Text Extraction from Video and Converted into Video Frame

Megha Khokhra
Department of Electronics & Communication Engineering
KIRC, Gujarat, India

Abstract— The goal of this paper is to detect the text from video. With the increasing popularity of practical vision systems and smart phones, text detection in natural scenes becomes a critical yet challenging task. In a society driven by visual information and with the drastic expansion of low-priced cameras, text recognition is nowadays a fast changing field. The goal of this paper is develop a robust algorithm that extracts text from video with cluttered background. We propose a simple but efficient methodology for text detection in video frames. The method is based on the gradient information and edge map selection. In the proposed method we first find the gradient of the image and then enhance the gradient information. We later binarized the enhanced gradient image and select the edges by taking the intersection of the edge map with the binary information of the enhanced gradient image. We use the edge detector for generating the edge map. The selected edges are then morphologically dilated and opened using suitable structuring elements and used for text regions. We then perform the projection profile analysis to identify the boundary of the text region. At the end, we implement a false positive elimination methodology to improve the text detection results. Then we make the video from image frames which contain the detection of text.

Key words: Text Extraction, Text Detection, Edge detection, Projection profile analysis, Dilated image, Gradient Image, Video text Detection

I. INTRODUCTION

The term text extraction refers to extraction of characters in the form of strings from images of different variation and degradations. In video frames, temporal information redundancy, depending on frame rate, brings additional data enabling statistical methods to work more efficiently. Still images are snapshots of a scene providing only their own information for further analysis. The text contained in the frames may be a good key to describe the video contents as the existing texts are closely related to the current content of the frames. The text in these video frames may range from caption text appearing in news video, scene text, and text in advertisement and so on. Moreover, document image analysis is now not only confined within the scanned document, it can also be extended to any camera based image. So Text Information Extraction (TIE) from images (still or video frames) is one of the important aspects of document image analysis. TIE from video frames is one of the difficult tasks in image analysis as it needs to deal with images from various backgrounds, font colors, sizes etc. Yet it gains huge interest of today's researchers.

A. Basic steps for text understanding system

- Geometry: Size, Alignment, Inter-character distance
- Colour
- Motion
- Edge
- Compression

The video image spitted into number of frames, each frame maintains the text. Then the Image is converted into Gray Scale to avoid the text color variation. A single value is corresponding to gray value and detecting the edge. Detecting the edge process is the boundary between two regions with relatively distinct gray-level properties. One is the horizontal direction of the image. Another is the vertical direction of the image.

Over the Survey text information extractions (TIE) from video have been proposed for special applications including page segmentation; address block location, license plate location, and content-based image/video indexing. In spite of such extensive studies, it is still not easy to design a general-purpose TIE system. This is because there are so many possible sources of variation. When extracting text from a shaded or textured background, from low-contrast or complex images, or from images having variations in font size, style, color, orientation, and alignment. These variations make the problem of automatic TIE extremely difficult.

II. BACKGROUND

A. Text detection from image

In March 2006, a new company entitled 'Scene Reader'2 was launched, whose aim is to recognize NS text in different situations with dedicated software. The algorithms of this text may improve results. Comparatively, a visual search engine is born is 2005. Its name is Riya3 and enables the search of similar images by directly uploading images or the search of pictures containing the given text, among many other kinds of searches. Results are really impressive!

B. Text detection from video

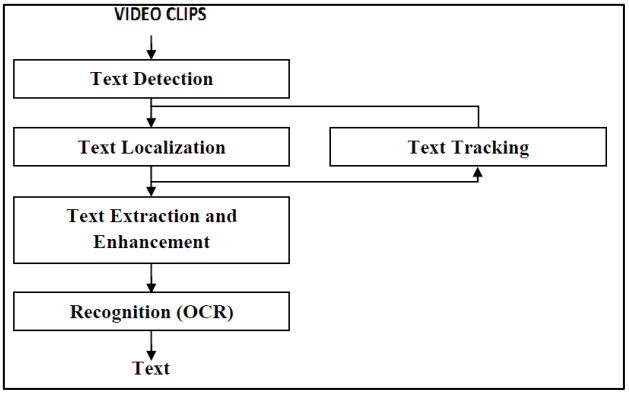

Fig. 1: Block Diagram for Extraction in Video

Though video images often suffer in degradation during transmission through various media, text portions can always be distinguished due to its discriminative pixel values with respect to the background. Text portions in an image always have distinct intensity values with respect to its background. The differences in the pixel values of an image are noted in the gradient of that image. Based on

this observation our proposed method performs using gradient information and text edge map selection. Generally the approach to detect text and caption in videos consists of the above steps:

III. THE PROPOSED METHOD

We describe the entire proposed methodology subsequently in the order where sub-section A describes gradient based procedure; edge based procedure is described in sub-section B; uniform color based procedure are discussed in C.[1]

A. An Overview

We here propose an improved method for creating gradient image because our intention was to create more and more edge pixels in the text portion of the image. In this method we first create horizontal and vertical gradient image of the original image and then merge them to get the resultant gradient image. To get the horizontal gradient image (HGI) of the original image (I) we consider the corresponding gray image (GI) of that image. The gradient value of a particular pixel (i, j) of HGI is calculated by taking the difference of the immediate lower pixel (i+1, j) with the current pixel (i, j) of the gray image (GI). In this stage we first apply canny edge detector to identify the edges in the image. We then select edges of the image by taking the intersection of the binarized information of the enhanced gradient image with the edge map. We next perform projection profile analysis (horizontal projection followed by vertical projection) to determine the boundary of the text region. We also applied the projection profile analysis for the second time within the rectangle (text boxes determined after the first projection profile) to make the boundary more accurate which further helps us to reduce number of inaccurate text boundary.

B. Text Localization

Text recognition is generally divided into four steps: detection, localization, extraction, and recognition. The detection step roughly classifies text regions and non-text regions. The localization step determines the accurate boundaries of text strings. The relative contrast between text and their background is an important feature for text detection. Thus, this paper proposes a morphology-based scheme for extracting the high contrast feature for locating all possible boundaries of text region. Strategies for picking the best arrangements exist.

Fig. 2: Localized Text Extraction in Video

C. Text Extraction

The extraction step filters out background pixels in the text strings, so that only the text pixels are left for the recognition. By using the Region based approach method, to extract all the text region areas can be well recovered. Since, the above two steps generate a binary text image, the recognition step can be done by commercial document OCR software. Therefore, only the two steps are discussed. However, the text localization and extraction in images present many more difficulties, due to complex background, unknown text color, degraded text quality caused by the lossy compression, and different language characteristics.

D. Text Region Localization

The extraction step filters out background pixels in the text strings, so that only the text pixels are left for the recognition. By using the Region based approach method, to extract all the text region areas can be well recovered. Since, the above two steps generate a binary text image, the recognition step can be done by commercial document OCR software. Therefore, only the two steps are discussed. However, the text localization and extraction in images present many more difficulties, due to complex background, unknown text color, degraded text quality caused by the lossy compression, and different language characteristics.

OCR is the acronym for Optical Character Recognition. This technology allows a machine to automatically recognize characters through an optical mechanism. Human beings recognize many objects in this manner our eyes are the "optical mechanism." But while the brain "sees" the input, the ability to comprehend these signals varies in each person according to many factors. By reviewing these variables, we can understand the challenges faced by the technologist developing an OCR system.

IV. EXPERIMENTS & RESULTS

Result of this Paper is As Shown below Figures.

Fig. 3: Text Detected in Image Frames

Fig. 4: Text Detected in Image Frames

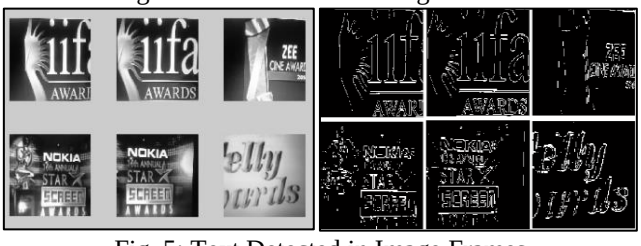

Fig. 5: Text Detected in Image Frames

Fig. 6: Text Detected in Image Frames

V. CONCLUSION

We have developed a video text and caption detection system. Viewing the corner points as the fundamental feature of character and text in visual media, the system detects video text with high precision and efficiency. We built up several discriminative features for text detection on the base of the corner points. These features can be used flexibly to adapt different applications. We also presented a novel approach to detect moving captions from video shots. Optical flow based motion feature is combined with the text features to detect the moving caption. Over 90% detection ratio is attained. The results are very encouraging. Most of the algorithms presented in this paper are easy to implement and can be straightforwardly applied to caption extraction in video programs with different languages. Our next focus will be on the word segmentation and text recognition based on the results of text detection.

REFERENCES

[1] Chong Yu, Yonghong Song , Quan Meng, Yuanlin Zhang, Yang Liu "Text Detection And Recognition In Natural Scene With Edge Analysis"[2015]
[2] Liang Wu, Palaiahnakote Shivakumara, Tong Lu, Member, IEEE, and Chew Lim Tan "A New Technique For Multi-Oriented Scene Text Line Detection And Tracking In Video"[2015]
[3] A. Dutta, U. Pal, Shivakumara, Ganguli, Bandyopadhya and L. Tan, "Gradient based Approach for Text Detection in Video Frames", CVPR Unit, Indian Statistical Institute, Kolkata, India,[2009]
[4] B.H.Shekar and Smitha M.L. "Morphological Gradient based Approach for text localization in video/scene images"[2014]
[5] Vladimir Y. Mariano Rangachar Kasturi "Locating Uniform- Colored Text In Video Frames"[2000]
[6] Xiaodong huang,"A novel of approach to detecting scene text in video"[2011]
[7] Ping Hu, Weiquiang Wang, Ke Lu,"Video text detection with text edges and convolutional neural network".[2013]
[8] Xiangrong Chen and Hongjiang Zhang "Text Area Detection from Video Frames", Microsoft Research China, [2000]
[9] Ze-Yu Zuo, Shu tian, Wei- Yi Pei, Xu-Cheng Yin , " Multi –Strategy Tracking Based Text Detection in Scene Videos."[2015]
[10] Christian Wolf AND Jean-Michel Jolion, "Model based text detection in images and videos," in Proc. IEEE Conf. Comput. Vis. Pattern Rec 2004.
[11]] Xu-Cheng Yin, Senior Member,IEEE,Ze-Yu Zuo,Shu Tian, and Cheng-Lin Liu,Fellow,IEEE." Text Detection,Tracking and Recognition in Video : A comprehensive Survey."
[12] Guozhu Liang, Palainhnaktoke Shivakumara, Tong lu, Member IEEE and Chew lim tan, Senior Member,IEEE."Multi-spectral Fusion based Approach for Arbitrarily-Oriented Scene Text Detection in Video Images".

