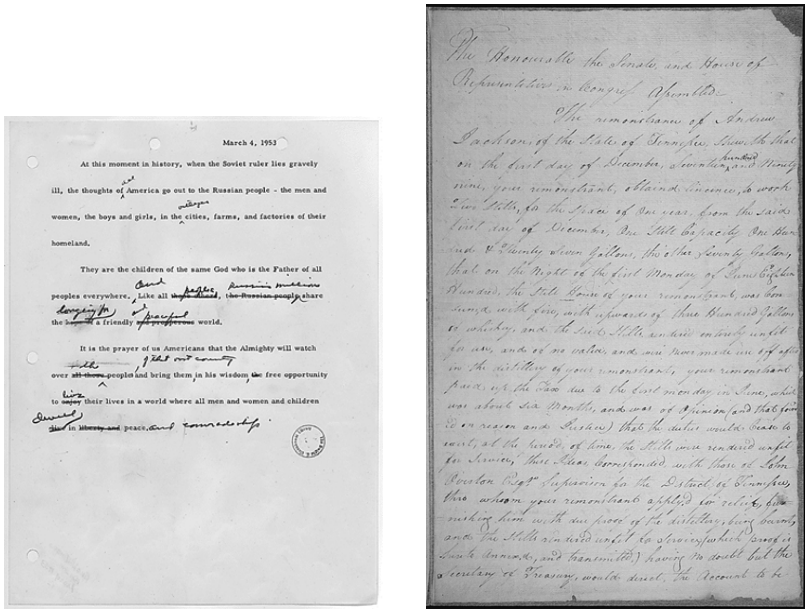
Image Set II

This set comprises six specially created words which contain a total of 72 characters with different backgrounds and different grey scales (non-uniform text). Examples of this set can be seen in *Figure 5.5*.



(a) Example low contrast image (b) Corresponding gray level histogram

Figure 5.3: Example Low Contrast Image of Set I



(a) Example bright image (b) Example low contrast image

Figure 5.4: Example Images of Set I

Image Set III

This set comprises 54 scanned handwritten documents. The documents were prepared by scanning the handwriting of nine different persons, using three different writing tools (pen, pencil and board-marker) on two different paper types (white paper and yellow envelope paper). Examples of the handwritten documents can be seen in *Figure 5.6*.

In summary, the above three sets, which contain multi-application document images with different levels of noise and contrast, non-uniform illumination (background), signal-dependent noise, smears and non-uniform foreground (text) will be used for the

(a) non-uniform text on uniform background (b) non-uniform text on non-uniform background

Figure 5.5: Example Images of Set II

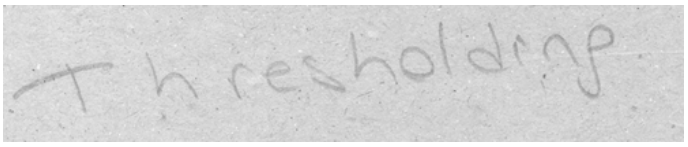
implementation of the considered thresholding methods.

5.3.2 Evaluation Procedure

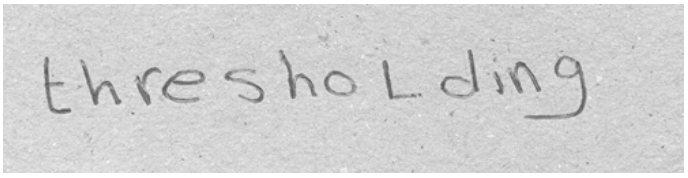
In order to evaluate the obtained results when applying the considered methods using the proposed database, visual inspection of the enhanced documents was used, in addition to three metrics which were derived using the Peak Signal-to-Noise Ratio (PSNR) of the enhanced images.

Visual Inspection

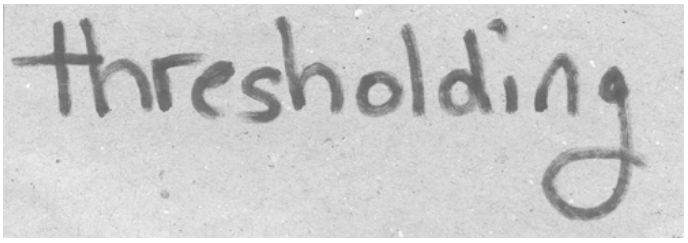
Visual inspection of all enhanced documents in the three sets was performed by 15 independent human analyzers, who were asked to consider the clarity and readability of the words within Set I documents, noise occurrence and continuity of characters to determine clear characters within Set II documents, and clearly recognized readable characters within handwritten words in Set III documents. The general results of visual inspection were categorized as: recognized or unrecognized words for Set I, clear or unclear characters for Set II, and recognized or unrecognized characters for Set III. We believe that this method of evaluation is necessary as one of the objectives is to provide clearly enhanced document images that can improve human readability of degraded documents. *Figure 5.7* shows the examples of readability evaluation of visual inspection procedure in



(a) pencil handwriting on yellow envelope paper



(b) pen handwriting on yellow envelope paper



(c) white board marker handwriting on yellow envelope paper

Figure 5.6: Example Images of Set III

all test images.

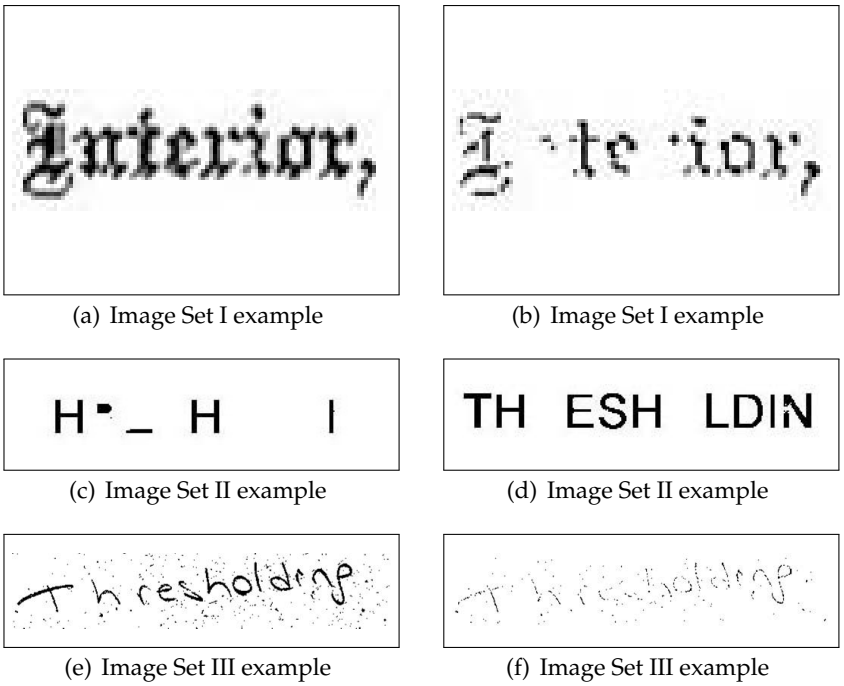
Figure 5.7 (a) shows the recognized word in image set I by all reviewers, (b) shows the unrecognized word in image set I by all reviewers (c) shows recognition of 3 characters out of 12 in image set II, (d) shows recognition of 9 characters out of 12 in image set II, (e) shows recognition of all characters of image in image set III and (f) shows the recognition of 5 characters out of 12 in image set III.

Computed Noise Analysis

Computed noise analysis [2] of all enhanced images in Set I is also applied as the second part in the evaluation procedure. The application of this noise analysis was considered unnecessary with images from Set II and III, as satisfactory and clear evaluation results can be easily obtained by only using visual inspection with these two sets, which is not the case with Set I.

Using PSNR of enhanced images three metric parameters for each of the 14 binarization methods used in experiments was derived. These parameters are the Average PSNR

[2] An article based on the content of this subsection is accepted to be published in the proceedings of the third international IEEE conference on signal image technology and internet based systems in December 2007.



(a) Image Set I example (b) Image Set I example

(c) Image Set II example (d) Image Set II example

(e) Image Set III example (f) Image Set III example

Figure 5.7: Readability evaluation of visual inspection procedure

Accuracy Rate (APAR), the Average PSNR Deviation (APD) of binarized images, and the Combined Performance Rate (CPR) of a thresholding method. PSNR (in dB) is defined as [7]:

$$PSNR = 10\, log_{10}\left(\frac{255^2}{MSE}\right) \tag{5.1}$$

MSE is the mean squared error as defined as [7]:

$$MSE = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{i,j} - \hat{X}_{i,j})^2 \tag{5.2}$$

where $(M \times N)$ is the size of the image, $Y_{i,j}$ and $\hat{X}_{i,j}$ represent the pixel values at location (i, j) of original and enhanced image, respectively.

The average PSNR accuracy rate (APAR) for a particular method is calculated by considering the maximum PSNR value obtained using the 14 thresholding methods, and the PSNR value obtained using only that particular method for a test image. The higher the APAR value is, the more efficient the thresholding method is. The average PSNR

deviation (APD) for a particular method is calculated by taking the difference between the maximum PSNR value obtained using the 14 thresholding methods, and the PSNR value obtained using only that particular method for a test image; the difference is then divided by the total number of test images. The lower the APD value is, the more efficient the thresholding method is. APAR and APD are defined as follows:

$$APAR_m \;=\; \left(\sum_{i=1}^{x} \left((PSNR_{mi} \times 100) \,/\, max(PSNR_i) \right) \right) \,/\, x \tag{5.3}$$

$$APD_m \;=\; \left(\sum_{i=1}^{x} \left((max(PSNR_i)) - PSNR_{mi}) \right) \right) \,/\, x \tag{5.4}$$

where $APAR_m$ is the average PSNR accuracy rate for method m, APD_m is the average PSNR deviation for method m, $PSNR_mi$ denotes the obtained PSNR value of enhanced image i using enhancement method m, $max(PSNR_i)$ denotes the maximum PSNR value of the enhanced image i obtained using the fourteen methods, and x is the total number of test images.

The Combined Performance Rate (CPR) is the third proposed parameter. CPR indicates the final performance of a thresholding method, where the higher the CPR value is the more superior the method is. CPR is defined as:

$$CPR_m \;=\; RW_m \times (APAR_m \,/\, 100) \tag{5.5}$$

where CPR_m is the Combined Performance Rate for method m, RW is the number of recognized words in Set I document images using visual inspection, and $APAR$ is the average PSNR accuracy rate for all document images in Set I.

Considering only PSNR values of the binarized enhanced images when evaluating the result is not always effective when comparing various thresholding methods and using such a diverse document image database. This is because a particular method may produce high PSNR values for a significant number of images, while producing low

PSNR values for the rest of the images, thus making it difficult to determine stability and efficiency of the methods. Therefore, APAR, APD, and CPR metrics were derived, which provide a uniform indication of the efficiency of the compared methods.

5.4 Results and Comparisons

Experiments involved software implementation of the fourteen methods to binarize all documents in database. The C-programming language was used in implementation. The methods were: Otsu, Kittler and Illingworth, Yanni and Horne, Ramesh et al., Mass-Difference (MD), Kapur et al., and Albuquerque et al. entropy thresholding as global methods, and Niblack, Sauvola et al., Mean-Gradient, Pattern Average Thresholding (PAT), Adaptive Logical Thresholding (ALT), Water Flow Model (WFM), and Bernsen thresholding as local methods.

Throughout the experiments, the parameters and segment sizes of the locally adaptive methods were chosen to provide the superior performance of the methods; as suggested by previous works and comparisons. In their evaluation, Trier and Jain [2] noticed that 15×15 segments with $k = -0.2$ produces superior results and this was considered by Sezgin and Sankur [4] in their evaluation. Also Trier and Jain suggest 15×15 segment size for Bernsen method with $l = 15$. Sauvola method is the improvement of Niblack method by adapting standard deviation thus 15×15 segment size with suggested values $k = 0.5$ and $R = 40$ were used [4]. While comparing some methods and proposing Mean-Gradient Method, Leedham et al. [3] notice that 15×15 segment size is suitable for their method with $k = -1.5$ and $R = 40$. After some experiments, segment sizes for PAT, ALT and WFM were decided as 15×15, 8×8 and 5×5 respectively with $w = 17$ in WFM.

Table 5.1 shows the segment sizes and chosen parameters for the implemented local thresholding methods.

Table 5.1: Segment Sizes and Parameters for Locally Adaptive Methods

Parameters	Niblack	Sauvola et al.	Mean-Gradient	PAT	ALT	WFM	Bernsen
Segment Size	15x15	12x12	15x15	15x15	8x8	5x5	15x15
k [2]	-1	0.5	-1.5	-	-	-	15
R [2]	-	128	40	-	-	-	-
w [2]	-	-	-	-	-	17	-

[3] k and R values are constants and w is the amount of water in WFM.

5.4.1 Image Set I Experiments

The first evaluation was performed on Set I images by visual inspection and noise analysis, with different results obtained for the different groups (bright, low-contrast and dark images) in this set.

Bright Images Group

After visual inspection of Bright images group (see *Table 5.2–Table 5.4*), MD thresholding method achieved higher recognition rate within global methods and WFM achieved higher recognition rate within local methods. Considering overall visual inspection results for bright image group shows that global methods produce better results than local methods; with MD thresholding method being superior to the other methods.

However, when applying computed analysis using two PSNR-derived metrics (see *Table 5.5–Table 5.7*), again MD achieved higher APAR and APD results within global methods and WFM achieved higher APAR and APD results within local methods. Considering overall computed noise analysis for bright image group shows WFM achieved higher rank and followed by MD and ALT.

Table 5.2: Visual Inspection Results of Global Methods for $Bright\ Images$ Group of Set I

Thresholding Method	Category[4]	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Otsu	G	3	8300	7533	767	90.75%
Kit. and Ill.	G	5	8300	4691	3609	56.51%
Yanni and Horne	G	6	8300	2932	5368	35.32%
Ramesh et al.	G	4	8300	7397	903	89.12%
MD	G	1	8300	8097	203	97.55%
Kapur et al.	G	2	8300	7939	361	95.65%
Albuquerque et al	G	2	8300	7939	361	95.65%

Table 5.3: Visual Inspection Results of Local Methods for $Bright\ Images$ Group of Set I

Thresholding Method	Category [4]	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Niblack	L	2	8300	7352	948	88.57%
Sauvola et al.	L	3	8300	7239	1061	87.21%
Mean-Gradient	L	6	8300	6518	1782	78.53%
PAT	L	5	8300	6833	1467	82.32%
ALT	L	4	8300	7194	1106	86.67%
WFM	L	1	8300	7917	383	95.38%
Bernsen	L	7	8300	3608	4692	43.43%

[4] G is the global and L is the local.

Table 5.4: Overall Visual Inspection Results for $Bright\ Images$ Group of Set I

Thresholding Method	Category	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Otsu	G	5	8300	7533	767	90.75%
Kit. and Ill.	G	12	8300	4691	3609	56.51%
Yanni and Horne	G	14	8300	2932	5368	35.32%
Ramesh et al.	G	6	8300	7397	903	89.12%
MD	G	1	8300	8097	203	97.55%
Kapur et al.	G	2	8300	7939	361	95.65%
Albuquerque et al	G	3	8300	7939	361	95.65%
Niblack	L	7	8300	7352	948	88.57%
Sauvola et al.	L	8	8300	7239	1061	87.21%
Mean-Gradient	L	11	8300	6518	1782	78.53%
PAT	L	10	8300	6833	1467	82.32%
ALT	L	9	8300	7194	1106	86.67%
WFM	L	4	8300	7917	383	95.38%
Bernsen	L	13	8300	3608	4692	43.43%

Low Contrast Images Group

In Low-Contrast images group of Set I, Kittler and Illingworth method achieved highest rank within global methods and Sauvola et al. method achieved highest rank within local methods using the visual inspection criterion (see *Table 5.8*,*Table 5.9* and *Table 5.10*). However, after performing noise analysis, Kittler and Illingworth, and Sauvola et al. methods seemed to add additional noise to binarized image, and MD and ALT methods achieved higher APAR and APD results in their groups. By considering overall results, it was shown that ALT is the superior method for this group which is followed by MD and WFM methods.

Table 5.5: APD and APAR results of global methods for all Set I groups

Thresholding Method	Category	Image Set I								
		Bright			Low Contrast			Dark		
		APD	APAR	R[5]	APD	APAR	R[5]	APD	APAR	R[5]
Otsu	G	2.952	86.561	3	2.769	81.189	3	4.254	71.16	1
Kittler and Illingworth	G	8.238	65.369	6	3.904	75.767	5	4.678	67.79	2
Yanni and Horne	G	6.699	69.500	5	6.144	62.723	6	-	-	-
Ramesh et al.	G	8.338	61.828	7	9.456	47.783	7	-	-	-
MD	G	1.029	95.744	1	1.106	89.176	1	4.705	67.62	3
Kapur et al.	G	1.801	91.834	2	2.719	79.268	2	8.714	41.39	4
Albuquerque et al.	G	4.967	76.438	4	3.065	77.585	4	-	-	-

Dark Images Group

In Dark images group of Set I, MD achieved highest rank within global methods by using visual inspection, and Niblack, ALT and WFM methods shared a similar rank producing

[5]R is the rank.

Table 5.6: APD and APAR results of local methods for all Set I groups

Thresholding Method	Category	Image Set I								
		Bright			Low Contrast			Dark		
		APD	APAR	R	APD	APAR	R	APD	APAR	R
Niblack	L	3.680	84.664	4	17.56	85.878	3	4.369	70.09	3
Sauvola et al.	L	9.539	59.093	7	6.327	65.064	6	0.241	98.08	1
Mean-Gradient	L	2.786	87.639	3	23.22	82.678	4	-	-	-
PAT	L	8.151	65.177	6	4.416	73.464	5	3.420	77.14	2
ALT	L	1.101	95.319	2	0.865	90.144	1	4.725	67.49	5
WFM	L	0.365	98.519	1	1.082	88.799	2	4.620	68.28	4
Bernsen	L	7.333	68.003	5	6.595	64.142	7	7.629	50	6

Table 5.7: Overall APD and APAR results for all Set I groups

Thresholding Method	Category	Image Set I								
		Bright			Low Contrast			Dark		
		APD	APAR	R	APD	APAR	R	APD	APAR	R
Otsu	G	2.952	86.561	6	2.769	81.189	7	4.254	71.16	3
Kittler and Illingworth	G	8.238	65.369	12	3.904	75.767	9	4.678	67.79	6
Yanni and Horne	G	6.699	69.500	9	6.144	62.723	11	-	-	-
Ramesh et al.	G	8.338	61.828	13	9.456	47.783	14	-	-	-
MD	G	1.029	95.744	2	1.106	89.176	3	4.705	67.62	7
Kapur et al.	G	1.801	91.834	4	2.719	79.268	6	8.714	41.39	10
Albuquerque et al.	G	4.967	76.438	8	3.065	77.585	8	-	-	-
Niblack	L	3.680	84.664	7	17.56	85.878	4	4.369	70.09	4
Sauvola et al.	L	9.539	59.093	14	6.327	65.064	12	0.241	98.08	1
Mean-Gradient	L	2.786	87.639	5	23.22	82.678	5	-	-	-
PAT	L	8.151	65.177	11	4.416	73.464	10	3.420	77.14	2
ALT	L	1.101	95.319	3	0.865	90.144	1	4.725	67.49	8
WFM	L	0.365	98.519	1	1.082	88.799	2	4.620	68.28	5
Bernsen	L	7.333	68.003	10	6.595	64.142	13	7.629	50	9

superior results within local methods using the visual inspection criterion.

By using computed noise analysis, it was seen that Otsu and Sauvola et al. method achieved highest rank in their groups. However, considering overall results showed that Sauvola et al. method is superior in noise removal and MD, Niblack, ALT and WFM methods are superior in visual inspection (see *Table 5.11– Table 5.13*).

Table 5.8: Visual Inspection Results of Global Methods for $Low\ Contrast$ Group of Set I

Thresholding Method	Category	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Otsu	G	4	1901	1216	685	63.96%
Kit. and Ill.	G	1	1901	1431	470	75.27%
Yanni and Horne	G	7	1901	95	1806	4.99%
Ramesh et al.	G	6	1901	114	1787	5.99%
MD	G	3	1901	1235	666	64.96%
Kapur et al.	G	2	1901	1388	513	73.01%
Albuquerque et al	G	5	1901	154	1747	8.10%

Table 5.9: Visual Inspection Results of Local Methods for *Low Contrast* Group of Set I

Thresholding Method	Category	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Niblack	L	6	1901	1102	799	88.57%
Sauvola et al.	L	1	1901	1510	391	87.21%
Mean-Gradient	L	7	1901	912	989	78.53%
PAT	L	3	1901	1349	552	82.32%
ALT	L	5	1901	1140	761	86.67%
WFM	L	2	1901	1359	542	95.38%
Bernsen	L	4	1901	1235	666	43.43%

Table 5.10: Overall Visual Inspection Results for *Low Contrast* Group of Set I

Thresholding Method	Category	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Otsu	G	7	1901	1216	685	63.96%
Kit. and Ill.	G	2	1901	1431	470	75.27%
Yanni and Horne	G	13	1901	95	1806	4.99%
Ramesh et al.	G	12	1901	114	1787	5.99%
MD	G	6	1901	1235	666	64.96%
Kapur et al.	G	3	1901	1388	513	73.01%
Albuquerque et al	G	11	1901	154	1747	8.10%
Niblack	L	9	1901	1102	799	88.57%
Sauvola et al.	L	1	1901	1510	391	87.21%
Mean-Gradient	L	10	1901	912	989	78.53%
PAT	L	5	1901	1349	552	82.32%
ALT	L	8	1901	1140	761	86.67%
WFM	L	4	1901	1359	542	95.38%
Bernsen	L	6	1901	1235	666	43.43%

Table 5.11: Visual Inspection Results of Global Methods for *Dark Images* Group of Set I

Thresholding Method	Category	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Otsu	G	3	90	55	35	61.11%
Kit. and Ill.	G	2	90	85	5	94.44%
Yanni and Horne	G	-	90	0	90	0%
Ramesh et al.	G	-	90	0	90	0%
MD	G	1	90	88	2	97.77%
Kapur et al.	G	3	90	55	35	61.11%
Albuquerque et al	G	0	90	0	80	0%

Table 5.12: Visual Inspection Results of Local Methods for *Dark Images* Group of Set I

Thresholding Method	Category	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Niblack	L	1	90	88	2	97.77%
Sauvola et al.	L	2	90	75	15	83.33%
Mean-Gradient	L	-	90	0	90	0%
PAT	L	3	90	58	32	64.44%
ALT	L	1	90	88	2	97.77%
WFM	L	1	90	88	2	97.77%
Bernsen	L	4	90	10	80	11.11%

Table 5.13: Overall Visual Inspection Results for $Dark\ Images$ Group of Set I

Thresholding Method	Category	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Otsu	G	3	90	55	35	61.11%
Kit. and Ill.	G	2	90	85	5	94.44%
Yanni and Horne	G	-	90	0	90	0%
Ramesh et al.	G	-	90	0	90	0%
MD	G	1	90	88	2	97.77%
Kapur et al.	G	3	90	55	35	61.11%
Albuquerque et al	G	0	90	0	80	0%
Niblack	L	1	90	88	2	97.77%
Sauvola et al.	L	2	90	75	15	83.33%
Mean-Gradient	L	-	90	0	90	0%
PAT	L	3	90	58	32	64.44%
ALT	L	1	90	88	2	97.77%
WFM	L	1	90	88	2	97.77%
Bernsen	L	4	90	10	80	11.11%

Combined Group Results

Set I, by considering all results, it is clear that, MD is superior in visual inspection (see and WFM is superior in noise analysis (see *Table 5.14–Table 5.16*).

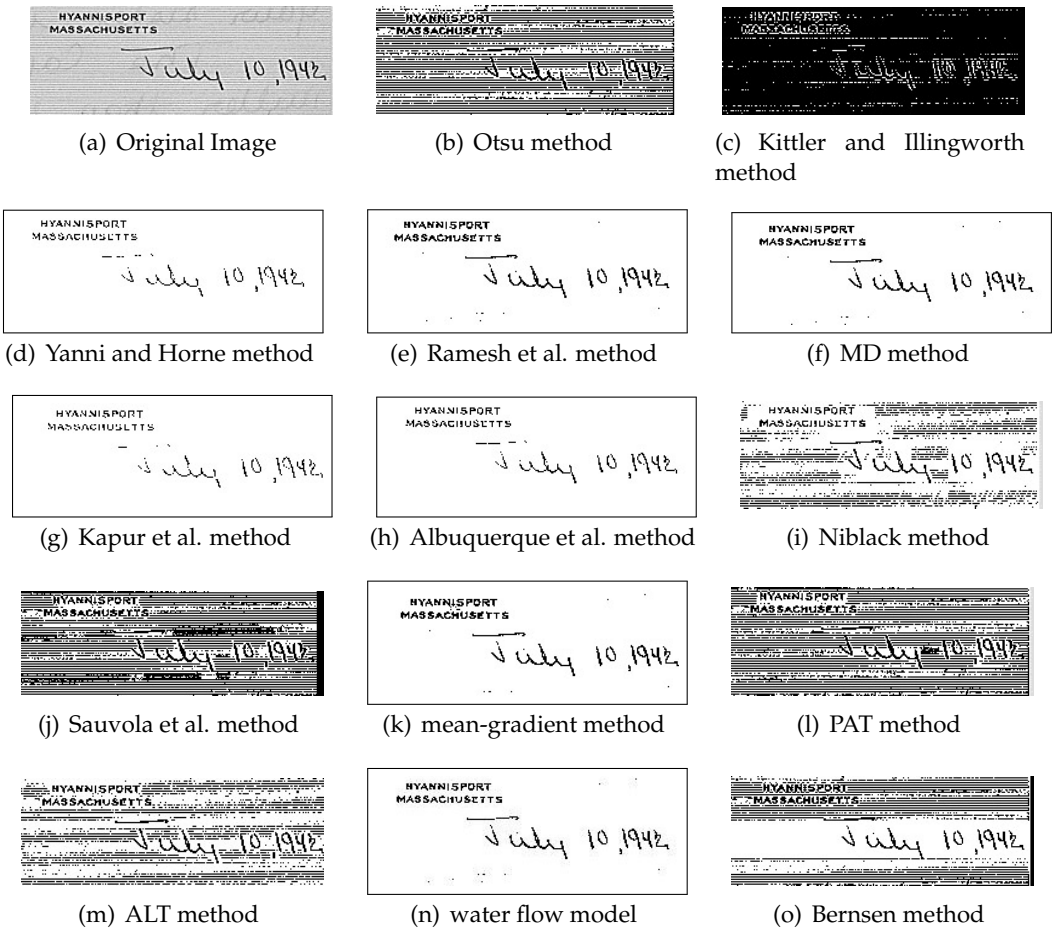
This results were supported by Final Performance Evaluation, which most successful results were achieved by WFM and MD respectively (see *Table 5.17–Table 5.22*). Examples of obtained resuts for Set I can be seen in *Figure 5.8*, *Figure 5.9*, and *Figure 5.10*.

MD was designed to extract dominant pixels within document which are categorized as 'text/foreground' and this was achieved by using maximum brightness point as the background limit of document and mean value as 'fuzzy layer'. Difference of maximum point to mean value shifts 'fuzzy layer' to left on image histogram to detect details of foreground while discarding relatively deviated pixels within image.

Water flow model simulates water which falls down to lower regions of 3D terrain of images. This yields effective extraction of dominant pixel values which are 'text/foreground' of image and unnecessary information eliminated. Considering global and local properties of images makes difference between WFM and other local methods such as Niblack, Sauvola and Bernsen which considers local information within segments. Thus, behavior of WFM mostly acts as global methods which are more effective for noise removal.

These simple and effective properties of MD and WFM, provides efficient achievement of desired goals in Set I document images binarization. MD achieved superior recognition of images by visual inspection and WFM achieved most successful clearance

of noises during binarization process. Thus, relatively similar results of WFM and MD make them superior for binarization of documents either in bright, low contrast or dark images.



(a) Original Image (b) Otsu method (c) Kittler and Illingworth method

(d) Yanni and Horne method (e) Ramesh et al. method (f) MD method

(g) Kapur et al. method (h) Albuquerque et al. method (i) Niblack method

(j) Sauvola et al. method (k) mean-gradient method (l) PAT method

(m) ALT method (n) water flow model (o) Bernsen method

Figure 5.8: Example result of low contrast image of Set I

Table 5.14: Overall Visual Inspection Results of Global Methods for Set I

Thresholding Method	Category	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Otsu	G	3	10291	8504	1787	82.63%
Kit. and Ill.	G	6	10291	6207	4084	60.31%
Yanni and Horne	G	7	10291	3027	7264	29.41%
Ramesh et al.	G	5	10291	7511	2780	72.98%
MD	G	1	10291	9420	871	91.53%
Kapur et al.	G	2	10291	9382	909	91.16%
Albuquerque et al	G	4	10291	8093	2198	78.64%

Table 5.15: Visual Inspection Results of Local Methods for Set I

Thresholding Method	Category	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Niblack	L	4	10291	8542	1749	83.00%
Sauvola et al.	L	2	10291	8824	1467	85.74%
Mean-Gradient	L	6	10291	7507	2784	72.94%
PAT	L	3	10291	8601	1690	83.57%
ALT	L	5	10291	8422	1869	81.83%
WFM	L	1	10291	9364	927	90.99%
Bernsen	L	7	10291	4853	5438	47.15%

Table 5.16: Overall Visual Inspection Results for Set I

Thresholding Method	Category	Rank	Total Words	Recognized Words	Unrecognized Words	Recognition Rate
Otsu	G	7	10291	8504	1787	82.63%
Kit. and Ill.	G	12	10291	6207	4084	60.31%
Yanni and Horne	G	14	10291	3027	7264	29.41%
Ramesh et al.	G	10	10291	7511	2780	72.98%
MD	G	1	10291	9420	871	91.53%
Kapur et al.	G	2	10291	9382	909	91.16%
Albuquerque et al	G	9	10291	8093	2198	78.64%
Niblack	L	6	10291	8542	1749	83.00%
Sauvola et al.	L	4	10291	8824	1467	85.74%
Mean-Gradient	L	11	10291	7507	2784	72.94%
PAT	L	5	10291	8601	1690	83.57%
ALT	L	8	10291	8422	1869	81.83%
WFM	L	3	10291	9364	927	90.99%
Bernsen	L	13	10291	4853	5438	47.15%

Table 5.17: General APD and $APAR$ Results of Global Methods for All Groups in Set I

Thresholding Method	Category	APD	$APAR$	Rank
Otsu	G	2.943863	86.37631	3
Kit. and Ill.	G	6.922652	70.1334	5
Yanni and Horne	G	6.821943	67.14557	6
Ramesh et al.	G	8.968075	57.24856	7
MD	G	1.120711	95.0755	1
Kapur et al.	G	2.198561	88.99776	2
Albuquerque et al.	G	4.572706	77.16784	4

Table 5.18: General APD and $APAR$ Results of Local Methods for All Groups in Set I

Thresholding Method	Category	APD	$APAR$	Rank
Niblack	L	3.148074	86.42465	7
Sauvola et al.	L	8.489479	62.98983	13
Mean-Gradient	L	2.865630	86.30933	3
PAT	L	7.019589	69.32447	5
ALT	L	1.098199	95.07637	2
WFM	L	0.653009	96.9005	1
Bernsen	L	7.173098	68.03938	6

Table 5.19: Overall APD and $APAR$ Results for All Groups in Set I

Thresholding Method	Category	APD	$APAR$	Rank
Otsu	G	2.943863	86.37631	6
Kit. and Ill.	G	6.922652	70.1334	9
Yanni and Horne	G	6.821943	67.14557	10
Ramesh et al.	G	8.968075	57.24856	14
MD	G	1.120711	95.0755	3
Kapur et al.	G	2.198561	88.99776	4
Albuquerque et al.	G	4.572706	77.16784	8
Niblack	L	3.148074	86.42465	7
Sauvola et al.	L	8.489479	62.98983	13
Mean-Gradient	L	2.865630	86.30933	5
PAT	L	7.019589	69.32447	11
ALT	L	1.098199	95.07637	2
WFM	L	0.653009	96.9005	1
Bernsen	L	7.173098	68.03938	12

Table 5.20: Final Performance Results of Global Methods for Set I

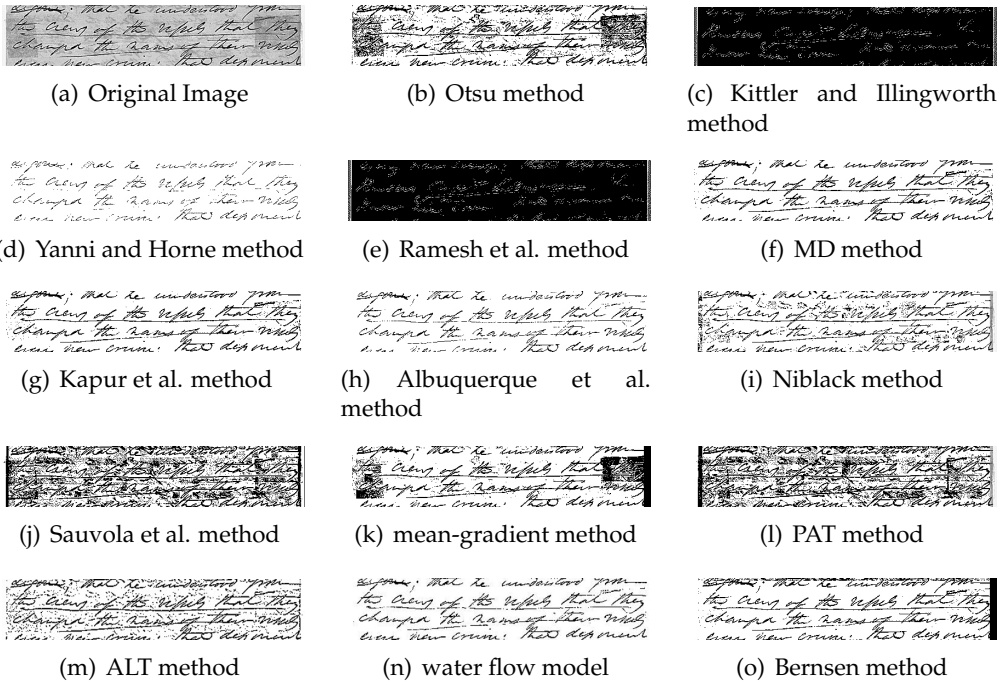
Thresholding Method	Category	CPR	Rank
Otsu	G	7344	3
Kit. and Ill.	G	4352	6
Yanni and Horne	G	4877	5
Ramesh et al.	G	4299	7
MD	G	8955	1
Kapur et al.	G	8349	2
Albuquerque et al.	G	6244	4

Table 5.21: Final Performance Results of Local Methods for Set I

Thresholding Method	Category	CPR	Rank
Niblack	L	7381	3
Sauvola et al.	L	5557	6
Mean-Gradient	L	6478	4
PAT	L	5962	5
ALT	L	8006	2
WFM	L	9072	1
Bernsen	L	3301	7

Table 5.22: Final Performance Results of All Methods for Set I

Thresholding Method	Category	CPR	Rank
Otsu	G	7344	6
Kit. and Ill.	G	4352	12
Yanni and Horne	G	4877	11
Ramesh et al.	G	4299	13
MD	G	8955	2
Kapur et al.	G	8349	3
Albuquerque et al	G	6244	8
Niblack	L	7381	5
Sauvola et al.	L	5557	10
Mean-Gradient	L	6478	7
PAT	L	5962	9
ALT	L	8006	4
WFM	L	9072	1
Bernsen	L	3301	14



(a) Original Image (b) Otsu method (c) Kittler and Illingworth method

(d) Yanni and Horne method (e) Ramesh et al. method (f) MD method

(g) Kapur et al. method (h) Albuquerque et al. method (i) Niblack method

(j) Sauvola et al. method (k) mean-gradient method (l) PAT method

(m) ALT method (n) water flow model (o) Bernsen method

Figure 5.9: Partial result of bright image of Set I

5.4.2 Image Set II Experiments

In Set 2, due to the uniform illumination of the background within the images in this set, global methods generally causes more loss of information than local ones.

Visual inspection and recognition of characters within binarized images showed (see *Table 5.23–Table 5.25*) that Sauvola method is the superior for created text documents and followed by PAT, ALT, Niblack and WFM methods. Examples of obtained results for Set II can be seen in *Figure 5.11*.

The aim of Set II was the performance observation of binarization methods in extreme non-uniform conditions for both background and characters. Although none of the methods could achieve more than 60% of cleared characters, it was seen that, local methods produced better results than global methods.

Sauvola et al. method was the improvement of Niblack method by adapting standard deviation to lower the threshold value. This property provides most successful binarization of specially created extreme non-uniform texts and background by lowering threshold value to detect more characters within documents. However, this property of Sauvola et al. method causes extreme noise addition to Set I images and to decrease the recognition of texts within document by visual inspection.

Also PAT, ALT and Niblack methods performed better results than other methods by considering local information of segments but they could not achieve the recognition rate as Sauvola method.

Table 5.23: Overall Visual Inspection Results of Global Methods for Set II

Thresholding Method	Category	Rank	Total Characters	Clear Characters	Unclear Characters	Recognition Rate
Otsu	G	3	72	13	59	18.05%
Kit. and Ill.	G	4	72	10	62	13.88%
Yanni and Horne	G	6	72	6	66	8.33%
Ramesh et al.	G	5	72	9	63	12.50%
MD	G	2	72	24	48	33.33%
Kapur et al.	G	1	72	28	44	38.88%
Albuquerque et al	G	5	72	9	63	12.50%

Table 5.24: Overall Visual Inspection Results of Local Methods for Set II

Thresholding Method	Category	Rank	Total Characters	Clear Characters	Unclear Characters	Recognition Rate
Niblack	L	3	72	36	36	50.00%
Sauvola et al.	L	1	72	43	29	59.72%
Mean-Gradient	L	5	72	22	50	30.55%
PAT	L	2	72	40	32	55.55%
ALT	L	2	72	40	32	55.55%
WFM	L	4	72	29	43	40.27%
Bernsen	L	6	72	12	60	16.66%

Table 5.25: Overall Visual Inspection Results for Set II

Thresholding Method	Category	Rank	Total Characters	Clear Characters	Unclear Characters	Recognition Rate
Otsu	G	8	72	13	59	18.05%
Kit. and Ill.	G	10	72	10	62	13.88%
Yanni and Horne	G	12	72	6	66	8.33%
Ramesh et al.	G	11	72	9	63	12.50%
MD	G	6	72	24	48	33.33%
Kapur et al.	G	5	72	28	44	38.88%
Albuquerque et al	G	11	72	9	63	12.50%
Niblack	L	3	72	36	36	50.00%
Sauvola et al.	L	1	72	43	29	59.72%
Mean-Gradient	L	7	72	22	50	30.55%
PAT	L	2	72	40	32	55.55%
ALT	L	2	72	40	32	55.55%
WFM	L	4	72	29	43	40.27%
Bernsen	L	9	72	12	60	16.66%

5.4.3 Image Set III Experiments

Set III consists of 6 different competitive image groups which were Pencil on White Paper (WP), Pen on WP, White Board Marker (WBM) on WP, Pencil on Yellow Envelope Paper (YP), Pen on YP, and WBM on YP. White Paper is used to performance evaluation

of thresholding methods on uniform background by using different writing materials as pen, pencil and white board marker which causes different illuminations and stroke widths of characters. Yellow envelope paper is used to performance evaluation of thresholding methods on non-uniform and noisy background which yellow envelope papers consists.

Table 5.26: Visual inspection results of global methods for Set III

Thresholding Method	Category	Recognized Characters								
		Pen			Pencil			WB Marker		
		WP	YP	Av.	WP	YP	Av.	WP	YP	Av.
Otsu	G	107	107	107	107	103	105	108	107	107.5
Kittler and Illingworth	G	108	83	95.5	108	47	77.5	108	108	108
Yanni and Horne	G	22	15	18.5	55	5	30	50	65	57.5
Ramesh et al.	G	26	22	24	49	6	27.5	66	58	62
MD	G	108	106	107	108	59	83.5	108	108	108
Kapur et al.	G	95	85	90	40	32	36	10	5	7.5
Albuquerque et al.	G	11	26	18.5	12	8	10	5	2	3.5

Table 5.27: Visual inspection results of local methods for Set III

Thresholding Method	Category	Recognized Characters								
		Pen			Pencil			WB Marker		
		WP	YP	Av.	WP	YP	Av.	WP	YP	Av.
Niblack	L	105	106	105.5	106	103	104.5	73	88	80.5
Sauvola et al.	L	95	97	96	89	100	94.5	103	105	104
Mean-Gradient	L	21	27	24	0	0	0	65	73	69
PAT	L	108	102	105	105	79	92	103	102	102.5
ALT	L	108	107	107.5	61	95	78	97	93	95
WFM	L	102	107	104.5	102	103	102.5	94	104	99
Bernsen	L	88	76	82	10	66	38	25	68	46.5

Table 5.28: Overall visual inspection results for Set III

Thresholding Method	Category	Recognized Characters								
		Pen			Pencil			WB Marker		
		WP	YP	Av.	WP	YP	Av.	WP	YP	Av.
Otsu	G	107	107	107	107	103	105	108	107	107.5
Kittler and Illingworth	G	108	83	95.5	108	47	77.5	108	108	108
Yanni and Horne	G	22	15	18.5	55	5	30	50	65	57.5
Ramesh et al.	G	26	22	24	49	6	27.5	66	58	62
MD	G	108	106	107	108	59	83.5	108	108	108
Kapur et al.	G	95	85	90	40	32	36	10	5	7.5
Albuquerque et al.	G	11	26	18.5	12	8	10	5	2	3.5
Niblack	L	105	106	105.5	106	103	104.5	73	88	80.5
Sauvola et al.	L	95	97	96	89	100	94.5	103	105	104
Mean-Gradient	L	21	27	24	0	0	0	65	73	69
PAT	L	108	102	105	105	79	92	103	102	102.5
ALT	L	108	107	107.5	61	95	78	97	93	95
WFM	L	102	107	104.5	102	103	102.5	94	104	99
Bernsen	L	88	76	82	10	66	38	25	68	46.5

For uniform background (white paper), it was observed that global methods, Otsu, Kittler and Illingworth and MD, and local methods PAT and WFM performed suitable

results for all kind of writing materials.

For non-uniform background (yellow envelope paper), it was observed that almost same methods produced better results than others, however, global methods except Otsu, failed to binarize pencil on yellow envelope paper because converting RGB to gray-scale causes extreme harmonization of text and background at low level contrast point. Hence, Otsu, MD, PAT, Niblack and WFM achieved better results than other methods.

Evaluation of whole images in Set III shows that, Otsu which is superior one followed by WFM, PAT, MD and Sauvola et al. methods which recognition rates are higher than 90% (see *Table 5.29–Table 5.31*). Examples of obtained results for Set III can be seen in *Figure 5.12*.

Table 5.29: Overall Visual Inspection Results of Global Methods for Set III

Thresholding Method	Category	Rank	Total Characters	Clear Characters	Unclear Characters	Recognition Rate
Otsu	G	1	648	639	9	98.61%
Kit. and Ill.	G	3	648	472	176	72.83%
Yanni and Horne	G	6	648	212	436	32.71%
Ramesh et al.	G	5	648	227	421	35.03%
MD	G	2	648	597	51	91.12%
Kapur et al.	G	4	648	267	381	41.20%
Albuquerque et al	G	7	648	64	584	9.87%

Table 5.30: Overall Visual Inspection Results of Local Methods for Set III

Thresholding Method	Category	Rank	Total Characters	Clear Characters	Unclear Characters	Recognition Rate
Niblack	L	4	648	581	67	89.66%
Sauvola et al.	L	3	648	589	59	90.89%
Mean-Gradient	L	7	648	186	462	28.70%
PAT	L	2	648	599	49	92.43%
ALT	L	5	648	561	87	86.57%
WFM	L	1	648	612	36	94.44%
Bernsen	L	6	648	333	315	51.38%

During the obtaining binarization results, processing time of each method were calculated. *Table 5.32* shows the average processing time of each method.

5.5 Summary

In this chapter, performed experiments on document image binarization, design of experiments, evaluation criteria and obtained results were explained in details by performing

Table 5.31: Overall Visual Inspection Results for Set III

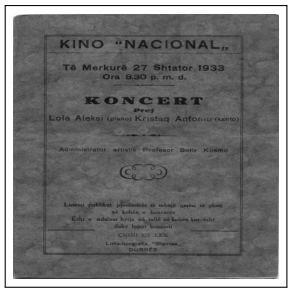
Thresholding Method	Category	Rank	Total Characters	Clear Characters	Unclear Characters	Recognition Rate
Otsu	G	1	648	639	9	98.61%
Kit. and Ill.	G	8	648	472	176	72.83%
Yanni and Horne	G	12	648	212	436	32.71%
Ramesh et al.	G	11	648	227	421	35.03%
MD	G	4	648	597	51	91.12%
Kapur et al.	G	10	648	267	381	41.20%
Albuquerque et al	G	14	648	64	584	9.87%
Niblack	L	6	648	581	67	89.66%
Sauvola et al.	L	5	648	589	59	90.89%
Mean-Gradient	L	13	648	186	462	28.70%
PAT	L	3	648	599	49	92.43%
ALT	L	7	648	561	87	86.57%
WFM	L	2	648	612	36	94.44%
Bernsen	L	9	648	333	315	51.38%

Table 5.32: Average Processing Time of the Methods

Thresholding Method	Average Time [6]
Otsu	0.68 s.
Kit. and Ill.	0.66 s.
Yanni and Horne	0.71 s.
Ramesh et al.	1.06 s.
MD	0.085 s.
Kapur et al.	0.75 s.
Albuquerque et al	0.75 s.
Niblack	0.013 s.
Sauvola et al.	0.11 s.
Mean-Gradient	0.29 s.
PAT	0.062 s.
ALT	0.42 s.
WFM	4.3
Bernsen	0.11

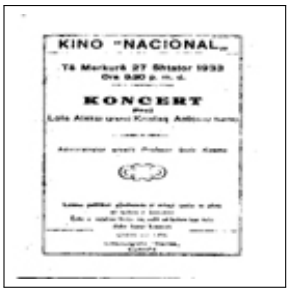
comparative evaluation of considered 14 local and global thresholding methods.

[6] using Pentium IV, 2 GHz. CPU and 512 MB RAM

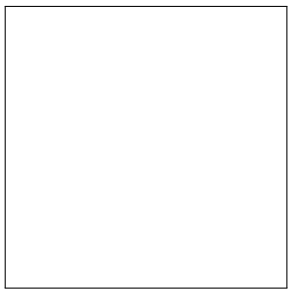


(a) Bernsen method

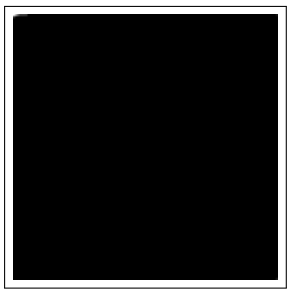
(b) Original Image



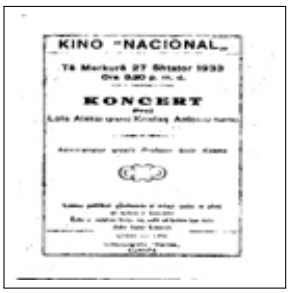
(c) Otsu method



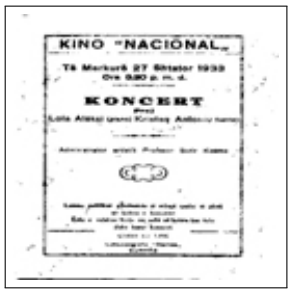
(d) Kittler and Illingworth method



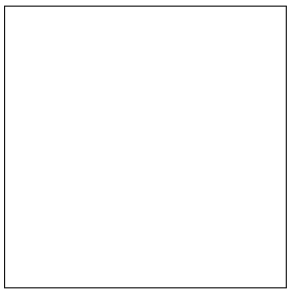
(e) Yanni and Horne method



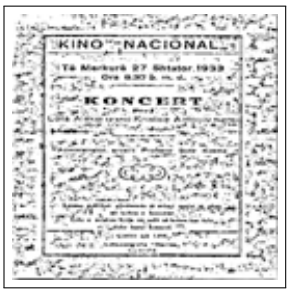
(f) Ramesh et al. method



(g) MD method

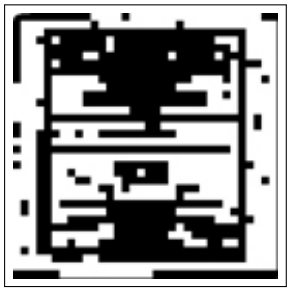


(h) Kapur et al. method

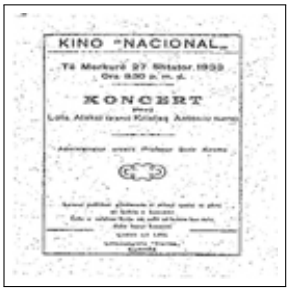


(i) Albuquerque et al. method

(j) Niblack method

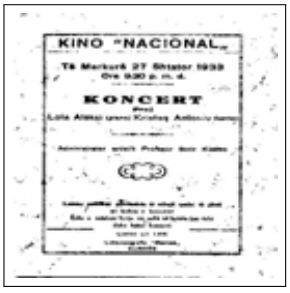


(k) Sauvola et al. method



(l) mean-gradient method

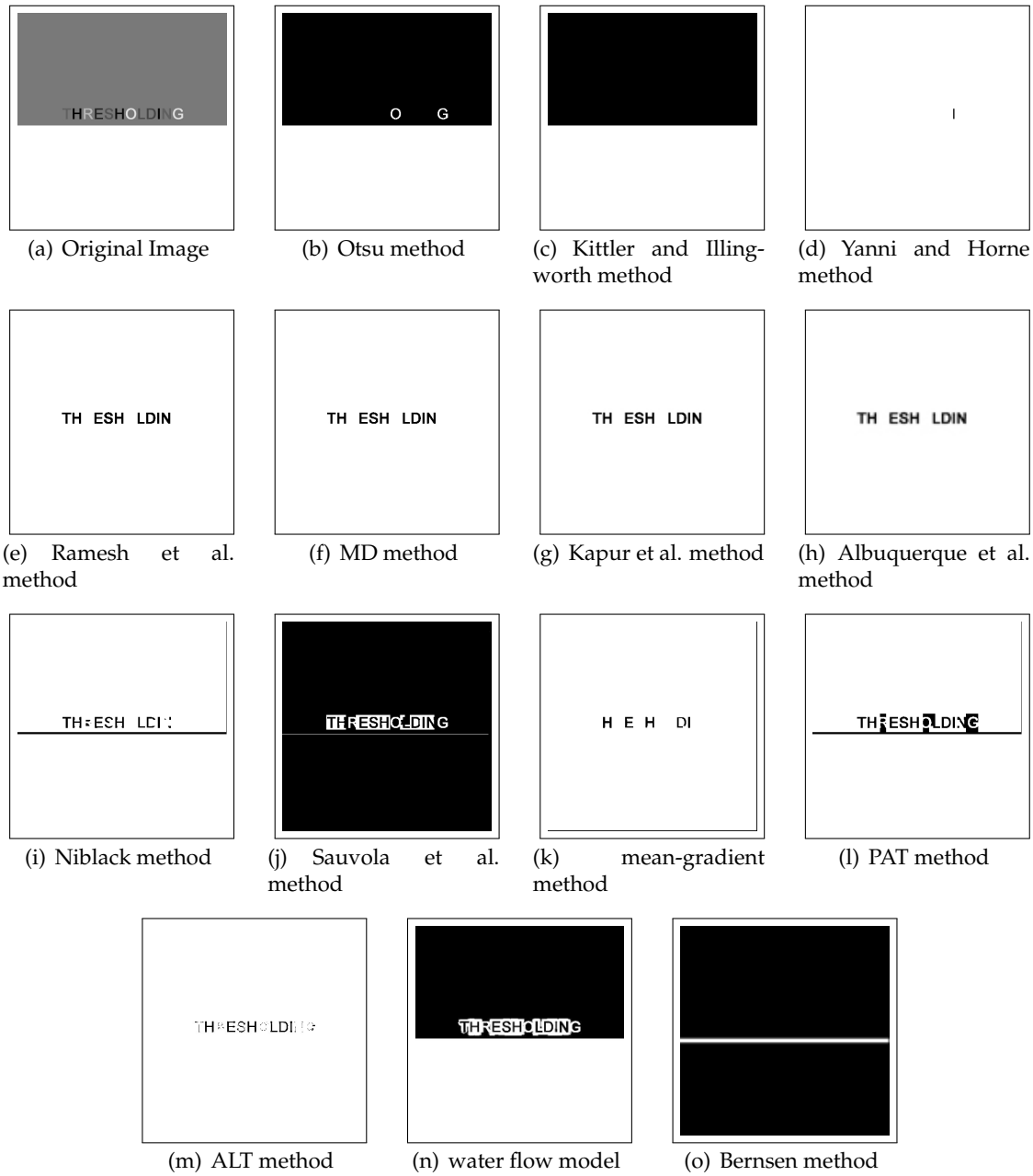
(m) PAT method



(n) ALT method

(o) water flow model

Figure 5.10: Example result of dark group of Set I



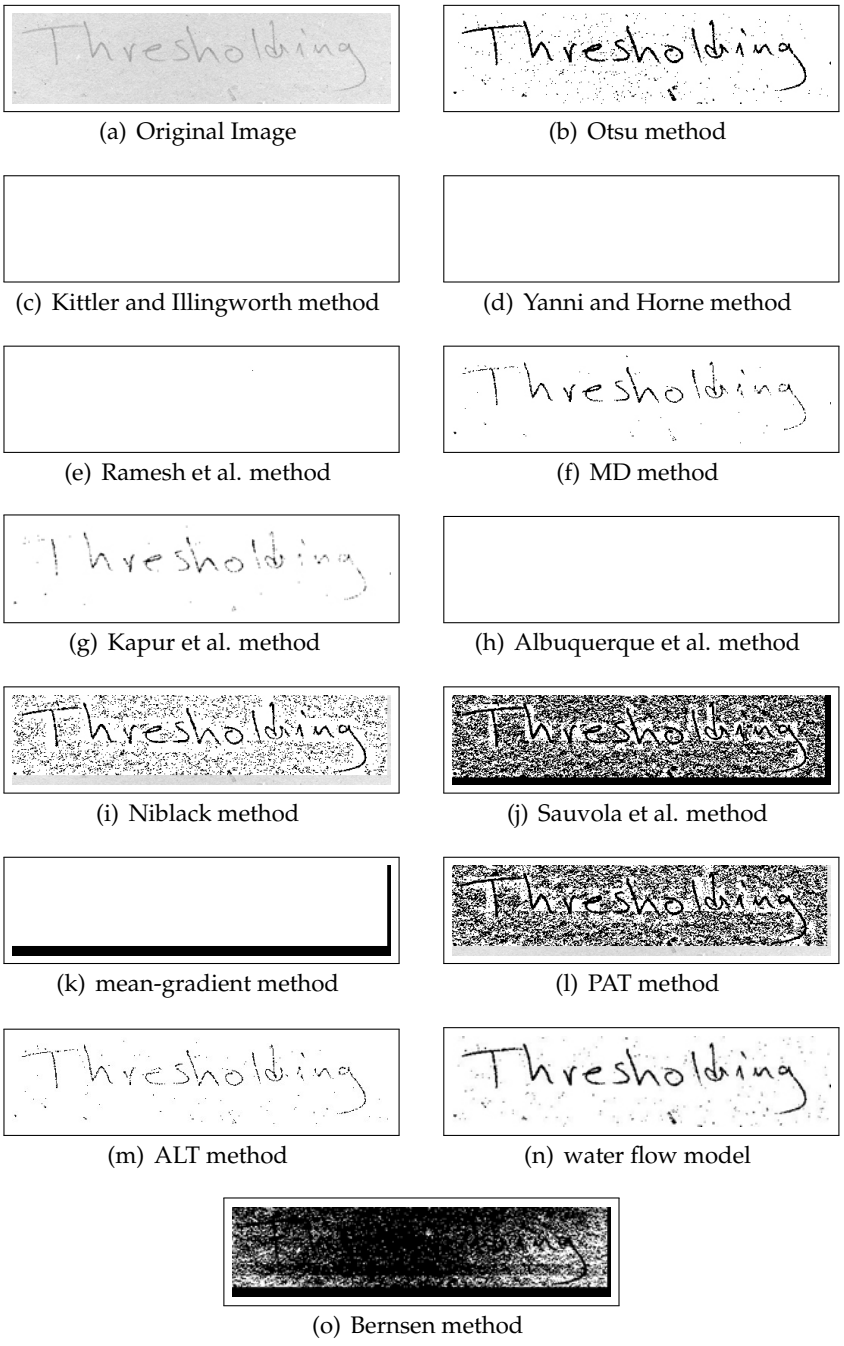
(a) Original Image (b) Otsu method (c) Kittler and Illingworth method (d) Yanni and Horne method

(e) Ramesh et al. method (f) MD method (g) Kapur et al. method (h) Albuquerque et al. method

(i) Niblack method (j) Sauvola et al. method (k) mean-gradient method (l) PAT method

(m) ALT method (n) water flow model (o) Bernsen method

Figure 5.11: Example result of created word image of Set II



(a) Original Image
(b) Otsu method
(c) Kittler and Illingworth method
(d) Yanni and Horne method
(e) Ramesh et al. method
(f) MD method
(g) Kapur et al. method
(h) Albuquerque et al. method
(i) Niblack method
(j) Sauvola et al. method
(k) mean-gradient method
(l) PAT method
(m) ALT method
(n) water flow model
(o) Bernsen method

Figure 5.12: Example result of handwritten image of Set III

CHAPTER 6

CONCLUSIONS

In this thesis, a novel single-stage global binarization method which is designed to enhance and separate dominant pixels of images was proposed and a local binarization method which is designed to enhance the alternative pixels of images, was investigated.

The global method which was named Mass Difference (MD) uses luminance value and mean of intensity to shift mean value of an image to the position that is determined as a threshold point. Variety of experiments were performed to demonstrate the efficiency of MD method and the results of these experiments suggest that MD thresholding method provided superior results when compared to the other considered methods in two different preliminary experiments under different conditions. Also statistical experiments were performed to show that the selected MD threshold point is the appropriate point.

The local method which was named as Pattern Averaging Threshold (PAT) uses mean value as a threshold point. Two experiments were performed to test the success and efficiency of the proposed local method in the visualization of the hidden data within the images and in the data preparation of a intelligent system.

Furthermore, twelve benchmark or recently developed thresholding methods which can be used for document image binarization were categorized into global and local thresholding groups and were implemented using the C language. The global thresholding group comprised: Otsu, Kittler and Illingworth, Yanni and Horne, Ramesh et al., Kapur et al., and Albuquerque et al. entropy methods. The local thresholding group comprised: Niblack, Sauvola et al., Mean-Gradient, Adaptive Logical Thresholding (ALT), Water Flow Model (WFM), and Bernsen methods.

A comprehensive multi-applications database was also created to comprise different challenging tasks such as historical documents, artificially created words and handwriting document images that simulate real life implementation in extreme and noisy conditions. The diversity of the document types within this database provides a larger database and multi-application document images with different levels of noise and contrast, non-uniform illumination (background), signal-dependent noise, smears, and non-uniform foreground (text). The use of multi-application database was suggested in order to provide a more objective evaluation and, consequently, the determination of an superior thresholding method.

The strategy for evaluating the 14 methods was based on two criteria: visual inspection and computed noise analysis of the binarized images. The sole use of one of these criteria does not provide an objective conclusion on which method is superior, thus the need for using both criteria. For this purpose, three PSNR-derived metric parameters that were used for noise analysis were introduced; these parameters were called: the Average PSNR Accuracy Rate (APAR), the Average PSNR Deviation (APD) of binarized images, and the Combined Performance Rate (CPR) of a thresholding method. The later determines the superior performance for a thresholding method and combines obtained results from both visual inspection and noise analysis criteria.

A comparative evaluation of the thresholding methods was performed using the multi-application image database and using the proposed evaluation strategy. The overall evaluation results of comparative evaluation suggest that MD thresholding method, which is a global method, provided superior results when compared to the other methods, and thus can be considered as the superior global thresholding method. On the other hand, WFM thresholding method, which is a local thresholding method, showed superior performance, and thus can be considered as the superior local thresholding method. By comparing MD and WFM, it was shown that these two methods achieved similar results in historical documents. MD was superior in visual inspection that was used to measure human readability with a minimal noise addition and WFM was superior in noise removal with a little loss of information. When using the especially created word documents, the recognition rate of both these methods decreased, whereas using handwriting document images both methods produced similar successful results.

However, the determination of mask size of WFM is a serious drawback that sometimes causes a huge loss of information during binarization and computational cost. MD has faster execution time which makes marginal difference in computational cost when binarizing a huge collection of documents. Hence, MD thresholding method can be considered as the superior threshold method.

Future work will focus on implementing MD thresholding for applications to images with similar characteristics to artificially created words and the implementing of PAT for security and forensic applications.

REFERENCES

[1] O. D. Trier and T. Taxt, "Evaluation of binarization methods for document images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 312–315, 1995. ◇ pp: 3, 75, 76, 77

[2] O. D. Trier, , and A. Jain, "Goal-directed evaluation of binarization methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 1191–1201, 1995. ◇ pp: 3, 75, 76, 77, 85

[3] G. Leedham, C. Yan, K. Takru, J. H. N. Tan, and L. Mian, "Comparison of some thresholding algorithms for text/background segmentation in difficult document image," in *Seventh International Conference on Document Analysis and Recognition*, pp. 859–864, 2003. ◇ pp: 3, 30, 32, 39, 41, 75, 76, 85

[4] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, pp. 146–165, 2004. ◇ pp: 3, 75, 76, 77, 85

[5] J. He, Q. D. M. Do, A. C. Downton, and J. H. Kim, "A comparison of binarization methods for historical archive documents," in *Eight International Conference on Documents Analysis and Recognition*, vol. 1, pp. 538–542, 2005. ◇ pp: 3, 75, 77

[6] S. E. Umbaugh, *Computer Imaging : Digital Image Analysis and Processing*. CRC Press, 2005. ◇ pp: 6

[7] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice Hall, 2008. ◇ pp: 6, 7, 8, 10, 12, 13, 19, 21, 27, 78, 83

[8] J. K. Kim, J. M. Park, K. S. Song, and H. W. Park, "Adaptive mammographic image enhancement using first derivatives and local statistics," *IEEE Transactions on Medical Imaging*, vol. 16, no. 5, pp. 495–502, 1997. ◇ pp: 24

[9] I. Koren, A. Laine, and F. Taylor, "Enhancement via fusion of mammographic features," in *1998 International Conference on Image Processing*, vol. 3, pp. 722–726, 1998. ◇ pp: 24

[10] M. Wirth, M. Fraschini, and J. Lyon, "Contrast enhancement of microcalcifications in mammograms using morphological enhancement and non-flat structuring elements," in *17th IEEE Symposium on Computer-Based Medical Systems*, 2004. ◇ pp: 24

[11] U. Numburi, G. P. Chatimavroudis, A. E. Uber, and J. P. Kalafut, "Modelling of contrast enhancement for cardiac multi-detector row ct," in *IEEE International Workshop on Imaging Systems and Techniques*, pp. 125–129, 2005. ◇ pp: 24

[12] M. Jiang, Q. Ji, and B. F. McEwen, "Enhancement of microtubules in em tomography," in *IEEE International Symposium on Biomedical Imaging*, pp. 1123–1126, 2004. ◇ pp: 24

[13] C. H. Lo, Y. Guo, and C. C. Lu, "A binarization approach for ct-mr registration using normalized mutual information," in *5th IASTED International Conference on Signal and Image Processing*, 2003. ◇ pp: 24

[14] E.W.Abel, C.A.Wigderowitz, and D.I.Rowley, "Frequency domain imaging of cancellous bone," in *18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1182–1184, 1996. ◇ pp: 24

[15] D. Sale, R. R. Schultz, and R. J. Szczerba, "Super-resolution enhancement of night vision image sequences," in *International Conference on Systems, Man, and Cybernetics*, vol. 3, pp. 1633–1638, 2000. ◇ pp: 24

[16] V. Areekul, U. Watchareeruetai, K. Suppasriwasuseth, and S. Tantaratana, "Separable gabor filter realization for fast fingerprint enhancement," in *IEEE International Conference on Image Processing*, vol. 3, pp. 253–256, 2005. ◇ pp: 24

[17] S. Wang and Y. Wang, "Fingerprint enhancement in the singular point area," *IEEE Signal Processing Letters*, vol. 11, no. 1, 2004. ◇ pp: 24

[18] L. Jin, S. Satoh, and M. Sakauchi, "A novel adaptive image enhancement algorithm for face detection," in *17th International Conference on Pattern Recognition*, vol. 4, pp. 843–848, 2004. ◇ pp: 24

[19] B. Lunden, K. Wester, and G. Bax, "Satellite image enhancement for rock type separation," in *International Geoscience and Remote Sensing Symposium 'Remote Sensing: Global Monitoring for Earth Management*, vol. 4, pp. 2047–2050, 1991. ◇ pp: 24

[20] T. R. Randolph and M. J. T. Smith, "Enhancement of fax documents using a binary angular representation," in *2001 International Symposium on Intelligent Multimedia,Video and Speech Processing*, pp. 125–128, 2001. ◇ pp: 25

[21] G. A. Ware, D. M. Chabries, R. W. Christiansen, and C. E. Martin, "Multispectral document enhancement: Ancient carbonized scrolls," in *IEEE International Geoscience and Remote Sensing Symposium*, vol. 6, pp. 2486–2488, 2000. ◇ pp: 25

[22] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognition*, vol. 19, no. 4, pp. 41–47, 1986. ◇ pp: 26, 29, 31, 33, 63, 75

[23] T. Riddler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Transactions on Sytems, Man and Cybernetics*, vol. 8, pp. 630–632, 1978. ◇ pp: 29, 31

[24] N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Transactions on Sytems, Man and Cybernetics*, vol. 9, pp. 62–66, 1979. ◇ pp: 29, 31, 32, 63, 75

[25] Y. Nakagawa and A. Rosenfeld, "Some experiments on variable thresholding," *Pattern Recognition*, vol. 11, no. 3, pp. 191–204, 1979. ◇ pp: 29, 32

[26] C. Chow and T. Kaneko, "Automatic, boundary detection of the leftventricle from cine-angiograms," *Computational Biomedical Research*, vol. 5, pp. 388–410, 1972. ◇ pp: 29, 75

[27] T. Pun, "A new method for gray-level picture threshold using the entropy of the histogram," *Signal Processing*, vol. 2, no. 3, pp. 223–237, 1980. ◇ pp: 29, 31

[28] Y. Yasuda, M. Dubois, and T. S. Huang, "Data compression for check processing machines," *Proceedings of IEEE*, vol. 68, pp. 874–885, 1980. ◇ pp: 29, 32

[29] J. White and G. Rohrer, "Image Thresholding for Optical Character Recognition and Other App1ications Requiring Character Image Extraction," Tech. Rep. 27, IBM J. Research and Development, 1983. ◇ pp: 29, 32, 75

[30] A. Rosenfeld and P. D. la Torre, "Histogram concavity analysis as an aid in threshold selection," *IEEE Transactions on Sytems, Man and Cybernetics*, vol. 13, pp. 231–235, 1983. ◇ pp: 29, 31

[31] J. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics, and Image Processing*, vol. 29, pp. 273–285, 1985. ◇ pp: 29, 31, 35, 37, 63, 75

[32] D. E. Lloyd, "Automatic Target Classification Using Moment Invariant of Image Shapes," Tech. Rep. AW126, RAE IDN AW126, UK, 1985. ◇ pp: 29, 31

[33] W. Niblack, *An Introduction to Digital Image Processing*. Prentice Hall, 1986. ◇ pp: 29, 32, 39, 75

[34] J. Bernsen, "Dynamic thresholding of gray-level images," in *8th International Conference on Pattern Recognition*, pp. 1251–1255, 1986. ◇ pp: 29, 32, 39, 45, 75

[35] P. W. Palumbo, P. Swaminathan, and S. N. Srihari, "Document image binarization: Evaluation of algorithms," *Proc. SPIE*, vol. 697, pp. 278–285, 1986. ◇ pp: 29, 32

[36] S. Abutaleb, "Automatic thresholding of gray-level pictures using two dimensional entropy," *Computer Vision, Graphics, and Image Processing*, vol. 47, pp. 22–32, 1989. ◇ pp: 29, 31, 75

[37] S. Yanowitz and A. Bruckstein, "A new method for image segmentation," *Computer Vision, Graphics, and Image Processing*, vol. 46, pp. 82–95, 1989. ◇ pp: 29, 32, 75

[38] T. Taxt, P. Flynn, and A. Jain, "Segmentation of document images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 1322–1329, 1989. ◇ pp: 30, 32, 75

[39] L. Eikvil, T. Taxt, and K. Moen, "A fast adaptive method for binarization of document images," in *8th International Conference on Pattern Recognition*, pp. 435–443, 1991. ◇ pp: 30, 32, 75

[40] J. R. Parker, "Gray level thresholding in badly illuminated images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 813–819, 1991. ◇ pp: 30, 63, 75

[41] C. H. Li and C. K. Lee, "Minimum cross-entropy thresholding," *Pattern Recognition*, vol. 26, pp. 617–625, 1993. ◇ pp: 30, 31

[42] M. Kamel and A. Zhao, "Extraction of binary character/graphics images from grayscale document images," *CVGIP: Graphical Models Image Processing*, vol. 55, no. 3, pp. 203–217, 1993. ◇ pp: 30, 32, 44

[43] M. Yanni and E. Horne, "A new aproach to dynamic thresholding," in *9th European Conference on Signal Processing*, pp. 34–44, 1994. ◇ pp: 30, 31, 33

[44] N. Ramesh, J. Yoo, and I. Sethi, "Thresholding based on histogram approximation," *IEE Vision, Image and Signal Processing*, vol. 142, no. 5, 1995. ◇ pp: 30, 31, 35, 36

[45] J. C. Yen, F. J. Chang, , and S. Chang, "A new criterion for automatic multilevel thresholding," *IEEE Transactions on Image Processing*, vol. 4, pp. 370–378, 1995. ◇ pp: 30, 31

[46] N. R. Pal, "On minimum cross-entropy thresholding," *Pattern Recognition*, vol. 29, no. 4, pp. 575–580, 1996. ◇ pp: 30, 31

[47] P. Sahoo, C. Wilkins, and J. Yeager, "Threshold selection using renyi's entropy," *Pattern Recognition*, vol. 30, pp. 71–84, 1997. ◇ pp: 30, 31

[48] I. A. Esquef, A. R. G. Mello, M. P. de Albuquerque, and M. P. de Albuquerque, "Image thresholding using tsallis entropy," *Pattern Recognition Letters*, vol. 25, pp. 1059–1065, 2004. ◇ pp: 30, 31, 36, 38

[49] W. Oh and B. Lindquist, "Image thresholding by indicator kriging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 590–602, 1999. ◇ pp: 30, 32

[50] J. Sauvola and M. Pietikainen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, pp. 225–236, 2004. ◇ pp: 30, 32, 39, 41

[51] Y. Solihin and C. G. Leedham, "Integral ratio: A new class of global thresholding techniques for handwriting images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 761–768, 1999. ◇ pp: 30, 31, 63, 76

[52] Y. Yang and H. Yan, "An adaptive logical method for binarization of degraded document images," *Pattern Recognition*, vol. 33, pp. 787–807, 2000. ◇ pp: 30, 32, 39, 44

[53] C. Wolf and J. Jolian, "Extraction and Recognition of Artificial Text in Multimedia Documents," Tech. Rep. RFV RR, RFV RR, 2002. ◇ pp: 30, 32

[54] E. Badekas and N. Papamarkos, "A system for document binarization," in *3rd International Symposium on Image and Signal Processing and Analysis*, pp. 909–914, 2003. ◇ pp: 30, 32

[55] M. Sezgin and B. Sankur, "Image multi-thresholding based on sample moment function," in *International Conference on Image Processing*, vol. 2, pp. 415–418, 2003. ◇ pp: 30, 31

[56] I. K. Kim, D. W. Jung, and R. H. Park, "Document image binarization based on topographic analysis using a water flow model," *Pattern Recognition*, vol. 35, pp. 265–277, 2002. ◇ pp: 30, 32, 39, 46

[57] E. Kavallieratou, "A binarization algorithm specialized on document images and photos," in *Eight International Conference on Document Analysis and Recognition*, vol. 1, pp. 463–467, 2005. ◇ pp: 30, 31

[58] E. Kavallieratou and H. Antonopoulou, "Cleaning and enhancing historical document images," *Lecture Notes in Computer Science*, vol. 3708, pp. 681–688, 2005. ◇ pp: 30

[59] Y. Chen and G. Leedham, "Decompose algorithm for thresholding degraded historical document images," *IEE Proc. Vis. Image Signal Process.*, vol. 152, no. 6, pp. 702–714, 2005. ◇ pp: 31, 32

[60] H.-S. Don, "A noise attribute thresholding method for document image binarization," in *Third International Conference on Document Analysis and Recognition*, vol. 1, pp. 231–234, 1995. ◇ pp: 31

[61] W. Ng and C. Lee, "Comment on using the uniformity measure for performance measure in image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 933–934, 1996. ◇ pp: 32

[62] T. Luo, K. Kramer, D. B. Goldgof, L. O. Hall, S. Samson, A. Remsen, and T. Hopkins, "Recognizing plankton images from the shadow image particle profiling evaluation recorder," *IEEE Transactions on Systems, Man, And Cybernetics-Part B: Cybernetics*, vol. 34, no. 4, pp. 1753–1762, 2004. ◇ pp: 48

[63] F. Zhao, X. Tang, F. Lin, S. Samson, and A. Remsen, "Binary plankton image classification using random subspace," in *IEEE International Conference on Image Processing*, vol. 1, pp. 357–360, 2005. ◇ pp: 48

[64] T. Luo, K. Kramer, D. B. Goldgof, L. O. Hall, S. Samson, A. Remsen, and T. Hopkins, "Learning to recognize plankton," in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 1, pp. 889–893, 2003. ◇ pp: 48

[65] D. Haverkamp, L. E. Soh, and C. Tsatsoulis, "A dynamic local thresholding technique for sea ice classification," in *International Geoscience and Remote Sensing Symposium*, vol. 2, pp. 638–640, 1993. ◇ pp: 48

[66] R. Bremananth, B. Balaji, M. Sankari, and A. Chitra, "A new approach to coin recognition using neural pattern analysis," in *IEEE Indicon Conference*, pp. 366–370, 2005. ◇ pp: 48

[67] A. Khashman, B. Sekeroglu, and K. Dimililer, "Icis: A novel coin identification system," *Lecture Notes in Computer Science*, vol. 3708, 2006. ◇ pp: 48, 73

[68] A. Khashman, B. Sekeroglu, and K. Dimililer, "Intelligent rotation-invariant coin identification system," *WSEAS Transactions on Signal Processing*, vol. 2, no. 5, pp. 781–786, 2006. ◇ pp: 48, 73

[69] A. Khashman, B. Sekeroglu, and K. Dimililer, "Intelligent coin identification system," in *IEEE International Symposium on Intelligent Control*, 2006. ◇ pp: 48, 73

[70] A. Khashman, B. Sekeroglu, and K. Dimililer, "Rotated coin recognition using neural network," *Advances in Soft Computing*, vol. 41, pp. 290–297, 2007. ◇ pp: 48

[71] V. Onnia and M. Tico, "Adaptive binarization method for fingerprint images," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 3692–3695, 2002. ◇ pp: 48

[72] P. Meenen and R. Adhami, "Approaches to image binarization in current automated fingerprint identification systems," in *Thirty-Seventh Southeastern Symposium on System Theory*, pp. 276–281, 2005. ◇ pp: 48

[73] Y. Zhang and Q. Xiao, "An optimized approach for fingerprint binarization," in *International Joint Conference on Neural Networks*, pp. 391–395, 2006. ◇ pp: 48

[74] J. Gao and M. Xie, "The layered segmentation, gabor filtering and binarization based on orientation for fingerprint preprocessing," in *The 8th International Conference on Signal Processing*, vol. 4, 2006. ◇ pp: 48

[75] X. J. amd D. Mojon, "Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 131–137, 2003. ◇ pp: 48

[76] H. Zhu, F. H. Y. Chan, F. K. Lam, and K. Y. Lam, "Segmentation of pathology microscopic images," in *19th IEEE/EMBS International Conference*, pp. 580–581, 1997. ◇ pp: 48

[77] A. Ruggeri, S. Pajaro, and A. Vita, "Classification of corneal layers in confocal microscopy," in *22. Annual EMBS International Conference*, pp. 1030–1032, 2002. ◇ pp: 48

[78] E. Espinoza, G. Martinez, J. G. Frerichs, and T. Scheper, "Cell cluster segmentation based on global and local thresholding for in-situ microscopy," in *IEEE International Symposium on Biomedical Imaging: Macro to Nano*, pp. 542–545, 2006. ◇ pp: 48

[79] T. Sund and K. Eilertsen, "An algorithm for fast adaptive image binarization with applications in radiotherapy imaging," *IEEE Transactions on Medical Imaging*, vol. 22, no. 1, pp. 22–28, 2003. ◇ pp: 48

[80] J. Baudewig, P. Dechent, K. Merboldt, and J. Frahm, "Thresholding in correlation analyses of magnetic resonance functional neuroimaging," *Magnetic Resonance Imaging*, vol. 21, pp. 1121–1130, 2003. ◇ pp: 48

[81] Z. Zhoua and Z. Ruana, "Multicontext wavelet-based thresholding segmentation of brain tissues in magnetic resonance images," *Magnetic Resonance Imaging*, vol. 25, pp. 381–385, 2007. ◇ pp: 48

[82] D. Y. Kim and J. W. Park, "Connectivity-based local adaptive thresholding for carotid artery segmentation using mra images," *Image and Vision Computing*, vol. 23, pp. 1277–1287, 2005. ◇ pp: 48

[83] A. F. Obrist, A. Flisch, , and J. Hofmann, "Point cloud reconstruction with sub-pixel accuracy by slice-adaptive thresholding of x-ray computed tomography images," *NDT & E International*, vol. 37, pp. 373–380, 2004. ◇ pp: 48

[84] Y. Ebrahim, "Entropy based thresholding of cross-dissolved ultrasound images," in *IEEE Canadian Conference on Electrical and Computer Engineering*, vol. 3, pp. 1477–1480, 2003. ◇ pp: 48

[85] N. Hiransakolwong, K. A. Hua, K. Vu, and P. S. Windyga, "Segmentation of ultrasound liver images: An automatic approach," in *2003 International Conference on Multimedia and Expo*, vol. 1, pp. 573–576, 2003. ◇ pp: 48

[86] A. Konkachbaev, M. Casanova, J. Graham, and A. Elmaghraby, "Automated recursive segmentation of large neocortical images using standard deviation as termination criteria," in *27th IEEE Annual Conference on Engineering in Medicine and Biology*, pp. 2531–2534, 2005. ◇ pp: 49

[87] H. J. Jeong, T. Y. Kim, H. G. Hwang, H. J. Choi, H. S. Park, and H. K. Choi, "Comparison of thresholding methods for breast tumor cell segmentation," in *7th International Workshop on Enterprise networking and Computing in Healthcare Industry*, pp. 392–395, 2005. ◇ pp: 49

[88] C. Varelaa, P. G. Tahoces, A. J. Mendez, M. Souto, and J. J. Vidal, "Computerized detection of breast masses in digitized mammograms," *Computers in Biology and Medicine*, vol. 37, no. 2, pp. 214–226, 2007. ◇ pp: 49

[89] N. Sang, H. Li, W. Peng, and T. Zhang, "Knowledge-based adaptive thresholding segmentation of digital subtraction angiography images," *Image and Vision Computing*, vol. 25, pp. 1263–1270, 2007. ◇ pp: 49

[90] R. Rodriguez, "A strategy for blood vessels segmentation based on the threshold which combines statistical and scale space filter application to the study of angiogenesis," *Computer Methods and Programs in Biomedicine*, vol. 82, pp. 1–9, 2006. ◇ pp: 49

[91] J. Mandel, *The Statistical Analysis of Experimental Data*. Dover Publications Inc., 1984. ◇ pp: 62

[92] J. R. Taylor, *An Introduction to Error Analysis*. University Science Books, 1997. ◇ pp: 62

[93] M. Bulmer, *Principles of Statistics*. Dover Publications Inc., 1979. ◇ pp: 62

[94] E. L. Crow, F. A. Davis, and M. W. Maxfield, *Statistics Manual*. Dover Publications Inc., 2007. ◇ pp: 62

[95] A. Khashman and B. Sekeroglu, "Enhancement of unclear patterns using pattern average thresholding," in *3rd International Symposium on Electrical, Electronic and Computer Engineering*, vol. 1, pp. 253–257, 2006. ◇ pp: 68

[96] K. V. Mardia and T. J. Hainsworth, "A spatial thresholding method for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 6, pp. 919–927, 1988. ◇ pp: 75

[97] O. D. Trier and T. Taxt, "Improvement of 'integrated function algorithm' for binarization of document images," *Pattern Recognition Letters*, vol. 16, pp. 277–283, 1995. ◇ pp: 75

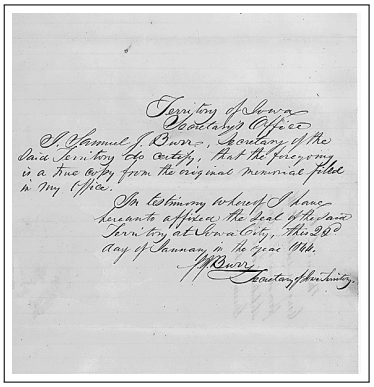
[98] M. Junker and R. Hoch, "On the evaluation of document analysis components by recall, precision and accuracy," in *Fifth International Conference on Documents Analysis and Recognition*, pp. 713–716, 1999. ◇ pp: 76

APPENDICES

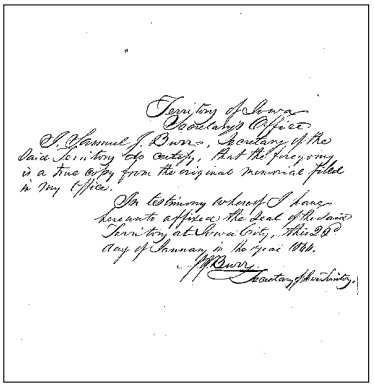
APPENDIX A

Example Document Image Binarization Results

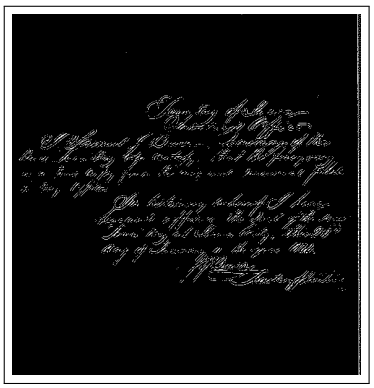
Appendix A presents the example document image binarization process results (see *Chapter 5*) that were obtained using 14 global and local thresholding methods.



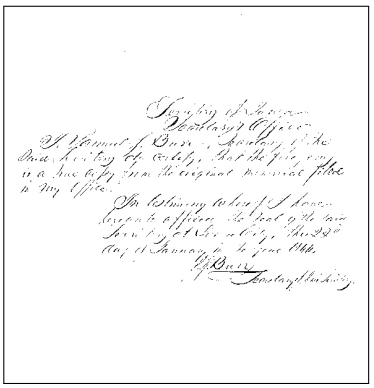
(a) Original low contrast image



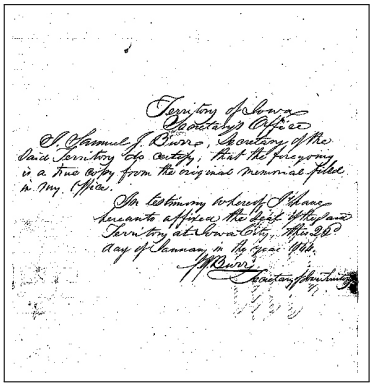
(b) Otsu method



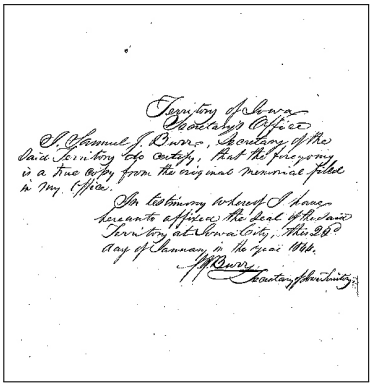
(c) Kittler and Illingworth method



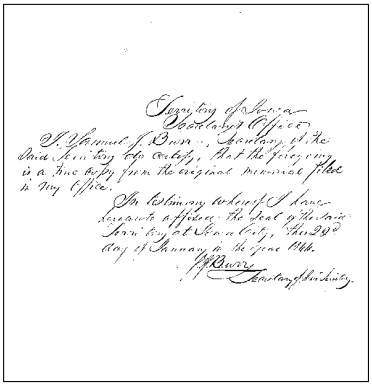
(d) Yanni and Horne method



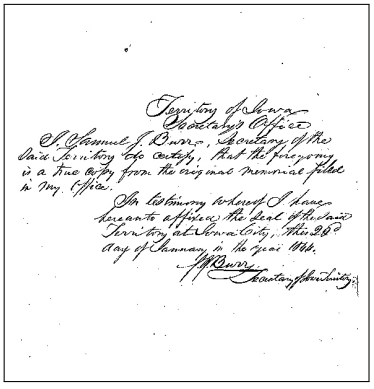
(e) Ramesh et al. method



(f) MD method

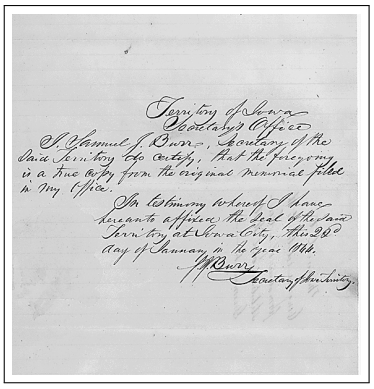


(g) Kapur et al. method

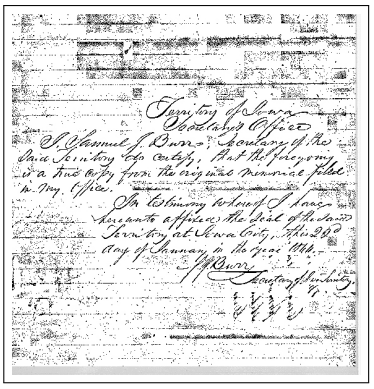


(h) Albuquerque et al. method

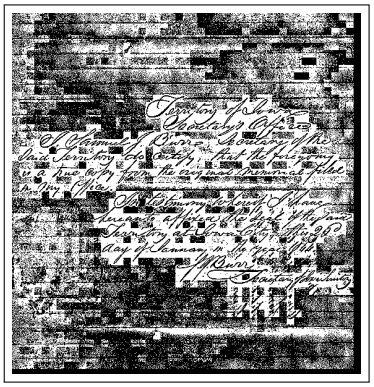
Figure A.1: Example result of low contrast image of Set I by global methods



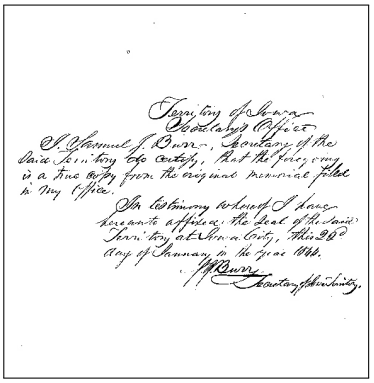
(a) Original low contrast image



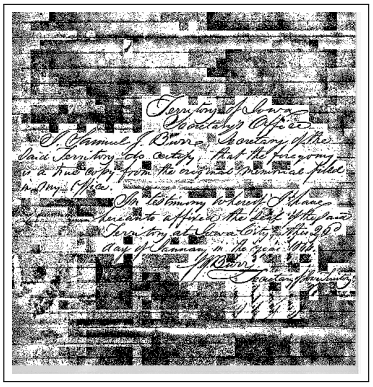
(b) Niblack method



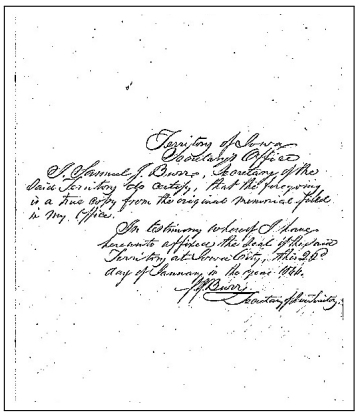
(c) Sauvola et al. method



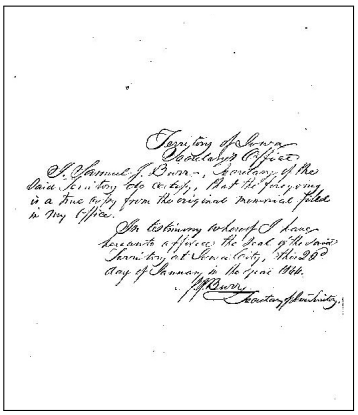
(d) mean-gradient method



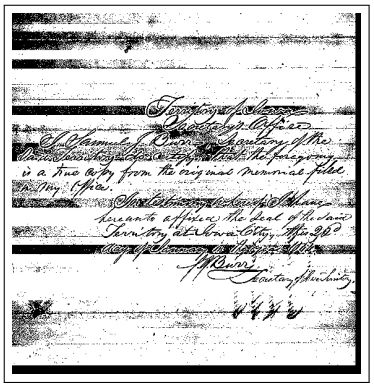
(e) PAT method



(f) ALT method

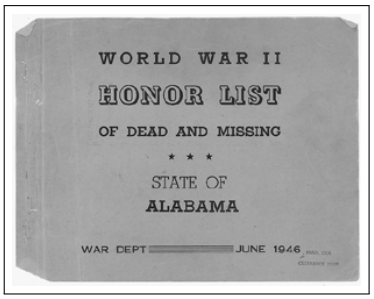


(g) water flow model

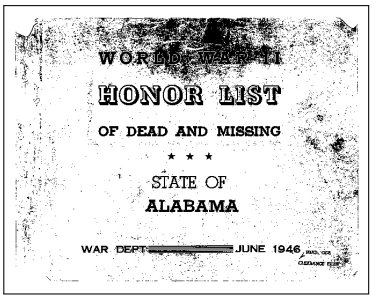


(h) Bernsen method

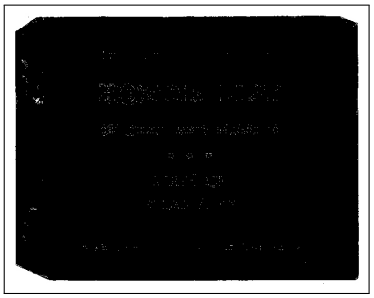
Figure A.2: Example result of low contrast image of Set I by local methods



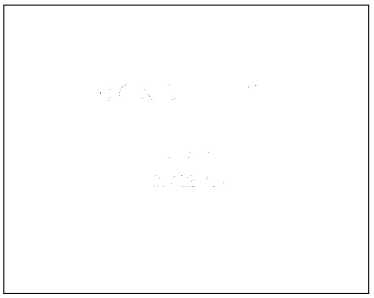
(a) Original low contrast image



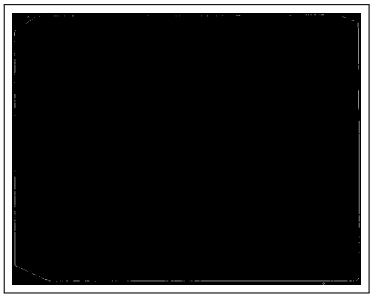
(b) Otsu method



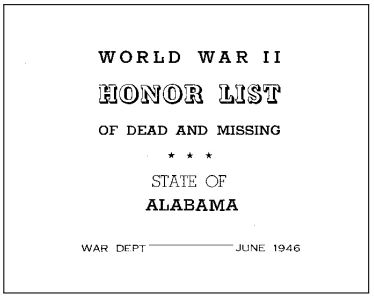
(c) Kittler and Illingworth method



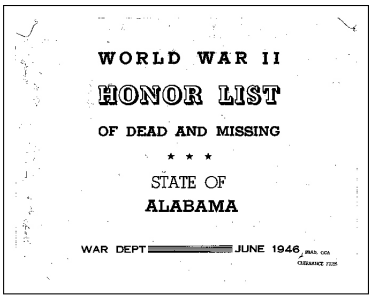
(d) Yanni and Horne method



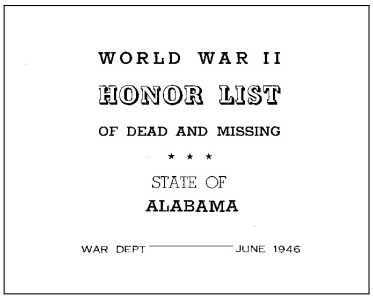
(e) Ramesh et al. method



(f) MD method

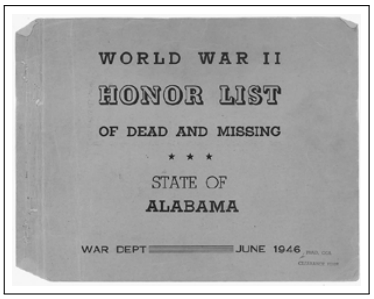


(g) Kapur et al. method

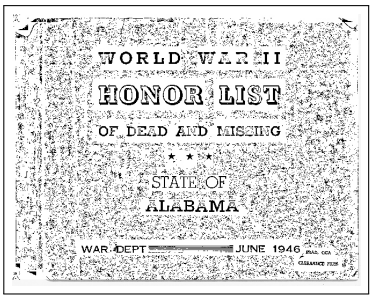


(h) Albuquerque et al. method

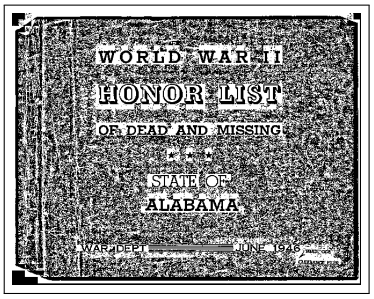
Figure A.3: Example result of low contrast image of Set I by global methods



(a) Original low contrast image

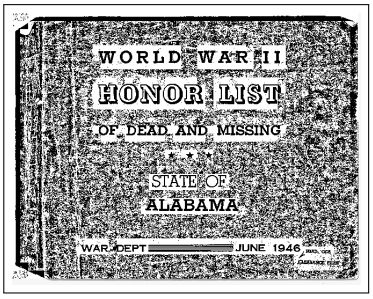


(b) Niblack method

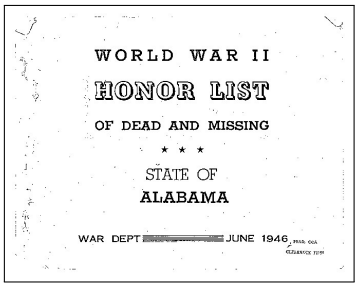


(c) Sauvola et al. method

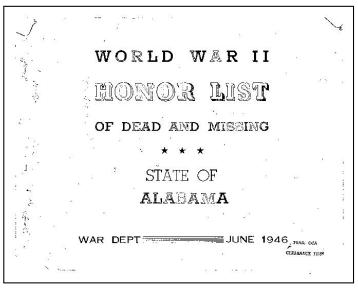
(d) mean-gradient method



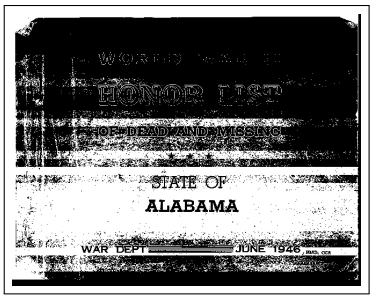
(e) PAT method



(f) ALT method



(g) water flow model

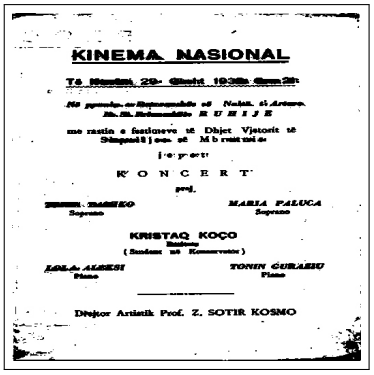


(h) Bernsen method

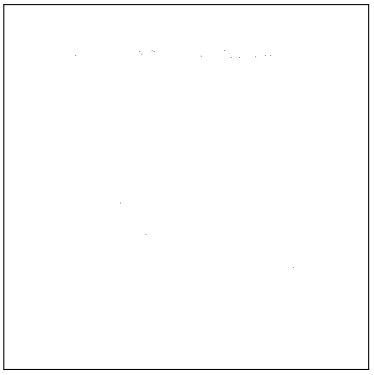
Figure A.4: Example result of low contrast image of Set I by local methods

(a) Original bright image

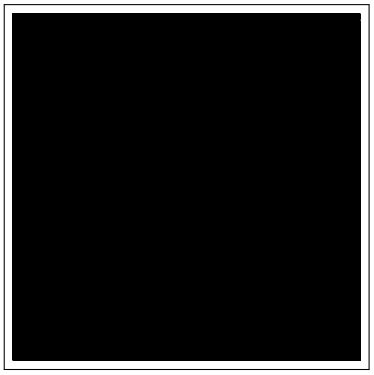
(b) Otsu method



(c) Kittler and Illingworth method



(d) Yanni and Horne method



(e) Ramesh et al. method

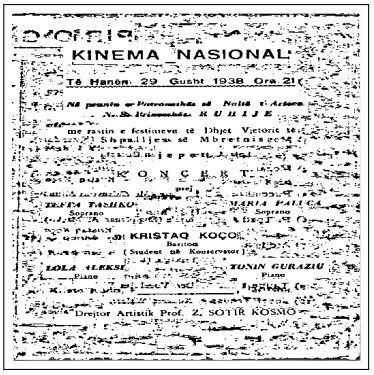
(f) MD method

(g) Kapur et al. method

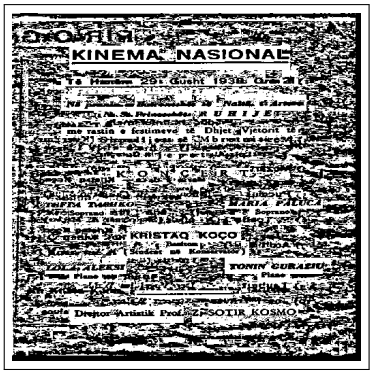
(h) Albuquerque et al. method

Figure A.5: Example result of low contrast image of Set I by global methods

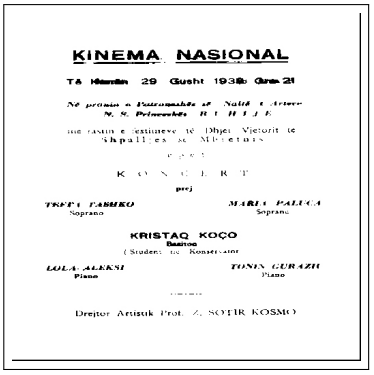
(a) Original bright image



(b) Niblack method

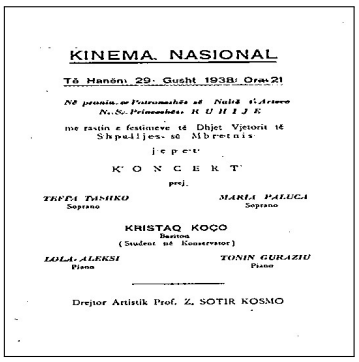


(c) Sauvola et al. method



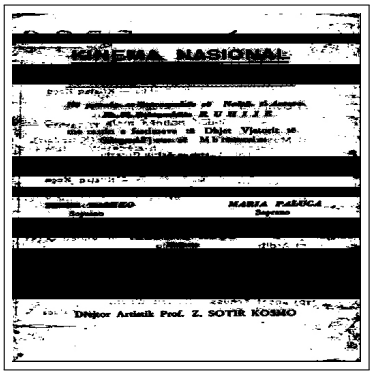
(d) mean-gradient method

(e) PAT method



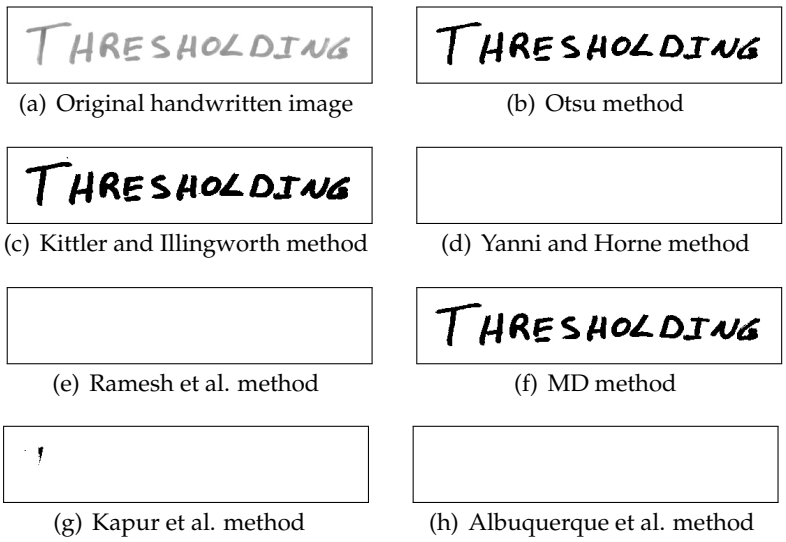
(f) ALT method

(g) water flow model



(h) Bernsen method

Figure A.6: Example results of bright image of Set I by local methods



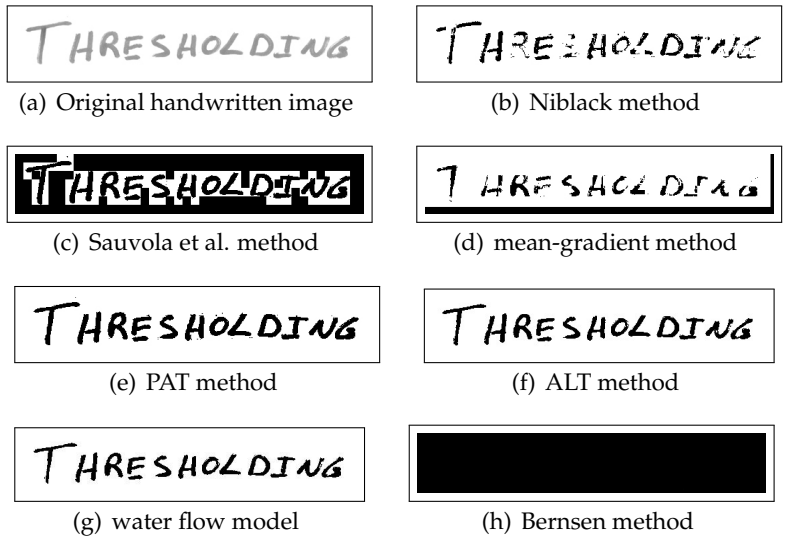
(a) Original handwritten image (b) Otsu method

(c) Kittler and Illingworth method (d) Yanni and Horne method

(e) Ramesh et al. method (f) MD method

(g) Kapur et al. method (h) Albuquerque et al. method

Figure A.7: Example results of handwritten image on white paper by white board marker - global methods



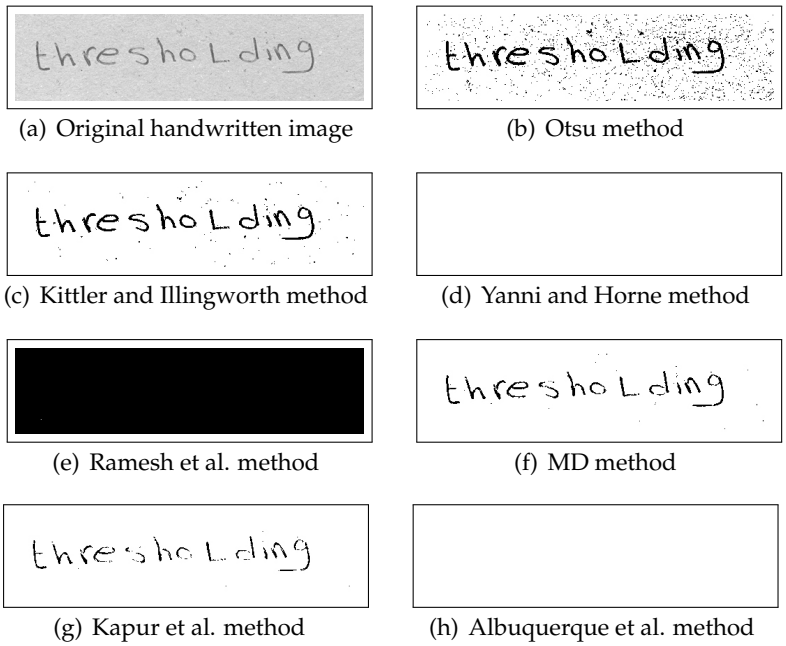
(a) Original handwritten image (b) Niblack method

(c) Sauvola et al. method (d) mean-gradient method

(e) PAT method (f) ALT method

(g) water flow model (h) Bernsen method

Figure A.8: Example results of handwritten image on white paper by white board marker in image set III - local methods



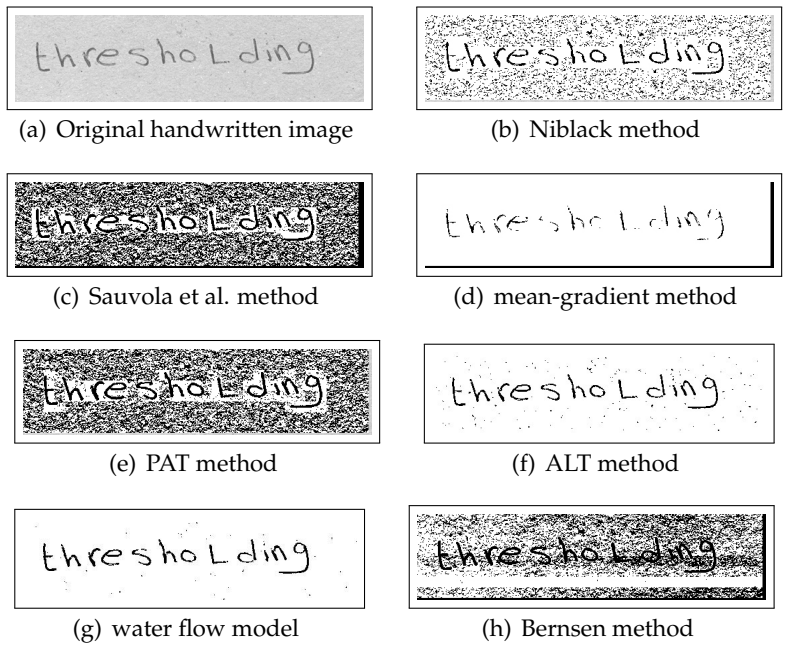
(a) Original handwritten image (b) Otsu method

(c) Kittler and Illingworth method (d) Yanni and Horne method

(e) Ramesh et al. method (f) MD method

(g) Kapur et al. method (h) Albuquerque et al. method

Figure A.9: Example results of handwritten image on yellow envelope paper by pen in image set III - global methods



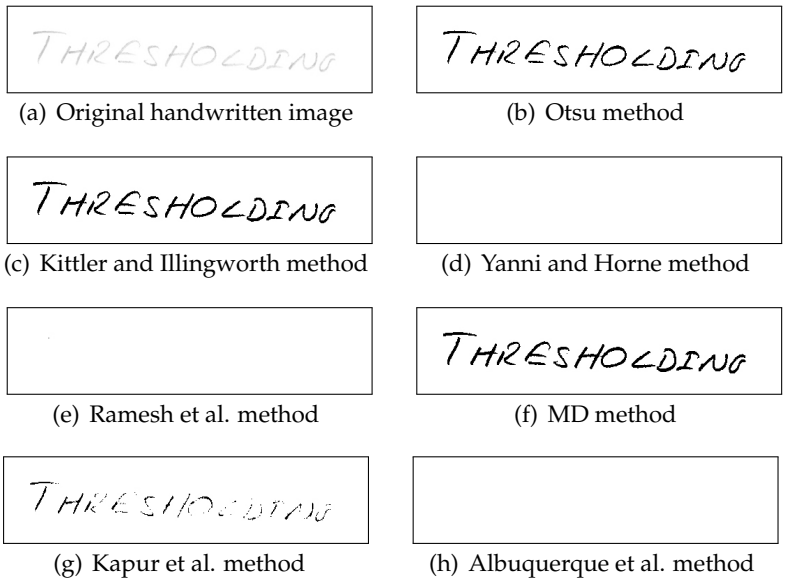
(a) Original handwritten image (b) Niblack method

(c) Sauvola et al. method (d) mean-gradient method

(e) PAT method (f) ALT method

(g) water flow model (h) Bernsen method

Figure A.10: Example results of handwritten image on yellow envelope paper by pen in image set III - local methods



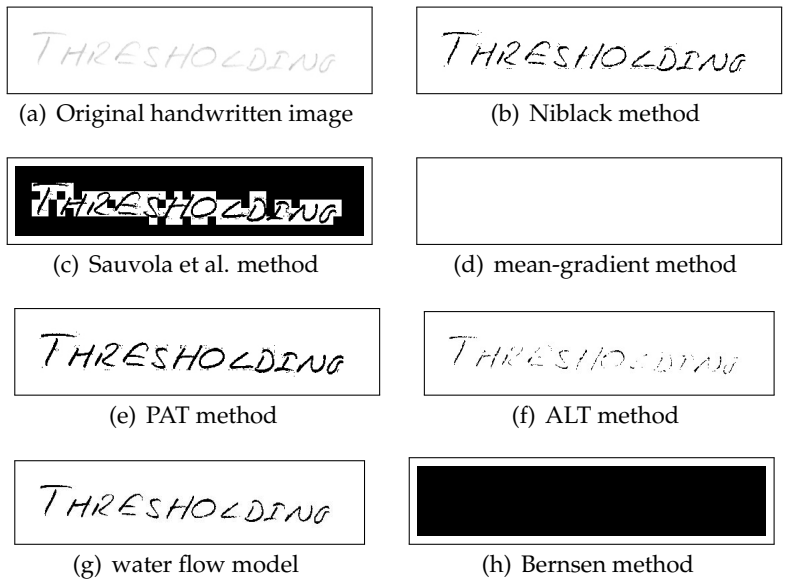
(a) Original handwritten image (b) Otsu method

(c) Kittler and Illingworth method (d) Yanni and Horne method

(e) Ramesh et al. method (f) MD method

(g) Kapur et al. method (h) Albuquerque et al. method

Figure A.11: Example result of pencil on white paper in image set III - global methods



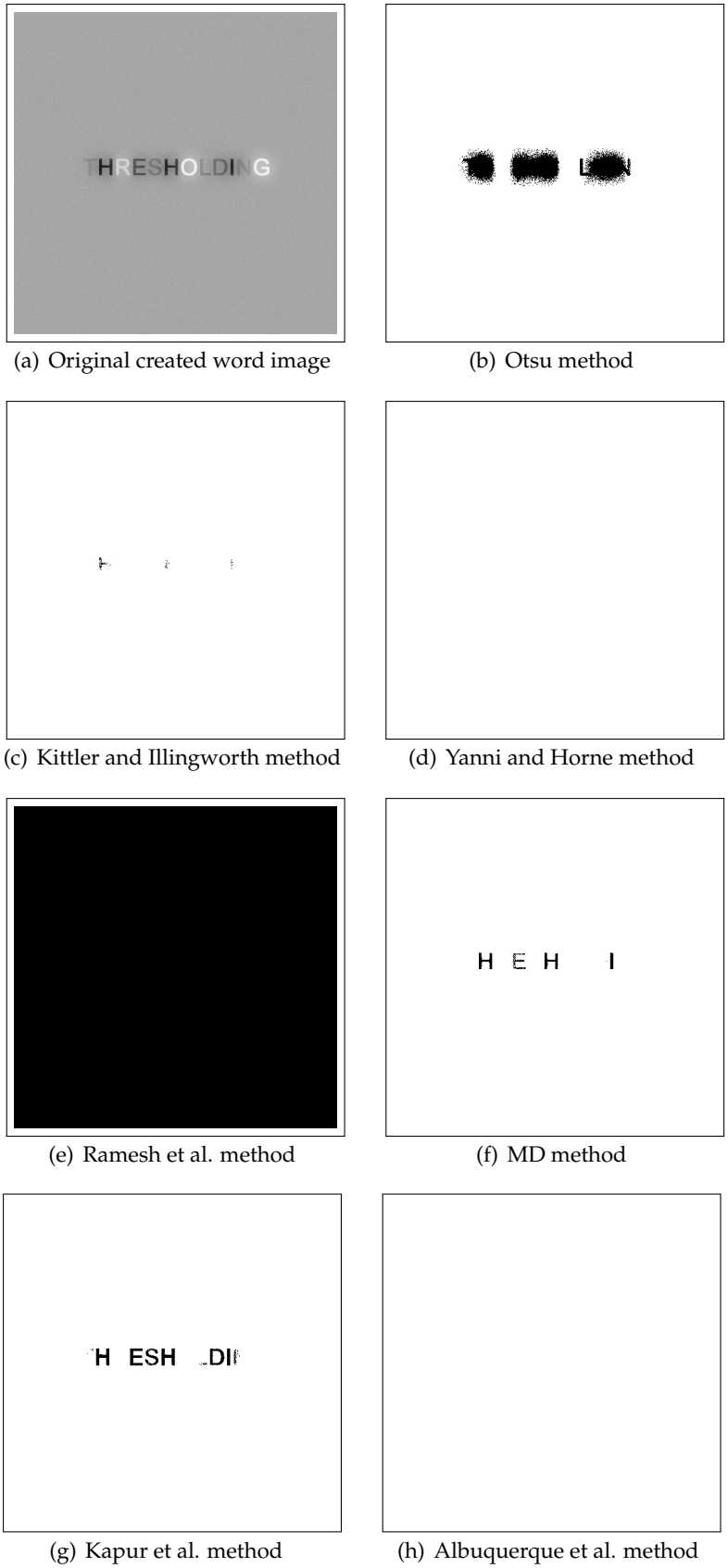
(a) Original handwritten image (b) Niblack method

(c) Sauvola et al. method (d) mean-gradient method

(e) PAT method (f) ALT method

(g) water flow model (h) Bernsen method

Figure A.12: Example result of pencil on white paper in image set III - local methods



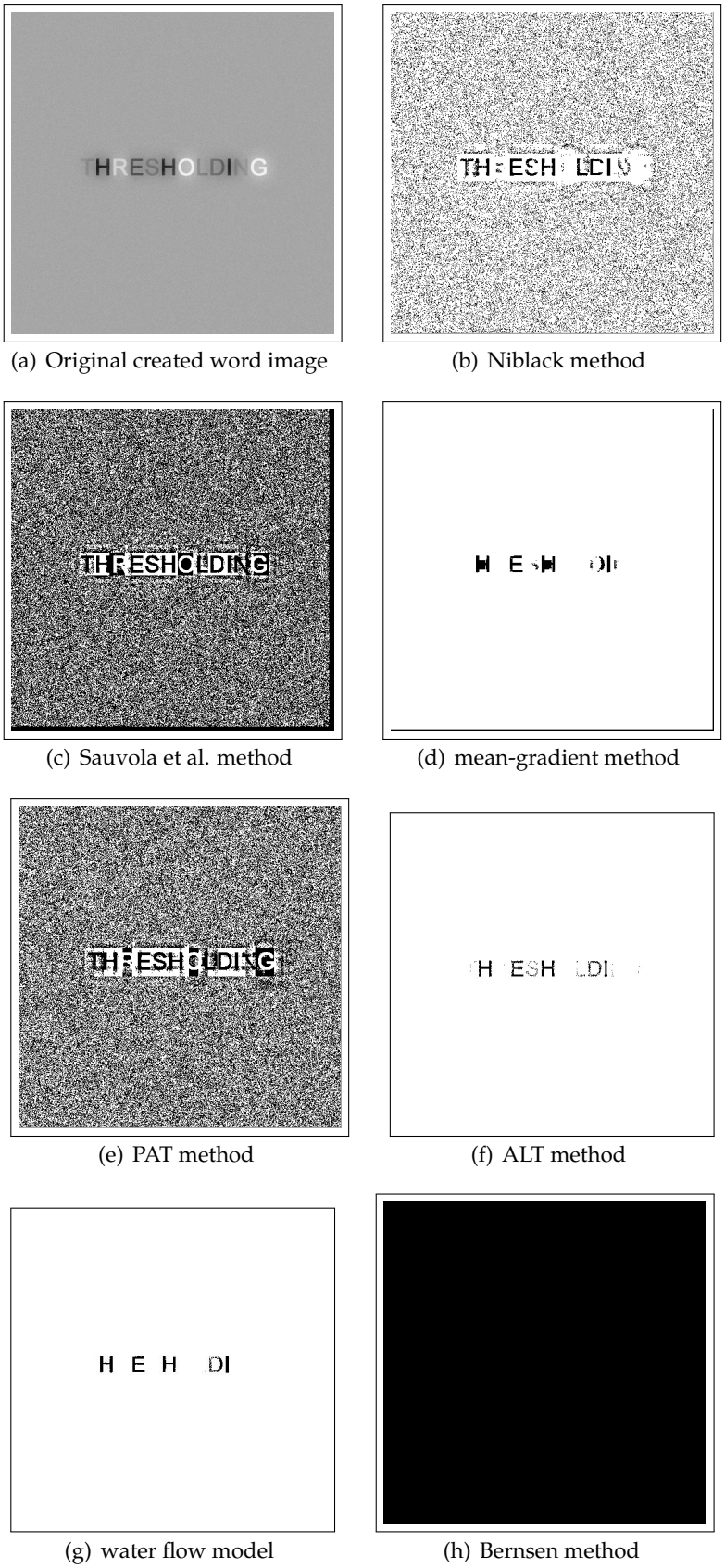
(a) Original created word image (b) Otsu method

(c) Kittler and Illingworth method (d) Yanni and Horne method

(e) Ramesh et al. method (f) MD method

(g) Kapur et al. method (h) Albuquerque et al. method

Figure A.13: Example result of artificially created text in image set II- global methods



(a) Original created word image (b) Niblack method

(c) Sauvola et al. method (d) mean-gradient method

(e) PAT method (f) ALT method

(g) water flow model (h) Bernsen method

Figure A.14: Example result of artificially created text in image set II- local methods

APPENDIX B

Flowcharts and Program Codes of MD and PAT Methods

Appendix B presents the source codes and the basic flowcharts of the proposed MD and PAT methods.

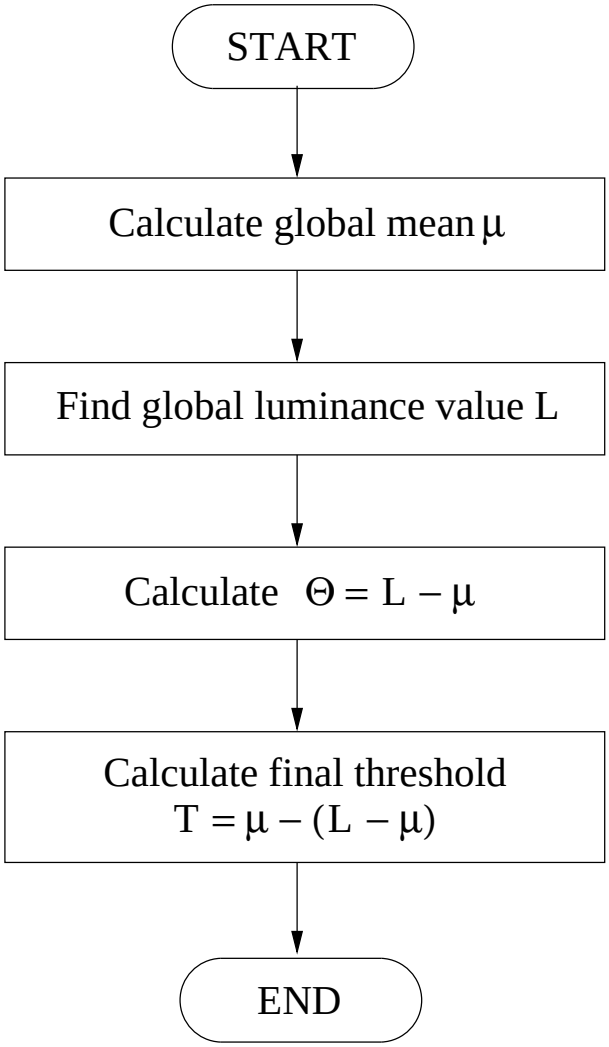
Table B.1: C Code for MD

```
for( y=0; y< ypixels; y=y+1)
{
   for(x=0;x<xpixels;x=x+1)
   {
      mean+= image[x][y];
      if(L<image[x][y])
      L=image[x][y];
   }
}
mean=mean/(ypixels*xpixels);
TH =(mean-((L-mean)));
for(y=0;y<ypixels;y=y+1)
{
   for(x=0;x<xpixels;x=x+1)
   {
      if(image[x][y]>abs(TH))
      image2[x][y]=255;
      else
      image2[x][y]=0;
   }
}
for (y=0;y<ypixels;y++)
{
   for (x=0;x<xpixels;x++)
   {
      fputc(image2[x][y],out3);
   }
}
```

■

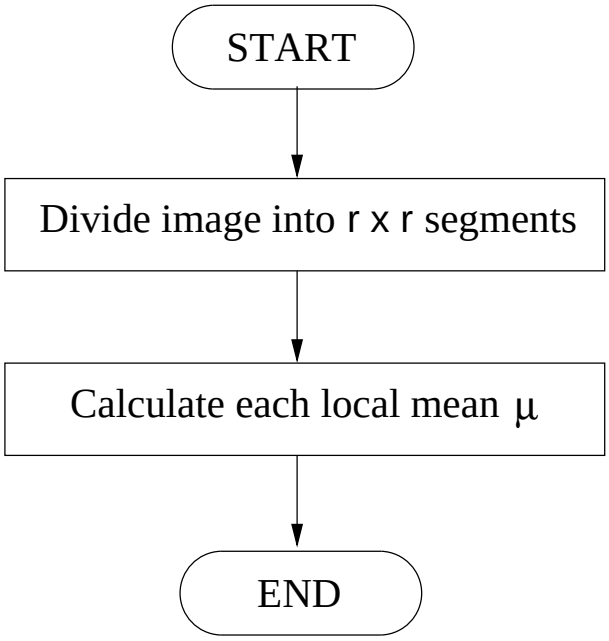
Table B.2: C Code for PAT

```
for( y=0; (y < ypixels - mask); y+=mask )
{
   for( x=0; (x < xpixels - mask); x+=mask )
   {   mean[i]=0;
      for( t=y; (t < y + mask); t++ )
      {
         for( z=x; (z < x + mask); z++ )
         {
            mean[i] += image[z][t];
         }
         mean[i] /= mask*mask;
         i++;
      }
   }
}
i=0;
for( y=0; (y < ypixels - mask); y+=mask )
   {
   for( x=0; (x < xpixels - mask); x+=mask )
   {
      for( t=y; (t < y + mask); t++ )
      {
         for( z=x; (z < x + mask); z++ )
         {
            if (image[z][t] < mean[i])
            {
               mean[z][t]=0;
            }
            else {
               image[z][t]=255;
            }
         i++;
      }
   }
}
for( y=0; (y < ypixels); y++ )
{
   for( x=0; (x < xpixels); x++ )
   {
      fputc( image[x][y], out );
   }
}
```



(a)

Figure B.1: MD Thresholding Method Flowchart



(a)

Figure B.2: PAT Method Flowchart