A Transition Map Method to Find Overlay Text

H. MOHAMED SHAJAHAN
College of Engineering and Information Technology,
University of Business and Technology, Jeddah, Saudi Arabia

Dr. MUNIR M. ALHADDAD
College of Engineering and Information Technology,
University of Business and Technology, Jeddah, Saudi Arabia

A transition method find overlay text brings important semantic clues in video content analysis such as video information retrieval and summarization, since the content of the scene or the editor's intention can be well represented by using inserted text. The main aim of the research is to propose a novel framework to detect the Overlay text information in video frames. This method produces better than the previous methods. Resultant accuracy is highly improved.

Keywords: Knowledge, Intelligence, Information, Image and Communication.

Introduction

Broadly, image processing may be subdivided into the following categories: enhancement, restoration, coding, and understanding. The goal in the first three categories is to improve the pictorial information either in quality (for purposes of human interpretation) or in transmission efficiency. In the last category, the objective is to obtain a symbolic description of the scene, leading to autonomous machine reasoning and perception with the development of video editing technology, there are growing uses of overlay text inserted into video contents to provide viewers with better visual

understanding. Most broadcasting videos tend to increase the use of overlay text to convey more direct summary of semantics and deliver better viewing experience. For example, headlines summarize the reports in news videos and subtitles in the documentary drama help viewers understand the content. Sports videos also contain text describing the scores and team or player names. In general, text displayed in the videos can be classified into scene text and overlay text. Scene text occurs naturally in the background as a part of the scene, such as the advertising boards, banners, and so on. In contrast to that, overlay text is superimposed on the video scene and used to help viewers understanding. As a preliminary preparation, data will be collected as part of this research. The main aim of the research is to propose a novel framework to find the Overlay text information in video frames.

Implementation

Lot of methods are already implemented for overlay text detection. Color based methods are not working properly because of un-uniform color distribution. Most of existing video text detection methods has been proposed on the basis of color, edge, and texture-based feature. The method proposed by Agnihotri [13], concentrates on the red color component, instead of all the 3 color components. Some methods used the high contrast video frames to extract the texts. Kim et al. [14] uses RGB color space and clustering concept. But no methods are fully efficient for clustering. So text detection is not so better in this case. The edge-based methods are not made success because of complex background. Modified edge map is introduced by Lyu et al. [15]. This is providing some improvement in overlay text detection.

Methodology

Transition Map Generation

As a rule of thumb, if the background of overlay text is dark, then the overlay text tends to be bright. On the contrary, the overlay text tends to be dark if the background of overlay text is bright. Therefore, there exists transient colors between overlay text and its adjacent background due to color bleeding, the intensities at the boundary of overlay text are observed to have the logarithmical change. The intensities of three consecutive pixels are decreasing logarithmically at the boundary of bright overlay text due to color bleeding by the lossy video compression. It is also observed that the intensities of three consecutive pixels increases exponentially at the boundary of dark overlay text. To find the intensity change in the transition region three steps are adopted. They are as follows:

- Saturation calculation
- Modified Saturation calculation
- Transition map generation

If a pixel satisfies the logarithmical change constraint, three consecutive pixels centered by the current pixel are detected as the transition pixels and the transition map is generated.

Video Frames

The difference of the previous frame's Transition map and current frame's transition map, decides whether to process the current frame or neglect the current frame. A threshold is used here for decision making.

Candidate Map Region Detection

The transition map can be utilized as a useful indicator for the overlay text region. To

generate the connected components, first generate a linked map [5]. If a gap of consecutive pixels between two nonzero points in the same row is shorter than 7% of the image width, they are filled with 1s. Next the Hole filling algorithm is used to fill the small gaps and to maintain the connectivity. Then each connected component is reshaped to have smooth boundaries. Since it is reasonable to assume that the overlay text regions are generally in rectangular shapes, a rectangular bounding box is generated by linking four points, which correspond to (min_x, min_y), (max_x, min_y),(min_x, max_y), (max_x, max_y)taken from the link map and candidate regions.

Overlay Text Region Determination

In this subsection, we introduce a texture-based approach for overlay text region

$$LBP_P = \sum_{i=0}^{P-1} s(g_i - g_c) 2^i$$

determination. Based on the observation that intensity variation around the transition pixel is big due to complex structure of the overlay text, we employ the local binary pattern (LBP) introduced in [6] to describe the texture around the transition pixel. LBP is a very efficient and simple tool to represent the consistency of texture using only the intensity pattern. LBP forms the binary pattern using current pixel and it's all square neighbor pixels and can be converted into a decimal numbers as follows:

Where

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

P denote the user's chosen number of square neighbor pixels of a specific pixel.

g_i ->neighbor pixels intensity.

g_c ->intensity of current pixel.

Overlay Text Marking

The rectangle bounding box is projected around the extracted overlay text region. Using the four corner points of candidate region we can mark the Text data.

Results and Discussions

Most of existing video text detection methods has been proposed on the basis of color, edge, and texture-based feature. Color-based approaches assume that the video text is composed of a uniform color. However, it is rarely true that the overlay text consists of a uniform color due to degradation resulting from compression coding and low contrast between text and background. Edge-based approaches are also considered useful for overlay text detection since text regions contain rich edge information. The commonly adopted method is to apply an edge detector to the video frame and then identify regions with high edge density and

strength. This method performs well if there is no complex background and it becomes less reliable as the scene contains more edges in the background. Texture-based approaches, such as the salient point detection and the wavelet transform, have also been used to detect the text regions. However, since it is almost impossible to detect text in a real video by using only one characteristic of text, some methods take advantage of combined features to detect video text.

Sample Output

Fig 1: Original Frame (Sample Frame)

Fig 2: Transition map generation

Conclusion

The various processes on overlay text detection from complex videos are proposed in this paper. The main concept of the work is based on the observation that there exist transient colors between inserted text and its adjacent background. We compute the density of transition pixels and the consistency of texture around the transition pixels to distinguish the overlay text regions

from other candidate regions. The local binary pattern is used for the intensity variation around the transition pixel in the proposed method. The boundaries of the detected overlay text regions are localized accurately using the projection of overlay text pixels in the transition map. This research is well adopted in video data processing.

Reference

Wonjun Kim and Changick Kim, "A New Approach for Overlay Text Detection and Extraction From Complex Video Scene ", IEEE transactions on image processing, vol. 18, no. 2, February 2009.

L.Agnihotri and N. Dimitrova, "Text detection for video analysis," in Proc. IEEE Int.Workshop on Content-Based Access of Image and Video Libraries, , pp. 109–113, Jun. 1999.

K. C. K. Kim et al., "Scene text extraction in natural scene images using hierarchical feature combining and verification," in Proc. Int.Conf. Pattern Recognition, vol. 2, pp. 679–682, Aug. 2004.

M. R. Lyu, J. Song, and M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," IEEE Trans. Circuit and Systems for Video Technology, vol. 15, no. 2, pp. 243–255, Feb. 2005.

M. Bertini, C. Colombo, and A. D. Bimbo, "Automatic caption localization in videos using salient points," in Proc. Int. Conf. Multimedia and Expo, Aug. 2001, pp. 68–71. (ICME), pp. 1721–1724, Jul. 2006