

A Critical review on video scene segmentation and Recognition

Dr. Ramesh R. Manza

Department of CS and IT

Dr. Babasaheb Ambedkar Marathwada University

Aurangabad (MS), India

manzaramesh@gmail.com

Bharatratna P. Gaikwad

Department of CS

Vivekanand College of Science

Aurangabad (MS), India

bharat.gaikwad08@gmail.com

Abstract:-The review of all the ten papers on the basis of video scene segmentation for text detection and recognition, among all these papers one is best paper “Robust Wavelet Transform Based Technique for Video Text Detection” they got the text detection and recognition rate is 96.7%.In this paper, they propose a new method based on wavelet transform, statistical features and central moments for both graphics and scene text detection in video images.

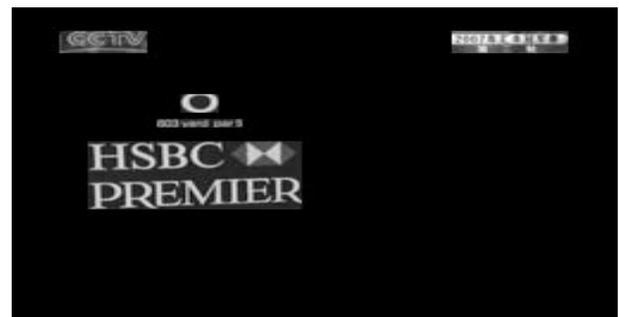
In contrast to scene text, artificial text is not only an important source of information but also a significant entity for indexing and retrieval purposes. Localization of text and simplification of the background in images is the main objective of automatic text detection approaches [3]. Hence, there is a need to develop a method which gives better detection rate with fewer false alarms irrespective of size, type, font and background of the texts in the video image[2]

INTRODUCTION

With hundreds of thousands of hours of archival videos, digital video now plays an important role in entertainment, education, and other multimedia applications. there is an urgent need for tools that allow efficient browsing and retrieving of video data . To achieve fast retrieving, and a simple and efficient text detection method With this intention, many methods have been proposed in the literature [2]. Many efforts have been made for text detection and recognition in videos and images. Indexing images or videos requires information about their content. This content is often strongly related to the textual information appearing in them, which can be divided into two groups:

- Text appearing accidentally in an image that usually does not represent anything important related to the content of the image. Such texts are referred to as scene text.
- Text produced separately from the image is in general a very good key to understand the image. In it is called artificial text.

Video scene Segmentation and Detection



Jie Xi and et.al. has work on Text detection, tracking and recognition to extract the text information in news and commercial videos. he has used Techniques morphological opening procedure on the smoothed edge map. Applying the horizontal and vertical projection. It Use SSD (sum of square difference) based matching method to track text blocks in consecutive frames, MSE(Mean Square Error) is taken to measure the dissimilarity. This OCR module takes the binarized text images as input and yields ASCII string of the recognized characters. They got the text detection rate is 94.7% and the recognition rate is 67.5%.The limitation of their work it is difficult to recognize small size characters that appear frequently in news videos. According to them this work may extends in future.

1. Apply some supervised classification methods in detection to reduce false alarms.
2. Integrate OCR module and character segmentation to make correct intra-word segmentation.
3. Use post processing to increase word recognition rate[1].

Palaiahnakote Shivakumara and et.al. has work on elimination of non significant edges from the segmented text portion of a video frame to detect accurate boundary of the text lines in video images. He has introduces candidate text block selection from a given image. Heuristic rules are formed based on combination of filters and edge analysis for identifying a candidate text block in the image, new edge based method for detecting both graphic text and scene text in video images efficiently. He has used Techniques Arithmetic mean Filter (AF) and Median Filter(MF) for remove noise in images. Canny edge and Sobel edge detector detects more edges when text is present than when there is no text present. They got percentage 93%. The Limitation of their work is the method fails to segment the complete text portion because of the complex background or isolated texts. According to them this work may extends in future to detect text lines with arbitrary direction and the robustness of the method handling complex backgrounds will be improved [2].

Qixiang Ye and et.al.has worked on the detection and verification of English Text and Chinese text from images and video fames. He has used Techniques for detection is based on Sobel edges feature and the verification uses the wavelet-based features and svm classifier. They got percentage detection rate English 93.9% and for Chinese 92.4% . According to them this work may extends in future to text detection in videos, the performance can be further improved by integrating the video temporal information[3].

Rainer Lienhart and et.al. has worked on the text localizing and segmenting text in complex images and videos, It is able to track each text line with sub-pixel accuracy over the entire occurrence in a video. He has used Techniques for text lines are identified by using a complex-valued multilayer feed-forward network trained to detect

text at a fixed scale and position, split-and-merge algorithm and high-frequency wavelet coefficients . They got percentage text recognition 69.9%. The Limitation of their work is the text heights are below 10 points that text can be not detected , some text are lost or missed in the segmentation stage. According to them this work may extends in future to globally adaptive threshold used in the binarization of the text bounding boxes to a locally adaptive threshold[4].

Huiping Li and et.al. has worked on the system can detect graphical text and scene text with different font sizes and can track text that undergoes complex motions. He has used Techniques as a hybrid wavelet/neural network based method is used to detect text regions and the tracking module uses sum of squared difference (SSD) -based image matching to find an initial position, followed by contour-based stabilization to refine the matched position. They got percentage 92.8%. The Limitation of their work is the tracker has difficulties when text moves too abruptly or keeps moving on a complex background. According to them this work may extends in future to making use of detected and tracked text to build a text-based video indexing and retrieval system[5].

Marios Anthimopoulos and et.al. has worked on the text detection from video and images , The detect and extract the text from T-shirts or road signs . He has used Techniques as Heuristic methods use empirical rules and thresholds in order to distinguish text from non text areas. A refinement stage uses an SVM classifier trained on features .A system for object detection based on Local Binary Patterns (LBP) . the edge Local Binary Pattern (eLBP) that is a modified LBP operator which actually describes the local edge patterns appeared in an image . They got percentage 76.42%. The Limitation of their work is the system detects barely readable scene text that was not considered as text in the ground-truthing procedure[6].

Palaiahnakote Shivakumara and et.al. has worked on detecting both graphic text and scene text with different fonts, size, contrast and backgrounds. He has used Techniques a Haar wavelet decomposition for detecting text in the video image, The edge and gradient based methods have been developed to reduce the number of computations in detecting text in the images. They got percentage 96.7%. According to them this work may extends the method can be extended to fix the bounding boxes for text lines with arbitrary direction[7].

Min Cai and et.al. has worked on the detecting the text from complex background and is robust for font size ,font color and languages . He has used Techniques to applies edge detection and uses a low threshold to filter out definitely non-text edges. Then a local threshold is selected to both keep low-contrast text and simplify complex

background of high-contrast text, edge-strength smoothing (ESS) operator and edge-clustering power (ECP) operator are used to highlight those areas with either high edge strength or high edge density. They got percentage 93.6%. The Limitation of their work is the missing some the text from video[8].

Y. Zhong and et.al. has worked on to automatically localize captions in JPEG compressed images and the I-frames (intra-frames) of MPEG compressed videos. Caption text regions are segmented from background images. He has used Techniques discrete cosine transform (DCT) to use the coefficients directly from compressed images and video as texture features to localize text regions. The texture features are extracted directly from domain using the quantized DCT coefficients and morphological operation are used to remove isolated noisy block and merge disconnected text blocks. The Limitation of their work is the font size of characters or the gap between the characters in the text is too big such that there is no strong texture present in a MPEG block and the contrast between the background and the text is too weak so that the text energy is not sufficiently high. According to them this work may extends in future to the compressed domains of the two color frames (Cr and Cb) to extract text with high color contrast. If segment text with a larger font size to compute texture features at a larger scale[9].

Yu Zhong and et.al. has worked on to locate text in compact disc (CD) and book cover images, automatically finding text in a complex color images and to extract characters from a more complex scene . He has used Techniques (OCR) optical character recognizer can be applied to only those regions of the image which contain text. Heuristic can be applied to a more complex image .To detect the any edge used Canny edge detector. The Limitation of their work is to locate text which is not well separated from the background means that the color of the text and background is similar [10].

CONCLUSION

All these papers by using the different techniques for scene segmentation for text detection method to detecting both graphic text and scene text with different fonts, size, contrast and backgrounds. Furthermore, the method can be extended to fix the bounding boxes for text lines with arbitrary direction and plan to implement a hybrid system

REFERENCE

- [1] Xi Jie, Xian-Sheng Hua, Xiang-Rong Chen, Liu Wenyin, HongJiang Zhang, "A Video Text Detection And Recognition System", IEEE International
- [2] P. Shivakumara, W. Huang and C. L. Tan," Efficient Video Text Detection Using Edge Features" The Eighth IAPR Workshop on Document Analysis Systems (DAS2008), Nara, Japan, September 2008, pp 307-314.
- [3] Q. Ye, W. Gao, W. Wang and W. Zeng. "A Robust Text Detection Algorithm in Images and Video Frames". IEEE ICICS-PCM, pp. 802-806, 2003.
- [4] R. Lienhart and A. Wemicke, "Localizing and segmenting text in images and videos" ,IEEE Trans. on Circuits and Systems for Video Technology, V01.12, No.4, April, 2002.
- [5] H. Li, D. Doermann, and O. Kia, "Automatic text detection and tracking digital video", IEEE Trans. Image Processing, vol. 9, pp. 147-156,Jan. 2000.
- [6] M. Anthimopoulos, B. Gatos and Ioannis Pratikakis ." A Hybrid System for Text Detection in Video Frames", The Eighth IAPR Workshop on Document Analysis Systems (DAS2008), Nara, Japan, September 2008, pp 286-293
- [7] Palaiahnakote Shivakumara, Trung Quy Phan and Chew Lim Tan "A Robust Wavelet Transform Based Technique for Video Text Detection" , 2009 10th International Conference on Document Analysis and Recognition.
- [8] M. Cai, J. Song and M. R. Lyu." A New Approach for Video Text Detection", In Proc. of International Conference On Image Processing, Rochester, New York, USA, pp. 117-120, 2002.
- [9] Y. Zhong, H. Zhang, and A.K. Jain. "Automatic caption localization in compressed video", IEEE transactions on pattern analysis and machine intelligence, vol. 22, no. 4, april 2000.
- [10] Y. Zhong, K. Karu, and A.K. Jain, "Locating Text in Complex Color Images ", Pattern Recognition, vol. 28, no. 10, pp. 1,523±1,536, Oct. 1995.