DEGRADED DOCUMENT IMAGE ENHANCEMENT USING GLOBAL THRESHOLDING

Niti Kamboj¹, Vinay Kumar²
¹department of Computer Science and Engineering
Ssgi, Yamunanagar, Kurukshetra University, Haryana, India
²Deputy Engineer, Department Naval System Bel (Bengaluru), Karnataka, India

Abstract: Recent years have witnessed the rapid growth of degraded images due to the increasing power of computing and the fast development of Internet. Because of this tremendous increase of quality of degraded images, there is an urgent need of image content description to facilitate automatic retrieval. Image is described by several low level image features, such as color, texture, shape or the combination of these features. Shape is an important low level image feature. In this thesis, enhancement of quality of degraded image using Global Thresholding. This thesis is primarily concerned with the extracting numerical features from binary images. Image processing is very vast field & one of the most important part of image processing is thresholding. Thresholding, which is an important pre-processing steps for the degraded image to enhance their quality & has been studies in relation to various images. There are different algorithms that have been used & studied for various factors of image analysis. The value of thresholding is based on which segmentation has been performed.

Keywords: Binarization, Noisy Documents, Threshold

I. INTRODUCTION

Degraded documents are archived and preserved in large quantities worldwide. Electronic scanning is a common approach in handling such documents in a manner which facilitates public access to them. Such document images are often hard to read, have low contrast, and are corrupted by various artifacts. Thus, given an image of a faded, washed out, damaged, crumpled or otherwise difficult to read document, one with mixed handwriting, typed or printed material, with possible pictures, tables or diagrams, it is necessary to enhance its readability and comprehensibility. Documents might have multiple languages in a single page and contain both handwritten and machine printed text. Machine printed text might have been produced using various technologies with variable quality. It is common for libraries to provide public access to historical and ancient document image collections. For such type of document images require specialized processing in order to remove background noise and become more legible. In this paper, we define a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. Firstly, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is reapplied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

Dealing with these types of problems, which are extensively present in historical documents, is important to improve their legibility and accessibility. This has special relevance in the context of digital libraries to provide public access to historical and ancient documents. In this paper, we define a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. Firstly, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is reapplied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

Dealing with these types of problems, which are extensively present in historical documents, is important to improve their legibility and accessibility. This has special relevance in the context of digital libraries to provide public access to historical and ancient documents. In this paper, we define a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. Firstly, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is reapplied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

Dealing with these types of problems, which are extensively present in historical documents, is important to improve their legibility and accessibility. This has special relevance in the context of digital libraries to provide public access to historical and ancient documents. In this paper, we define a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. Firstly, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is reapplied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

Dealing with these types of problems, which are extensively present in historical documents, is important to improve their legibility and accessibility. This has special relevance in the context of digital libraries to provide public access to historical and ancient documents. In this paper, we define a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. Firstly, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is reapplied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.
II. METHODOLOGY

A. Intermean Method
Let \( I = \text{gray level (0-255)} \) & \( y_i = \text{no. of pixels with gray level } I \) = height of level \( I \) in the histogram.

Three partial sums are calculated as:

\[
A_i = \sum_{l=0}^{j} y_i \\
B_j = \sum_{l=0}^{j} (i * y_i) \\
C_j = \sum_{l=0}^{j} (i * l * y_i)
\]

In this method \( t \) can be calculated as an integral part of \( Bn/An \). This is an initial guess for \( t \).

The mean gray levels in the two classes defined by the threshold below & above it are calculated as:

\[
\mu_t = Bt/At \\
\sqrt{t} = \frac{(Bn-Bt)}{(An-At)}
\]

\( t \) is calculated as:

\[
\sqrt{\mu} = \frac{(\mu_t - \sqrt{t})}{2}
\]

then & are recalculated & a new value of \( t \) is obtained. This process is repeated until convergence.

B. Minerror Method

In this method, histogram can be viewed as an estimate of the probability density function \( p(g) \) of the mixture of the population comprising the gray levels of objects & background pixels.

\( P(g/I) = \text{prob. Density function} \)
Two components \( 1 - 1, 2 \)
\( \mu \) = mean of mixture
\( \sigma \) = standard deviation
\( p_i = \text{a priori probability} \)

Supposing gray level data is thresholding at \( T \) then:

\[
P_i(T) = \sum_{g=a}^{b} h(g)
\]

\[
\mu_i(T) = \sum_{g=a}^{b} g * h(g)/p_i(T)
\]

\& \( \bar{\chi} \) & \( \bar{\sigma}(T) = \sum_{i=0}^{T+1} (g - \mu_i(T)) * h(g)/p_i(T) \)

Where \( a = 0 \) \( i = 0 \)
\( T+1 \) \( i = 2 \)
\( b = T \) \( i = 1 \)
\( n \) \( i = 2 \)

C. Moment Preserving Threshold

Color image processing by using binary quaternion-moment-preserving thresholding technique. This paper presents a new moment-preserving thresholding technique, called the binary quaternion moment preserving (BQMP) thresholding, for color image data. Based on representing color data by the quaternions, the statistical parameters of color data can be expressed through the definition of quaternion moments. Analytical formulas of the BQMP thresholding can thus be determined by using the algebra of the quaternions. The computation time for the BQMP thresholding is of order of the data size. By using the BQMP thresholding, quaternion-moment-based operators are designed for the application of color image processing, such as color image compression, multiclass clustering of color data, and subpixel color edge detection. The experimental results show that the proposed operator for proposed edge operator can detect the color edge at the subpixel level. Therefore, the proposed BQMP thresholding can be used as a tool for color image processing.

Mathematical Calculations: Let \( p_0 \) & \( p_1 \) denote the fractions of the below & above threshold pixels in \( f \) respectively.
The first three moments of \( g \) are just:

\[
\sum_{i=1}^{3} \sum_{j} z_j < 1
\]

\[
\mu_i = \sum p_j[z_j] \quad i = 1, 2, 3
\]

\[
\mu = \sum_{i=0}^{n} \sum_{g=a}^{b} g * h(g)/p_i(T)
\]

\[
Z_i = \text{gray values} \& \text{preserving the first three moments:}
\]

\[
m_i = \mu_i \quad \mu = 1, 2, 3
\]

\[
p_0 + p_1 = 1
\]

After solving the above we get:

\[
P_0 \text{ is obtained } \& \text{ corresponding value of } T \text{ is:}
\]

\[
(1/n) \sum p_i
\]

\[
z_i < 1
\]

D. Otsu Method:

Otsu’s method is used in computer vision to perform thresholding. If each pixel \( p \) of an image has been assigned an ‘interest’ score, \( f(p) \) then Otsu creates a histogram off over the image; and selects a threshold to maximize the between class variance. In statistics, a histogram is a graphical display of tabulated frequencies; a histogram is the graphical version of a table which shows what proportion of cases fall into each of several or many specified categories. The categories are usually specified as no overlapping intervals of some variable. The categories (bars) must be adjacent. The histogram is one of the seven basic tools of
quality control, which include the histogram, Pareto chart, check sheet, control chart, cause and effect diagram, flowchart and scatter diagram.

E. Flow Chart

III. RESULTS AND DISCUSSIONS

The propose work compare four algorithms which used to improvement the quality of degraded images using global threshold. The methods are Intermean method gives the result of improve the quality used threshold value 5, and Minerror method gives threshold value 8, and Moment preserving method improve the quality at threshold value 9, and last method Otsu method improve the quality at threshold value 12. Otsu’s method is used in computer vision to perform thresholding. If each pixel p of an image has been assigned an ‘interest’ score, I(p) then Otsu creates a histogram off over the image; and selects a threshold to maximize the between-class variance. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is reapplied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Certain kinds of histograms, however, help decide an appropriate thresholding method that offers good result for that type of images.

A. Intermean Method

Output 1
Enter the name of file with path and extension as BMP : c:\imagel.bmp
open file.bmp
size of image : 37526
offset1 : 0
offset2 : 0
offset start of image : 1078
offset of bmp header structure : 40
width is : 288
height is : 195
no of plane in image : 1
bytes per pixel : 0

Output 2

Output 3

Output 4

Output 5

www.ijtre.com Copyright 2014 All rights reserved.
B. Moment Preserving Thresholding

Output 1
Enter the name of file with path and extension as BF : c:\image1.bmp
open file.bin
size of image : 37526
offset : 0
offset_start : 8
offset_end : 15
width is : 256
height is : 196
no of pixel in image : 1
byte per pixel : 0

Output 2

Output 3

Output 4

Output 5

IV. CONCLUSIONS
The discussion of the various aspects has studies together to conclude which of these four algorithms considered is better the enhancement of the image. The algorithm used in this project solves the purpose of segmentation & enhance the quality of the image but different algorithms have different computational time. The user can use any of the algorithms depending on his ease to understand the concept. all the algorithms are applied on each part to increase its quality & in the output the gray level & the original image is shown with their respective histograms.otsu method best of blood images moment preserving best for noisy images.

REFERENCES