Abstract—In the process of improving degraded printed document images, segmentation is one of the challenging tasks due to various document degradations. Here we propose a new approach for the enhancement of degraded document image. It consists of an adaptive image contrast for document image binarization. Adaptive image contrast can be constructed using local image gradient and local image contrast. Further, the text stroke edge pixels are detected through the combination of binarized adaptive contrast map and canny’s edge map. The text within the document is then segmented by a local thresholding technique. Then a post processing is applied to improve the quality of degraded images. It works for different types of degraded document images. Then an artificial neural network is used for optical character recognition purpose.

Index Terms—Adaptive image contrast, Canny edge detection, Document image analysis, Document image binarization, Text stroke edge pixel

I. INTRODUCTION

Document binarization is the pre-processing step in most document analysis systems. The aim of document binarization is to convert the input grey scale or color document into a bi-level. Binarizing an image represents converting the image into black and white representation. i.e., intensity information will be reduced to two values respectively '0' and '1'. This representation is commonly used because most of the documents in practice have one color for text, and a different color for background [1].

As more and more text documents are scanned, accurate and fast document image binarization is becoming increasingly important. Though document image binarization has been studied for several years, the thresholding of degraded document images is still a challenging problem due to the high inter/intra variation between the text stroke edges in different document images [2]. Usually, it distinguishes text areas from background [3]. When processing degraded document images, binarization is not an easy job.

![Fig 1. Some examples of documents images of our test collection](image)

Thresholding plays a major role in binarization of document images. Thresholding can be classified into global thresholding and local thresholding [4]. Global binarization methods such as one proposed by Otsu is used to find a single threshold for the whole document [5]. Then each pixel is assigned to foreground or background based on this value [6]. Global binarization methods are very fast and less time consuming [7]. For many years, the binarization of grayscale images was based on the global thresholding algorithms [8]. If the illumination over the document is not uniform, global binarization methods tend to produce noise along the page borders.

Local thresholding binarization algorithms employ methods based on clustering procedures, entropy, and neighbourhood information [9]. Some techniques are Niblack’s that uses local mean and standard deviation [10], Bernsen’s which calculates local thresholds using neighbours [11], and Sauvola’s which applies two different algorithms to determine a different threshold for each image and finally binarizes the document image [12].

II. PROPOSED APPROACH

![Fig. 2 Block diagram of Image binarization and character recognition](image)

Fig 2. Block diagram of proposed method

A. Adaptive Contrast Image

An adaptive image contrast can be constructed by using local image contrast and local image gradient. The local image gradient is widely used for edge detection and it can be used to detect the text stroke edges of the effectively that have a uniform document background. It often detects many non stroke edges from the background of degraded document that often contains certain image variations due to noise, bleed through, uneven lighting etc [13]. To extract only the stroke
edges properly, the image gradient should be normalized to compensate the image variation within the document background. The local image gradient is defined as follows:

$$C_g(i,j) = I_{max}(i,j) - I_{min}(i,j)$$  \hspace{1cm} (1)

Where \(I_{max}(i,j)\) and \(I_{min}(i,j)\) refer to the maximum and minimum intensities within the neighborhood window and \(C_g(i,j)\) denotes the local image gradient.

The image contrast is widely used for edge detection and it can be used to detect the text stroke edges of the document images effectively that have a uniform document background.

$$C(i,j) = I_{max}(i,j) - I_{min}(i,j)$$  \hspace{1cm} (2)

Where \(I_{max}(i,j)\) and \(I_{min}(i,j)\) refers to the maximum and minimum intensities within the neighborhood window and \(C(i,j)\) denotes the local image contrast. In the implemented system, the neighborhood window size is 3 * 3. The term \(\epsilon\) is a positive but infinitely small number, which is added in case if the local maximum is equal to 0.

To overcome the over-normalization problem, combine the local image contrast with local image gradient and derive an adaptive local image contrast as follows:

$$C_a(i,j) = \alpha C(i,j) + (1 - \alpha)C_g(i,j)$$  \hspace{1cm} (3)

Where \(C_g(i,j)\) refers to the local image gradient and \(C(i,j)\) denotes the local contrast and is normalized to \([0, 1]\. The local windows size is set to 3 * 3. The term \(\epsilon\) is a positive but infinitely small number, which is added in case if the local maximum is equal to 0.

The value of \(\alpha\) can be expressed by a power function as follows:

$$\alpha = \left(\frac{\text{Std}}{120}\right)^\gamma$$  \hspace{1cm} (4)

Where \(\gamma\) is a pre-defined parameter and \(\text{Std}\) denotes the document intensity standard deviation. \(\gamma\) can be selected from \([0, 1]\, where the power function becomes a linear function when \(\gamma = 1\. Therefore, the local image gradient will play the major role in Equation 3 when \(\gamma\) is large and the local image contrast will play the major role when \(\gamma\) is small.

**B. Text Stroke Edge Pixel Detection**

Image contrast construction is used to detect the stroke edge pixels properly. Constructed contrast image has a clear bimodal pattern. So the text stroke edge pixels are detected using Otsu’s global thresholding method. Then the combined edge map can be constructed through the combination of binarized adaptive contrast map and Canny’s edge detector [1]. Canny’s edge detector has a good localization property that it can mark the edges close to real edge locations.

Canny edge detector uses a list of criteria to improve current methods of edge detection. The most important is low error rate. The second criterion is good localization property. Third criterion is to have only one response to a single edge [14].

**C. Local Threshold Estimation**

The document image text can be extracted based on the detected text stroke edge pixels as follows

$$R(x, y) = \begin{cases} 1 & (x,y) \leq E_{\text{mean}} + \frac{E_{\text{std}}}{2} \\ 0 & \text{otherwise} \end{cases}$$  \hspace{1cm} (5)

Where \(E_{\text{std}}\) and \(E_{\text{mean}}\) are the standard deviation and mean of the intensity of the detected text stroke edge pixels within a neighbourhood window W.
Further, it is an addition over basic roulette selection.

Fig 7. Local threshold estimation

The window should be at least larger than the stroke width in order to contain stroke edge pixels. The size of the neighbourhood window W can be set based on the stroke width of the document image. In the edge width estimation algorithm, first the edge image is scanned horizontally row by row. Then select the edge pixels, which are labelled 0 and the pixels next to them are labelled to 1 in the edge map. So improperly detected edge pixels are removed. In the remaining edge pixels in the same row, the two adjacent edge pixels are matched to pairs and the distance between them are calculated. Then the stroke edge width EW can then be estimated by using the most frequently occurring distances of the adjacent edge pixels.

D. Post-Processing

The binarization result can be further improved by using certain domain knowledge. First, the foreground pixels that do not connect with other foreground pixels are filtered out. Second, the neighbourhood pixel pair that lies on symmetric sides of a text stroke edge should belong to different classes. One pixel of the pixel pair is therefore labelled to the other category if both of the two pixels belong to the same class.

Further, it is an addition over basic roulette selection.

Fig 8. Post processing output

E. Optical character recognition

Optical Character Recognition is a process of converting printed document or scanned page to ASCII character. The document image itself can be either machine printed on handwritten, or the combination of two. Computer system using such an OCR system can improve the speed of input operation. Recognition of printed characters is itself a challenging problem since there is a variation of the same character due to change of fonts or introduction of different types of noises [14].

Artificial neural networks are commonly used to perform character recognition due to their high noise tolerance. In optical character recognition feature extraction step is very important. At present the software does perform well either in terms of speed or accuracy but not better.

Neural Network is formed by interconnecting a number of elements called neurons in different layers together. Consider a system in which a neural network is formed using three layers – Input, Hidden and Output. where Input layer is used to fed all the input data into the network, which is followed by a hidden layer of neurons for further processing of data and finally the output layer of neurons for calculating the desired response of the Neural network.

Fig 9. Training window of neural network

Neural network is used for the classification of the character. First we create the neural network, then train the network with training data. A feed-forward back propagation network is created. newff() function is used to create a feed forward network. net = newff(Input, target, [hi, TF, BTF], creates a neural network. TF means the transfer function and BTF is back propagation training function. Then we simulate the network with test data which we obtained from post processing procedure.

Fig 10. Optical Character Recognition

III. Conclusion

The method is aimed at the development of a binarization model to improve the quality of a degraded document images. This proposed technique is simple and robust, only few parameters are used. It works for different kinds of degraded document images. It makes use of the local image contrast that is evaluated based on the local maximum and minimum. The method has been tested on the various printed images and it provide substantial improvement. Our method decreases the background degradation effect and improves the quality of the image. Then an artificial neural network is used for optical character recognition purpose.
ACKNOWLEDGMENT

First and foremost, praises and thanks to the God, the Almighty, for His showers of blessings throughout my project work to complete it successfully. First of all I would like to acknowledge with much appreciation the Pivotal role of my guide, Miss Manju Thomas, who gave his full support to complete my work. This work would not have been possible without his guidance, support and encouragement. Under his guidance I successfully overcame many difficulties and learned a lot.

REFERENCES


Anju Gopi received the degree in Electronics and Communication Engineering from Mahatma Gandhi University, Kerela, India.

Manju Thomas received the graduation in Electronics and Communication Engineering from MG university, kerela, india and post graduation in Signal Processing from Cochin University Of Science And Technology (CUSAT), Kerela, India. She now works as an assistant professor at CUSAT.