ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 5, Issue 12, December 2016

Extracting text from natural scene images by HOG Character Descriptor

P. Shiva Reddy, Dr.M.N.Giri Prasad

Abstract--- Text information in scene images can provide valuable information for many applications such as automatic geocoding, assistive navigation, scene image understanding and context based image retrieval. However extracting text directly from complex back ground with multiple shapes and colours is a difficult task because of cluttered backgrounds with noise, non - text outliers and diverse text patterns such as fonts, sizes andcharacter types. The main focus of this system is scene text recognition from detected text regions. In text detection, previously proposed algorithms such as layout analysis of colour decomposition and horizontal alignment are used to obtain text regions from scene images. Proposed system uses character descriptor histogram of oriented gradients (HOG) to extract representative and discriminative text features from character patches. It is able to detect text strings from complex background and recognize characters in the text regions.

Keywords: Scene text detection, scene text recognition, character descriptor, stroke configuration, text understanding, text retrieval.

I. INTRODUCTION

Camera-based text informationcan provide valuable information for many applications such as media analysis, assistant navigation, content retrieval and scene understanding. In scene images and videos, text characters, numbers(0 to 9) and strings appear in nearby sign boards and hand-held tools and provide useful information about surrounding environment and objects.To extract information from scene images and videos, an efficient scene text detection and scene text recognition algorithms are essential. However, extracting text directly from complex back ground with multiple shapes and colours is a difficult task because of the following two factors. 1. Complex back grounds with noise and non-text outliers. 2. Distinct text patterns such as fonts, sizes and character types. Text detection and recognition in scene images is difficult task than recognizing text from scanned documents. To solve these problems extracting the text from scene images is divided into two categories [19], text detection and text recognition. We design two recognition schemes firstly, a character descriptor is used to foretell the category of a character in an image patch. It combines different feature detectors (Harris Detector (HD), MSER Detector (MD), Dense Detector (DD) and Random Detector(RD)) and Histogram of oriented Gradients (HOG) descriptor[5] and [1]. Secondly, we model a binary classifier for each character class by designing stroke configuration from character boundary.

II. RELATED WORK

In this section, we present a general overview of previous methods on scene text recognition. Optical character recognition(OCR) systems [2] can achieve almost perfect recognition rate on printed text in scanned documents but cannot accurately recognize text information directly from camera-captured scene images and videos. And they are sensitive to fonts, sizes and character types. Even though some OCR systems started to support scene character recognition, the recognition accuracy rate is still much lower than the recognition for scanned documents. Scale-invariant feature transform (SIFT) is an algorithm adopted to recognize text characters in different languages. Voting and geometric verification algorithm [13] was used to filter out false positive matches. Character structure was modelled by HOG features and cross correlation analysis of character similarity for text recognition and detection.

III. LAYOUT-BASED SCENE TEXT DETECTION

In natural scene images and videos, text characters, numbers (0 to 9) and strings appear in nearby signboards and hand-held tools. They are composed of characters in uniform colour and aligned in a line, while non-text background layers are in the form of disorganized layouts. Layout-Based scene text detection is divided into two schemes. They are Layout analysis of colour decomposition and Layout analysis of horizontal alignment [17]. By using colour uniformity and horizontal alignment, we can localize text regions in camera-based images.

A. Layout Analysis of Colour Decomposition

According to our observations the text characters and numbers (0 to 9) on nearby sign boards and hand-held tools in general appear in uniform colour. So we can identify text information by extracting pixels with similar colours. To divide a camera-based image into multiple colour-based layers, we have created a boundary clustering algorithm based on bigram colour uniformity [17]. Text information is generally attached to a plane surface with uniform colours. We model the uniformity of their colour difference as bigram colour uniformity. Colour difference is related to character boundary, which provides as a boundary between text edges and attachment surface. Then we design colour difference by a vector of colour pair which is obtained by cascading the RGB colours of text and surfaces. Each boundary can be described by a colour pair, and then we

All Rights Reserved © 2016 IJARECE

cluster the boundaries with similar colour pairs into the same layer. Boundaries of text characters are separated from the complex background outliers.

B. Layout Analysis of Horizontal Alignment

According to our observation, text information generally appear in strings composed of different characters in similar size and nearly horizontal alignment. We model an adjacent character grouping algorithm [18] to identify for image region containing text information. To design the boundary size and location of a text information, a bounding box is assigned to each boundary in a colour layer. For each bounding box, we search for its siblings in similar size and vertical locations. If many sibling bounding boxes are obtained on its left and right then we group all these bounding boxes into a single region. This region contains an imperfect part of text strings. Then we repeat above procedure to find all text information in this colour layer, and this process is called as adjacent character grouping.

IV. Structure-Based Scene Text Recognition

From the detected text locations, character recognition is performed to extract text information. We design two character recognition schemes. In text understanding, character recognition is a multi-class classification problem. For each of the 62 characters, we train a binary classifier to distinguish a character from the other characters. In text retrieval, binary classifier distinguishes character class from other classes or back ground outliers. The specified characters are defined as queried characters. In text recognition, to better design character structure, we model stroke configuration for each character based on specific partitions of character boundary and skeleton.

A. Character Descriptor

We design a character descriptor to model character structure for effective character recognition. Figure shows the flow chart of our proposed character descriptor.it uses four types of detectors, they are Harris Detector(HD), MSER Detector(MD), Dense Detector(DD) and Random Detector(RD). Harris Detector is used to extract key points from junctions and corners. MSER Detector is used to extract the key points from stroke components. Dense Detector is used to extract key points uniformly and Random Detector is used to extract key points in a random pattern. The extracted key points at each detector, the HOG descriptor [1] and [5] is calculated as an observed feature vector x in feature space. Each character patch is normalized into size $128 * 128$. In the process of feature quantization, the BOW model and GMM [1] model are used to aggregate the extracted features. Bag of Words(BOW) model is applied to key points from all the four detectors but Gaussian Mixture Model(GMM) is applied to those only from dense detector and random detector, why because, GMM-Model requires fixed number and locations of the key points. In both model, character patch is mapped into characteristic histogram. By the aggregate of BOW model and GMM model feature representations. We derive the character descriptor with significant discriminative capability for recognition.

B. BOW Model

The Bag of Words Model [1] conveys a character patch from the training set as a frequency histogram of visual words. The BOW model is computationally efficient and resistant to intra-class variations. For each four detectors HD,MD,DD and RD, we develop a vocabulary of 256 visual words. The four detectors are applied to character patch to extract their respective key points, and then their corresponding HOG features are mapped into the respective vocabularies, obtaining four frequency histogram of visual words.

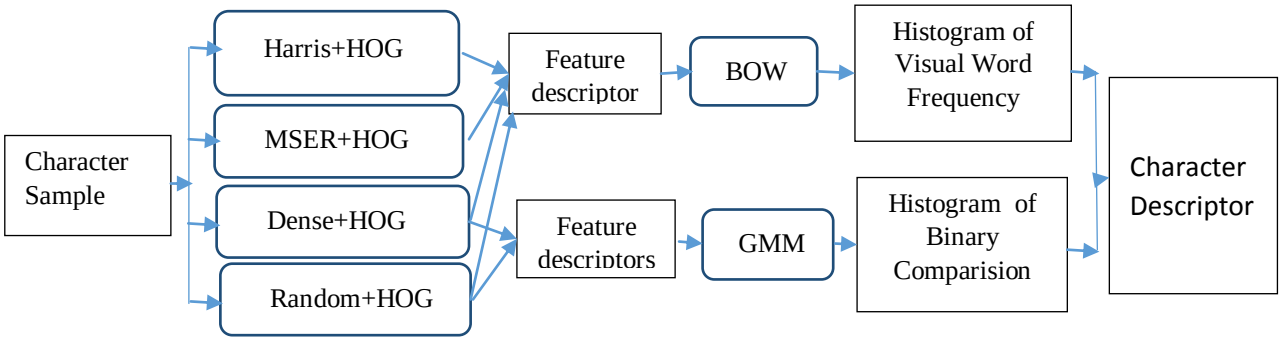

Figure 1.Flow chart of character descriptor

C. Gaussian Mixture Model

To describe the local feature distributions, we develop a GMM [1] contains 8 Gaussian Distributions. In the process of developing GMM, K-means Clustering (K=8) is first applied to calculate K centers of the HOG descriptor, where the sth ($1 \leq s \leq K$) center is used as initial mean μ_s of the s-th Gaussian in GMM. Then the initial weights w_s and covariances σ_s are calculated from the means. A likelihood vector from all Gaussians is represented by equation…..(1)

All Rights Reserved © 2016 IJARECE

$$P_x = \sum_{s=1}^{K} w_s p_s(x|\mu_s, \sigma_s) \dots\dots\dots\dots\dots\dots\dots\dots (1)$$

$$p_s(x|\mu_s, \sigma_s) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Where x denotes a HOG –based feature vector at a key point, P_x denotes the likelihood vector of feature vector x, and $p_s(x|\mu_s, \sigma_s)$ denotes the probability value of x at the sth Gaussian. For likelihood vectors (P_x, P_y), where

$$P_x = \sum_{s=1}^{K} w_s p_s(x|\mu_s, \sigma_s) \text{ and}$$

$$P_y = \sum_{s=1}^{K} w_s p_s(y|\mu_s, \sigma_s)$$

GMM-based feature representation by histogram of binary comparisions.

$$F_{x,y} = \sum_{s=1}^{k} 2^{s-1} * (p^s) \dots\dots\dots\dots\dots\dots\dots\dots (2)$$

$$p^s = 1 \,; if\ w_s p_s(x|\mu_s, \sigma_s) \geq w_s p_s(y|\mu_s, \sigma_s)$$

$$p^s = 0 \,; if\ w_s p_s(x|\mu_s, \sigma_s) < w_s p_s(y|\mu_s, \sigma_s)$$

Character stroke configuration:

Stroke width consistency is employed to detect scene text in complex back ground and achieve better performance. Stoke is a region bounded by two parallel boundary segments. Their orientation is regarded as stroke orientation and the distance between them is regarded as stroke width. The stroke configuration is estimated by synthesized characters generated from computer software. Character boundary and character skeleton are obtained by applying discrete contour evolution [9] (DCE).

We estimate the stroke width and orientation on sample points of character boundary 'n' points are sampled evenly from the polygon character boundary. In our method, we assign 128 to 'n'. The no of points to be sampled on each side of the polygon boundary is proportional to its length. Stroke width and orientation at each boundary sample point 'b' is estimated. We take 'b' and its two neighbouring sample points to fit a line when they are approximately collinear. And then the slope or tangent direction at 'b' is used as stroke orientations. The stroke width is obtained from exploring length along the normal vector at 'b' until another point is occurred. And then we evaluate the skeleton-based stroke map from the consistency of stroke width and orientation. The values of stroke width and orientations are compared with its neighbouring points at each boundary sample point. And these parameters are compatible with the synthesized character patches with size$128 * 128$. The sample points satisfying the stroke related features construct the stroke sections of character boundary. While remaining samples points compose junction section of a character boundary as shown in figure.

Stroke alignment method:

The basic structure of a character class can be described by the mean value of all stroke configurations from class character samples. We employ a stroke alignment method to estimate a mean value of stroke configuration. Hence, it is able to handle various fonts, styles and sizes. Equation (3) indicates objective function of stroke alignment.

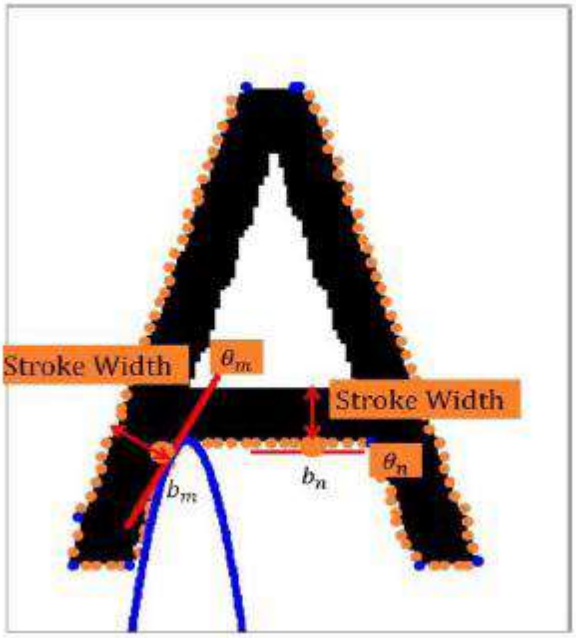

Figure 2.Stroke Orientations and stroke width denoted by red line and red double arrows respectively.,$b_n$is approximately collinear ,while $b_m$fits a quadratic curve with neighbouring point

$$E = \sum_{i} (D(\bar{\mu}, T_i(\mu_i)) + g(T_i)) \dots\dots\dots\dots\dots (3)$$

$$D(\mu_m, \mu_n) = \sum_{j} ||\mu_m(j) - \mu_n(j)||^2$$

Where D is distance between stroke configuration of samples, μ is mean values of stroke configurations, T_i is transformations applied on strokes i-th stroke configuration, $g(T_i)$ is amplitude of transformation.

All Rights Reserved © 2016 IJARECE

V. Experimental Results

To evaluate the performance of the proposed algorithm gray scale image with dimensions $512 * 512$ were chosen. The only parameter in this algorithm is the Accuracy Rate (AR), which isdefined as ratio between number of correctly recognized text characters and the total number of characters. The experimentalresults in first table shows that our proposed descriptor performs better than SYNTH+FERNS with AR 0.47 and comparable with NATIVE+FERNS having AR 0.54 by using Chars74 samples.

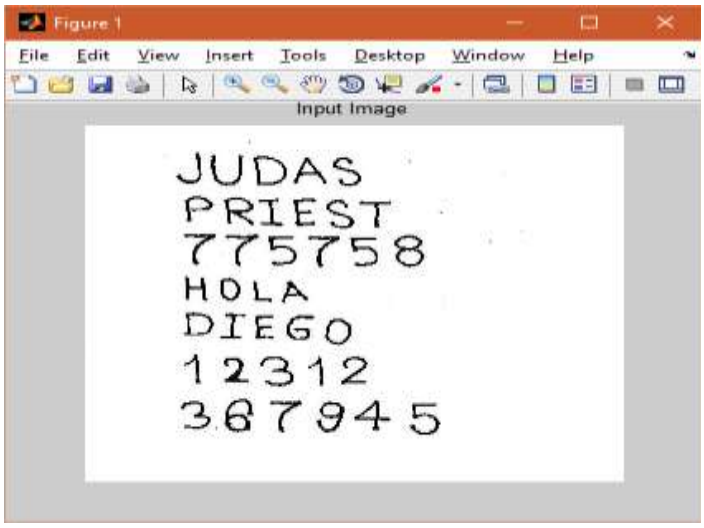

Figure 3. Input Image

Text Detection:

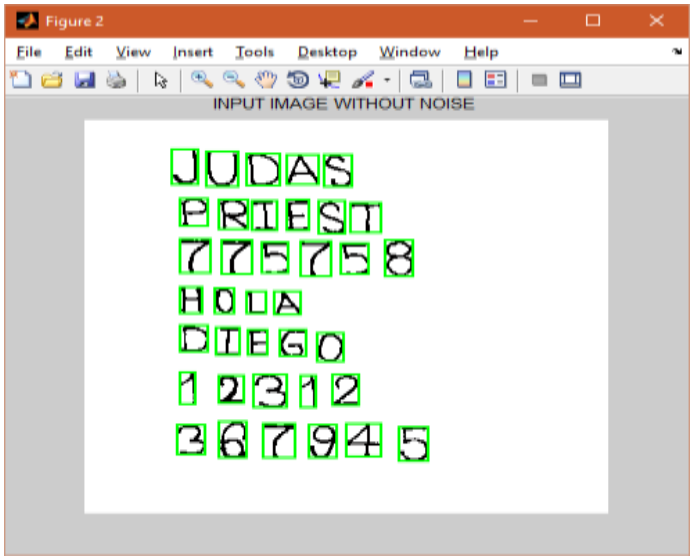

Figure 4. The adjacent character grouping process. The green box denotes bounding box of a boundary in a colour layer.

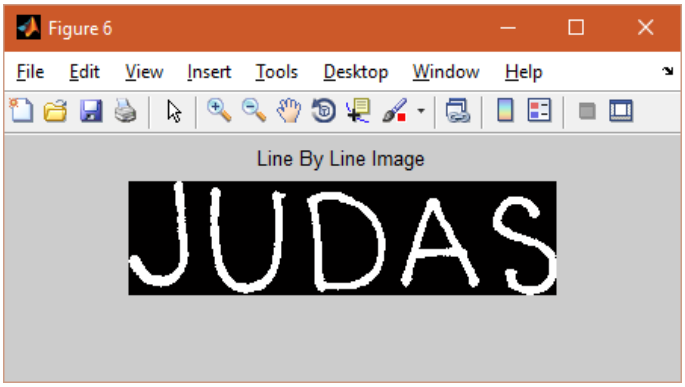

Figure 5. Extracted line one from detected text region

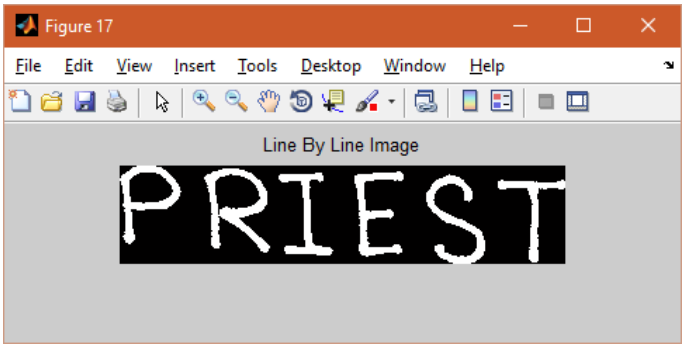

Figure 6.Extracted line two from detected text region

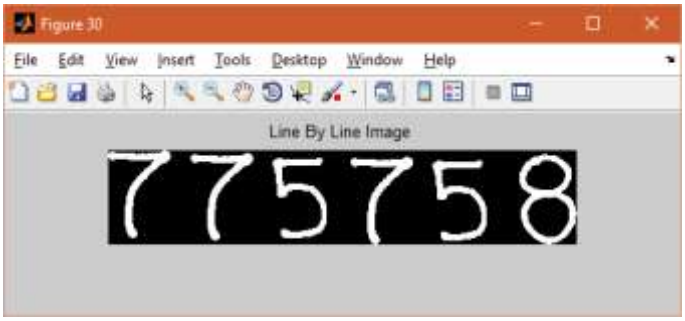

Figure 7. Extracted line three from detected text region

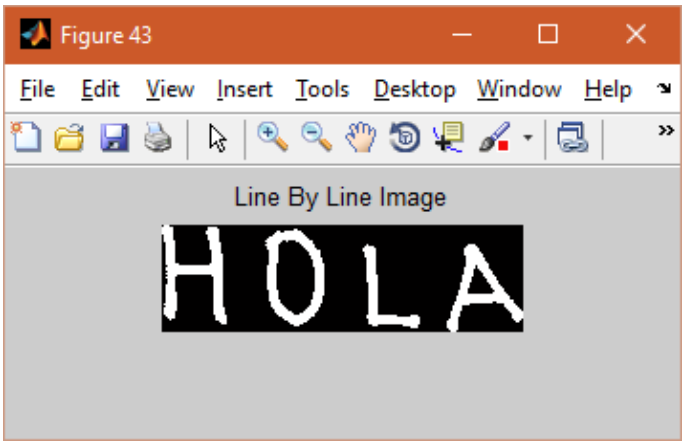

Figure 8. Extracted line four from detected text region

All Rights Reserved © 2016 IJARECE

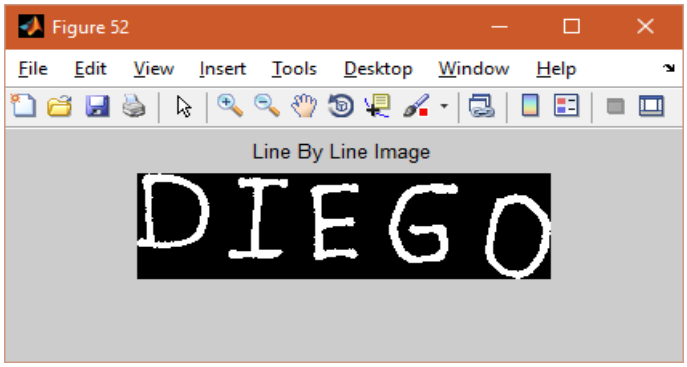

Figure 9. Extracted line five from detected text region

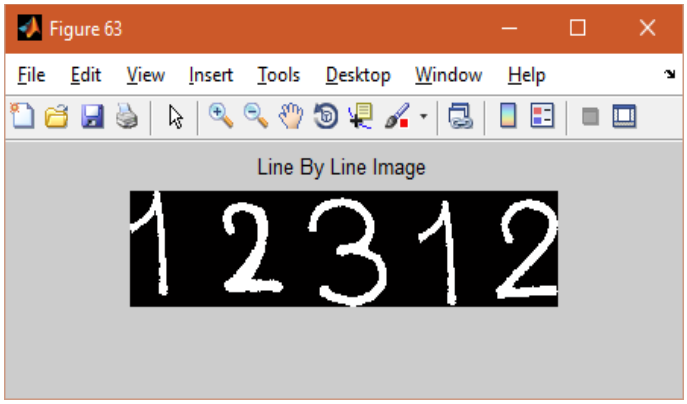

Figure 10. Extracted line six from detected text region

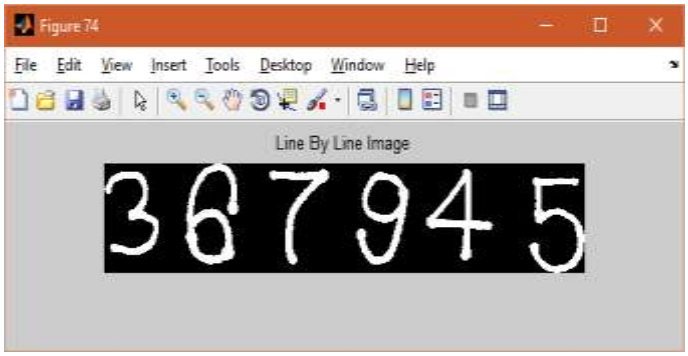

Figure 11.Extracted line seven from detected text region

Text Recognition:

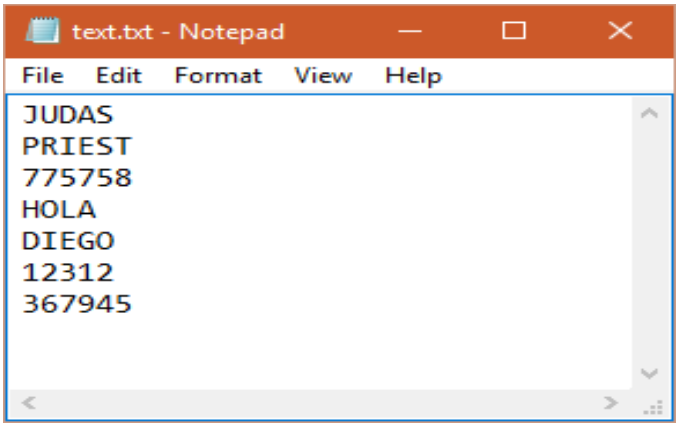

Figure 12.Extracted text information from Detected text region

TABLE I
ACCURACY RATES OF SCENE CHARACTER RECOGNITIONIN CHARS74K DATASET, COMPARED WITH PREVIOUSLY PUBLISHED RESULTS [20]

Chars74K Dataset	AR
Ours	0.60
BOW-based representation only	0.53
GMM-based representation only	0.47
ABBYY	0.31
SYNTH+FERNS	0.47
NATIVE+FERNS	0.54

VI. Conclusion and Future Scope

We have designed a method of scene text recognition from detected text regions. It detects text information regions from the detected text regions. In text detection, layout analysis of colour decomposition and horizontal alignment is performed to search for image regions of text information. In text recognition, character descriptor is effective to extract representative and discriminative text features for recognition schemes. In future work, we will improve the accuracy rate of text detection and lexicon analysis to extend our system to world-level recognition.

REFERENCES

[1] Chucai Yi, Student Member, IEEE, and Yingli Tian, Senior Member, IEEE."Scene Text Recognition in Mobile Applications by Character Descriptor and Structure Configuration". IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 7, JULY 2014

All Rights Reserved © 2016 IJARECE

[2] R.Beaufort and C. Mancas-Thillou, "A weighted finite-state framework for correcting errors in natural scene OCR," in Proc. 9th Int. Conf. Document Anal. Recognit., Sep. 2007, pp. 889–893.

[3] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes," IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99, Jan. 2004.

[4] A. Coates et al., "Text detection and character recognition in scene images with unsupervised feature learning," in Proc. ICDAR, Sep. 2011, pp. 440–445.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2005, pp. 886–893.

[6] T. de Campos, B. Babu, and M. Varma, "Character recognition in natural images," in Proc. VISAPP, 2009.

[7] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. CVPR, Jun. 2010, pp. 2963–2970.

[8] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Johsi, "Text extraction and document image segmentation using matched wavelets and MRF model," IEEE Trans. Image Process., vol. 16, no. 8, pp. 2117–2128, Aug. 2007.

[9] L. J. Latecki and R. Lakamper, "Convexity rule for shape decomposition based on discrete contour evolution," Comput. Vis. Image Understand., vol. 73, no. 3, pp. 441–454, 1999.

[10] N. Nikolaou and N. Papamarkos, "Color reduction for complex document images," Int. J. Imag. Syst. Technol., vol. 19, no. 1, pp. 14–26, 2009.

[11] P. Shivakumara, W. Huang, and C. L. Tan, "An efficient edge based technique for text detection in video frames," in Proc. IAPR Workshop Document Anal. Syst., Sep. 2008, pp. 307–314.

[12] R. Smith, "An overview of the tesseract OCR engine," in Proc. Int. Conf. Document Anal. Recognit., Sep. 2007, pp. 629–633.

[13] K. Wang and S. Belongie, "Word spotting in the wild," in Proc. Eur. Conf. Comput. Vis., 2010.

[14] K. Wang, B. Bbenko, and S. Belongie, "End-to-end scene text recognition," in Proc. Int. Conf. Comput. Vis., Nov. 2011, pp. 1457–1464.

[15] J. J. Weinman, E. Learned-Miller, and A. R. Hanson, "Scene text recognition using similarity and a lexicon with sparse belief propagation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 10, pp. 1733–1746, Oct. 2009.

[16] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu, "Detecting texts of arbitrary orientations in natural images," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 1083–1090.

[17] C. Yi and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping," IEEE Trans. Image Process., vol. 20, no. 9, pp. 2594–2605, Sep. 2011.

[18] C. Yi and Y. Tian, "Localizing text in scene images by boundary clustering, stroke segmentation, and string fragment classification," IEEE Trans. Image Process., vol. 21, no. 9, pp. 4256–4268, Sep. 2012.

[19] J. Zhang and R. Kasturi, "Extraction of text objects in video documents: Recent progress," in Proc. 8th IAPR Int. Workshop DAS, Sep. 2008, pp. 5–17.

[20] T. de Campos, B. Babu, and M. Varma, "Character recognition in natural images," in Proc. VISAPP, 2009.

Mr. P. Shiva Reddy Received B Tech Degree in (ECE) From K.S.R.M college of engineering, Kadapa in 2011. Currently he is doing M Tech in JNTU College of Engineering Anantapur. His areas of Interest are Image processing and Digital electronics.

Dr. M.N. Giri Prasad completed M.Tech and Ph.D. Working as Professor & Head, ECE Department JNTU College of Engineering, Anantapur. Fields of Interest are IMAGE PROCESSING AND BIO-MEDICAL SIGNAL PROCESSING

All Rights Reserved © 2016 IJARECE