Chapter 1: Introduction

The recent technological boom and development of digital devices has made multimedia such as image and video, one of the major carriers of information. More and more researchers are attracted to design an automated system for text extraction as the need to index, browse and retrieve the information present in an image or video is increasing. One, such as created application is the portable managing of bank accounts using the mobile banking application provided by the banking institutions that allows that permits the clients to play out the trades even on transferring the picture of the check to the server. These sorts of uses rely on upon a Textual Information Extraction framework which can recognize, limit and concentrate the content data contained in an image. An interpretation camera is another such application that obtains the picture, recognize content from it and afterward decipher it in required dialect the local language.

Numerous text detection algorithms have been proposed on text detection and localization. All these algorithms broadly belong to either edge-based techniques or connected-component based techniques or texture based techniques. Distinctive methods were expressed in the past for perceiving content information in a picture. Every technique suits for some predefined set of images like images with low light, images with sharp edges, and so forth. For such different types of images the output produced by algorithm may vary due to dimensions of image, colour, contrast, and orientation. In natural images, detection of text has some challenges because of variety in content size and alignment

In Human society, the dominant information medium is document. Hence document image analysis is a considerable research task in the area of image processing and pattern recognition. The text extraction is a segmentation of text from degraded background which is normally used to asset boundaries lines, curves etc of document image. This segmentation of degraded text document image have certain issues to solve such as how to extract clear text strings on the leading side of page from the seriously sipping, dominating, overlapping and interfering images originating from the back side. This document degradation can be restored using image binarization technique.

In general, scanned documents include text, line-drawings and graphic regions. It can also be considered as mixed type documents. In most of the practical applications, we need to recognize or improve the content of the document. In such cases, it is preferable to convert the documents into a binary form. Thus Image binarization plays a key role in the field of Image Processing. Document images may be degraded due to the poor quality of paper, ink plot, fading, document aging. Visibility of document degrades due to the scanning and printing that means it is very difficult to understand them. Degraded document image restoration plays a very important role in enhancing the degraded noise in images

Document binarization is typically executed in the pre-processing phase of several document image processing associated fields such as optical character recognition (OCR) and document vision retrieval. Image binarization exchanges a picture up to 256 gray levels to a black and white picture. The simplest approach to apply image binarization is to prefer a threshold value and organize all pixels with values above this threshold as white and all other pixels as black. Examples of thresholding applications are document image analysis, where the goal is to extract printed characters, logos, graphical content,: map processing, where lines, legends, and characters are to be found scene processing, where a target is to be detected and quality inspection of materials where defective parts must be delineated.

In document binarization process, thresholding technique is well-known technique. Local and Global Thresholding are two important types of the thresholding process. Global thresholding has good performance because there is a good separation between the foreground and the background. On the other hand, local area information may guide the threshold value for each pixel in local (adaptive) thresholding techniques. Due to high inter/intra change enclosed in the text stroke and document background across different document images, a single thresholding technique is not suitable to solve the different issues. On a large scale Binarization techniques are classified as global and local thresholding.

Global thresholding is preferred when the images are having equal contrast spreading on background and foreground. Local adaptive thresholding is used for restoring the foreground pixel from document image with background large contrast variation and noise. Earlier binarization is carried out by using global thresholding algorithms. In 1979 Otsu proposed adaptive global thresholding algorithm. This method, evaluate a distinct threshold value for the entire document, the gray scale value is use to assign each pixel of document image as foreground or background [1]. Global thresholding methods are fast and give great results but it tends to produce noise along the page borders and it takes too much time for multilevel threshold selection. The local thresholding techniques are use to overcome this problem. In this techniques, on the basis of grayscale value of neighbouring pixels, select a distinct threshold value for each pixel and the threshold value is calculated using the statistical approach. These techniques have been widely used in document image analysis because they have a effective performance in extracting the character strokes from an image that has spatially uneven gray levels due to degradations. To enhance the degraded document, its image is binarized before processing it. It is nothing but segmenting the document background and the foreground text. For the document image processing task the perfect document image binarization technique is the best option. After many types of work done in the binarization technique, the thresholding of degraded document images is found to be a challenging task because of the high inter/intra variation between the text stroke and the document background across various document images. The stroke brightness, document background, stroke width and stroke connection vary in the handwritten text within the degraded documents

Adaptive image binarization is required where an optimal thresholding is selected for each picture area. Thresholding is the easiest technique of image segmentation, from gray scale image. Thresholding can be used to generate binary images. Document images frequently experience from dissimilar variety of degradation that makes the document binarization a challenging task.

1.1 Importance of Binarization

Binarization is a significant phase in all systems of images processing and analysis. It has objective to reduce the amount of information present in the image and to lookout only applicable data which allows us to apply straightforward analysis technique performance of document analysis depends on binarization algorithm. It should on the one hand: conserve the most of information and details present in the entered picture and then it should remove the noise superposed to the original picture.[15]

1.2 Problem Definition:

Image binarization as our main objective we present here the major problem that we address in the thesis.

1. First order derivative operators are discrete differentiation operator which functions by calculating the gradient of the image intensity function and makes use of the maximum directional gradient. The Prewitt masks are simpler to implement but are very sensitive to noise also Prewitt operator is more sensitive to **horizontal** and **vertical edges**. In general, the first order derivative operators are very sensitive to noise and produce thicker edges and are

not efficient for localization of the edge. On the other hand, Laplacian of Gaussian belongs to a class of second order derivatives, i.e. the second derivative of Gaussian. The zero-crossings of the second-order derivatives are good for localization of the edge. The DoG function returns a nearby estimate the scale-normalized LoG, as reviewed by Lindeberg (1994). Lindeberg demonstrated that the normalization of the Laplacian is required for genuine scale invariance. Also the use of LoG is computationally expensive [23]. But processing time while using DoG is less and produce a strong response for poorly detected edges because they have fixed characteristics in all directions.

Our aim is to obtain the text region from the natural scene pictures which may include slanted and curved text with minimum computational cost using Difference of Gaussian method. We developed a novel approach of text extraction which uses modified Difference of Gaussian as edge detector, Niblack's Thresholding for binarization [17] [18] and morphological operations for post processing and also compares the proposed method with the two other edge detectors namely, Gaussian Edge Detector and Prewitt Edge Detector [18] by evaluating the performance parameters.

2. Generally the text and back ground is separated by collating the intensity of each pixel with the mean grey value and fixing a threshold for separating text and background. If the intensity of text is low (due to fading or bleed through), the threshold technique fails. Also, Color to gray conversion allows the application of single-channel algorithms on color images. Unfortunately, since the color-to-gray conversion is to convert a 3D vector to a 1D gray-scale value, it is essentially a dimensionality reduction process, hence inevitably suffers from information loss. The baseline and widely used method in decolorizing an input image is to extract its luminance channel (e.g., CIE Y). If the original image is in the RGB format, the luminance can be calculated by linearly summing its R, G and B channels with a defined weight (e.g., the rgb2gray function in Matlab). However, using luminance channel image alone cannot faithfully represent structures and contrasts in some color images such as those having iso-luminant regions. [11]

To filters noise and enhances the text without introducing any unwanted noise and we used $p_i \log p_i$ as compared to p_i in the Nick's algorithm which helps in the filtering of the noise. For lower values of p_i [12], the function $p_i \log p_i$ exhorts a non-linear nature due to which a more appropriate value is assigned to the threshold which indeed helps to differentiate black pixels or text better as compared to p_i. Also to address the problem of low contrast on iso-luminant regions we used Gradient Correlation Similarity for Efficient Contrast Preserving Decolorizationin place of conventional rgb to gray conversion in pre-processing step. [11]

1.3 Overview of work done

1.3.1 Comparison of various standard binarization techniques

In the first section of the thesis work we have performed some standard binarization techniques namely Nick's Thresholding method, Niblack's Thresholding algorithm, Sauvola's binarization method, Otsu's method and Bernsen's binarization method on document images. All these methods were programmed and experimented on standard DIBCO-2011 data set by help of MATLAB tool. Various performance parameters are calculated from the resultant images of these methods e.g. Precision Rate, Recall Rate, DRD, NRM, fMeasure, Sensitivity, Selectivity etc which are shown in the table given in Results section.

1.3.2 Novel Approach for Text Extraction from Natural Scene Images

In this part of thesis work we developed a method to extract text regions or binarize Natural Scene Images. Distinctive methods were expressed in the past for perceiving content information in a picture. Every technique suits for some predefined set of images like images with low light, images with sharp edges, and so forth. For such different types of images the output produced by algorithm may vary due to dimensions of image, colour, contrast, and orientation. In natural images, detection of text has some challenges because of variety in content size and alignment. Our aim is to obtain the text region from the natural scene images which may include slanted and curved text with minimum computational cost using Difference of Gaussian method. This section proposes a novel approach of text extraction which uses modified Difference of Gaussian as edge detector, Niblack's Thresholding for binarization [17] [18] and morphological operations for post processing and also compares the proposed method with the two other edge detectors namly, Gaussian Edge Detector and Prewitt Edge Detector [18] by evaluating the performance parameters. Experiments on the dataset show that the promising result has been achieved by our proposed method.

1.3.3 A Novel Binarization Technique by Application of Gradient Correlation Similarity Measure for Contrast Preserving and Log based Binarization

Generally the text and back ground is separated by collating the intensity of individual pixel with the mean grey value and fixing a threshold for separating text and background. If the intensity of text is low (due to fading or bleed through), the threshold technique fails. Also, Color to gray conversion allows the application of single-channel methods on color images. Unfortunately, since the color-to-gray conversion is to convert a 3D vector to a 1D gray-scale value, it is essentially a dimensionality reduction process, hence suffers from information loss. The baseline and widely used technique for decolorizing an input image is to extract its luminance channel (e.g., CIE Y) [11]. If the original image is in the RGB format, the luminance can be obtained by linearly summing its R, G and B channels with a defined weight (e.g., the rgb2gray function in Matlab). However, using luminance channel image alone cannot faithfully represent structures and contrasts in some color images such as those having iso-luminant regions.

To filters noise and enhances the text without introducing any unwanted noise and we used $p_i \log p_i$ as compared to p_i in the Nick's algorithm which helps in the filtering of the noise. For lower values of p_i, the function $p_i \log p_i$ exhorts a non-linear nature due to which a more appropriate value is assigned to the threshold which indeed helps to differentiate black pixels or text better as compared to p_i [12]. Also to address the problem of low contrast on iso-luminant regions we used Gradient Correlation Similarity for Efficient Contrast Preserving Decolorizationin place of conventional rgb to gray conversion in pre-processing step.[11]

In the proposed method we used GCS based decolorization for Color to gray conversion in pre-processing step and log based Nick's Thresholding for the Binarization of document images.

The proposed method is programmed and experimented on standard DIBCO-2011 data set by help of MATLAB tool. Various performance parameters are calculated from the resultant images of these methods e.g. Precision Rate, Recall Rate, DRD, NRM, fMeasure, Sensitivity, Selectivity etc which are shown in the table given in Results section.

1.4 Thesis Layout

1.4.1. Introduction:

This chapter contains an introduction to the problem explored in the thesis. We provide the motivation behind exploring the problem of binarization in Natural scene and Document images and outline the contributions of the dissertation.

1.4.2. Related Work:

This chapter includes the work done in past and We also highlight the inadequacies and deficiencies in the current techniques.

1.4.3. Comparison of Various Standard Binarization Techniques:

In this chapter of the thesis work we have explained some standard binarization techniques namely Nick's Thresholding method, Niblack's Thresholding algorithm, Sauvola's binarization method, Otsu's method and Bernsen's binarization method on document images. Various performance parameters are calculated from the resultant images of these methods e.g. Precision Rate, Recall Rate, DRD, NRM, fMeasure, Sensitivity, Selectivity etc which are shown in the tables 1-5 given in Results section. Also merits and demerits of each algorithm is stated in subsequent table.

1.4.4 A Novel Approach for Text Extraction from Natural Scene Images:

Digital images acquired by the camera have useful information of interest. Image retrieval systems provide a platform for retrieving the information for content based retrieval systems and image search engines. This chapter addresses the problem of text region extraction for slanted images and presents a novel way to deal with concentrate content areas in a picture by application of Difference of Gaussian as a second order derivative edge detector combined with Niblack's Thresholding. Detailed comparison of existing algorithms using edge detectors is done with the stated method. The proposed algorithm is tested on an image with various content size, textual style styles and content dialect. Execution is assessed in view of precision rate and recall rate for proposed strategy on sample data set and ICDAR 2015 data set.

1.4.5 A Novel Binarization Technique by Application of Gradient Correlation Similarity Measure for Contrast Preserving and Log based Binarization

In this chapter we proposed a binarization method incorporating on GCS decolorization for Color to gray conversion in pre-processing step and log based Nick's Thresholding for the Binarization of document images. The proposed method is programmed and experimented on standard DIBCO-2011 data set by help of MATLAB tool. Various performance parameters are calculated from the resultant images of these methods e.g. Precision Rate, Recall Rate, DRD, NRM, fMeasure, Sensitivity, Selectivity etc which are shown in the table given in Results section.

1.4.6 Conclusion:

In this chapter we present the summary of the work done. We also outline the future scope of this work.

Chapter 2: Related Work

Binarization is an essential step for document image analysis. In general, different available binarization techniques are implemented for different types of binarization problems.

Many thresholding techniques have been reported for document image binarization. Adaptive thresholding [6] estimates a local threshold for each document image pixel and it is often a better approach to deal with different variations within degraded document images. the early window-based adaptive thresholding techniques [3], [4] estimate the local threshold by using the mean and the standard variation of image pixels within a local neighbourhood window. The thresholding performance depends heavily on the window size and hence the character stroke width and this is the main drawback of window-based thresholding technique. Other approaches have also been reported, including background subtraction, texture analysis, recursive method, decomposition method, contour completion, Markov Random Field , matched wavelet, cross section sequence graph analysis , self-learning, Laplacian energy user assistance and combination of Binarization techniques. These methods combine different types of image information and domain knowledge and are often complex. The local image contrast and the local image gradient are very useful features for segmenting the text from the document background because the document text usually has certain image contrast to the neighbouring document background. They are very effective and have been used in many document image binarization techniques.

In [7], a learning framework for the optimization of the binarization methods is introduced, which is designed to determine the optimal parameter values for a document image. The framework works with any binarization method performs three main steps: extracts features, estimates optimal parameters, and learns the relationship between features and optimal parameters. An approach is proposed to generate numerical feature vectors from 2D data. The statistics of various maps are extracted and then combined into a final feature vector, in a nonlinear way. The optimal behaviour is learned using support vector regression (SVR). The experiments are done using grid-based Sauvola's method and Lu's method on the DIBCO2009 and DIBCO2010 datasets.

A pixel-based binarization evaluation methodology for historical handwritten/machine-printed document images is presented in [10]. In the evaluation scheme in [6], the recall and precision evaluation measures are properly modified using a weighting scheme that diminishes any potential evaluation bias. Additional performance metrics of the proposed evaluation scheme consist of the percentage rates of broken and missed text, false alarms, background noise, character enlargement and merging. The validity of the method is justified by several experiments conducted in comparison with other pixel-based evaluation measures.

An image binarization technique is proposed in [27] for degraded document images that takes into consideration the adaptive image contrast. The adaptive image contrast is a combination of the local image contrast and the local image gradient that is tolerant to text and background variation caused by different types of document degradations. An adaptive contrast map is first constructed for an input-degraded document image. The contrast map is then binarized and combined with Canny's edge map to identify the text stroke edge pixels. The document text is further segmented by a local threshold that is estimated based on the intensities of detected text stroke edge pixels within a local window. It has been tested on three public datasets achieving accuracies of around 90 %. There are many challenges addressed in handwritten document image binarization, such as faint characters, bleed-through, and large background ink stains. Usually, binarization methods cannot deal with all the degradation types effectively. Motivated by the low detection rate of faint characters in binarization of handwritten document images, a combination of a global and a local adaptive binarization method at connected component level is proposed in [17] that aims in an improved overall performance. Initially, background estimation is applied along with image normalization based on background compensation. Afterward, global binarization is performed on the normalized image. In the binarized image, very small components are discarded and representative characteristics of a document image such as the stroke width and the contrast are computed. Furthermore, local adaptive binarization is performed on the normalized image taking into account the aforementioned characteristics. Finally, the two binarization outputs are combined at connected component level. Authors report good performance after extensive testing on the DIBCO series datasets which include a variety of degraded handwritten document images.

The majority of binarization techniques are complex and are compounded from filters and existing operations. However, the few simple thresholding methods available cannot be applied to many binarization problems. In [18], a local binarization method is presented based on a simple, novel thresholding method with dynamic and flexible windows. The method is tested on selected samples of DIBCO 2009 benchmark dataset.

Prewitt edge detection and Gaussian edge detection belong to the class of first order derivative. The Prewitt operator is a discrete differentiation operator which functions by calculating the gradient of the image intensity function and makes use of the maximum directional gradient. The Prewitt masks are simpler to implement but are very sensitive to noise also Prewitt operator is more sensitive to horizontal and vertical edges. In general, the first order derivative operators are very sensitive to noise and produce thicker edges and are not efficient for localization of the edge[25][26]. In [24], U. Bhattacharya, S. K. Paruri & S. Mondal et al extricated associated segments (both highly contrasting) from the twofold picture. At that point, morphological opening operation alongside an arrangement of criteria is connected to concentrate features of Devanagari or Bangla text. Further, in [28], H. Raj & R. Ghosh et al used different geometrical properties for identifying the location of the whole-text part related to the detected headlines form morphological operations.

On the other hand, Laplacian of Gaussian belongs to a class of second order derivatives, i.e. the second derivative of Gaussian. The zero-crossings of the second-order derivatives are good for localization of the edge. The DoG function returns a nearby estimate the scale-normalized LoG, as reviewed by Lindeberg (1994). Lindeberg demonstrated that the normalization of the Laplacian is required for genuine scale invariance. Also the use of LoG is computationally expensive. But processing time while using DoG is less and produce a strong response for poorly detected edges because they have fixed characteristics in all directions.[25][26] [27]

Chapter 3: Comparison of Various Standard Binarization Techniques

In this chapter of the thesis work we have explained some standard binarization techniques namely Nick's Thresholding method, Niblack's Thresholding algorithm, Sauvola's binarization method, Otsu's method and Bernsen's binarization method on document images. Various performance parameters are calculated from the resultant images of these methods e.g. Precision Rate, Recall Rate, DRD, NRM, fMeasure, Sensitivity, Selectivity etc which are shown in the table given in Results section. Also merits and demerits of each algorithm is stated in subsequent table.

3.1 Thresholding Methods:

3.1.1. Nick's Thresholding Technique:

Nick algorithm (6) calculates threshold using the following equation.

$$T_{Nick} = \mu + k\sqrt{\left[\frac{1}{N}\left(\sum {p_i}^2 - \mu\right)\right]} = \mu + k\sigma_Y \quad (1)$$

This equation gives a substantial threshold value for differentiating lower intensity pixels from background noise. But it introduces unwanted noise in the image. Also there is a lack of connectivity in the text.[6][7]

Nick's binarization algorithm is an improvement of Niblack method where it works very well for most degraded document. This technique solves the issue of noise in white pages and low contrast problem.

3.1.2. Nibkack's Thresholding Algorithm:

The Niblack's algorithm estimates a threshold value to individual pixel by scanning a rectangular window over the grayscale image [3]. The size of rectangular window can be adjusted. The threshold value is ascertained utilizing the local mean and standard deviation of the considerable number of pixels in the window and is computed by taking after conditions :

$$T_{Niblack} = M + k * SD \quad (2)$$

$$T_{Niblack} = M + k * \sqrt{\frac{\sum(p_i - m)^2}{NP}} \quad (3)$$

Here NP is the total no of pixels a grayscale image has, T represents the threshold value, M is the mean of gray levels of pixels and k is fixed depending upon the noise in the background and is -0.2

.

3.1.3. Sauvola's Binarization Algorithm:

The Sauvola algorithm is alteration of niblack algorithm. It asserts to advance niblack's technique by calculating the threshold using the forceful variety of picture gray value standard deviation R:

$$T_{Sauvola} = m * (1 - k * (1 - \frac{S}{R}) \qquad (4)$$

Sauvola's method wins over the background noise problems but creates thinned and broken characters [2][6][31]. This technique is an advancement of Niblack's method. This technique solves the issue of presence of large amount of noise in the background areas. Threshold value can be computed by using the dynamic range of images gray-scale standard deviation. However, this method has some problems also, when the text pixel value close to the foreground image.

3.1.4. Bernsen's Thresholding Algorithm:

It is an adaptive local technique of which the threshold is designed for each pixel of the image. For individual pixel of coordinates(x, y), the threshold is given by[9][13]:

$$T(x, y) = \frac{z_{low}+z_{high}}{2} \qquad (5)$$

z_{low} and z_{high} are the minimum and the maximum gray level in a squared window r*r centered more than the pixel (x, y).If the distinction quantity is lesser than a threshold 1, then the neighbourhood contains a single class: background or text.

3.4.5. Otsu's Global Thresholding Method:

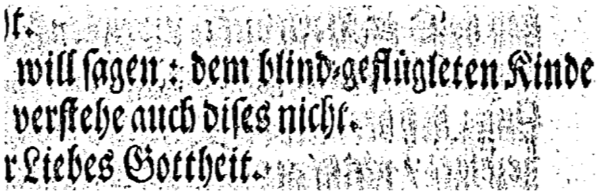
Otsu method is solitary of the well-known global methods. This technique discovered the threshold T which split the gray level histogram into two segments. The computation of inter-classes or intra classes variances is based on the normalized histogram of the image H=[h0.......h255]where {hi=1}.Otsu method is apply to routinely execute clustering-based image thresholding. In this we thoroughly look for the threshold that minimizes the intra-class variance distinct as a weighted sum of variances of the two classes:

$$\sigma^2 prb(t) = prb_1(t)\sigma_1^2 + prb_2(t)\sigma_2^2 \qquad (6)$$

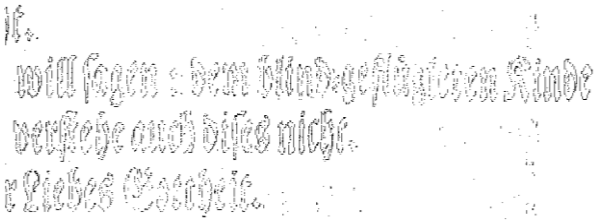
Here prb are the probabilities of the two classes divided by threshold and variances of the classes. The class probability and class means can be computed iteratively.[1][13][31]

3.2 Result and Discussion:

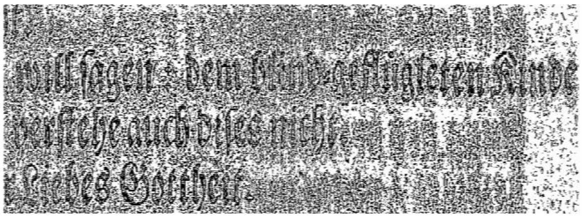
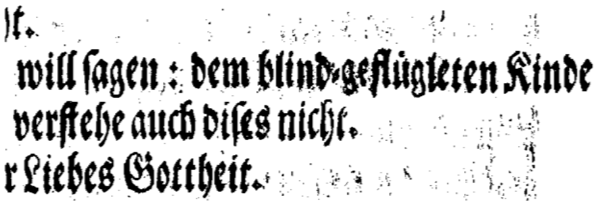
The experiment is done on more than 100 images and DIBCO2011 dataset. The result for some test images is shown below. The performance of the algorithm is validated based on Precision rate, Recall rate, DRD, NRM, fMeasure, Specificity, Selectivity etc which are shown in table below.

(a) (b)

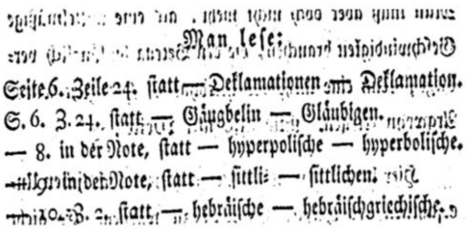



(c) (d)

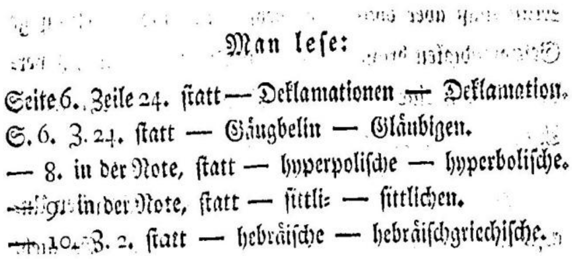
(e) (f)

Fig.1 (a) Original Image “PR1.jpg” (b) Nick’s Thresholding Result (c) Niblack’s Algorithm Result (d) Sauvola Thresholding Result (e) Bernsen’s Binarization Result (f) Otsu’s Thresholdig Result

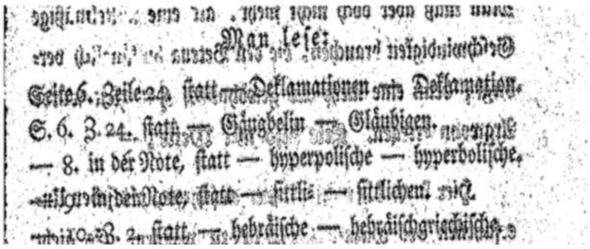



(a) (b)

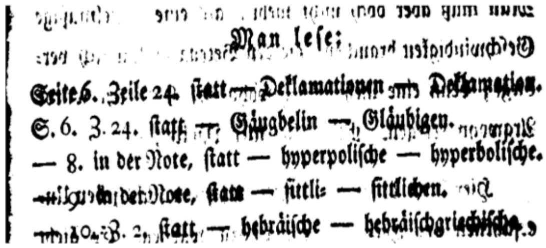
(c)



(d)

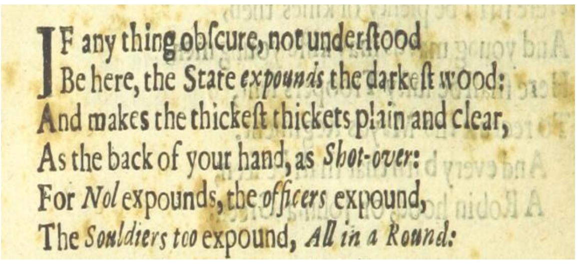


(e)

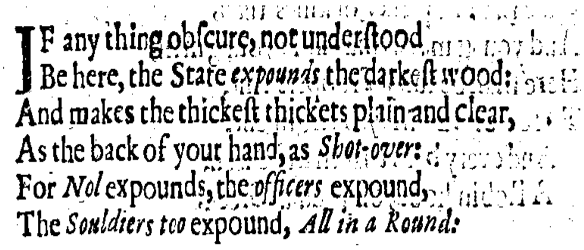


(f)

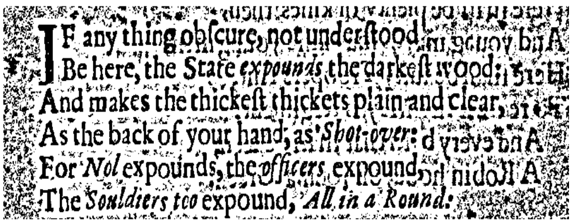
Fig.2 (a) Original Image “PR2.jpg” (b) Nick’s Thresholding Result (c) Niblack’s Algorithm Result (d) Sauvola Thresholding Result (e) Bernsen’s Binarization Result (f) Otsu’s Thresholdig Result



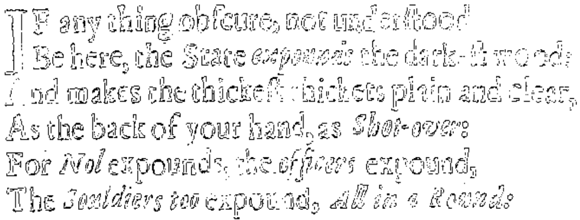
(a)



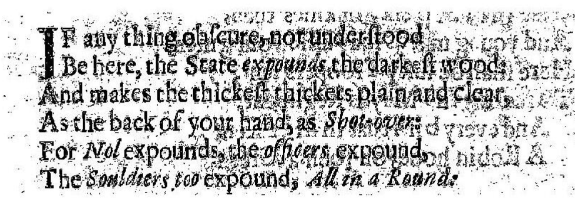
(b)



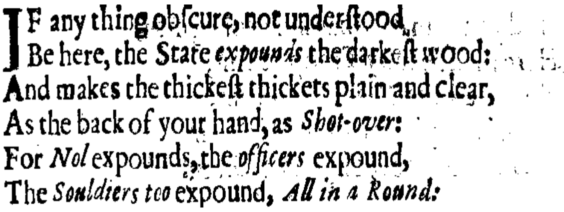
(c)



(d)



(e)



(f)

Fig.1 (a) Original Image “PR3.jpg” (b) Nick’s Thresholding Result (c) Niblack’s Algorithm Result (d) Sauvola Thresholding Result (e) Bernsen’s Binarization Result (f) Otsu’s Thresholdig Result

(a)

(b)

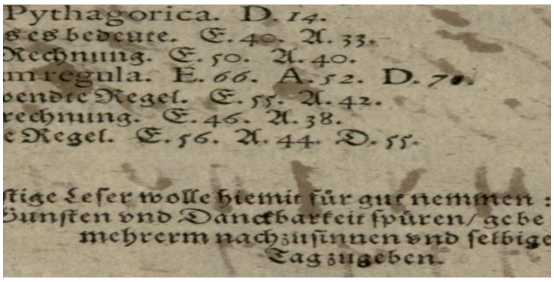
(c)

(d)

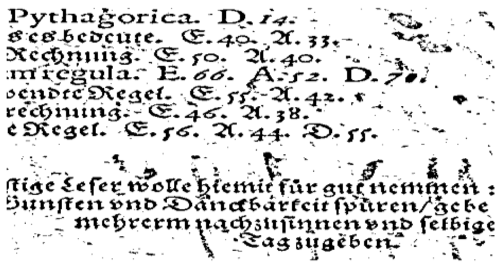
(e)

(f)

Fig.4 (a) Original Image “PR4.jpg” (b) Nick’s Thresholding Result (c) Niblack’s Algorithm Result (d) Sauvola Thresholding Result (e) Bernsen’s Binarization Result (f) Otsu’s Thresholdig Result

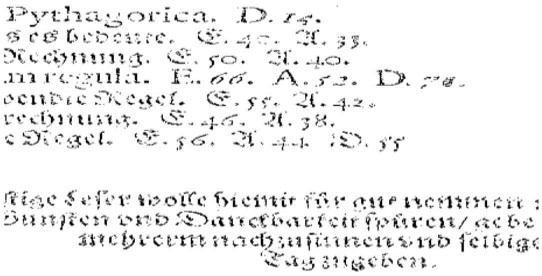


(a)

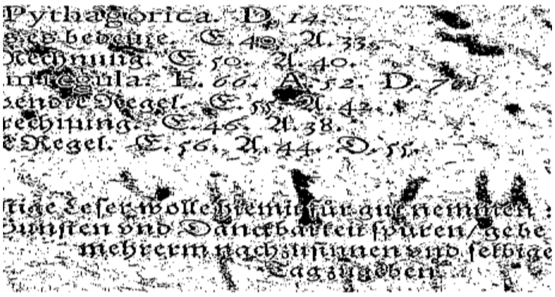


(b)

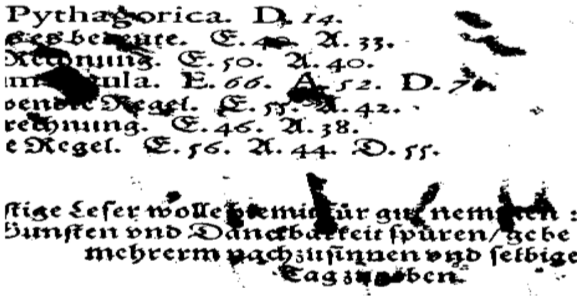
(c)



(d)

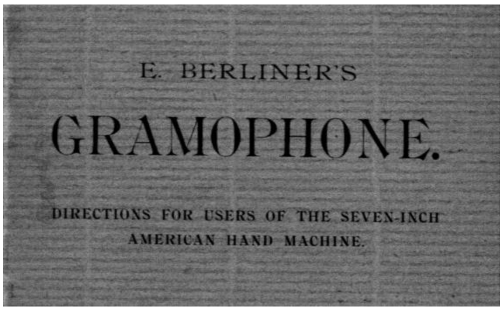


(e)

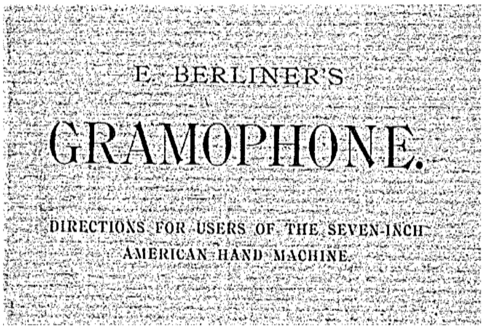


(f)

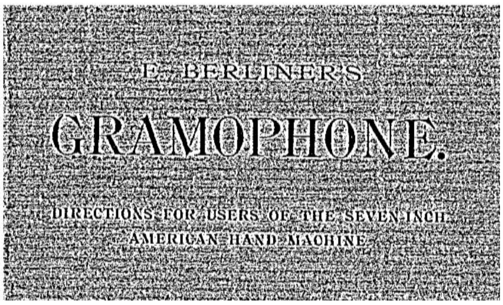
Fig.5 (a) Original Image “PR5.jpg” (b) Nick’s Thresholding Result (c) Niblack’s Algorithm Result (d) Sauvola Thresholding Result (e) Bernsen’s Binarization Result (f) Otsu’s Thresholdig Result



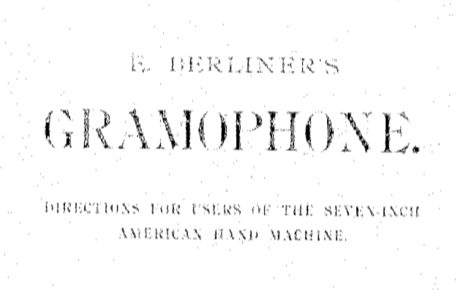
(a)



(b)

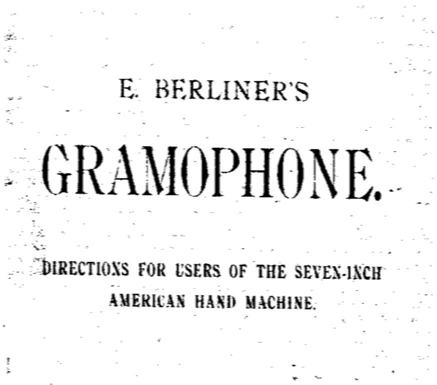


(c)



(d)

(e)



(f)

Fig.6 (a) Original Image “PR5.jpg” (b) Nick’s Thresholding Result (c) Niblack’s Algorithm Result (d) Sauvola Thresholding Result (e) Bernsen’s Binarization Result (f) Otsu’s Thresholdig Result

Image name	PR1	PR2	PR3	PR4	PR5	PR6
Precision	0.71530	0.65303	0.87958	0.78864	0.80273	0.24908
Recall	0.81281	0.80034	0.77754	0.76995	0.80476	0.88797
F-measure (%)	76.09427	71.92167	82.54169	77.91816	80.37477	38.90359
Sensitivity	0.81281	0.80034	0.77754	0.76995	0.80476	0.88797
Specificity	0.92910	0.94355	0.97284	0.96923	0.96321	0.86035
BCR	0.87095	0.87194	0.87519	0.86959	0.88398	0.87416
BER (%)	12.90452	12.80566	12.48121	13.04112	11.60153	12.58397
F-measure of sens/spec (%)	86.70730	86.60633	86.42921	85.81710	87.68851	87.39420
Geometric Accuracy	0.86901	0.86900	0.86972	0.86386	0.88043	0.87405
pFMeasure (%)	82.02089	78.16647	92.04676	87.04493	87.41451	39.67281
NRM	0.12905	0.12806	0.12481	0.13041	0.11602	0.12584
PSNR	10.37138	11.35277	11.74821	12.46977	12.10079	8.59234
DRD	13.49466	13.39973	6.18022	9.03052	6.49703	89.41494
MPM (x1000)	45.83018	30.52530	15.78847	20.62062	23.05287	75.41286

Table 1: Experimental Results of Nick’s Thresholding Algorithm

Image name	PR1	PR2	PR3	PR4	PR5	PR6
Precision	0.39702	0.27979	0.52599	0.29161	0.39656	0.10286
Recall	0.84297	0.87500	0.83181	0.82106	0.86827	0.91556
F-measure (%)	53.98026	42.39989	64.44584	43.03655	54.44579	18.49398
Sensitivity	0.84297	0.87500	0.83181	0.82106	0.86827	0.91556
Specificity	0.71942	0.70099	0.80874	0.70260	0.75419	0.58341
BCR	0.78119	0.78799	0.82027	0.76183	0.81123	0.74949
BER (%)	21.88062	21.20051	17.97267	23.81725	18.87718	25.05116
F-measure of sens/spec (%)	77.63090	77.83895	82.01111	75.72223	80.72175	71.26890
Geometric Accuracy	0.77875	0.78318	0.82019	0.75952	0.80922	0.73086

pFMeasure (%)	56.40659	43.63775	68.34232	44.93107	56.39091	18.62488
NRM	0.21881	0.21201	0.17973	0.23817	0.18877	0.25051
PSNR	5.87756	5.54998	7.29151	5.49701	6.42224	3.97812
DRD	42.30214	58.22489	21.93836	55.42503	30.00728	262.37434
MPM (x1000)	184.01239	192.60359	166.29828	199.90015	188.71620	224.32260

Table 2 Parameters of Niblack's Thresholding Algorithm

	PR1	PR2	PR3	PR4	PR5	PR6
Precision	0.93664	0.98179	0.99651	0.99866	0.99779	0.96578
Recall	0.18157	0.05257	0.28881	0.32913	0.38557	0.41385
F-measure (%)	30.41776	9.98019	44.78333	49.50918	55.62073	57.94120
Sensitivity	0.18157	0.05257	0.28881	0.32913	0.38557	0.41385
Specificity	0.99731	0.99987	0.99974	0.99993	0.99984	0.99924
BCR	0.58944	0.52622	0.64428	0.66453	0.69271	0.70654
BER (%)	41.05599	47.37782	35.57224	33.54680	30.72949	29.34582
F-measure of sens/spec (%)	30.72120	9.98937	44.81592	49.52481	55.65251	58.52901
Geometric Accuracy	0.42554	0.22927	0.53734	0.57368	0.62089	0.64306
pFMeasure (%)	17.47340	9.11657	36.72217	52.79622	74.38014	63.77025
NRM	0.41056	0.47378	0.35572	0.33547	0.30729	0.29346
PSNR	8.25857	9.54100	8.39299	10.59940	10.15396	15.25949
DRD	22.19744	18.95464	14.46888	13.78818	8.94436	15.48480
MPM (x1000)	5.37811	1.87670	4.15973	1.27055	1.17358	0.53183

Table 3 Parameters of Sauvola's Thresholding Algorithm

	PR1	PR2	PR3	PR4	PR5	PR6
Precision	0.30250	0.44370	0.53474	0.27254	0.48778	0.49915
Recall	0.64707	0.71739	0.74298	0.64338	0.71927	0.70856
F-measure (%)	41.22669	54.82914	62.18945	38.28821	58.13281	58.57010
Sensitivity	0.64707	0.71739	0.74298	0.64338	0.71927	0.70856
Specificity	0.67301	0.88060	0.83506	0.74393	0.85948	0.88649
BCR	0.66004	0.79900	0.78902	0.69366	0.78937	0.79752
BER (%)	33.99562	20.10049	21.09776	30.63433	21.06281	20.24769
F-measure of sens/spec (%)	65.97890	79.06610	78.63358	69.00125	78.31459	78.75994
Geometric Accuracy	0.65992	0.79482	0.78768	0.69183	0.78625	0.79255
pFMeasure (%)	41.0739	57.96832	66.36903	40.06462	64.15129	62.37271

NRM	0.33996	0.20100	0.21098	0.30634	0.21063	0.20248
PSNR	4.79320	8.58458	7.35997	5.70064	7.89099	8.60088
DRD	54.39009	26.7691 0	20.81478	51.73635	20.01697	19.54054
MPM (x1000)	159.86192	53.4449 8	70.91107	130.39470	70.26797	47.55075

Table 4: Parameters of Bernsen's Binarization Algorithm

	PR1	PR2	PR3	PR4	PR5	PR6
Precision	0.88595	0.59233	0.95839	0.98340	0.72391	0.88454
Recall	0.87740	0.93527	0.79946	0.77434	0.88286	0.91914
F-measure (%)	88.16549	72.53088	87.17411	86.64354	79.55210	90.15058
Sensitivity	0.87740	0.93527	0.79946	0.77434	0.88286	0.91914
Specificity	0.97525	0.91455	0.99114	0.99805	0.93735	0.99374
BCR	0.92632	0.92491	0.89530	0.88619	0.91011	0.95644
BER (%)	7.36757	7.50891	10.46985	11.38056	8.98939	4.35616
F-measure of sens/spec (%)	92.37405	92.47948	88.50414	87.20757	90.92903	95.49836
Geometric Accuracy	0.92503	0.92485	0.89016	0.87911	0.90970	0.95571
pFMeasure (%)	93.41869	74.21667	96.48163	97.73783	83.20282	92.24764
NRM	0.07368	0.07509	0.10470	0.11381	0.08989	0.04356
PSNR	13.73226	10.80802	13.20375	15.08948	11.47554	20.01837
DRD	5.08545	15.29938	3.80525	3.74881	7.81348	4.70503
MPM (x1000)	9.14362	37.97265	2.78050	0.64502	19.83551	1.74935

Table 5: Parameters of Otsu's Thresholding Results

3.3 Conclusion:

Various standard binarization techniques were applied on document images and results for DIBCO 2011 data set are attached in the chapter. From the results on images it can be concluded that Otsu's method is a global thresholding technique, which provide good results but it need more time for multilevel threshold selection. In Niblack's thresholding method, the output binary image shows some problem in background noise particularly in considering size of the windows (i.e. empty window). For large size window size it produces better results, because it covers maximum portion of text and for that user have to set parameters manually. Nick's binarization algorithm is an improvement of Niblack method where it works very well for most degraded document. This technique solves the issue of noise in white pages and low contrast problem. For Sauvola's binarization method, by neglecting the size of window it gives better result with minimum binarization noise. However, method fails to identify low-quality text like highly illuminated text, text in low contrast or thin pen stroke text. Bersen's algorithm is simple, but do not work effectively on degraded document images with a complex background.

Chapter 4: A Novel Approach for Text Extraction from Natural Scene Images:

Our aim is to search the text region from the natural scene pictures which may include slanted and curved text with minimum computational cost using Difference of Gaussian method. A short description of the methods used in our proposed algorithm is given below.

4.1 Gaussian Filter and Difference of Gaussian:

Gaussian filter is a filter having Gaussian function as impulse response. Gaussian filters have properties demonstrating no overshoot to a step function excitation at the same time minimizing the rise and fall time. By mathematical definition, a Gaussian filter changes the input signal by convolving with a Gaussian function [16]. As a gray image is a two dimensional object, two dimensional Gaussian functions is characterized underneath.

$$g(u,v)=\frac{1}{2\pi\sigma^2}e^{-\frac{u^2+v^2}{2\sigma^2}} \tag{7}$$

Where v represents the distance from the origin in the vertical axis, u denotes the distance from the origin in the horizontal axis and σ is the standard deviation of the Gaussian distribution [16][21]. In image processing, difference of Gaussians (DoG) is an element improvement calculation that incorporates the subtraction of two blurred versions of the image under observation with different standard deviations thus subsequently sparing the edge information in a picture.

For a grayscale image, the blurred images are produced by convolution of the original grayscale image with a Gaussian function [16] with distinctive standard deviations. Obscuring an image using a Gaussian kernel smothers just high frequency spatial information. Subtraction of one image from other protects the spatial information that falls between the frequency sets that are preserved in the two blurred images. In this manner, the DoG is a band pass filter that disposes the modest bunch of spatial frequencies that are existent in the basic grayscale image.

4.2 Niblack's Adaptive Thresholding Algorithm:

The Niblack's algorithm estimates a threshold value for individual pixel by scanning a rectangular window over the grayscale image [23][24]. The size of rectangular window can be adjusted. The threshold value is ascertained utilizing the local mean and standard deviation of the considerable number of pixels in the window and is computed by taking after conditions [3]:

$$T_{Niblack} = M + k * SD \tag{8}$$

$$T_{Niblack} = M + k * \sqrt{\frac{\sum(p_i - m)^2}{NP}} \tag{9}$$

Here NP is the total no of pixels in a grayscale image, T represents the threshold value, M is the mean of gray levels of pixels and k is fixed depending upon the noise in the background and is -0.2.

The algorithm is summarized below and each step is explained in detail in sub-sections.

i. Convert the image into a grayscale image.
ii. Filter the image for supressing noise using a Gaussian low pass filter.
iii. Use Niblack's Thresholding algorithm to binaries the image
iv. Apply modified Difference of Gaussian to detect the edges of the filtered image.
v. Apply morphological operations.
vi. Multiply the resultant image with binary image of the input image to obtain text in contrast with Plain background.

4.3 Pre-processing:

Pre-processing encourages us to enhance the execution and make the procedure productive. The major steps are converting images to grayscale images and then binarization of images and noise elimination using some filtering.

a) Gray Scaling: The input image is an RGB image comprises of intensities R (Red), G (Green) and B (Blue). The direct application of an algorithm might not split up the text from the background, hence the input image is transformed to grayscale image which results intensities of each pixel in the range of [0,255].

b) Filtering: To enhance picture quality and for further handling, Gaussian filter which is rationally symmetric is used to eliminate noises such as blurred high frequency noise and white noise by smoothening the image.

c) Binarization: Binarization results in a black and white image consisting of pixel values either 0 or 1. In the proposed algorithm the binarization is performed using Niblack's thresholding algorithm. Binarization helps in separating the text from the background up to some extent, but alone binarization in not enough to get proper text information from the data.

4.4 Edge Detection using modified Difference of Gaussian:

The difference of Gaussian calculation wipes out the high frequency elements that often include random noise removal. It reduces noise in the image by blurring and also preserves the edge information in the image. The DoG also eliminates the low frequency background blotches. Finally the resultant image is normalized to [0, 1].In our method the modified equation for DoG is given by:

$$\Gamma(x, y) = I * (\mathcal{G}_{\sigma 1} - \alpha * \mathcal{G}_{\sigma 2}) \qquad (10)$$

Here $\Gamma(x, y)$ is the output image after application of DoG, I is the input gray scale image, $\mathcal{G}_{\sigma 1}$ and $\mathcal{G}_{\sigma 2}$ are Gaussian function provided σ2>σ1 and α is a constant which lies between 1 and 2.

4.5 Morphological Operations:

The previous described processes results in text regions. Morphological dilation is used to form a cluster of text regions. A reasonable estimated organizing component ought to be picked such that only least non-text area should be clustered within. Morphological opening is utilized to expel non-content objects.

4.6 Character Extraction*:*

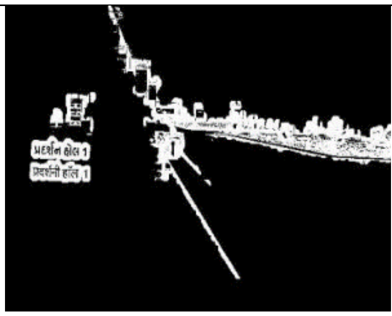
Extraction of the characters as is done in this step by multiplying the resultant image with black and white version of the original image.

4.7 Experimental Results:

The experiment is done on more than 100 images. The result for some test images is shown below. Results are compared with the output using the Prewitt Edge Detector and Gaussian Edge Detector. Also the proposed strategy is validated on distinct images of the training data set form ICDAR 2015, and demonstrated that the technique can enhance the text detection performance. The performance of the algorithm is validated based on Precision rate and Recall rate. The stated parameters are the average of 100 test images. Overall, 74.8% accuracy is obtained from this algorithm on precision value, 76.2 % on recall value and an average processing time of 0.9 seconds. Performance comparison of the proposed method with some known method is given below in Table 1. However, if size of slanted or curved text is very small then the second order edge detection methods does not produce better text extraction which is also mentioned in[29].

(a)

(b)

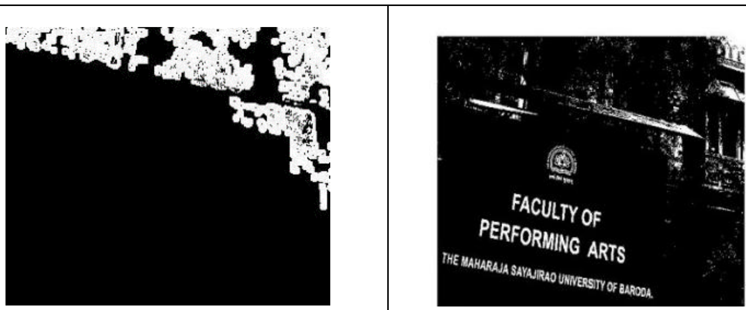


(c)

(d)

(e)

(f)



(g) (h)

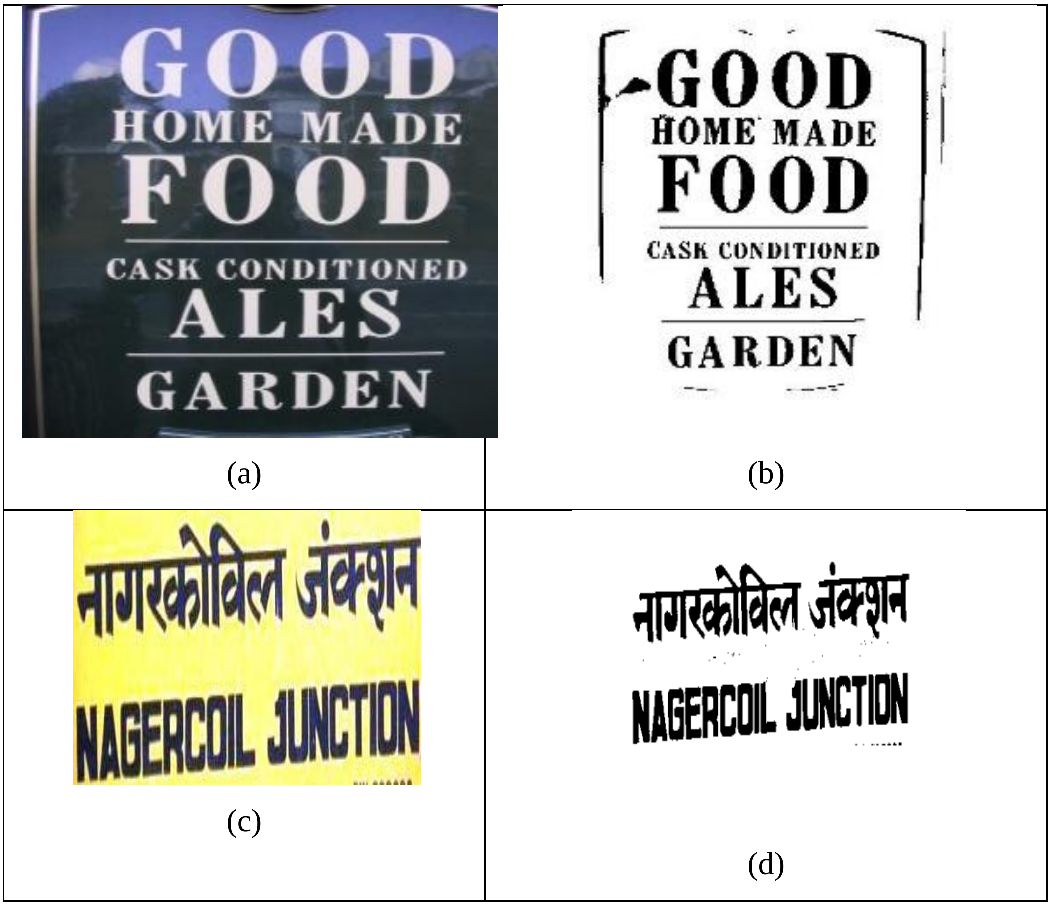
Fig 7: (a) (e) Input Image; (b) (f) Output of Prewitt Edge Detection Algorithm; (c) (g) Output of Gaussian Edge Detection Algorithm; (d) (h) Output of Proposed Method;

(a) (b)

Fig 8: (a) Input Image form ICDR2015 Training Dataset; (b) Output of Proposed Algorithm

(a) (b)

Fig 9: (a) Input Image form [25] (b) Output of Proposed Algorithm



(a) (b)

(c) (d)

Fig 10: (a) Input Image form ICDR2015[33] Training Dataset (b) Output of Proposed Algorithm (c) Input image from [26] (d) Output of Proposed Algorithm

Table 6: Comparison of evaluated results

Method	No. of Images	Precision rate (%)	Recall rate (%)
U.Bhattacharya, S. K. Paruri & S. Mondal[25]	100	68.8	71.2
Neumann & Matas[30]	-	73.1	64.7
TH-TextLoc System[30]	-	66.97	57.68

H. Raj & R. Ghosh[28]	62	72.8	74.2
Proposed Method	100	74.8	76.2

4.8 Conclusion

A simple method for extraction of text regions is proposed. The algorithm is quite robust in locating text regions. The proposed strategy has created promising results for natural scene pictures consist of content of various size, text style, and arrangement with changing foundations. The methodology additionally distinguishes nonlinear text areas. The framework proposed in this study can be executed in content acknowledgment programs or significant software.

The proposed algorithm gives robust result for slanted text components as shown in Fig 9 and Fig 7(h). However, the small size of curved or slanted text limits the performance of the algorithm as bias in the localization of the text edges for second order derivative.

Chapter 5: A Novel Contrast Preserving Binarization Technique

Generally the text and back ground is separated by comparing the intensity of each pixel with the mean grey value and fixing a threshold for separating text and background. If the intensity of text is low (due to fading or bleed through), the threshold technique fails. Also, Color to gray conversion enables the application of single-channel algorithms on color images. Unfortunately, since the color-to-gray conversion is to convert a 3D vector to a 1D gray-scale value, it is essentially a dimensionality reduction process, hence inevitably suffers from information loss. The baseline and widely used method in decolorizing an input image is to extract its luminance channel (e.g., CIE Y). If the source image is in the RGB format, the luminance can be obtained by linearly summing its R, G and B channels with a fixed weight (e.g., the rgb2gray function in Matlab). However, using luminance channel image alone cannot faithfully represent structures and contrasts in some color images such as those having iso-luminant regions.

To filters noise and enhances the text without introducing any unwanted noise and we used $p_i \log p_i$ as compared to p_i in the Nick's algorithm which helps in the filtering of the noise. For lower values of p_i , the function $p_i \log p_i$ exhorts a non-linear nature due to which a more appropriate value is assigned to the threshold which indeed helps to differentiate black pixels or text better as compared to p_i. Also to address the problem of low contrast on iso-luminant regions we used Gradient Correlation Similarity for Efficient Contrast Preserving Decolorizationin place of conventional rgb to gray conversion in pre-processing step.

In the proposed method we used GCS based decolorization for Color to gray conversion in pre-processing step and log based Nick's Thresholding for the Binarization of document images.

The proposed method is programmed and experimented on standard DIBCO-2011[32] data set by help of MATLAB tool. Various performance parameters are calculated from the resultant images of these methods e.g. Precision Rate, Recall Rate, DRD, NRM, fMeasure, Sensitivity, Selectivity etc which are shown in the table given in Results section.

5.1 GcsDecolor: Gradient Correlation Similarity for Efficient Contrast Preserving Decolorization

Considering that the input colour image is in the RGB format, where indexes *r, g, b* stand for the RGB channels. Let $\delta_{x,y}$ ($|\delta_{x,y}| = \sqrt{\sum_{c=\{r,g,b\}}(I_{c,x} - I_{c,y})^2}$ be the colour contrast having a signed value denoting the difference of a colour pair and $g_x - g_y$ indicate gray difference value between pixels, then the classical L2-norm based energy function is defined as:

$$\min_{g} \sum_{(x,y)\in P} \left(g_x - g_y - \delta_{x,y}\right)^2$$

P stands for a pixel pair pool which contains the local and non-local candidates.

The GCS measure employs one order multivariate polynomial function set c={r,g,b} and limit the sum of weight to be one. The measure finds gradient correlation adaptively for each channel rather than whole channels at one time. Pixel wise similarity is computed between the gradient magnitudes in each channel of original colour image and the resulting gray-scale image i.e.

$$\min_{w_c} - \sum_{(x,y)\in P} \sum_{c\in\{r,g,b\}} \frac{2|I_{c,x}-I_{c,y}||\nabla g_{x,y}|}{|I_{c,x}-I_{c,y}|^2+|\nabla g_{x,y}|^2} \tag{11}$$

$$\text{s.t } g = \sum_{c=\{r,g,b\}} w_c I_c \; ; \; \sum_{c=\{r,g,b\}} w_c = 1$$

The motive is to optimize above model with respect to weight coefficients $\{w_c \mid c=r,g,b\}$ by using parametric formulation.

The GCS-derived model (11) poses two qualities. First, calculating the correlation between two absolute values eliminates the difficulty of determining the sign of *δx,y*. Second, by dividing the number of pixel pairs in *P*, the value of GCS measure lies into the interval of [0, 3] for any colour-to-gray conversion. Hence it gives a universal value, which has the potential to be used for the evaluation of the conversion between different images.

The novelty of GCS model (11) lies in two aspects: first, the structure preservation is described by the gradient correlation, instead of the commonly used gradient error; second, the similarity is computed between the resulting gray image and the original image in each channel of the RGB space, while it is done between the gray image and the summation of the gradients of the three channels of RGB space in previous works.

To solve GCS model Augmented Langrangian and alternative direction method based solver is used. The Al method is well suited optimization algorithm for solving the constrained problem in GCS model. AL method transforms the original unconstrained minimization problem to an equivalent constrained problem and then alternating minimization strategy is used to iteratively find solution of the sub problem.

AL scheme targets the problem given below:

$$\min_{w\in\Omega,s,t,r} \sum_i \sum_{c\epsilon\{r,g,b\}} -2|s_i| r_{c,i}$$

s.t. $s = \sum_{c\in\{r,g,b\}} w_c \nabla I_c$, $\left(t_{c,i}|s_i| + \nabla I_{c,i} + \varepsilon_1\right) r_{c,i} = 1$, $\left(\left|\nabla I_{c,i}\right| + \varepsilon_1\right) t_{c,i} = |s_i|$

Where r and t are auxiliary variables and $\Omega = \{w | \sum_{c=\{r,g,b\}} w_c = 1\}$

This problem () can be evaluated by standard augmented Langrangian method. Starting from $\lambda^{1,0}$=0, $\lambda_{c,i}^{2,0} = 0$ and $\lambda_{c,i}^{3,0} = 0$, it solves

$$\min_{w\epsilon\Omega,s,t,r} \sum_i \sum_c -2|s_i| r_{c,i} + \sum_c (\langle \lambda_{c,i}^{3,k}, |s_i| - (\nabla I_{c,i} + \varepsilon_1) t_{c,i} \rangle + \frac{\gamma^3}{2} \left\| |s_i| - (\left|\nabla I_{c,i}\right| + \varepsilon_1) t_{c,i} \right\|_2^2 +$$

$$\sum_c (\langle \lambda_{c,i}^{2,k}, \left(t_{c,i}|s_i| + \left|\nabla I_{c,i}\right| + \varepsilon_1\right) r_{c,i} - 1 \rangle) + \frac{\gamma^2}{2} \left\| \left(t_{c,i}|s_i| + \left|\nabla I_{c,i}\right| + \varepsilon_1\right) r_{c,i} - 1 \right\|_2^2 +$$

$$\langle \lambda^{1,k}, \sum_c w_c \nabla I_c - s \rangle + \frac{\gamma^1}{2} \|s - \sum_c w_c \nabla I_c\|_2^2 \qquad (12)$$

At the k-th iteration and for w^{k+1}, s^{k+1},t^{k+1}, r^{k+1}, then update the multipliers λ^1,$\lambda_{c,i}^2$ and $\lambda_{c,i}^3$ by formula

$$\lambda^{1,k+1} = \lambda^{1,k} + \gamma^1 \sum_c w_c^{k+1} \nabla I_c - s^{k+1} \qquad (13)$$

$$\lambda_{c,i}^{2,k+1} = \lambda_{c,i}^{2,k} + \gamma^2 [\left(t_{c,i}^{k+1} \left|s_i^{k+1}\right| + \left|\nabla I_{c,i}\right| + \varepsilon_1\right) r_{c,i}^{k+1} - 1] \qquad (14)$$

$$\lambda_{c,i}^{3,k+1} = \lambda_{c,i}^{3,k} + \gamma^3 \left(\left|s_i^{k+1}\right| - \left(\left|\nabla I_{c,i}\right| + \varepsilon_1\right) t_{c,i}^{k+1}\right) \qquad (15)$$

At each iteration, solving the augmented Lagrangian function (11) for *w*, *s*, *t* and *r* simultaneously is difficult, an alternative choice is to minimize it with respect to each block variable *w*, *s*, *t* and *r* one at a time while fixing the other three blocks at their latest values. The corresponding four sub-problems can be solved as follows:

For the *w*-sub problem, it needs to solve a constraint least square problem.

$$w^{k+1} = \arg\min_{w\epsilon\Omega} \frac{\gamma^1}{2} \left\| \sum_c w_c \nabla I_c - (s^k - \frac{\lambda^{1,k}}{\gamma^1} \right\|_2^2 \qquad (16)$$

The determination of variable *w* can be attained first by the classical least square solution and then the projection operation

$$w_c^{k+1} = \frac{w_c^{k+1}}{\sum_{c\epsilon\{r,g,b\}} w_c^{k+1}} \qquad (17)$$

Algorithm: GCS Decolor

1: Initialize $\lambda^{1,0}$=0, $\lambda_{c,i}^{2,0} = 0$ and $\lambda_{c,i}^{3,0} = 0$

2: for k=0 to k-1 do

3: update w^{k+1} according to equation (16)

4: update s^{k+1} according to equation (18)

5: update t^{k+1} according to equation (19)

6: update r^{k+1} according to equation (20)

7: update $\lambda^{1,k+1}$,$\lambda_{c,i}^{2,k+1}$ and $\lambda_{c,i}^{3,k+1}$ according to equation (13) (14) (15); end for

For the *s*-sub problem, the objective function with respect to variable *s* is as follow:

$$\min_{s} \sum_i \sum_c \begin{matrix} -2r_{c,i}^k|s_i| + \frac{\gamma^2}{2}\left\|r_{c,i}^k\right\|_2^2\left\|t_{c,i}^k\right\|_2^2\|s_i\|_2^2 + \langle\lambda_{c,i}^{2,k} + \gamma^2[r_{c,i}^k(|\nabla I_{c,i}| + \varepsilon_1 - 1], |s^i|t_{c,i}^k r_{c,i}^k\rangle + \\ \sum_i\left(\frac{\gamma^1}{2}\|s_i\|_2^2 - \langle\gamma^1\sum_c w_c^{k+1}\nabla I_{c,i} + \lambda^1, s_i\rangle\right) + \sum_i\sum_c\frac{\gamma^3}{2}\|s_i\|_2^2 - [\gamma^3(|\nabla I_{c,i}| + \varepsilon_1)t_{c,i}^k - \lambda_{c,i}^{3,k}]|s_i| \end{matrix}$$

Letting $b_i^k = \sum_c \gamma^1 + \gamma^2\left\|r_{c,i}^k\right\|_2^2\left\|t_{c,i}^k\right\|_2^2 + \gamma^3$ and

$d_i^k = \sum_c -2r_{c,i}^k + [\lambda_{c,i}^{2,k} + \gamma^2(r_{c,i}^k(|\nabla I_{c,i}| + \varepsilon_1) - 1)]r_{c,i}^k t_{c,i}^k - [\gamma^3|\nabla I_{c,i}| + \varepsilon_1)t_{c,i}^k - \lambda_{c,i}^{3,k}]$, it yield s_i^{k+1}

$$= \arg\min_{s_i} d_i^k|s_i| + \frac{b_i^k}{2}\|s_i\|_2^2 - \langle\gamma^1\sum_c w_c^{k+1}\nabla I_{c,i} + \lambda^{1,k}, s_i\rangle$$

$$= d_i^k|s_i| + \frac{b_i^k}{2}[s_i - (\gamma^1\sum_c w_c^{k+1}\nabla I_{c,i} + \lambda^{1,k})/b_i^k]^2$$

$$= Shrink[\frac{(\lambda^1\sum_c w_c^{k+1}\nabla I_{c,i} + \lambda^{1,k})/b_i^k, d_i^k}{b_i^k}] \quad (18)$$

Similarly for t-sub problem objective function is:

$$\min_{t_{c,i}} \begin{matrix} 2\lambda_{c,i}^{2,k}|s_i^{k+1}|r_{c,i}^k t_{c,i} + \gamma^2\left\||s_i^{k+1}|r_{c,i}^k t_{c,i} + \left(\left((|\nabla \mathrm{I}_{c,i}| + \varepsilon_1)\mathrm{r}_{c,i}^k\right) - 1\right)\right\|_2^2 \\ -2\lambda_{c,i}^{3,k}(\nabla I_{c,i} + \varepsilon_1)t_{c,i} + \gamma^3((|\nabla I_{c,i}| + \varepsilon_1)t_{c,i} - |s_i^{k+1}|)^2 \end{matrix}$$

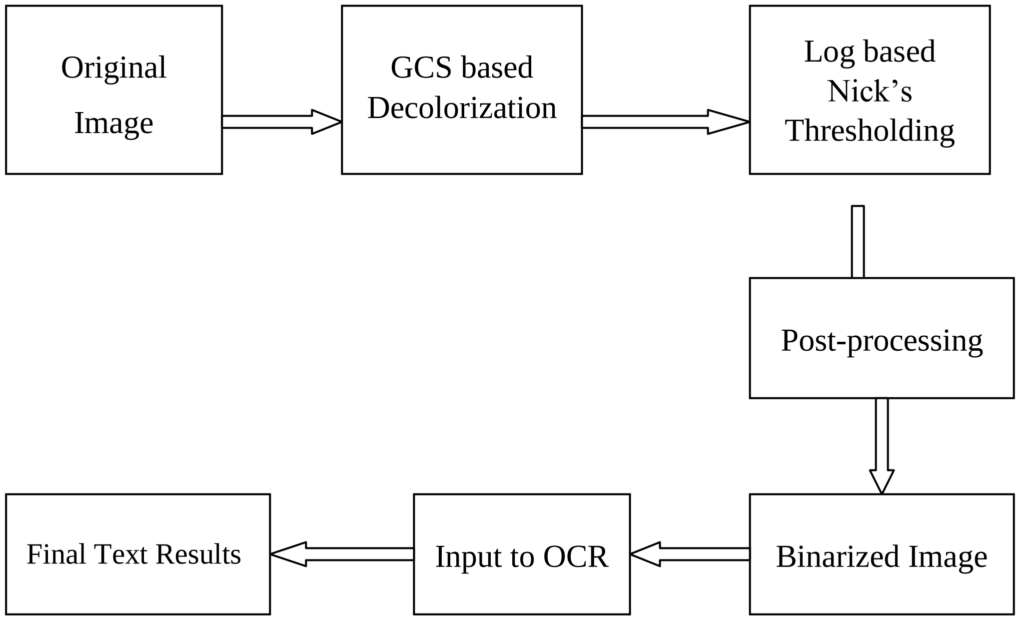
$$t_{c,i}^{k+1} = \frac{\gamma^3\left|s_i^{k+1}\right|(|\nabla I_{c,i}| + \varepsilon_1) + (|\nabla I_{c,i}| + \varepsilon_1)\lambda_{c,i}^{3,k} - \lambda_{c,i}^{2,k}\left|s_i^{k+1}\right|r_{c,i}^k - \gamma^2\left|s_i^{k+1}\right|r_{c,i}^k((|\nabla I_{c,i}| + \varepsilon_1)r_{c,i}^k - 1)}{\gamma^3(|\nabla I_{c,i}| + \varepsilon_1)^2 + \gamma^2(\left|s_i^{k+1}\right|r_{c,i}^k)^2} \quad (19)$$

The *r* -subproblem can also be computed analytically. Specifically, we obtain the optimal solution by minimizing the following objective function.

$$\min_{r_{c,i}} -2|s_i^{k+1}|r_{c,i} + \lambda_{c,i}^2(t_{c,i}^{k+1}|s_i^{k+1}| + |\nabla I_{c,i}| + \varepsilon_1)r_{c,i} + \frac{\gamma^2}{2}\left\|(t_{c,i}^{k+1}|s_i^{k+1}| + |\nabla I_{c,i}| + \varepsilon_1)r_{c,i} - 1\right\|_2^2$$

$$r_{c,i}^{k+1} = \frac{2\left|s_i^{k+1}\right| - \lambda_{c,i}^{2,k}\left(t_{c,i}^{k+1}\left|s_i^{k+1}\right| + \left|\nabla I_{c,i}\right| + \varepsilon_1\right) + \gamma^2\left(t_{c,i}^{k+1}\left|s_i^{k+1}\right| + \left|\nabla I_{c,i}\right| + \varepsilon_1\right)}{\gamma^2\left(t_{c,i}^{k+1}\left|s_i^{k+1}\right| + \left|\nabla I_{c,i}\right| + \varepsilon_1\right)^2} \tag{20}$$

5.2 Proposed Method:



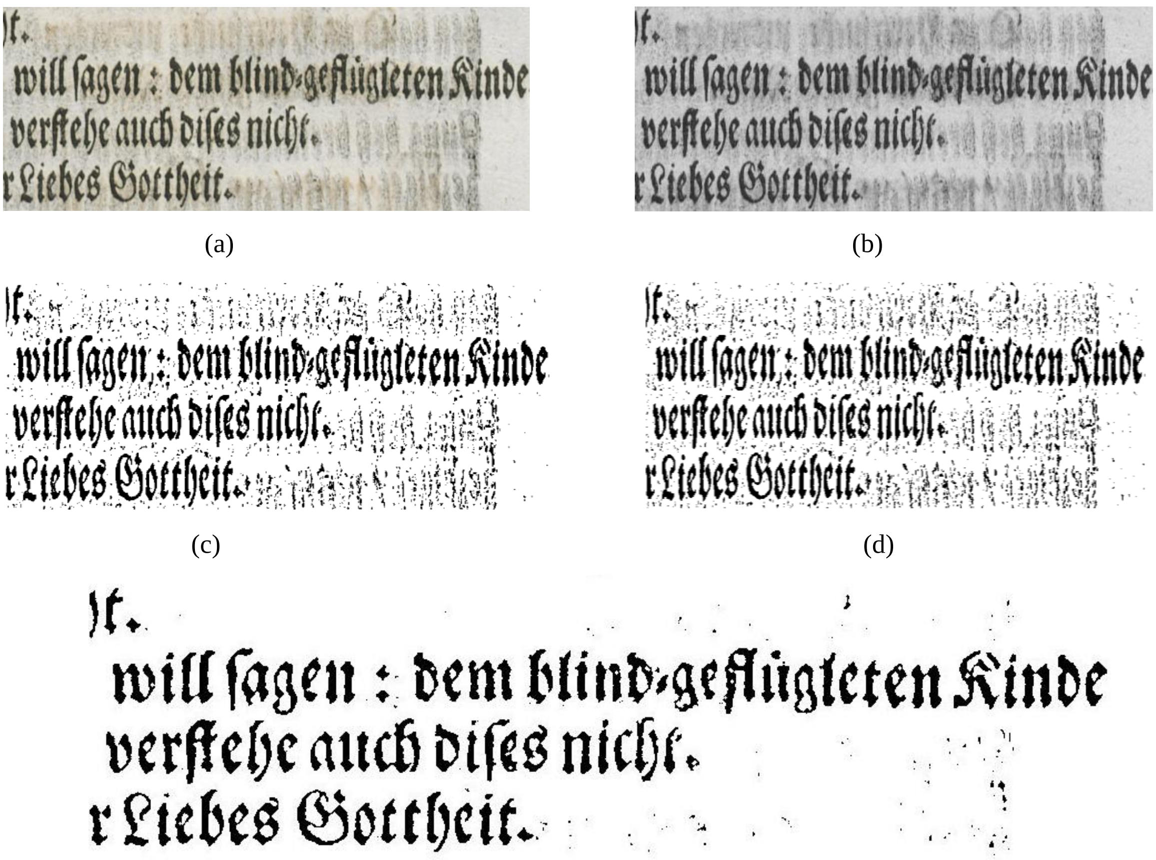
To filters noise and enhances the text without introducing any unwanted noise and we used $p_i \log p_i$ as compared to p_i in the Nick's algorithm which helps in the filtering of the noise. For lower values of p_i, the function $p_i \log p_i$ exhorts a non-linear nature due to which a more appropriate value is assigned to the threshold which indeed helps to differentiate black pixels or text better as compared to p_i. Also to address the problem of low contrast on iso-luminant regions we used Gradient Correlation Similarity for Efficient Contrast Preserving Decolorizationin place of conventional rgb to gray conversion in pre-processing step.

In the proposed method we used GCS based decolorization for Color to gray conversion in pre-processing step and log based Nick's Thresholding for the Binarization of document images.

The main aim of binariztion is to differentiate text of lower intensity from back ground and avoid unwanted noise. We used the same equation of Nick by replacing p_i with $p_i \log p_i$. The use of $p_i \log p_i$ as compared to p_i helps in the filtering of the noise [12]. For lower values of $p_{i,}$ the function $p_i \log p_i$ exhorts a non-linear nature due to which a more appropriate value is assigned to the threshold which indeed helps to differentiate black pixels or text better as compared to p_i.[]

5.3 Experimental Results:

The proposed method is programmed and experimented on standard DIBCO-2011 [32] data set by help of MATLAB tool. Various performance parameters are calculated from the resultant images of these methods e.g. Precision Rate, Recall Rate, DRD, NRM, fMeasure, Sensitivity, Selectivity etc which are shown in the table given below. Also the proposed method is compared with standard thresholding methods e.g. Nick's thresholding method, Niblack's thresholding method, Sauvola Thresholding method, Bernsen's thresholding method, otsu's binarization method and GCS decolorization based Nick's thresholding method in terms of performance measures.

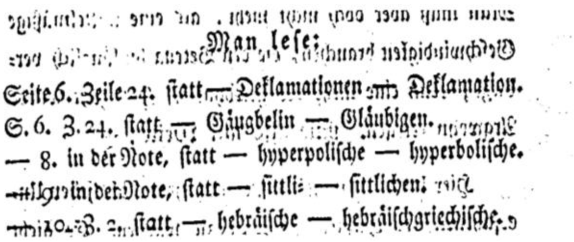


(a) (b) (c) (d) (e)

Fig. 11 (a) Original Image "PR1.jpg" (b) Decolorized Image (c) Result of Nick's Thresholding after Decolorization (d) Nick's Thresholding Output (e) Proposed Method Output

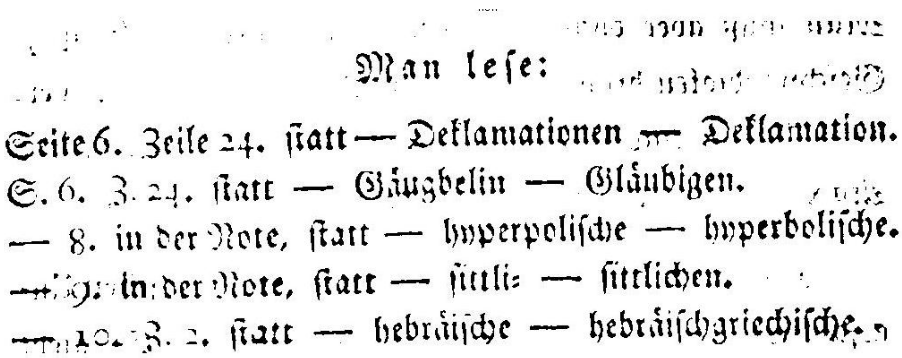
(a)

(b)



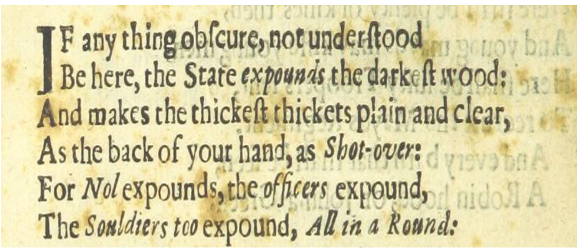
(c)

(d)

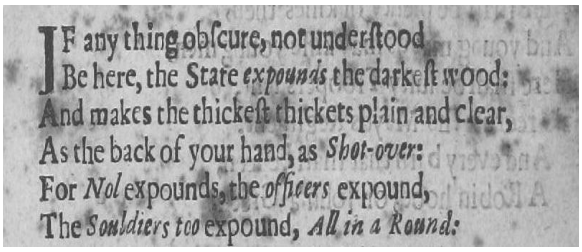


(e)

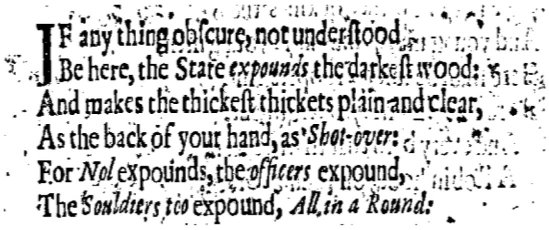
Fig. 12 (a) Original Image "PR2.jpg" (b) Decolorized Image (c) Result of Nick's Thresholding after Decolorization (d) Nick's Thresholding Output (e) Proposed Method Output



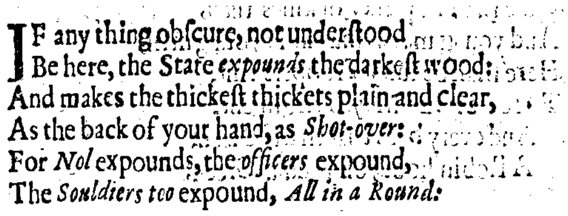
(a)



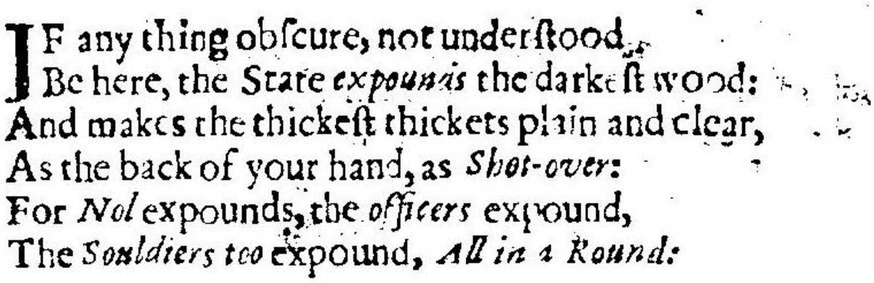
(b)


IF any thing obscure, not understood
Be here, the State expounds the darkest wood:
And makes the thickest thickets plain and clear,
As the back of your hand, as Shot-over:
For Nol expounds, the officers expound,
The Souldiers too expound, All in a Round:

(c)


IF any thing obscure, not understood
Be here, the State expounds the darkest wood:
And makes the thickest thickets plain and clear,
As the back of your hand, as Shot-over:
For Nol expounds, the officers expound,
The Souldiers too expound, All in a Round:

(d)


IF any thing obscure, not understood
Be here, the State expounds the darkest wood:
And makes the thickest thickets plain and clear,
As the back of your hand, as Shot-over:
For Nol expounds, the officers expound,
The Souldiers too expound, All in a Round:

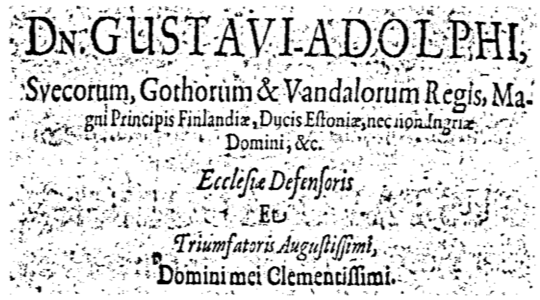
(e)

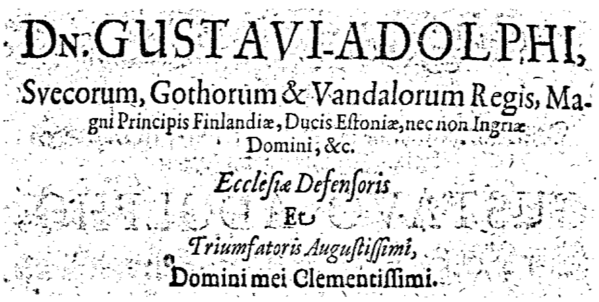
Fig. 13 (a) Original Image "PR2.jpg" (b) Decolorized Image (c) Result of Nick's Thresholding after Decolorization (d) Nick's Thresholding Output (e) Proposed Method Output


DN.GUSTAVI-ADOLPHI,
Svecorum, Gothorum & Vandalorum Regis, Magni Principis Finlandiæ, Ducis Estoniæ, nec non Ingriæ
Domini, &c.
Ecclesiæ Defensoris
Et
Triumfatoris Augustissimi,
Domini mei Clementissimi.

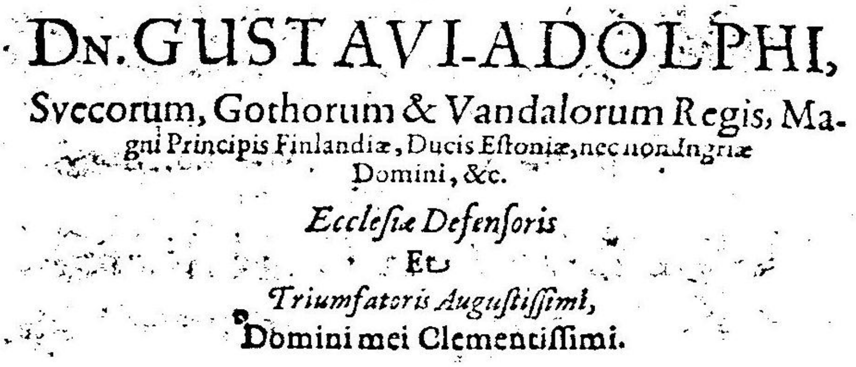

DN.GUSTAVI-ADOLPHI,
Svecorum, Gothorum & Vandalorum Regis, Magni Principis Finlandiæ, Ducis Estoniæ, nec non Ingriæ
Domini, &c.
Ecclesiæ Defensoris
Et
Triumfatoris Augustissimi,
Domini mei Clementissimi.

(a) (b)


DN.GUSTAVI-ADOLPHI,
Svecorum, Gothorum & Vandalorum Regis, Magni Principis Finlandiæ, Ducis Estoniæ, nec non Ingriæ
Domini, &c.
Ecclesiæ Defensoris
Et
Triumfatoris Augustissimi,
Domini mei Clementissimi.

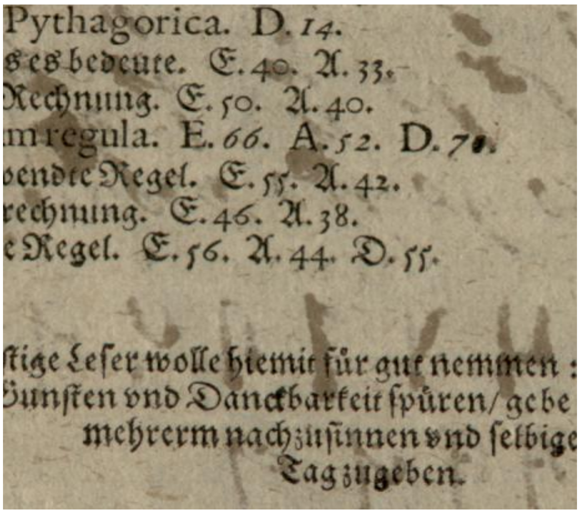

DN.GUSTAVI-ADOLPHI,
Svecorum, Gothorum & Vandalorum Regis, Magni Principis Finlandiæ, Ducis Estoniæ, nec non Ingriæ
Domini, &c.
Ecclesiæ Defensoris
Et
Triumfatoris Augustissimi,
Domini mei Clementissimi.

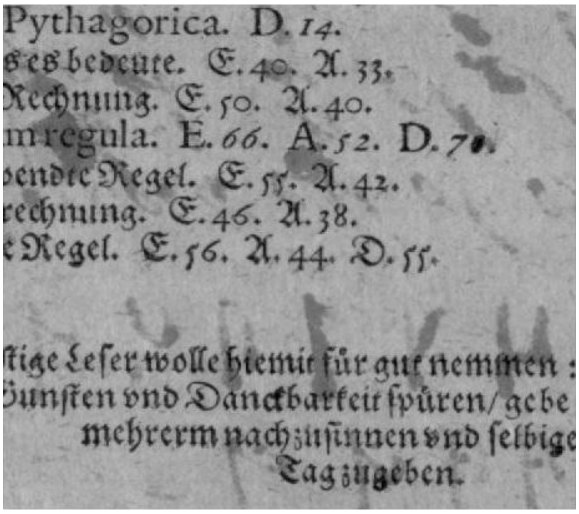
(c) (d)


DN. GUSTAVI ADOLPHI,
Svecorum, Gothorum & Vandalorum Regis, Magni Principis Finlandiæ, Ducis Estoniæ, nec non Ingriæ Domini, &c.
Ecclesiæ Defensoris
Et
Triumfatoris Augustissimi,
Domini mei Clementissimi.

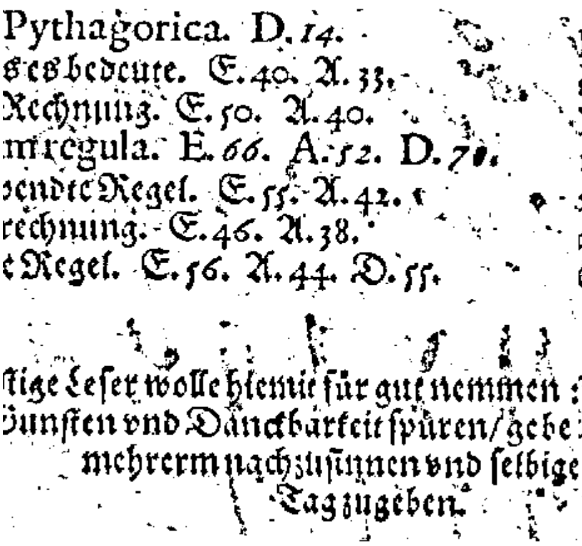
(e)

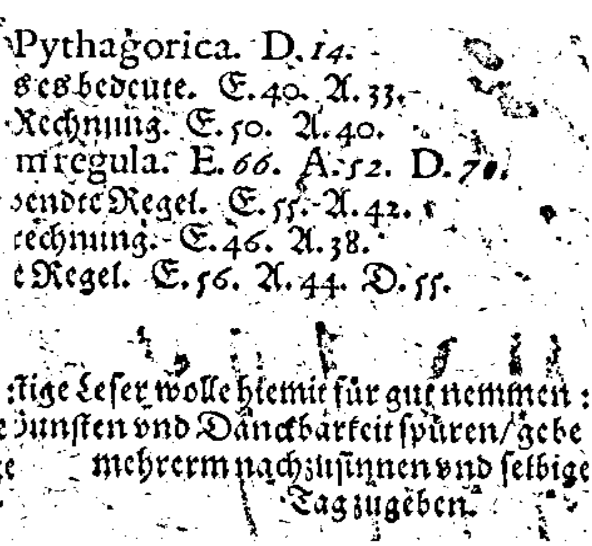
Fig. 14 (a) Original Image "PR4.jpg" (b) Decolorized Image (c) Result of Nick's Thresholding after Decolorization (d) Nick's Thresholding Output (e) Proposed Method Output


Pythagorica. D. 14.
ſes bedeute. E. 40. A. 33.
Rechnung. E. 50. A. 40.
m regula. E. 66. A. 52. D. 70.
dende Regel. E. 55. A. 42.
rechnung. E. 46. A. 38.
e Regel. E. 56. A. 44. D. 55.
ſtige Leſer wolle hiemit für gut nemmen:
Gunſten vnd Danckbarkeit ſpüren/ gebe
mehrerm nachzuſinnen vnd ſelbige
Tag zugeben.

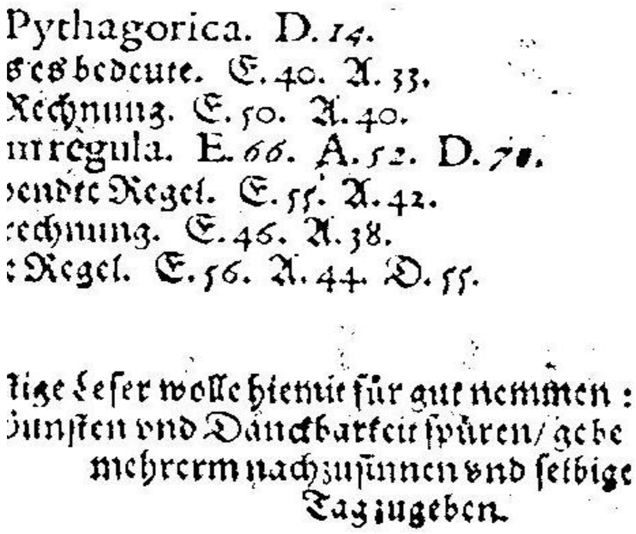


(a) (b)





(c) (d)


Pythagorica. D. 14.
ees bedeute. E. 40. A. 33.
Rechnung. E. 50. A. 40.
mregula. E. 66. A. 52. D. 70.
ende Regel. E. 55. A. 42.
rechnung. E. 46. A. 38.
Regel. E. 56. A. 44. D. 55.

tige Leser wolle hiemit für gut nemmen :
unsten und Danckbarkeit spüren / gebe
mehrerm nachzusinnen und selbige
Tag zugeben.

(e)

Fig. 15 (a) Original Image "PR5.jpg" (b) Decolorized Image (c) Result of Nick's Thresholding after Decolorization (d) Nick's Thresholding Output (e) Proposed Method Output

	PR1	PR2	PR3	PR4	PR5
Precision	0.71530	0.65303	0.87958	0.78864	0.80273
Recall	0.81281	0.80034	0.77754	0.76995	0.80476
F-measure (%)	76.09427	71.92167	82.54169	77.91816	80.37477
Sensitivity	0.81281	0.80034	0.77754	0.76995	0.80476
Specificity	0.92910	0.94355	0.97284	0.96923	0.96321
BCR	0.87095	0.87194	0.87519	0.86959	0.88398
BER (%)	12.90452	12.80566	12.48121	13.04112	11.60153
F-measure of sens/spec (%)	86.70730	86.60633	86.42921	85.81710	87.68851
Geometric Accuracy	0.86901	0.86900	0.86972	0.86386	0.88043
pFMeasure (%)	82.02089	78.16647	92.04676	87.04493	87.41451
NRM	0.12905	0.12806	0.12481	0.13041	0.11602
PSNR	10.37138	11.35277	11.74821	12.46977	12.10079
DRD	13.49466	13.39973	6.18022	9.03052	6.49703
MPM (x1000)	45.83018	30.52530	15.78847	20.62062	23.05287

Table 7: Parameters of Nick's Thresholding Results

	PR1	PR2	PR3	PR4	PR5
Precision	0.71091	0.57917	0.80930	0.69011	0.76671
Recall	0.78126	0.70815	0.80217	0.79464	0.74845
F-measure (%)	74.44268	63.71950	80.57159	73.86951	75.74674
Sensitivity	0.78126	0.70815	0.80217	0.79464	0.74845

Specificity	0.93038	0.93169	0.95177	0.94679	0.95763
BCR	0.85582	0.81992	0.87697	0.87072	0.85304
BER (%)	14.41830	18.00809	12.30312	12.92826	14.69618
F-measure of sens/spec (%)	84.93215	80.46821	87.05885	86.40704	84.02141
Geometric Accuracy	0.85256	0.81226	0.87377	0.86739	0.84660
pFMeasure (%)	81.64612	70.42052	88.19599	80.79836	83.49864
NRM	0.14418	0.18008	0.12303	0.12928	0.14696
PSNR	10.15788	10.24538	11.04351	11.36981	11.23875
DRD	13.86569	16.92871	7.93286	12.51750	7.42475
MPM (x1000)	41.97335	30.80609	27.62239	35.21861	18.51041

Table 8: Parameters of Nick's Thresholding Results after GCS based Decolorization

	PR1	PR2	PR3	Pr4	PR5
Precision	0.92264	0.98094	0.99202	0.94576	0.99948
Recall	0.56342	0.71854	0.64722	0.65622	0.40526
F-measure (%)	66.87893	82.22478	77.70416	76.45629	57.66917
Sensitivity	0.56342	0.71854	0.64722	0.65622	0.40526
Specificity	0.98387	0.99359	0.99525	0.98877	0.99997
BCR	0.77365	0.85607	0.82123	0.82249	0.70261
BER (%)	22.63533	14.39332	17.87686	17.75060	29.73858
F-measure of sens/spec (%)	71.65204	83.39729	78.43586	78.88802	57.67721
Geometric Accuracy	0.74454	0.84495	0.80258	0.80551	0.63659
pFMeasure (%)	78.47213	94.32041	95.30023	89.02787	76.83957
NRM	0.22635	0.14393	0.17877	0.17751	0.29739
PSNR	11.84415	12.52720	11.22042	11.97931	10.86654
DRD	10.29961	6.59340	6.05374	5.35269	8.70630
MPM (x1000)	6.96427	2.54271	3.37251	1.40425	0.72633

Table 9: Parameters of Proposed Method

	Nick's Method	Nick+Decolor	Proposed Method
Precision	0.767856	0.71124	0.968168
Recall	0.79308	0.766934	0.598132
F-measure (%)	77.77011	73.67	72.18667
Sensitivity	0.79308	0.766934	0.598132
Specificity	0.955586	0.943652	0.99229
BCR	0.87433	0.855294	0.79521
BER (%)	12.56681	14.47079	20.47894
F-measure of sens/spec (%)	86.64969	84.57753	74.01008
Geometric Accuracy	0.870404	0.850516	0.766834
pFMeasure (%)	85.33871	80.91193	86.79204
NRM	0.12567	0.144706	0.20479
PSNR	11.60858	10.81107	11.68752
DRD	9.720432	11.7339	7.401148
MPM (x1000)	27.16349	30.82617	3.002014

Table 10 Comparison of performance parameters

Table 10 shows the comparison of various performance parameters of the stated method with conventional Nick's thresholding algorithm and Nick's method after application of GCS based decolorization. The precision rate, specificity and pfmeasure are improved significantly and results show reduction in DRD and MPM values.

5.4 Conclusion:

The proposed thresholding algorithm produces better binarized images as compared to the standard contemporary algorithms. This argument is even more supported by the computation of the above performance parameters. The proposed algorithm removes some of the existing degradations such as ink stains, bleed-through, poor contrast due to age of paper and the ink and filters out the noise that the Nick Algorithm fails to address. The running time of this algorithm is less due to its simplicity as compared to other algorithms.

Chapter 6: Conclusion

In the above thesis first section of work covers application of some standard thresholding algorithms on Document Images which were tested on the DIBCO 2011 dataset and performance parameters are evaluated. Merits and demerits of each algorithm is also explained.

The second section of work covers a novel method of text extraction using Difference of Gaussian followed by Niblack's thresholding algorithm. The DoG is applied in the pre-processing stage which helps in reducing the noise in the image. The proposed method is best suited for slanted images. PSNR values for proposed method are compared with some standard and recently proposed methods which validate the superiority of the method.

The third section describes an algorithm for text extraction from Document images. The algorithm uses GCS based decolorization in the place of conventional RGB to Gray conversion followed by log based Nick's thresholding. The output images and the performance parameters clearly shows that the proposed method is superior to the standard Nick's algorithm and GCS based Nick's Thresholding algorithm.

Appendix-1

1. Precision: Precision *Nitrogiannis, Gatos and Pratikakis (2008)* is the percentage of estimated ground truth image EG which is identified in the binarized output image B.

$$Precision = \frac{\sum_{x=1,y=1}^{x=Ix,y=Iy} EG(x,y).B(x,y)}{\sum_{x=1,y=1}^{x=Ix,y=Iy} B(x,y)} \times 100\% \tag{23}$$

False alarms are the foreground areas of the binary image which is not identified during the estimation of precision.

2. Recall: Recall *Nitrogiannis, Gatos and Pratikakis (2008)* is defined as the percentage of the skeletonized ground truth image SG which is identified in the binarized image B. Recall is given by the following equation:

$$Precision = \frac{\sum_{x=1,y=1}^{x=Ix,y=Iy} SG(x,y).B(x,y)}{\sum_{x=1,y=1}^{x=Ix,y=Iy} B(x,y)} \times 100\% \tag{24}$$

3. F-Measure: It is the harmonic mean of precision and recall. It is given by *Gatos, Pratikakis and Perantonis (2006)*

$$FM = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{25}$$

4. Specificity: It is defined as the portion of actual negative which are predicted as negative.

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{26}$$

5. Balanced Classification Rate: BCR *Trier and Taxt (1995)* is the mean of specificity and recall.

$$BCR = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{27}$$

6. Pseudo F-measure (pFmeasure): pfmeasure *Leedham, Yan, Tan and Mian* introduced in *Gatos, Ntirogianni and Pratikakis(2016)* is defined by the following set of equations-

The skeletonized ground truth image is defined as:

$$SG(x.y) = \begin{cases} 0, background \\ 1, text \end{cases} \tag{28}$$

$$\text{p- Recall} = \frac{\sum_{x=1,y=1}^{x=M,y=N} SG(x,y).B(x,y)}{\sum_{x=1,y=1}^{x=M,y=N} SG(x,y)} \tag{29}$$

p-FM = $\frac{2 \text{ x p−Recall x Precision}}{\text{p−Recall+Precision}}$ (30)

7. (NRM): NRM *Gatos, Pratikakis and Perantonis (2006)* denotes the pixel wise mismatch between estimated ground truth image and the binarized image. It is a combined result of false positive and rate NR_{FP} false negative rate NR_{FN}.

$$NRM = \frac{NR_{FP}+NR_{FN}}{2} \quad (31)$$

Where $NR_{FM} = \frac{N_{FM}}{N_{FM}+N_{TP}}$, $NR_{FP} = \frac{N_{FP}}{N_{FP}+N_{TP}}$ (32)

8. Peak Signal to Noise Ratio (PSNR): PSNR *Gatos, Pratikakis and Perantonis (2006)* is a measure of how close 2 images are. Thus, the higher the value of PSNR, the higher the similarity between the two PxQ images. The difference between foreground and background is taken to be equal to C and MSE is the mean square error.

$$PSNR = 10\log\left(\frac{C^2}{MSE}\right) \quad (33)$$

$$MSE = \frac{\sum_{x=1}^{x=M}\sum_{y=1}^{y=N}\left(I(x,y)-I'(x,y)\right)^2}{PQ} \quad (34)$$

9. Misclassification Penalty Metric (MPM) : MPM *Leedham, Yan, Tan and Mian* evaluates prediction object by object against the Ground truth image.

$$MPM = \frac{MP_{FN} + MP_{FP}}{2} \quad (35)$$

Where $MP_{FN} = \frac{\sum_{i=1}^{N_{FN}} d_{FN}^{i}}{D}$, $MP_{FP} = \frac{\sum_{j=1}^{N_{FP}} d_{FP}^{j}}{D}$ (36)

d_{FN}^{i} and d_{FP}^{j} denote the distance of the i^{th} false negative and j^{th} false positive from the contour of the text in the GT image. The normalization factor D is the sum of over all the pixel to contour distances of the GT object.

10. Distance Reciprocal Distortion Metric (DRD): DRD *Trier and Taxt (1995)* is used to measure the visual distortion in binary images. It measures the distortion for all the S flipped pixels as follows

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN}\text{pixel} \quad (37)$$

DRD represents the distortion of k^{th} flipped pixel.

REFERENCES

[1]. Otsu, Nobuyuki. "A threshold selection method from gray-level histograms." *IEEE transactions on systems, man, and cybernetics* 9.1 (1979): 62-66.

[2]. Sauvola, Jaakko, and Matti Pietikäinen. "Adaptive document image binarization." *Pattern recognition* 33.2 (2000): 225-236.

[3]. W. Niblack, "An Introduction to Image Processing", Prentice-Hall, Englewood Cliffs, NJ, pp. 115-116, 1986.

[4]. P. K. Sahoo, S. Soltani, A. K. C. Wong, A survey of thresholding techniques, Computer Vision, Graphics Image Processing, 41 (2), 233&260, 1988

[5]. Gatos, B., Ntirogianni, K. and Pratikakis, I., "ICDAR 2009 document image binarization contest (DIBCO)" , Proc. Int. Conf. Document Anal. Recognit. , pp.1375 -1382 , 2009

[6]. Gatos, B., Ntirogianni, K. and Pratikakis, I., "H–DIBCO 2010—Handwritten document image binarization competition" , Proc. Int. Conf. Frontiers Handwriting Recognit. , pp.727 -732 , 2010

[7]. Gatos, B., Ntirogianni, K. and Pratikakis, I., "ICDAR 2011 Document Image Binarization Contest (DIBCO 2011)", Proc. 11th Int. Conf. Document Analysis and Recognition,2011.

[8]. Sauvola, J. and Pietikinen, M., "Adaptive document image binarization," Pattern Recognition, vol. 33, pp. 225–236, 2000

[9]. J. Bernsen, "Dynamic thresholding of gray-level images," in Proc. Int.Conf. Pattern Recognit., , pp. 1251–1255, Oct. 1986

[10]E. Zemouri, Y. Chibani, and Y. Brik, "Enhancement of historical document Images by combining global and local binarization technique," International Journal of Information and Electronics Engineering, vol. 4, pp. 1-5, 2014.

[11]Liu, Qiegen, et al. "GcsDecolor: gradient correlation similarity for efficient contrast preserving decolorization." *IEEE Transactions on Image Processing*24.9 (2015): 2889-2904.

[12] Natarajan, Jayanthi, and Indu Sreedevi. "Enhancement of ancient manuscript images by log based binarization technique." *AEU-International Journal of Electronics and Communications* 75 (2017): 15-22.

[13] He, Jingyu, et al. "A comparison of binarization methods for historical archive documents." *Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on*. IEEE, 2005.

[14]. Sharma S., "Extraction of Text Natural Images", IJARCSSE, Vol.3, [2007].

[15]. Agarwal P, Varma R "Text Extraction from Images", IJCSET, April 2012, Vol 2, Issue 4, 1083-1087

[16]. Gonzalez R.C., Woods R.E., Eddins S.L., Digital Image Processing Using MATLAB, Pearson Education [2004].

[17]. M. Swamy Das, B. HimaBindhu, A. Govardhan, "Evaluation of Text Detection and Localization Methods in Natural Images", IJETAE, ISSN 2250-2459, Volume 2, Issue 6, June 2012

[18]. S. Ferreira, V. Garin, and B. Gosselini, "A text detection technique applied in the framework of a mobile camera-based application," in Proc. Int. Workshop Camera-Based Doc. Anal. Recognit. 2005, pp. 133–139.

[19]. S. A. Angadi and M. M. Kodabagi, "Text region extraction from low resolution natural scene images using texture features," Advance Computing Conference (IACC), 2010 IEEE 2nd International, Patiala, 2010, pp. 121-128

[20]. Giri, P. S., "Text Information Extraction and analysis from Image Using Digital Image Processing Techniques". International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 11, November 2014 pg. 576-582

[21] Yusuke Matsuda, Shinichiro Omachi and Hirotomo Aso, “String Detection from Scene Images by Binarization and Edge Detection”, IEICE, D, Vol.J93-D, No.3, pp.336-344, 2010.

[22] P. K. Sahoo, S. Soltani, A. K. C. Wong, A survey of thresholding techniques, Computer Vision, Graphics Image Processing, 41 (2), 233&260, 1988

[23] B. Sankur and M. Sczgin. "Survey over image thresholding techniques and quantitative performance evaluation ", pp-148, Journal of Electronic Imaging, January 2004, Vol. 13(1)

[24] Bhattacharya, Ujjwal, Swapan Kumar Parui, and Srikanta Mondal. "Devanagari and bangla text extraction from natural scene images." 10th International Conference on Document Analysis and Recognition, pp. 171-175, IEEE, 2009.

[25] Albovik, “Handbook of Image and Video Processing”, Academic Press, 2000.

[26] R.M.Haralick, L.G.Shapiro, “Computer and Robot Vision”, Volume 1, Addition- Wesley Publishing Company Inc., 1992.

[27] Lindeberg, T. 1994. Scale-space theory: A basic tool for analysing structures at different scales. Journal of Applied Statistics, 21(2):224-270.

[28] Raj, H. and Ghosh, R., 2014, September. “Devanagari text extraction from natural scene images”. In Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on (pp. 513-517). IEEE.

[29] P.R.S. Mendonca, D. Padfield, J. Miller, and M. Turek, “Bias in the Localization of Curved Edges,” Proc. European Conf. Computer Vision, part 2, pp. 554-565, May 2004

[30] Yin, Xu-Cheng, et al. "Robust text detection in natural scene images." IEEE transactions on pattern analysis and machine intelligence 36.5 (2014): 970-983.

[31] Graham Leedham, Chen Yan, et al. "Comparison of some thresholding algorithms for text/background segmentation in difficult document images." *Proceedings of the seventh international conference on document analysis and recognition*. Vol. 2. 2003.

[32] DIBCO-2011 Dataset, http://utopia.duth.gr/~ipratika/DIBCO2011/benchmark/dataset/DIBCO11-machine_printed.rar

[33] DIBCO-2015 Dataset, http://utopia.duth.gr/~ipratika/DIBCO2015/benchmark/dataset/DIBCO15-natural_scene.rar