AUTOMATIC TEXT EXTRACTION FROM COMPLEX COLORED IMAGES USING GAMMA CORRECTION METHOD

Sumathi, C.P. and G. Gayathri Devi

Department of Computer Science, S.D.N.B. Vaishnav College for Women, Chennai-600 044, Tamil Nadu, India

ABSTRACT

The aim of this study is to propose a new methodology for text region extraction and non text region removal from complex background colored images. This study presents a new approach based on Gamma correction by determining a gamma value for enhancing the foreground details in an image. The approach also uses gray level co-occurrence matrices, texture measures, threshold concepts. The proposed method is a useful preprocessing technique to remove non text region and to show the text region in the image. Experiments were on various images from the datasets collected and tagged by the ICDAR robust reading dataset collection team. Experimental results show that the proposed method has a good performance on extracting text regions in an image.

Keywords: Gamma Correction, GLCM, Texture Measures, Thresholding, Text Region Extraction

1. INTRODUCTION

Text Extraction from image is concerned with extracting the relevant text data from a collection of images. Rapid development of digital technology has resulted in digitization of all categories of materials. Lot of resources are available in electronic medium. Many existing paper-based collections, historical manuscripts books, journals, scanned document, video images, maps, posters, broadsides, newspapers, micro facsimile, microfilms, university archives, book plates, graphic materials, coins, currency, stamps, business cards, advertisements, web pages are converted to images and these images present many challenging research issues in text extraction and recognition. Text extraction from images have many useful applications in document analysis, detection of vehicle license plate, analysis of article with tables, maps, charts, diagrams, keyword based image search, identification of parts in industrial automation, content based retrieval, object identification, street signs, text based video indexing, page segmentation, document retrieving, address block location.

Due to growing requirement for information many research work has been done on text extraction in images. Several techniques have been developed for extracting the text from an image. The existing methods were based on morphological operators, wavelet transform, artificial neural network, skeletonization operation, edge detection algorithm, histogram technique. All these techniques have their benefits and restrictions.

2. RELATED WORK

Strouthopoulos *et al.* (2002) method was based on a combination of an Adaptive Color Reduction (ACR) technique, Principal Component Analyzer (PCA) and Self-Organized Feature Map (SOFM) was used to achieve color reduction.

Zhan *et al.* (2006) proposed an algorithm that uses the multiscale wavelet features and the structural information to locate candidate text lines. Then a SVM classifier was used to identify true text from the candidate text lines. Kumar *et al.* (2007) proposed a scheme for the extraction of textual areas from an image using Globally Matched Wavelet Filters (GMW) filters with Fisher classifiers. To improve the

Corresponding Author: Sumathi, C.P., Department of Computer Science, S.D.N.B. Vaishnav College for Women,
Chennai-600 044, Tamil Nadu, India

result Markov Random Field (MRF) based post processing had been applied.

Liu and Sarkar (2008) proposed an algorithm that employs two novel filters and a basic component-based text detection framework. The framework uses the Niblack algorithm to threshold images and groups components into regions with commonly used geometry features. Pan *et al*. (2009) proposed a novel hybrid method where a Conditional Random Field (CRF), Minimum Classification Error (MCE) learning and graph cuts inferece algorithm were followed.

Babu *et al*. (2010) proposed a new text extraction algorithm which was insensitive to noise, skewness and text orientation, color or intensity, layout and orientation from a text/graphics heterogeneous document images. Zhang and Kasturi (2010) proposed a new unsupervised text detection approach which is based on Histogram of Oriented Gradient and Graph Spectrum.

Zhang and Kasturi (2011) proposed a new text extraction approach based on character and link energies by analyzing the properties of single characters and text objects.

The article (Sumathi *et al*., 2012) discussed in detail about the various existing schemes on extracting the text from an image.

3. GAMMA CORRECTION

In a CRT monitor, the voltage V driving the electron flow is related to the intensity I of light emitted by the phosphor hit by the electron according the formula $I \approx V^{\gamma}$ where γ is a constant that depends on the phosphor Jayaraman *et al*. (2011). This non linearity causes serious distortions in the colors of the image displayed with some areas being too dark, while others are being saturated. To avoid this problem, an amplitude filter is required that corrects the intensity of each pixel before the information is passed to the CRT called gamma correction. The correction is given by $V \approx I^{1/\gamma}$ Gamma correction controls the overall brightness of an image. When the amount of γ is less than one, the transformed image becomes lighter than the original image; and when the amount of γ is greater than one, the transformed image becomes darker than the original image. The gamma curve is shown in **Fig. 1**.

To enhance the image, knowledge of gamma is required. A proper estimation of gamma value enhances the contrast of the image. This correction must be applied to the pixel intensities before the signal is converted into voltage in the CRT.

4. TEXTURAL ANALYSIS

Texture analysis refers to the characterization of regions in an image by their texture content. Texture analysis attempts to quantify intuitive qualities described by terms such as rough, smooth, silky, or bumpy as a function of the spatial variation in pixel intensities.

Texture is one of the significant characteristics used in identifying objects or regions of interest in an image. Texture contains important information about the structural arrangement of surfaces. Texture can be defined as a regular repetition of an element or pattern on a surface.
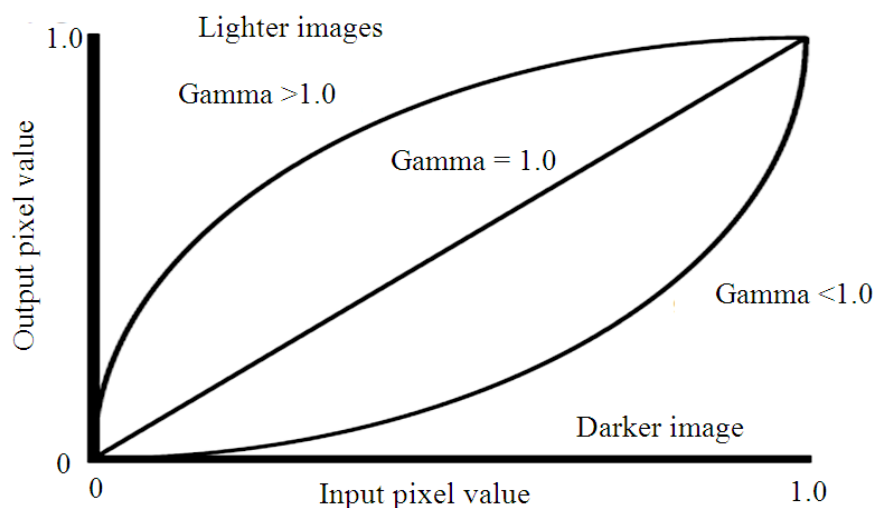

Fig. 1. Gamma curve

Spectral, textural and contextual features are three fundamental pattern elements used in human interpretation of images. Spectral features describe the average tonal variations in various bands of the visible and/or infrared portion of an electromagnetic spectrum. Textural features contain information about the spatial distribution of tonal variations within a band. Contextual features contain information derived from blocks of pictorial data surrounding the area being analyzed.

First order texture measures are statistics calculated from the original image values, like Mean, Variance, Skewness, Kurtosis and do not consider pixel neighborhood relationships. Spatial gray level co-occurrence estimates image properties related to second-order statistics which considers the relationship among pixels or groups of pixels (Srinivasan and Shobha, 2008).

4.1. Gray-Level Co-Occurrence Matrix

Haralick *et al*. (1973) suggested the use of Gray Level Co-occurrence Matrices (GLCM) often called gray tone spatial dependence matrix or Gray level Dependency Matrix in second order texture information extraction in an image. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM and then extracting statistical measures from this matrix.

GLCM is defined as "A two dimensional histogram of gray levels for a pair of pixels, which are separated by a fixed spatial relationship". GLCM of an image is computed using a displacement vector, defined by its radius δ and orientation θ.

4.1.1. Radius δ

Large displacement δ value would yield a GLCM that does not capture detailed textural information. Pixel is more likely to be correlated to other closely located pixel than the one located far away, so classification accuracies with $\delta = 1$ and $\delta = 2$ give best results. Displacement value equal to the size of the texture element improves classification.

4.1.2. Angle θ

Every pixel has eight neighboring pixels allowing eight choices for θ, which are 0, 45, 90, 135, 180, 225, 270 or 315°. However, taking into consideration the definition of GLCM, the co-occurring pairs obtained by choosing θ equal to 0° would be similar to those obtained by choosing θ equal to 180°. This concept extends to 0, 45, 90 and 135° as well. Hence, one has four choices to select the value of θ.

4.1.3. Quantized Gray Levels (G)

The dimension of a GLCM is determined by the maximum gray value of the pixel. Number of gray levels is an important factor in GLCM computation. More levels would mean more accurate extracted textural information, with increased computational costs.

4.2. Textural Features

Some of the texture features extracted from gray level co-occurrence matrix are discussed below (**Table 1**):

- Energy: This statistic also called Uniformity or Angular Second Moment. ASM describes image smoothness. It measures the textural uniformity that is pixel pair repetitions
- Contrast: Contrast evaluates local gray level variation. This statistic measures the spatial frequency of an image and is a difference moment of GLCM
- Correlation: The correlation feature is a measure of gray tone linear dependencies in the image
- Inverse Difference Moment: This statistic is also called as homogeneity. It measures image homogeneity as it assumes larger values for smaller gray tone differences in pair elements
- Entropy: This statistic measures the disorder or complexity of an image. The entropy is large when the image is not texturally uniform and many GLCM elements have very small values

Table 1. Texture features

Texture feature	Formula
Energy	$f_1 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)^2$
Contrast	$f_2 = \sum_{n=0}^{N_g-1} n^2 \sum_{i=1}^{N_g} \sum_{j=1, |i-j|=n}^{N_g} p(i,j)$
Correlation	$f_3 = \dfrac{\sum_i \sum_j (i,j)p(i,j) - u_x u_y}{\sigma_x \sigma_y}$
Homogenity	$f_5 = \sum_i \sum_j \dfrac{1}{1+(i-j)^2} p(i,j)$
Entropy	$f_9 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log(p(i,j))$

5. THRESHOLDS

Threshold is one of the widely used methods for image segmentation. It is useful in discriminating foreground from the background. By selecting an adequate threshold value T, the gray level image can be converted to binary image. The binary image should contain all of the essential information about the position and shape of the objects of interest (Al-Amri *et al.*, 2010). The advantage of obtaining first a binary image is that it reduces the complexity of the data and simplifies the process of recognition and classification. The most common way to convert a gray-level image to a binary image is to select a single threshold value (T). Then all the gray level values below this T will be classified as black (0) and those above T will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold T.

The thresholding methods are classified into six groups according to the information they are exploiting (Sezgin and Sankur, 2004). These categories are:

- Histogram shape-based methods, where, for example, the peaks, valleys and curvatures of the smoothed histogram are analyzed
- Clustering-based methods, where the gray-level samples are clustered in two parts as background and foreground object, or alternately are modeled as a mixture of two Gaussians
- Entropy-based methods result in algorithms that use the entropy of the foreground and background regions, the cross-entropy between the original and binarized image
- Object attribute-based methods search a measure of similarity between the gray-level and the binarized images, such as fuzzy shape similarity, edge coincidence
- Spatial methods use higher-order probability distribution and/or correlation between pixels
- Local methods adapt the threshold value on each pixel to the local image characteristics

5.1. Otsu's Method

This is used to automatically perform histogram shape-based image thresholding or the reduction of a gray level image to a binary image. The algorithm assumes that the image to be threshold contains two classes of pixels or bi-modal histogram (e.g., foreground and background) then calculates the optimum threshold separating those two classes so that their combined spread (intra-class variance) is minimal. The extension of the original method to multi-level thresholding is referred to as the Multi Otsu method.

Otsu's method (Otsu, 1979) exhaustively searches for the threshold that minimizes the intra-class variance (the variance within the class), defined as a weighted sum of variances of the two classes:

$$\sigma^2_w(t) = \omega_1(t)\sigma^2_1(t) + \omega2(t)\sigma^2_2(t)$$

Weights ω_i are the probabilities of the two classes separated by a threshold t and σ^2_i variances of these classes.

Otsu shows that minimizing the intra-class variance is the same as maximizing inter-class variance:

$$\sigma^2_b(t) = \sigma^2 - \sigma^2_w(t)$$
$$= \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2$$

Which is expressed in terms of class probabilities ω_i and class means μ_i.

The class probability $\omega_1(t)$ is computed from the histogram as t:

$$\omega_1(t) = \sum_o^t p(i)$$

While the class mean is $\mu_1(t)$:

$$\mu_1(t) = \left(\sum_o^t p(i)x(i)\right) / \omega_1$$

where, x(i) is the value at the center of the ith histogram bin. Similarly, $\omega_2(t)$ and $\mu_2(t)$ can be computed on the right-hand side of the histogram for bins greater than t. The class probabilities and class means can be computed iteratively. This idea yields an effective algorithm.

6. PROPOSED METHOD

Aim of this research is to suppress non-text background details from the image by applying appropriate gamma value and to remove non text region. The present research is to estimate the gamma value of an input image. To find a proper gamma value, a range of gamma values from 0.1 to 10, with interval of 0.1 is applied for the image resulting in 100 images. Gray Level co-occurrence matrix for each image is computed to extract the textural features contrast and energy after converting into gray image. Each image is processed for

four combinations of radius and angle that is 1 and 0°, 1 and 45°, 1 and 90° and finally 1 and 135°. Contrast and Energy measures of matrices of the four orientations are averaged and threshold value is calculated for each image by using Otsu's threshold algorithm and recorded in a table. The table will have the column of Gamma, Contrast, Energy and Threshold, **Table 2**.

6.1. Estimation of Gamma Value

The value of Energy and contrast has to be examined for the image of gamma value = 1 (Original Image) from the generated table to determine the value of gamma.

Rule 1: If the value of Energy >= 0.05, find an instance where threshold value is 0.5 from the table. If more than one instances are found, select an instance which has maximum value of Contrast and Energy >= 0.05. If there is no instance found, find an instance where threshold value is next nearer to 0.5. The corresponding gamma value of this selected instance is the estimated gamma value.

Rule 2: If the value of Energy <0.05 and the value of Contrast >= 1000, find an instance which has the value of Energy >= 0.1, the value of Contrast >= 1000 and threshold value of 0.5 from the table for the Gamma values 1 to 10. If more than one instance is found, select an instance which has the value of Energy maximum and the value of Contrast >1000. If there is no such instance found, find an instance in between gamma value of 0.1 and 0.9 such that value of the threshold should be nearer or next nearer value of 0.5. The corresponding gamma value of this selected instance is the estimated gamma value.

Rule 3: If the value of Energy <0.05 and the value of Contrast <1000, find an instance which has the value of Energy >= 0.1, the value of Contrast is maximum and the maximum contrast value should be greater than 100 for the Gamma values 1 to 10.If no such instance found, find an instance in between gamma value of 0.1 and 1 such that value of the threshold should be nearer or next nearer value of 0.5. The corresponding gamma value of this selected instance is the estimated gamma value.

By applying this estimated gamma value to an input image, the background suppressed image will be achieved. This Gamma corrected image will be converted to gray scale image. At last, Otsu's thresholding algorithm is used to calculate the threshold value and applied to create output image.

7. EXPERIMENTATION

ICDAR datasets which contains images are used for experiment. To find a proper gamma value, a range of gamma values from 0.1 to 10, with interval of 0.1 is applied to an input image (**Fig. 2**) and the values of Contrast, Energy, Threshold is recorded. **Table 2** is obtained for the input image. The value of contrast is 483.1 and Value of Energy is 0.000582 are found for the original Image. Gamma value is estimated by applying the Rules stated Rule1/Rule2/Rule3. Here, Rule 3 is applied for the input image as the value of Energy <0.05 and the value of Contrast <1000 and Gamma value estimated is 5.5. By applying this gamma value to the input image, the background suppressed image is achieved (**Fig. 2b**). The Gamma Corrected image is converted to gray scale image and a threshold value (Otsu's Algorithm) is evaluated and applied to create output image (**Fig. 2c**).

When a range of gamma values from 0.1 to 10, with interval of 0.1 is applied to an input image (**Fig. 6a**) **Table 3** is obtained. The value of contrast is 1870 and the value of Energy is 0.00676are found for the original Image. Rule 2 is applied for the input image as the value of Energy <0.05 and the value of Contrast >1000. From the table it is found that there is no instance which has the value of Energy >= 0.1, the value of Contrast >= 1000 and threshold value of 0.5 for the Gamma values 1 to 10. So, Gamma value must be in between 0.1 and 0.9. By applying the last part of Rule 2, Gamma value is determined as 0.7. The output image is shown in (**Fig. 6c**).

When an input image (**Fig. 7a**) is processed with the range of 0.1 to 10 gamma values **Table 4** is obtained. The value of contrast is 617.7 and value of Energy is 0.4858 are found for the original Image. Rule 1 is applied for the input image as the value of Energy >= 0.05. More than one instance is found, so an instance which has the value of Energy maximum and the value of Contrast >1000 has been selected. The corresponding value is of this instance is 3.1. Gamma Corrected image and output image is shown in the (**Fig. 7b-7c**) respectively.

Table 2. Evaluated values for the image **Fig. 2a**

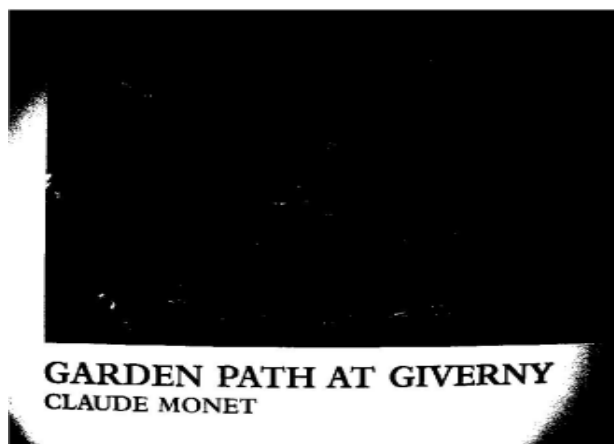| Gamma	Contrast	Energy	Threshold
0.1	20.45	0.005404	0.90588
0.2	67.96	0.001928	0.82353
….			
5.3	472.4	0.10082	0.25098
5.4	474.4	0.10321	0.25098
5.5	475.8	0.105	0.24706
5.6	466	0.1099	0.24706
….			
9.9	352.7	0.4772	0.32941
10	350.9	0.4889	0.33333

(a)

(b)

(c)

Fig. 2. Original Image Contrast = 483.1 and Energy = 0.000582, Estimated Gamma Value = 5.5, (a) Original Image (b) Gamma Corrected Image (c) Output Image

(a)

(b)

(c)

Fig. 3. Original Image Contrast = 61.39 and Energy = 0.001455, Estimated Gamma Value = 7, (a) Original image (b) Gamma corrected image (c) Output image

(a)

(b)

(c)

Fig. 4. Original image contrast = 33.31 and Energy = 0.001303, estimated gamma value = 0.6, (a) original image (b) gamma corrected image (c) output image

(a)

(b)

(c)

Fig. 5. Original image contrast = 32.6 and energy = 0.00109, estimated gamma Value = 9.2, (a) original image (b) gamma corrected image (c) output image

(a)

(b)

(c)

Fig. 6. Original image contrast = 1870 and Energy = 0.00676, estimated gamma value = 0.7, (a) original image (b) Gamma corrected image (c) output image
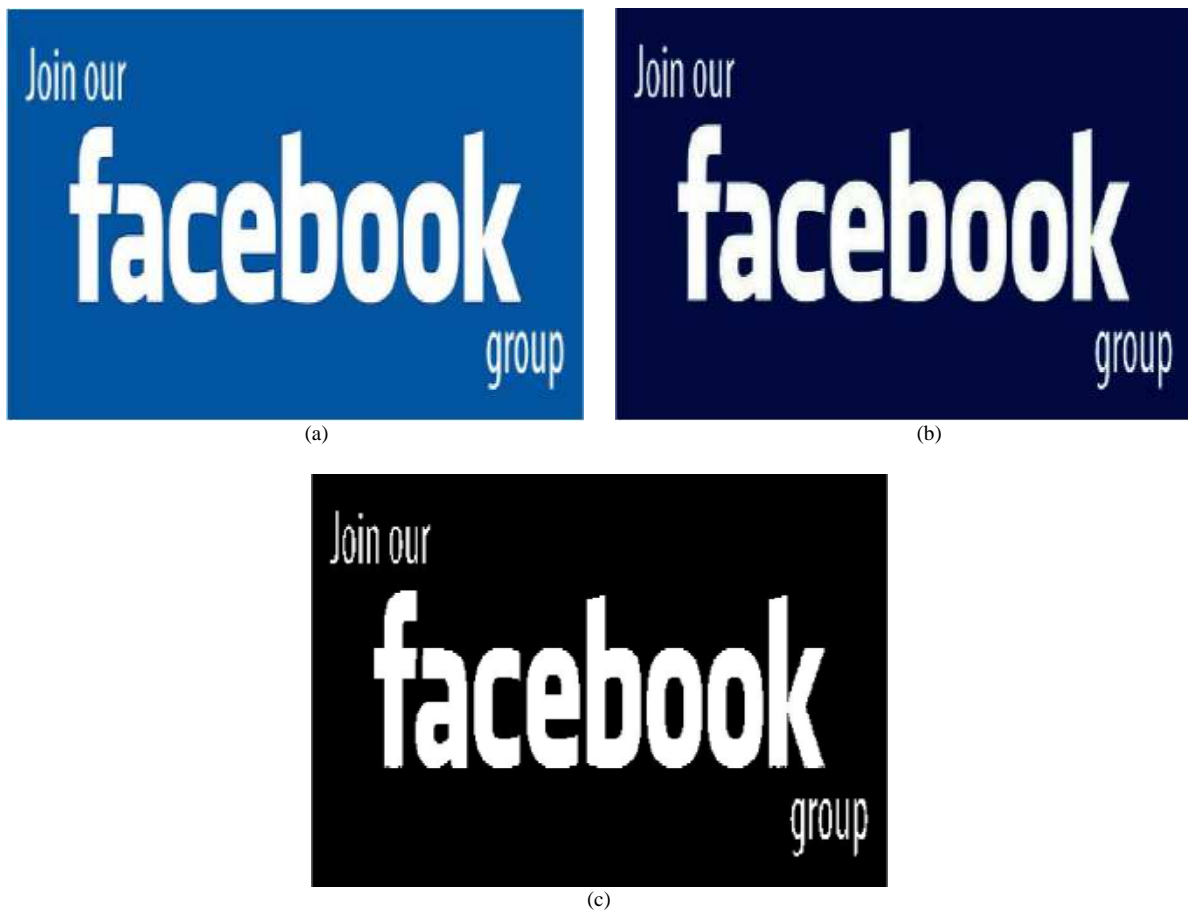

(a)

(b)

(c)

Fig. 7. Original image contrast = 617.7 and energy = 0.4858, estimated gamma value 3.1, (a) original image (b) gamma corrected image (c) output image

(a)   (b)

(c)

Fig. 8. Original image contrast = 458.2 and energy = 0.00089, estimated gamma value = 0.4, (a) Original image (b) gamma corrected image (c) output image

Table 3. Evaluated values for the image **Fig. 6a**

Gamma	Contrast	Energy	Threshold
0.1	784.7	0.053810	0.67843
….			
0.6	1487	0.010300	0.61176
0.7	1594	0.009676	0.58608
0.8	1692	0.009846	0.58839
…...			
9.9	3249	0.006462	0.42745
10	3245	0.006444	0.42745

Table 4. Evaluated values for the image **Fig. 8a**

Gamma	Contrast	Energy	Threshold
0.1	48.29	0.4872	0.89804
0.2	133.3	0.4847	0.81961
……..			
2.9	1003	0.4663	0.52941
3	1014	0.4823	0.50961
3.1	1023	0.4907	0.50784
3.2	1021	0.5042	0.50608
…			
10	1300	0.6358	0.42784

Table 5. Results on ICDAR Dataset

Algorithm	Precision	Recall
Ashida *et al*. (2003)	0.55	0.46
Wolf-Maier *et al*. (2003)	0.30	0.46
Liu and Sarkar (2008)	0.67	0.46
Zhang and Kasturi (2010)	0.67	0.46
Pan *et al*. (2009)	0.67	0.71
Zhang and Kasturi (2011)	0.73	0.62
Proposed Method	0.78	0.91

To evaluate the performance of the proposed method, ICDAR Robust Reading Dataset is used. Some of the Experiments results have been shown in the **Fig. 2-8**.

The results in this research show that the new proposed method removes the background of non text region and retains the text details of the image. Proposed method can detect most text region successfully, including text with different styles, size, font, orientations and color. **Table 5** shows the comparison of proposed algorithm with the other algorithms. The

proposed text extraction approach has the average precision rate of 78% and recall rate of 96%.

8. CONCLUSION

The paper presents a new algorithm for the extraction of text region information in an image. This proposed method suppresses the background non text region detail based on Gamma correction. The algorithm estimates the gamma value by using texture measure without any prior details of the imaging device. The proposed technique is an essential preprocessing stage for most of the object separation method. The algorithm is applied on several images with text of different styles, size, font, alignment and complex backgrounds taken from ICDAR datasets and shown promising results. The future work mainly concentrates on the next stage of developing an exact and fast algorithm to remove further unwanted details and separate alpha numeric character from the output image obtained from the new proposed technique.

9. REFERENCES

Al-Amri, S.S., N.V. Kalyankar and S.D. Khamitkar, 2010. Image segmentation by using threshold techniques. J. Comput.

Ashida, H., Y. Saito, C. Kojima, K. Kobayashi and N. Ogasawara *et al*., 2003. A functional link between RuBisCO-like protein of *Bacillus* and photosynthetic RuBisCO. Science, 302: 286-290. DOI: 10.1126/science.1086997

Babu, G., R. Mohan, P. Srimaiyee and A. Srikrishna, 2010. Text extraction from hetrogenous images using mathematical morphology. J. Theoretical Applied Inform. Technol., 16: 39-47.

Haralick, R., K. Shanmugam and I. Dinstein, 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybernet., 3: 610-621. DOI: 10.1109/TSMC.1973.4309314

Jayaraman, S., S. Easikkirajan and T. Veeraraman, 2011. Digital Image Processing. 1st Edn., Tata McGraw-Hill Education, New Delhi, ISBN-10: 1259081435, pp: 723.

Kumar, S., R. Gupta, N. Khanna, S. Chaudhury and S.D. Joshi, 2007. Text extraction and document image segmentation using matched wavelets and MRF model. IEEE Trans. Image Process., 16: 2117-2128. DOI: 10.1109/TIP.2007.900098

Liu, Z. and S. Sarkar, 2008. Robust outdoor text detection using text intensity and shape features. Proceedings of the 19th International Conference on Pattern Recognition, Dec. 8-11, IEEE Xplore Press, Tampa, FL., pp: 1-4. DOI: 10.1109/ICPR.2008.4761432

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man. Cyber., 9: 62-66.

Pan, Y.F., X. Hou and C.L. Liu, 2009. Text localization in natural scene images based on conditional random field. Proceedings of the 10th International Conference on Document Analysis and Recognition, Jul. 26-29, IEEE Xplore Press, Barcelona, pp: 6-10. DOI: 10.1109/ICDAR.2009.97

Sezgin, M. and B. Sankur, 2004. Survey over image thresholding techniques and quantitative performance evaluation. J. Electr. Imag., 13: 146-165. DOI: 10.1117/1.1631315

Srinivasan, G.N. and G. Shobha, 2008. Statistical texture analysis. Proc. World Acad. Sci. Eng. Technol., 36: 1264-1269.

Strouthopoulos, C., N. Papamarkos and A.E. Atsalakis, 2002. Text extraction in complex color documents. Patt. Recogn., 35: 1743-1758. DOI: 10.1016/S0031-3203(01)00167-4

Sumathi, C.P., T. Santhanam and G. Gayathri Devi, 2012. A survey on various approaches of text extraction in images. Int. J. Comput. Sci. Eng. Survey, 3: 27-42. DOI: 10.5121/ijcses.2012.3403

Wolf-Maier, K., R.S. Cooper, J.R. Banegas, S. Giampaoli and H.W. Hense *et al*., 2003. Hypertension prevalence and blood pressure levels in 6 European Countries, Canada, and the United States. J. Am. Med. Assoc., 289: 2363-2369. DOI: 10.1001/jama.289.18.2363

Zhan, Y., W. Wang and W. Gao, 2006. A robust split-and-merge text segmentation approach for images. Proceedings of the 18th International Conference on Pattern Recognition, (ICPR' 06), IEEE Xplore Press, Hong Kong, pp: 1002-1005. DOI: 10.1109/ICPR.2006.169

Zhang, J. and R. Kasturi, 2010. Text detection using edge gradient and graph spectrum. Proceedings of the 20th International Conference on Pattern Recognition (ICPR), Aug. 23-26, IEEE Xplore Press, Istanbul, pp: 3979-3982. DOI: 10.1109/ICPR.2010.968