Detecting Character in Video Frames for Traffic Vehicle

P. Mahalakshmi[1], T. Manigandan[2], D. Chitra[3]

[1]*Department of CSE, P. A. College of Engineering and Technology, Coimbatore, Tamil Nadu*
[2]*Principal, P. A. College of Engineering and Technology, Coimbatore, Tamil Nadu.*
[3]*Professor, Department of CSE, P. A. College of Engineering and Technology, Coimbatore, Tamil Nadu.*

Abstract

Reading text from video frames is a challenging problem that has received a significant amount of attention. The two key components of most systems are (i) text detection from images and (ii) character recognition and many recent methods have been proposed to design better feature representations and models for both. Detection of text and identification of characters in scene images is a challenging visual recognition problem. As in much of computer vision the challenges posed by the complexity of these images have been combated with hand designed features [1], [2], [3] and models that incorporate various pieces of high-level prior knowledge [4], [5]. The produce results from a system that attempts to learn the necessary features directly from the data as an alternative to using purpose-built text-specific features or models. Detecting text regions in natural scene images has become an important area due to its varies applications.

Scene text detection from video as well as natural scene images is challenging due to the variations in background, contrast, text type, font type, font size, etc. Besides arbitrary orientations of texts with multi-scripts add more complexity to the problem. Text Information Extraction (TIE) System involves detecting text regions in a given image, localizing it, extracting the text part and recognizing text using OCR. The extracted features are used by the trained SVM classifier to detect the text regions. After detecting text regions characters are extracted and finally displayed. Text in camera captured images contains important and useful information. Text in images can be used for identification indexing and retrieval. Detection and localization of text from camera captured images is still a challenging task due to high variability of text appearance detected text regions are merged and localized.

Keywords* – *Text detection, Character Recongnition, Computer Vision, Text Information Extraction, SVM, OCR.

I. INTRODUCTION

Scene text detection from video as well as natural scene images is challenging due to the variations in background, contrast, text type, font type, font size, etc. Besides arbitrary orientations of texts with multi-scripts add more complexity to the problem. Text detection and localization from camera captured images is still a challenging task due to high variability of text appearance.

Scene text shows the diversities in size, font, color, language and alignment.Texts present in camera captured images are considered as one of the important and strong source of information.That image and about the place or situation from the image was captured. Text detection and extraction from images have a lot of valuable and useful application. Texts present in an image or video can be classified as scene text and caption text. Scene text exists in the image naturally. Caption texts refer to those texts which are added manually and easy for user. Scene texts overlap with the background. Therefore scene text detection and extraction are difficult as compared to the detection of caption text. Compared to the scanned document images text extraction from the natural scenes are not easy because they exist in arbitrary orientation different sizes and background interference. Examples of scene texts include signs on streets, display boards on shops, texts on vehicles, advertisement boards etc.

Text detection and extraction from images and video sequences is more effectively. The natural scenes text detection is made difficult by high variation in character color, font, size, and orientation. The light sources introduce highlight, shadow, reflection and color offset in images and also cameras introduce additional noise, blurring, and viewing angle distortion. A video sequence is a succession of images.

The images store a video on a computing support means to store a sequence of images will have to be perfectly presented to the user at sufficient intervals in the videoframes.

The extraction of the text in a video sequence has become controversial since the emergence of data processing and represents a revolution in the world of the multimedia. This Detection and Extraction of the Text in a video sequence consists achieving a system that extracts the inclusive text in video and relying on hypotheses and in trying to take advantage of the previous research this field.

II. RELATED WORKS

For text detection in video a large number of approaches have been developed can be classified into three categories: connected component based [11, 12], edge and gradient based [8, 9, 26], and texture based approaches [6, 7, 20]. Connected component based approaches are fast and good for images that have high contrast texts and plain background just like methods in the document analysis field. Other side these approaches will not be suitable for text detection in video and natural scene images due to low video resolution and natural scene complexity. To improve the performance of text detection edge and gradient based approaches are developed. These approaches are good at recall but poor at precision because the proposed features are sensitive to background complexity leading to more false positives. For text detection in video, a large number of approaches have been developed and processed.

Then it proposes character segmentation based on vertical profiles before text extraction.The most common approach involves three key components [26]: character candidate extraction, character classification, and text grouping. Grouping text as a set of words or sentences depends on the objective of the algorithm and may involve text line estimation and validation. Existing scene text detection algorithms can be divided into two types based on their character candidate extraction method: (1) sliding window based methods that exhaustively scan windows at all possible locations and scales and (2) connected component based methods that utilizes character candidates extracted with particular constraints e.g., consistent stroke width or extremal region.

A variety of machine learning techniques have been used for text detection, including unsupervised feature learning, Convolutional Neural Networks deformable part based models [8] belief propagation and Conditional Random Fields.

The identify text regions using gradient local-correlation to find edge pairs and estimate stroke width. The relationship between different colors and shapes are fed into SVM classifiers to detect text. The natural scene images these are having shadows, non-uniform illumination, low contrast and large signal dependent noise. Connected component analysis is used to define the lastly binary images that mainly consist of text regions. One of their advantage is the ability to compute dense features efficiently over large images. They are best known for their applications to detection and recognition [14] but they have also been used for image segmentation, particularly for biological image segmentation[20].

III. METHODOLOGY

Texture based detection method can detect and localize accurately but speed is problem. Learning based methods give more accurate results but difficult to realize and storage is problem. Clustering based methods are faster but computational complexity is the bottleneck. Text understanding systems include three main phases: text detection, text localization and text extraction. The able to detect the location of text in the natural videos. Scene Images and then extract the text from the videoframes. Text string detection and extraction have a variety and useful applications.The people travel through different places for various purposes it will be difficult for them to understand the text present on display boards in the foreign countries.

SVM is a training algorithm for learning with a strong theoretical foundation in statistical learning theory. Training data set was generated by labeling the features extracted from the test file to recognize a character.

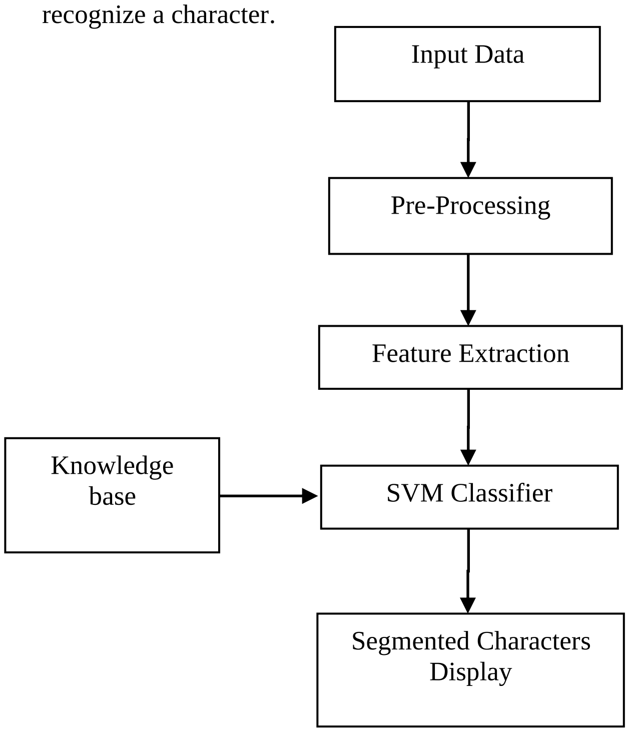


Fig 1: Architecture of Text Detection and Extraction System using SVM classifier

A. Pre-Processing

The detect the text in the image and the given image required to be preprocessed. First converting image into gray scale then binarization of image is done and later noise is removes and edge detection method is applied on the images.

B. Feature Extraction

Feature extraction is a type of dimensionality reduction that efficiently represents interesting parts of an image and those parts are called as compact feature vector. A reduced feature representation is required to quickly complete tasks such as image matching and retrieval. Feature detection feature extraction and matching are combined to solve common computer vision problems such as object detection and recognition texture classification and face detection.

C. Svm Classifier

Support vector machines SVM are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification. The training SVM classifier all features of the candidate region are stored in a file. Training of SVM is done for classification. The datasets will have text images and nontext images. Those will be pre-processed and after that features will be extracted and stored in the files. Later this knowledge will be used to test the input test image.

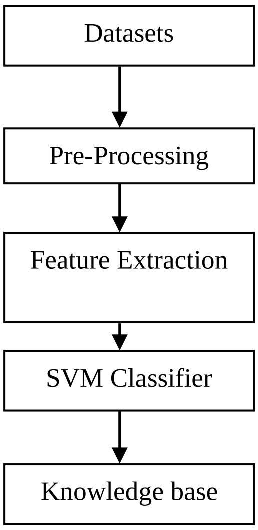
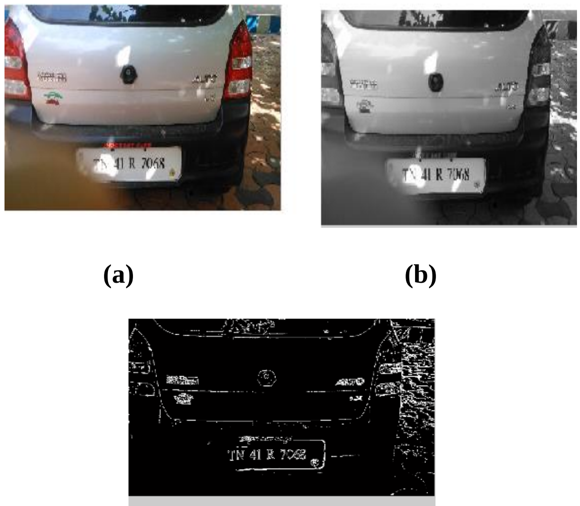


Fig 2: Architecture of training SVM classifier with Text and Non-text data

D. Segmented Character Display

After training of the SVM, it performs classification of the given test image. The test image will be preprocessed and there will be connected components. Features of connected components will be calculated and given input to SVM. So SVM will find the nearest match to the existing trained features. Text detected image will be then further processed in order to extract the characters from the image. The characters detected in the image will be cropped and all the characters will be displayed on the screen.



(a) **(b)**

(c)

E. SVM Classification

The training of SVM we first apply our connected component extraction clustering and feature extraction steps and we train a support vector machine classifier for the classification of

Fig.3 Two level wavelet decomposition.
(a) Original image (b) Grayscale image
(c) Segemented image

square block as text and nontext component. For a testing image, we do all the above steps and finally decision result of all the square blocks of a cluster is integrated. If the number square blocks are text is greater than the non text, then that cluster is classified as a text component.

F. Character Recognition

The identify machine printed characters in an automated manner has obvious applications in numerous fields. Optical character recognition (OCR), as this field is commonly known, has To improve the performance of text detection edge and gradient based approaches are developed. These approaches are good at recall but poor at precision because the proposed features are sensitive to background complexity leading to more false positives. For text detection in video, a large number of approaches have been developed and processed.

G. Text Detection

Even when sign locations are known in images, correctly detecting text on road signs is still not easy because of deformations, highlights, shadows, and other factors. To work around these changes in an image, we use an edge-based cascade text detection method that integrates edge detection, adaptive searching, color analysis, and geometry alignments analysis

IV. RESULTS

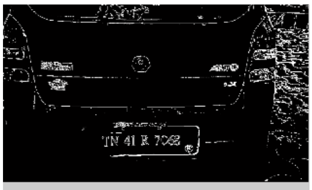
The performance of Text Detection and Character Extraction system is analyzed based on the recognition rate of Individual Images. Text detection and character extraction performance has many measurement standards. The most important and popular measures are:

$$\text{Accuracy} = \frac{\text{Number of correctly classified samples}}{\text{Number of samples}}$$

$$\text{Recall} = \frac{\text{Number of correctly detected text}}{\text{Number of character}}$$

(a)

(b)



(c)

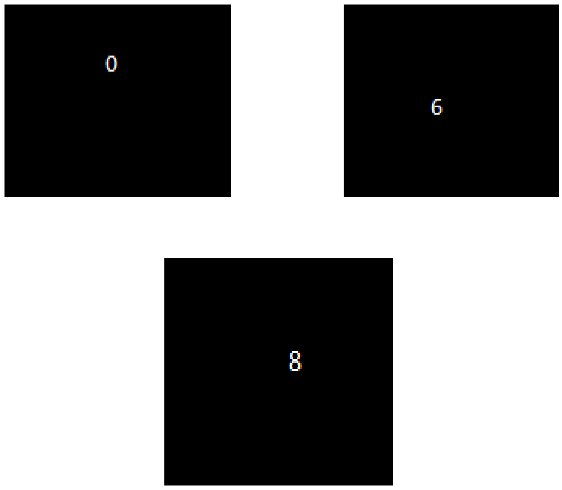


Fig. 4: Character Detection Result

V. CONCLUSION

Detection of text in Natural Scene Images is challenging for complex background. There are many methods available to perform the text detection and character extraction in natural scene images. The SVM classifier is used to detect the text. For training SVM eight features such as Average Entropy, Variance of Histogram, Relative smoothness, skewness, kurtosis, Local Energy of Gabor Filter and Horizontal and vertical projections. Experimental results demonstrated the effectiveness of the method by locating most text regions in test images. The selected images are pre-processed, features are extracted and classified using SVM classifier and characters are detect and extracted.

REFERENCES

[1] Arbelaez P. and Maire M. and Fowlkes C. and Malik J. (2011), 'Contour Detection and Hierarchical Image Segmentation', IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 898-916.

[2] Boykov Y. and Jolly M. P. (2001), 'Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images', In Proceedings of International Conference of Computer Vision (ICCV), volume 1.

[3] Carreira J. and Sminchisescu C. (2012), 'CPMC: Automatic Object Segmentation Using Constrained Parametric Min-Cuts', IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1312-1328.

[4] Ciresan D. and Meier U. and Masci J. and Schmidhuber J. (2011), 'A committee of neural networks for traffic sign classification', IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1918-1921.

[5] Clement Farabet M. and Camille Couprie Y. and Laurent Najman M. and Yann LeCun L. (2013), 'Learning Hierarchical Features for Scene Labeling', IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1915-1929.

[6] Farabet C. and Couprie C. and Najman L. and LeCun Y. (2011), 'Neuflow: A runtime reconfigurable dataflow processor for vision', In Proceedings of the Fifth IEEE Workshop on Embedded Computer Vision, pp. 109-116

[7] Farabet C. and Couprie C. and Najman L. and LeCun Y. (2012), 'Scene parsing with multiscale feature learning purity trees and optimal covers', In Proceedings of the International

Conference on Machine Learning (ICML), February, pp. 575-582.

[8] Farabet C. and Couprie C. and Najman L. and LeCun Y. (2012), 'Scene parsing with multiscale feature learning purity trees and optimal covers', In Proceedings of the International Conference on Machine Learning (ICML), pp.575-582.

[9] Felzenszwalb P. and Huttenlocher D. (2004), 'Efficient graph-based image segmentation', International Journal of Computer Vision.DOI. 10.1023.

[10] Fulkerson B. and Vedaldi A. and Soatto S. (2009), 'Class segmentation and object localization with superpixel neighborhoods', IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 670 – 677.

[11] Gould S. and Fulton R. and Koller D. (2009), 'Decomposing a scene into geometric and semantically consistent regions', IEEE International Conference on Computer Vision, pp. 1 – 8.

[12] Gould S. and Rodgers J. and Cohen D. and Elidan G. and Koller D. (2008), 'Multi-class segmentation with relative location prior', IEEE International Conference on Computer Vision, pp. 1480 – 1486.

[13] He X. and Zemel R. (2008), 'Learning hybrid models for image anno-tation with partially labeled data', Advances in Neural Information Processing Systems, pp. 625-632.

[14] Kavukcuoglu K. and Ranzato M. and Fergus R. and LeCun Y. (2008). 'Fast inference in sparse coding algorithms with applications to object recognition', Technical report Courant Institute of Mathematical Sciences, pp. 1010-3467.

[15] Kavukcuoglu K. and Ranzato M. and Fergus R. and LeCun Y. (2009), 'What is the best multi-stage architecture for object recognition', In Proceeding. International Conference on Computer Vision (ICCV'09). IEEE, 2009, pp. 2146 – 2153.

[16] Kavukcuoglu K. and Ranzato M. and Fergus R. and LeCun Y. (2010), 'Learning convolutional feature hierachies for visual recognition', In Advances in Neural Information Processing Systems , volume 23.

[17] Kumar M. and Koller D. (2010), 'Efficiently selecting regions for scene understanding', In Computer Vision and Pattern Recognition (CVPR), DOI. 10.1109.

[18] LeCun Y. and Boser B. and. Denker J. S and Henderson D. Howard R. E. Hubbard R. E. and Jackel L. D. (1990), 'Handwritten digit recognition with a back-propagation network'. In Computer Vision and Pattern Recognition (CVPR), pp. 396-404.

[19] LeCun Y. and Boser B. and Denker J. S and Henderson D. Howard R. E. Hubbard R. E. and Jackel L. D. (1998), 'Gradient-based learning applied to document recognition', In Computer Vision and Pattern Recognition (CVPR), DOI. 10.1109.

[20] LeCun Y. and Bottou L. and Orr G. and Muller K. (1998), 'Efficient backprop', In Computer Vision and Pattern Recognition (CVPR), pp. 9-48.

[21] Lempitsky V. and. Vedaldi A and Zisserman A. (2011), 'A pylon model for semantic segmentation', In Advances in Neural Information Processing Systems, pp. 171-179.

[22] Najman L. and Schmitt M. (1996), 'Geodesic saliency of watershed contours and hierarchical segmentation', IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1163 – 1173.

[23] Osadchy M. and LeCun Y. and Miller M. (2007), 'Synergistic face detection and pose estimation with energy-based models', Journal of Machine Learning Research, pp. 196-206.

[24] Pantofaru C. and Schmid C. and Hebert M. (2008), 'Object recognition by integrating multiple image segmentations', In ECCV 10th European Conference on Computer Vision, pp. 481-494.

[25] Ranzato M. and Huang F. and Boureau Y. and LeCun Y. (2007), 'Unsupervised learning of invariant feature hierarchies with applications to object recognition', In Proceeding of Computer Vision and Pattern Recognition, DOI. 10.1109.

[26] Russell B. and Torralba A. and Liu C. and Fergus R. and Freeman W. (2007), 'Object recognition by scene alignment', In Neural Advances in Neural Information, In Proceeding of Computer Vision and Pattern Recognition, DOI. 10.1167.

[27] Russell C. and Torr P. H. S. and Kohli P. (2009), 'Associative hierarchical CRFs for object class image segmentation', In Proceeding of Computer Vision and Pattern Recognition, DOI. 10.1109.

[28] Schulz H. and Behnke S. (2012), 'Learning object-class segmentation with convolutional neural networks', In 11th European Symposium on Artificial Neural Networks (ESANN), DOI. 10.1.1.307.2322

[29] Shotton J. and Winn J. M. and Rother C. and Criminisi A. (2006) , 'shape and context modeling for multi-class recognition and segmentation', In Proceeding of Computer Vision and Pattern Recognition, pp. 1-15.

[30] Socher R. and Lin C. C. and Ng A. Y. and Manning C. D. (2011), ' Parsing Natural Scenes and Natural Language with Recursive Neural Networks', In Proceedings of the 26th International Conference on Machine Learning (ICML), pp. 1675-1680.

[31] Turaga S. and Briggman K. and Helmstaedter M. and Denk W. and Seung H.(2009), 'Maximin affinity learning of image segmentation', In Proceedings of the 26th International Conference on Machine Learning (ICML), pp. 1865-1873.

[32] Vaillant R. and Monrocq C. and LeCun Y. (1994), 'Original approach for the localisation of objects in images', In Proceedings of the 26th International Conference on Machine Learning (ICML), DOI. 10.1049.