Text Extraction Using Natural Images

Preetam Shanbhag[1], Renuka Reddy[2], Abhishek Ballary[3], Arun Hiremath[4] Dr.J.D.Pujari[5], Prof. Varsha Jadhav[6]

[1234]*Bacheloreof Engineering, Department of Information Science and Engineering, SDM College of Engineering and Technology Dharwad, Karnataka*

[5,6]*Department of Information Science & Engineering, S.D.M. College of Engineering and Technology, Dharwad, Karnataka*

Abstract-Extraction and recognition of text present in images and videos have become a very popular research area in the last decade. Generally, text present in in image and video frames is of different size, orientation, style, etc. with complex backgrounds, noise, low resolution and contrast. These factors make the automatic text extraction and recognition in images and video frames a challenging task. The objective of this study is to propose a technique for text region localization and extraction in natural scene images .In this paper a hybrid methodology will be proposed which converts English language to other even in complex background. The proposed approach involves four steps. First, text regions in a natural image are extracted based on edge features using Tesser act package. In the second step, text regions are tested for text content or non-text using SVM classifier. In the third step, detection of multiple lines in localized text regions is being done and line segmentation is performed using horizontal profiles. In the fourth step, each character of the segmented line is extracted using vertical profiles. Looking at the growing popularity and the recent developments in the processing of text in images and video frames, this review imparts details of current trends and potential directions for further research activities to assist researchers.

Keywords: Text Detection, Text extraction and Text Recognition, Localization, Classification, Segmentation

I. INTRODUCTION

Now a days many hand held devices and scanners acquire images. Many digital cameras and phone cameras capture videos. Text is born as an explicit carrier of high level semantics. This unique property makes text different from other generic visual cues, such as contour, colour and texture. Text detection

and recognition has emerged as an important problem in the past few years. Advancements in the field of computer vision and machine learning as well as increase in the applications based on text detection and recognition has resulted in this trend. Various workshops and conferences like International Conference on Document Analysis and Recognition (ICDAR) are being organized on international level giving further rise to developments in field of text processing from imagery. Text in image will have important information. It would be very helpful if we recognize the text meaningfully and convert it to other languages which will make a design for automatic text detection and recognition system.

Now a days the digital cameras are very popular, portable and efficient and image capturing tools, which are commonly attached with various devices like tablets, mobiles phones and wrist watches pens and head mounted devices. Numerous useful technologies are embedded for fabricating these devices. Potential technologies may include recognition of texts in natural scene images, conversion of recognized text to other language, text-to-speech conversion and so on. Extraction and recognition of texts in natural scene images are suitable for foreigners with language impediment and persons with visual impairment.

Developing a robust system for extraction and recognition of texts from captured images is a great challenge due to several factors which include, spacing, distribution, colour, layout, light, and background complexity, variations of style, fonts and presence of multilingual scripts. Text appearing in images can provide very useful semantic evidence to describe the image content. In general, text existing in images can be categorized into caption text and scene text.

II. RELATED WORK

Extracting text from image is a difficult task. To perform this task various techniques have been implemented before. Cluster classification [1] is one of the techniques which have high accuracy in detecting text area and non-text area. There is new trend towards content based document image retrieval technique without going through OCR process [2]. There is another technique named as sliding window detection which has high accuracy of detecting text in natural scene. This paper uses different techniques for text extraction, web context search and web mining. Different techniques used are character descriptor and stroke configuration [4] for text detection and extraction, item ranking based on users interest [6] for web context search and semantic and synaptic web mining at low entropy for retrieving most relevant data from the web[7].

III. PROPOSED METHODOLOGY

The proposed methodology consists of five phases: pre-processing, potential text region detection, feature extraction, classification and character extraction. Text detection, tracking, character segmentation, recognition

and correction are important processing steps used in this approach. Neural networks based horizontal text detection is performed, followed by statistical intensity based shortest path algorithm for character segmentation. Convolutional neural classification is used for recognition along with language model. Ref. [3] also follows stepwise methods to detect texts of arbitrary orientation from natural images. Component extraction, component analysis, candidate linking and chain analysis are four stages through which proposed system proceeds. Connected components are extracted in component extraction stage using edge detection followed by SWT.

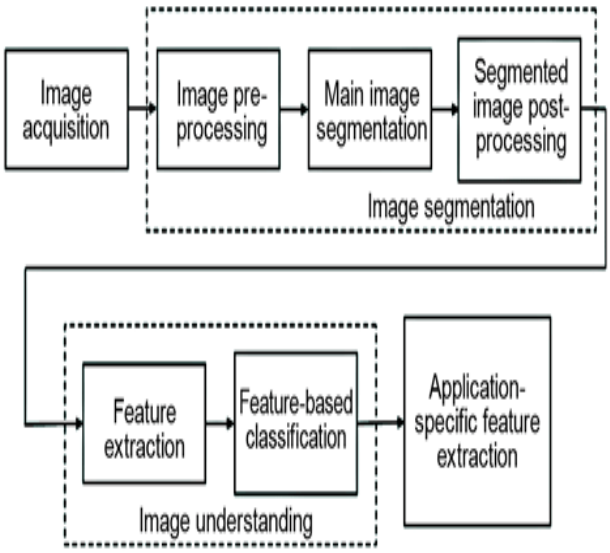


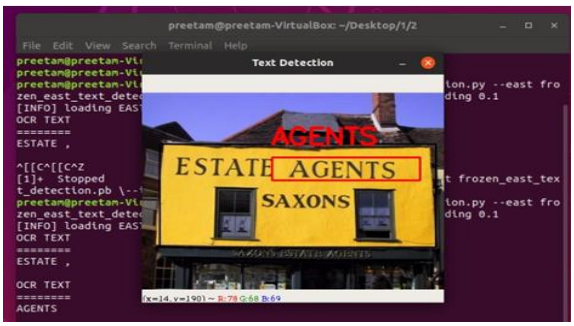
Fig. 1Block Diagram of Text Extraction

3.1 Pre-Processing

After the conversion of the input colour image into grayscale image, the median filter is applied to remove any noises present in the grayscale image and to obtain sharp edges in the image. The aim of pre-processing is an improvement of the image data that suppresses unwanted distortions or enhances some image features important for further processing .Grayscale images have range of shades of gray without apparent colour. These are used as less information needs to be provided for each pixel. In an 8 bit image, 256 shades are possible. The darkest possible shade black is represented as 00000000 and lightest possible shade white is represented as 11111111.

3.2 Potential Text Region Detection

To obtain the edge map containing strong edges contourlet transform is applied on pre-processed image. The morphological operations are applied on the transformed image, which yield segmented image. The connected components in the segmented image are filtered based on thresholding the geometric properties of the connected components. The filtered connected components are the potential text regions detected in the image, which will be classified as text or non-text regions using the trained SVM classifier in the classification phase.



Fig, 2 Text Region Detection

3.3 Feature Extraction

Gray level co-occurrence matrix (GLCM) has proved to be a popular statistical method of extracting textural feature from images. It is a 2D histogram of pair wise neighboring pixels with local textural uniformity in an image. It is denoted by C(i, j) and represents joint probability of occurrence of pixels with intensities i and j. It is computed using a displacement vector with radius δ and orientation θ. The values δ=1 and θ=0, 45, 90 and 135 degrees are chosen in the present work. From the co-occurrence matrix, Haralick defined fourteen textural features to extract the characteristics of texture statistics of images. The four important features, namely, Contrast, Homogeneity, Energy and Entropy, of potential text regions are extracted, which are then used to classify the regions into text and non-text classes.

3.4 Classification

The support vector machine (SVM) is trained using GLCM textural features and statistical features of text regions in training images, and is then used to classify the potential text regions extracted from natural scene test image as a text or non-text.

3.5 Character Extraction

Text localized image will be then further processed in order to extract the characters from the image. Horizontal projection is used to detect multiple lines in the localized text image and then each line is segmented. Further, each line segment is subjected to vertical projection to segment the characters.

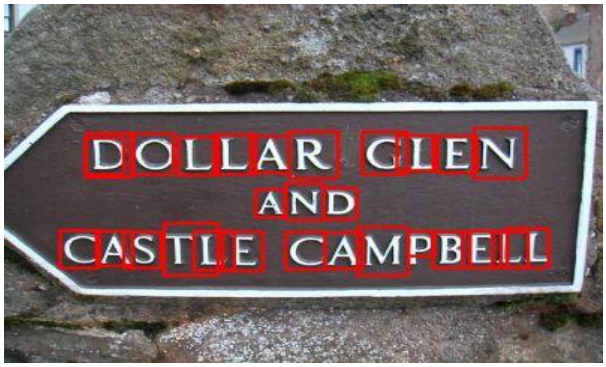


Fig. 3 Text Extraction and Segmentation

IV. CONCLUSION

The occurrences of text in a natural scene image pose a difficult challenging task in digital image processing due to complex background and illumination variations. The text may be multilingual with variations in font style, font size, scale, lighting and orientation. The proposed method is based on Contourlet transform which preserves the dominant edges of text in the image, and thus leads to more accurate localization of text regions.

ACKNOWLEDGMENT

This paper gives us an opportunity to convey our gratitude to all those who have helped us to reach a stage where we have immense confidence to launch our career in the competitive world of Information Science and Engineering . We would like to thank

Dr. Jagadeesh D. Pujariwho provided us the much needed guidance on every step.We will always be grateful to him who has been mentoring us throughout. Last but not the least we would like to thank our family and friends who encouraged us and advised us constantly.

FUTURE SCOPE

There are many future scopes related to this. There is need of developing algorithm that works efficiently even in very complex background and even if the outline of the text is not clear. Even further this can be extended to the algorithm that converts into many other languages. The proposed work can be extended to work on degraded text or broken characters.

REFERENCES

[1] Qixiang Ye, David Doermann, “Text Detection and Recognition in Imagery: A Survey” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014

[2] K. Elagouni, C. Garcia and P. Sbillot,” A Comprehensive Neural-Based Approach for Text Recognition in Videos using Natural Language Processing,” in Proc. ACM Conf. Multimedia Retrieval, 2011.

[3] C. Yao, X. Zhang, X. Bai, W. Liu, Y. Ma, and Z. Tu,” Detecting Texts of Arbitrary Orientations in Natural Images,” in Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, pp.1083-1090, 2012.

[4] K. Wang, S. Belongie,” Word Spotting in the Wild”, in Proc. European Conference on Computer Vision, pp. 591- 604, 2010.

[5] Shivananda V. Seeri, J. D. Pujari, P. S. Hiremath, Text Localization and Character Extraction in Natural Scene Images using Contourlet Transform and SVM Classifier, Published Online May 2016 in MECS

[6] Taufin M Jeeralbhavi, Dr. Jagadeesh D. Pujari, Shivananda V. Seeri, Text Extraction and Localization From Captured ImagesJune 2016.

[7] Vijayakumar, R. Nedunchezhianm, A Novel Method For Super Imposed Text Extraction In A Sports Video, International Journal Of Computer Applications, Volume 15– No. 1, 0975 – 8887, February 2011, pp. 1-6.

[8] Xiaopei Liu, Zhaoyang Lu, Jing Li, Wei Jiang, Detection and Segmentation Text from Natural Scene Images Based on Graph Model, WSEAS TRANSACTIONS on SIGNAL PROCESSING, E-ISSN: 2224-3488, Volume 10, 2014, pp. 124-135.

[9] Chitrakala Gopalan and Manjula, Text Region Segmentation from Heterogeneous Images, IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.10, October 2008, pp. 108-113.

[10] S. T. Deepa, S. P. Victor, Tamil Text Extraction, International Journal of Engineering Science and Technology (IJEST), ISSN: 0975-5462, Vol. 4 No.05 May 2012, pp. 2176-2179.