ISSN : 2231-0711

Tajinder Kaur et al | IJCSET(www.ijcset.net) | July 2015 | Vol 5, Issue 7,272-277

Text Extraction from Natural Scene using PCA

Tajinder Kaur

Department of Computer Engineering
Punjabi University
Patiala, India

Nirvair Neeru

Department of Computer Engineering
Punjabi University
Patiala, India

Abstract— **Many techniques and algorithm have been developed to solve the problem of text extracted from natural scenes. Text extraction is emerging and challenging era in the computer vision. Text which is embedded into the image contains semantic information which is used in many other applications such as information retrieval of complex images, robot navigation, useful for visually impaired persons, street signs, automatic read the sign board and use in so many other applications. Most of the research work in this area has been done only on printed text, a very few research is addressing the LED scene text. Scene text is difficult to extract due to blur image, variations in color, noise problem, complex background, discontinuity, poor lighting conditions, and variation in illumination. LED is Light Emitting Diode which is widely used in displaying the information in LED boards. Now days LED display that is natural scene is being widely used for displaying announcements, sign boards, banners for displaying information. To extract the text from the LED display is not an easy task, it is very complex due to its discontinuity. A matrix of segments is used to display the character of LED, which is combined together to generate an LED text. So, The aim of this paper to propose a technique to extract an LED text from natural scene image. In the preprocessing step, The RGB input image will be converted to a grayscale image, image is binarization and noise is removed. Then FFT and FFT shift is used to extract the text region because the text is generally found in higher frequency and it is the fastest method. The text or non-text region is classified. Finally, apply the template matching method is developed using PCA, which is used to recognize the extracted text and display in boundary boxes. The experimental results of the proposed method show the extraction rate is 73.25.**

Keywords— ***Connected Component method, FFT method, FFT shift, MSER, Morphological operations, PCA, Text extraction.***

I. INTRODUCTION

Text extraction method is a challenging task due to rapidly increase the digitization of all the material. It is complex because we need to find out where the text actually located in the image and how much part is not relevant to the recognition process. A text extraction in natural scene contains useful and valuable information and makes it easy which can be understood by human and computer. This research topic is very active and challenging task in computer vision applications. Text extraction process involves text detection, localization, extraction, segmentation and recognition of text [1]. Text extraction in natural scene image use in many applications such that mobile text recognition, scene understanding, automatic recognition sign board, supports for visually impaired persons, license plate detection, robot navigation, extract traffic sign board text that uses for intelligent transport system, navigational support for tourist guide, information retrieval etc..These are the variety of application which is develop with the use of mobile phone because mobile phone captured the scene text and directly convert into the recognition process. So, when we perform any text recognition, it is very important to extract the text region accurately. Text extraction in natural scene image use for so many applications, but still it is challenging task due complexity of its complex background, color variation, noise problem, image illumination changes, image distortion, blurring problem and lighting condition [2]. To extract the text from the LED display is not an easy task, it is very complex due to its discontinuity. A matrix of segments is used to display the character of LED, which is combined together to generate an LED text.The text which is superimposed into an image contains a useful text which represents the whole image information. Text is mainly classified into two categories:

a. Scene Text
b. Artificial Text.

The scene text is also known as ‘graphics text’. Scene text means the text which is shown in natural image .The artificial text is also known as ‘caption text’ which is superimposed on the image[3]. Following show the two kinds of images:

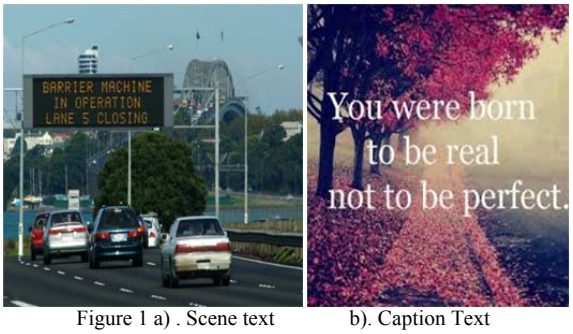

Figure 1 a) . Scene text b). Caption Text

Scene text (Figure. 1 (a)) is a natural type of text which is accidently happening when we capture the image. It

contains the useful information which helps to understand the whole idea about the image. There are some examples of scene text is like a vehicle number plate, street signboards, banners, traffic sign board and so on. This text is difficult to extract due to their various styles, font, color, contrast, complex background, low and high resolution, orientation, alignment, blurring and shadowing effects.

Artificial text (Figure 1 (b)) is also called the 'caption text' which is inserted in the image or video. This text could be segmented, detected and extracted using various techniques. The caption text is added into news channels, movies and videos where the subtitle is superimposed. Caption texts are rotating text, subtitle text, moving text. Artificial text may or not a fixed in position and shape and low resolution problem.

Various existing methods of text detection and extraction for natural scene can be roughly classified into two categories: region based technique, texture based techniques, hybrid technique and morphological based method [4-5].

A. Region based method

Region based method is also known as sliding window based method that uses a bounding box or sliding window to detect a text from a natural scene and use some heuristic technique to recognize text. This method uses a bottom up approach in which small component is successively combined into large until all the area are identified in the image. A morphological operation is used to merge the component and filter out the false region and mark the boundary around each text. In this approach a text region is identified from a complex background and removes the false or non-text region. This approach is based upon color, edge, shape, contour and geometry features [3] [4]. On the basis of these features separates text or non-text region. The speed of region based method is slow as compared to other techniques. Edge based and Connected Component is a further classification of the region based approach.

a. *Edge based Method:* Edge based method are an efficient method for text extraction.Its aim to find out the high contrast between text and the background. Edges are the most important feature of text character rather than orientation, color, layout, etc [5].For detecting the text in images we can use these operators i.e The Canny edge detector, Robert edge detector , Prewitt edge detector and Sobel edge detector .There are the main three properties of the text which is density, edge strength and orientation variance that is superimposed into the image[6].For detecting the boundary of text reliably these properties are mainly used.After that merge the boundary of text and separate the text or non text from the image.
b. *Connected Component based method:* Connected Component method directly extract the Candidate character from natural scene image by using color clusering and edge detection[6]-[7].Segment the connected component and then mege it after that the false positive components are removed using some classifier or heurstic method.CC method computation cost is low as compared to other techniques and the extracted candidate component directly use in recognition process.Without the prior information about the text position or scale the connected component mehod can't extract the candidate character effeciently.

B. Texture based method

The texture based method uses different texture properties to extract a text or to decide wheather or not the pixels belongs to the text of the image[8]. This technique uses the textual properties that separate them from complex background or non text region. Various methods are used in this approach to extract textual information like Wavelets, Fast Fourier Transform and Gabor filters, DCT Transform Wavelet and variance is used to find the textual properties of the text region in the image. A train classifier is used to extract the features of the target image region [9]-[10]. The main aim of train classifier is to distinguish the text or non-text region for a scene.

C. Hybrid technique

The hybrid technique uses a combination of both techniques, i.e. region based and texture based approach. In this, the first step region based approach is used to detect a text or character candidate using the CC method. The features are extracted from text region and use a classifier to decide which region contains a text or non-text on the basis of texture based method[11]-[12]. The main disadvantage of these approaches that the single method is not suitable for all the natural scene images due to size, color, font variation varies from one image to another image.

D. Morphological based Method

Morphological method is based upon geometrical and mathematics approach for character recognition and image analysis[13].It is used to extract the constrast feature of text in the input image.The feature of the geometrical image is never changed when a specfic transformation is applied on it the transformation like translation, scaling and rotation.The feature of the image is still maintain when the text color or even lighting conditions is changed of the image.

As mention above methods, we use a connected component method and PCA for recognition purpose in our proposed method.Our proposed method we work on LED text rather than printed text. LED is a Light Emitting Diode which is widely used in displaying the information in LED boards[14]-[15]. Now days LED display that is natural scene is being widely used for displaying announcements, sign boards, banners for displaying information.LED text is difficult to extract because of its discontinuous nature. A matrix of segments is used to display the character of LED, which is combined together to generate an LED text.The character of the LED text is displayed in matrix form which is coming in rectangular or circle shape.The size of the matrix of the LED character 3× 5, 4×6, 5×7, 5×8 and 6×7 dots [15].

II. PROPOSED METHODOLOGY

The objective of this proposed method to develop a system that will able to extract the LED display board text in natural scenes. As mentioned above, different methods they have their own advantages and limitations.The flow chart of the proposed method of the text extracted from natural scene image is shown in Figure 2. First, The input image is RGB then convert it into grayscale image and preproced the given image. Then apply FFT and FFT shift of the gray scale image.At last step PCA is used for recognition purpose and segment the text in the input image.First discusses the algorithm of the proposed system and after that describes the flow chart(Figure 2) of it.

Algorithm:

- Start
- Acquire the scene image
- Compute the FFT for center weightage
- Compute the center weightage using the pixel density in row & column
- Inverse FFT
- Compute the content weightage using thresholding

  $a \leq THR \leq b$ ----------Eq. 1

  $p = \sum_{i=1}^{n} p(i,j)$----------Eq.2

  $px = mean (\sum p(i,j) \leq THR)$--------Eq.3
- Minimum average is computed using Eq. 3

  $P = \sum p \leq px$
- Extract the character from P
- Apply PCA to recognize the character
- Make the boundary boxes or segment the detected text

Flow Chart of the Proposed Method:

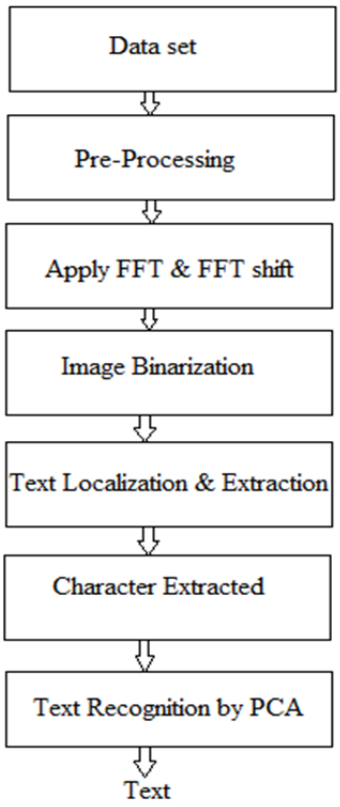

Figure 2. Flow Chart

A. *Preprocessing :*

The preprocessing is a very important step in text extraction method because it achieves a better result.First, we load the image into the database which is mostly a color image. Then perform the RGB to gray conversion on the preprocessed image and convert it into gray scale.We perform the conversion because the input image is RGB form every pixel of this convert it into a single value of output which reflects the higher brightness of that pixel.After that noise is removed using filters.

B. *FFT & FFT shift apply for grouping the pixels:*

Before applying FFT & FFT shift,

1) First, we check the frequency of pixels and grouping the pixels with nearest neighboring pixels with at least distance. Then use the Euclidean distance to calculate the weightage of column or row. The weighted column near the center of the image are emphasized in order to make the centrally located text component more prominent in image matrix.

   Euclidean Distance (Ed) = $\sqrt{(i-i')^2 + (j-j')^2}$ ----------Eq.1

   i & i' denote the x coordinate and j & j' the y coordinate.

   Edw = $\sum_{i=1-n}^{interval=100} Ed(i,j,interval)$-------Eq.2

   Edw is denotes weighted Euclidean distance function. It calculates Euclidean weight of 100 row or column.

   $TH = \sum Edw$ ----------Eq.3

   Here, TH denotes the threshold value. Then no. of structuring elements.

   The image is made prominent on the basis of centrally weighted rows or column & the row or column on edge are ignored or eroded to level best possible.

   $Ed_L = \int_0^N Edw$----------Eq.4
2) In Second step, apply FFT (Fast Fourier Transform) which is a fast and efficient method to extract the features or frequencies in the frequency domain.Convert the image datatype 'uint8' to 'double' before applying the FFT.Compute the 2-D FFT upon the image matrix to group the pixels in the nearest bindings.

   2-D FFT F(u,v) can be expressed following formula in the image f(x,y),

   $F(u,v) = \frac{1}{MN} \sum_{x=0}^{M} \sum_{y=0}^{N} f(x,y) e^{-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)}$ ---Eq.5

   Where x & y initial pixels coordinate and u & v are the coordinates of the transformed image.
3) After FFT then apply Fourier Shift in the computed FFT image.FFT shift is used to shift the computed binding weight to the center of the image according to Fast Fourier weightage.The distance & Fast Fourier is merged together to intensify the center pixel groups.
4) Inverse Fourier shift is applied to spread the weight across the neighbor pixels of the image,

which will make the whole central area visually prominent.

5) Apply inverse 2-D Fourier to dilute the visual effects computed by fft. Inverse FFT formula, f(x,y) = $\sum_{x=0}^{M}\sum_{y=0}^{N} F(u,v)e^{j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)}$ ……..Eq.6

C. Image Binarization:

To binarize the image first calculate the threshold using the following formula:

$$T = \frac{fmax-(fmax-fmin)}{3} \text{ ----------Eq.7}$$

Convert the image Pixel values into a binary value according to threshold T.

$$g(i,j) = \begin{cases} 1 & if\ p(i,j) \geq T \\ 0, & if\ p(i,j) < T \end{cases} \text{----------Eq.8}$$

D. Text Localized & Extracted:

Text localization means where the text in the image is accurately located.In text localization, we apply connected component method which is described above.The text is localized by rectangular bounding box. The centrally located features across the adaptive frequency area are marked as an ROI (Region of Interest).The candidate text region is white with the black background is extracted from the image. Filter out the text region from the non-text region.

E. Character Extracted:

The text is extracted, now extract the single –single character from the whole text The extracted character is segmented from each other with a labeling or a single bounding box around the character .

F. Text Recognized by PCA:

At the last step, we perform a template matching to generate a template for recognition purpose.Template matches use for classifying the input character of the given image with the template.The extracted characters or text is recognized by PCA i.e Principal Component Analysis.

Principal Component Analysis:

PCA is a powerful and useful method which preserves the information of the image while reducing the dimensionality and extract the useful features. Another name of the PCA is Karhunen Loeve expansion, which is used for feature extraction. It transforms the data into new coordinates by orthogonal linear transformation, i.e. is called PCA.It divides the data into coordinates, The data which a maximum variance comes under a first coordinate, other maximum variance of the second coordinates and so on. It reduces the dimensions of existing data such that greater variance in the input data is visible in compact, feature region.PCA is a mathematical system which converts a high dimension & correlated variables into lower dimension & uncorrelated variables is known as principal component. It is the use of compression and reduce the amount of redundant information.The computation is easier because the data is reduced in size. PCA is applied to calculate the covariance matrix from the image matrix and finding the Eigen values and Eigen vector. PCA is used to extract the principal component corresponding to the Eigen values which is selected in the covariance matrix.

PCA can work in two steps, i.e. Training & Testing steps.

In the training phase, The Eigen vector is found by calculating the covariance matrix of the training data set and the dataset is collected of the Eigen vector of the different images with a higher Eigen value. Other testing phase, tested image is mapped to same Eigen vectors and calculate the Euclidean distance for classification.

Algorithm of PCA:

Training Stage:

1. Obtain the data set containing N training image samples of dimension M×M : $X_1, X_2, X_3 \ldots\ldots, X_N$ is defined.
   where X_i column vector.
2. Convert these N images into vectors X_i,
   $1 \leq i \leq N$ of dimension M^2.
3. Obtain mean(m) image vector,
   $m = \frac{1}{N}\sum_{i=1}^{N} X_i$
4. Obtain the difference vector(ϕ_i) by subtracting the mean(m) image vector of the training image.
   $\phi_i = X_i$ - m
5. Obtain the covariance Matrix C having dimensions N×N and find its Eigen vector & Eigen values.

   $C = \frac{1}{M}\sum_{n=1}^{M} \phi_n {\phi_n}^T$
   $= AA^T$
   $A = [\phi_1, \phi_2, \phi_3, \ldots\ldots\ldots \phi_n]$ with dimension M× N.
6. Compute the Eigen vectors e_i of AA^T
   As the dimensions of AA^T are very large so computation of eigenvectors is impractical.
7. Obtain Eigen values u_i of A^TA [dimensions N × N.
   AA^T has M^2 Eigen vectors and Eigen values.
   A^TA has N Eigen vectors and Eigen values.
8. Obtain the best N eigenvectors of AA^T by following equation.
   $e_i = Au_i$
9. Take only V Eigen vectors corresponding to V largest Eigen values.

Represent the Training Data set using Eigen Vectors:

Weight of each training image is calculated as:

$w_j = {e_j}^T . (X_i$ - m) ,
where j =1,2,3,…….....V.

Weight vector is represented as:

$\mu = [w_1, w_2, w_3, \ldots\ldots, w_N]^T$;

Every image in training database is represented by weight vector:

$\mu_i = [w_1^i, w_2^i, w_3^i \ldots\ldots, w_N^i]^T$;

Testing Stage:

Let the image to be tested is q, its weight (w_i) is calculated by multiplying Eigen vector e_i with the difference image.

$w_i = {e_i}^T . (q$ - m)

Weight vector of unknown image is calculated as:

$\mu = [w_1, w_2, w_3, \ldots\ldots, w_N]^T$;

Compute:

$\varepsilon_j = \min \| \mu - \mu_j \|$; j = 1, 2,3… N

N is no. of training images.
So, ɖ is recognized as jth hand gesture from training database.

III. EXPERIMENTAL RESULTS

The proposed model is implementing MATLAB 2013a and tested on an intel core i_3 with 4GB RAM running on windows 7.

The proposed system is tested using 20 different images and found the number of characters which have been detected correctly and then calculate the character level accuracy for each image using the following formula:

$$\textbf{Accuracy} = \frac{\text{Correctly Detected Characters}}{Total\ Charactes} \times 100\text{-----Eq.1}$$

The result of the accuracy of each image is shown is a table and graph of accuracy is obtained.

Table 1: Showing performance of the proposed system on a per sample basis.

Sr. No.	Total Characters	Correctly Detected Characters	Accuracy in Percentage (%)
1	10	10	100
2	19	16	84.21
3	38	25	65.78
4	15	15	100
5	4	4	100
6	9	1	11.11
7	13	8	61.53
8	13	13	100
9	16	10	62.5
10	9	9	100
11	12	12	100
12	10	10	100
13	10	0	0
14	8	8	100
15	5	0	0
16	5	5	100
17	8	4	50
18	16	8	50
19	5	4	80
20	4	4	100
Average Accuracy			73.85

In the table 2, the measurable mistakes have been gotten from the after effects of proposed calculation. The aggregate number of characters, accurately and dishonestly recognized characters, not distinguished numbers, genuine positives, and genuine negatives, false positive and false negative properties have been gotten from the consequences of the proposed calculation and precision and recall is calculated.

Table 2: Statistical Errors

Category	Result
Total Characters	229
Correctly Detected Characters	166
False Detected Characters	41
Not detected Characters	22
True Positive	166
True Negative	0
False Positive	41
False Negative	22
Sensitivity (Recall)	88.30%
Precision	80.19%

$$\textbf{Precision:} \frac{\text{True Positive}}{\text{True Positive + False Positive}} \text{----------Eq.2}$$

$$\textbf{Recall:} \frac{\text{True Positive}}{\text{True Positive + False Negative}} \text{----------Eq.3}$$

Result:

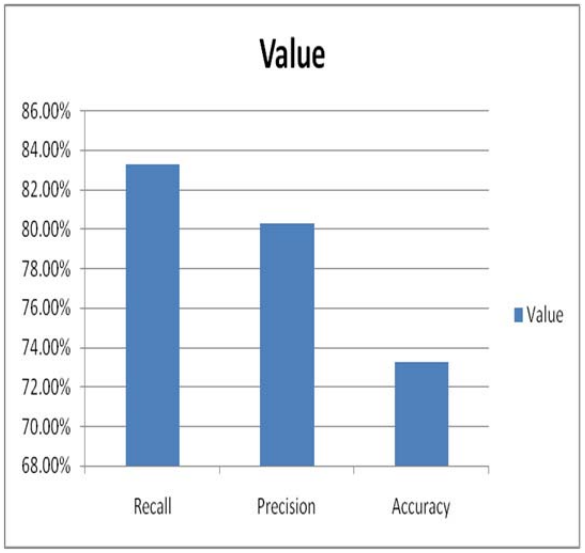

Figure 3. The statistical analysis of the character recognition results

IV. CONCLUSIONS

In this research paper, we represent a method of text extracted from natural scene using PCA.The text is extracted from LED display board which can widely use in street board, traffic sign board, parking space, information about the alternate route for intelligent transport system. The algorithm uses connected component method to select the candidate character in the image.Then the extracted text is recognized by PCA. The proposed method uses FFT and PCA, which is a computational fast method gives better accuracy result than the others.

The proposed method effectiveness is texted on 20 different images from the database which is collected from

different environment and internet.Test the accuracy of the proposed technique, calculate the Precision, Recall and F-measure(Accuracy).We found the number of characters which have been detected correctly and then calculate the character level accuracy for each image.Then the result of the proposed method show the detection rate of precision and recall i.e. 80.19% and 80.19%.The experimental result show the effectiveness and overall accuracy 73.85% of the proposed work which is done in LED images.

The proposed technique has some limitation which will be considered as a future work. This technique can't detect text which is continuous. Sometime the text is not detected and extract properly due to image is tilted, shadow area and complex background. The proposed method focus on centrally weighted text, detect the text easily. The algorithm discards the false region of the tilted or shadowing area. The FFT and PCA are used to make the system computationally and gives the better accuracy rate of the proposed technique. So the system can be further enhanced by removing the above limitations.

REFERENCES

[1] Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu, "A Hybrid Approach to Detect and Localize Texts in Natural Scene Images", IEEE Transactions on Image Processing, Vol. 20, No. 3, March 2011.

[2] Anhar Risnumawan, Palaiahankote Shivakumara, Chee Seng Chan and Chew Lim Tan, "A Robust Arbitrary Text Detection System For Natural Scene Images", Expert System with Application 41(2014) 8027-8048.

[3] Prachi R. Dussawar and Prof. Parul Bhanarkar Jha, "A Survey on Comparision and Performance Analysis of Text Extraction Techniques", IJARSE, Vol. No.4, Issue No.01, January 2015.

[4] H. Raj, R. Ghosh, "Devanagari Text Extraction from Natural Scene Images", 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2014, pp. 513-517.

[5] Xu-Cheng Yin, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao, "Robust Text Detection in Natural Scene Images", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 36, no. 5, May 2014.

[6] Shivananda V. Seeri, J. D. Pujari and P. S. Hiremath, "Multilingual Text Localization in Natural Scene Images using Wavelet based Edge Features and Fuzzy Classification", IJETTCS , Volume 4, Issue 1, January-February 2015.

[7] L. Neumann and J. Matas, "Real-time scene text localization and recognition," in Proceedings of International Conference on Computer Vision and Pattern Recognition , 2012, pp. 3538–3545.

[8] Cong Yao, Xiang Bai and Wenyu Liu, "A Unified Framework for Multi-oriented Text Detection and Recognition", IEEE Trans. On Image Processing, Vol. 23, No. 11, November 2014.

[9] Shraddha Naik and Sankhya Nayak, "Text Detection and Character Extraction in Natural Scene Images", International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 2, February 2015.

[10] Chucai Yiand Yingli Tian, "Localizing Text in Scene Images by Boundary Clustering, Stroke Segmentation, and String Fragment Classification", IEEE Trans. on Image Processing, Vol. 21, No. 9, September 2012.

[11] Neha Gupta, V.K Banga, "Localization of Text in Complex Images Using Haar Wavelet Transform", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-1, Issue-6, November 2012.

[12] Xu-Cheng Yin, Wei-Yi Pei , Jun Zhang and Hong-Wei Hao, "Multi-Orientation Scene Text Detection With Adaptive Clustering", IEEE Trans. On Pattern Analysis and Machine Intelligence, 2013.

[13] R.Chandrasekaran and RM. Chandrasekaran, "Morphology based Text Extraction in Images", IJCST Vol. 2, Issue 4, Oct. - Dec.

[14] Jack Greenhalgh and Majid Mirmehdi, "Recognizing Text-Based Traffic Signs", IEEE Trans. on Intelligent Transportation Systems, 2014.

[15] Wahyono and Kanghyun Jo, "LED Dot matrix text recognition method in natural scene" Neurocomputing 151, pp 1033–1041, 2015.

[16] Cunzhao Shi, Chunheng Wang, Baihua Xiao, Yang Zhang and Song Gao, "Scene text detection using graph model built upon maximally stable extremal regions" Pattern Recognition Letters 34 107–116, 2013.

[17] P.Subbuthai, Azha Periasamy and S.Muruganand, "Identifying the Character by Applying PCA Method using Matlab", International Journal of Computer Applications (0975 – 8887) Volume 60– No.1, December 2012.

[18] J.Ashok and Dr.E.G.Rajan, "Principal Component Analysis Based Image Recognition" , International Journal of Computer Science and Information Technologies, Vol. 1 (2) , 2010, 44-50.

[19] Pramod Kumar Pandey, Yaduvir Singh and Sweta Tripathi, "Image Processing using Principle Component Analysis", International Journal of Computer Applications , Volume 15– No.4, February 2011.