IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)
e-ISSN: 2278-1676,p-ISSN: 2320-3331, Volume 8, Issue 6 (Nov. - Dec. 2013), PP 61-64
www.iosrjournals.org

Detection and Localization of Text Information in Video Frames

Ankit Patel, Jignesh Prajapati, Punit Raninga
E & C Department, Parul Institute of Engineering And Technology PIET-Limda ,Vadodara,India

Abstract: *Video text detection plays an important role in semantic-based video analysis. In the field of information retrieval, text detection and extraction from video has become an emerging area to solve the fundamental problem of content-based image retrieval (CBIR) to fill in the semantic gap between low level and high level features. Text detection and extraction enables the understanding of video contents with the help of text recognition using optical character recognition techniques, to give a partial solution to bridge the semantic gap between the low level and high level features. In this paper we have proposed a method which detect & localize horizontal text with uniform background.*

I. Introduction

A video frame contains variety types of information. Among them, text is one of the most informative types. These text blocks provide valuable information such as scene locations, speaker names, program introductions, sports scores, special announcements, dates and time.

IN video databases, each video is manually tagged with a few keywords to allow for searching and retrieval. However, this process is laborious and inconsistent, i.e., two users may choose different keywords for the same video. An alternative approach is to generate the keywords from the text that appears in the frames.
The text in video frames can be classified broadly into two large categories - caption, artificial, or overlay text and scene text.

Caption text comprises of text strings that are generated by graphic titling machines and composited of text captions, but could also be graphical elements with text contained within them or graphic effects using text. Such text is placed intentionally by the program editor to provide information on the story being depicted, identifying the people, and/or location, etc. Examples of such text are found as credit titles, ticker tape news, information in commercials, etc.

Scene text, on the other hand, occurs naturally in the scene being imaged. The image of this text may be distorted by perspective projection, be subject the illumination conditions of the scene, be susceptible to occlusion by other objects, suffer from motion blurring etc. Scene text can also be on planar surfaces such as road signs, billboards, etc., as well as non-planar surfaces such as the text on soft drink cans.Text detection methods can be classified into three approaches: Region based and texture-based .Region-based methods use pixel’s information of a text region or their differences with the background. First, some features are extracted from local regions to determine if they contain texts. Then, clustering approaches are employed to localise text regions more accurately. Performances of the region-based methods depend on text orientation and cluster’s numbers. These methods can be divided into two sub-groups: edge-based and connected component (CC)-based approaches. Edge-based methods, which are robust to text size, focus on the high contrast between text and background. CC-based methods regard texts as a set of separate CCs with distinct intensity, colour distributions and enclosed contours. Without prior information about text size and position, accurate text segmentation is difficult using CC-based methods. The CC-based algorithms are relatively simple to implement, but they are not accurate for text localisation within complex background.

On the other hand, texture-based methods assume that text in images have distinct textural properties which can be used to discriminate them from non-text regions. Generally speaking, in complex background, texture-based algorithms are more successful than the CC-based algorithms.

Image characteristics of text

In this section we list the image characteristics of text that may be used to segment it from the rest of the imaged scene. Many of these follow from research in document image analysis, while others are from extensive observations of video data.

Size

Text appears in a variety of sizes in video data. Since text is intended to be readable at a range of viewing distances in a limited time, there is usually a minimum size to text characters. However, the upper bound on character sizes is much looser. Text can often be as large as half the frame height or more.

Colour and Intensity

Colour is a strong feature for use in visual information indexing. Text characters tend to have a perceptually uniform colour and intensity over the character stroke. While the character stroke appears to have the same colour, in reality it is usually composed a many different colours. In cases where the colour does vary across the caption, it varies in a gradual way so that adjacent characters or character segments have very similar colours.

Edges

Most text has edges between its boundaries and the background. Thus some detection of edges is necessary. Text extraction can be made even more effective by testing whether the presence of an edge is also accompanied by an edge of a colour connected component. Most scripts have strong edges in a particular direction, e.g. Latin scripts have strong vertical edges.

Contrast

The colour for artificial text are chosen so as to have contrast against the back ground. This is also true for a lot of scene text (signs, billboards, etc.). Thus colour connected components belonging to text can be segmented against the frame. Sometimes, contrast between text and the background may not be high. This is because the background over which text is composited is varying both spatially and temporally and it often happens that the colour of caption characters is similar to that of the surrounding frame region. However, even if this occurs it is usually true only for a portion of the caption or for a short duration, since otherwise, the editor would have chosen a different colour for the text.

Geometry

Characters belonging to an artificial text string are usually horizontally aligned. However, if the text has been produced by computer special effects, they may actually appear like non-planar scene text . The projection of scene text on the image plane will usually not be horizontal, so this particular geometric constraint cannot be applied to **scene** text. Also the aspect ratio of the individual characters as well as that of the entire caption lies in a certain range.

Non-uniformity of background

While the rest of the video frame may contain changes due to camera or objection motion, artificial text tends to remain constant. If it moves, it is generally in a Predictable fashion. Scene text also tends to move in a predictable fashion. Its movement can be coordinated with camera motion characteristics such as pans, zooms, etc., or associated scene object movement.

Inter-character gap

Most caption strings consist of a certain minimum and a maximum number of characters, i.e. at least one word but usually not more than a few. This is not true of single or double character logos or symbols. Since captions do not contain more than a few words, the characters are usually well separated. Thus, touching characters, a thorny issue in document character recognition, is not as important in this application. However, for noisy video or unusually large text, this may not be true. Also, usually the inter-character separation over multiple frames is nonzero and character positions relative to each other remain fixed. This characteristic is better qualified in the next item.

Rigidity

Text captions and stationary scene text tend to retain their shape, orientation, size and font over multiple frames. Infrequently, this may not be true; text may have special graphic effects added such as zooming up or down. But in general, this is good feature to use.

Motion

The position of artificial text can change, as in scrolling text for example, but does so in a very uniform way, either vertically or horizontally. Successive frames exhibit very small jitter in caption position. An important point is that the text available in video may be in many different languages and scripts. For different scripts, the nature of the imaged text may differ;

e.g., Arabic script may have inter-character characteristics very different from English. Therefore a robust text detection algorithm should be able to work even in the face of one or more of these features failing.

II. Proposed Method

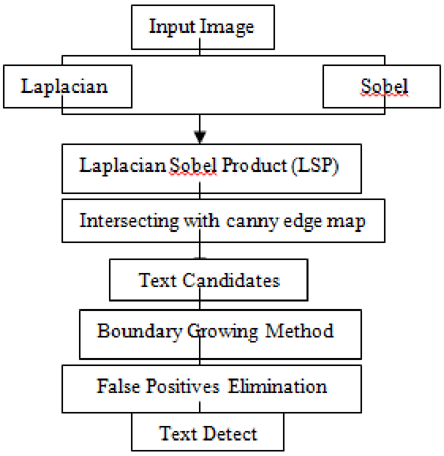

Input Image
Laplacian
Sobel
Laplacian Sobel Product (LSP)
Intersecting with canny edge map
Text Candidates
Boundary Growing Method
False Positives Elimination
Text Detect

Fig 1. Flow Chart Of Proposed Method

A. Text Enhancement

Text regions typically have a large number of discontinuities from text to background regions and background to text. Thus, this property gives a strong indication of the presence of text. In this section, we try to exploit this property by using Laplacian and Sobel masks and combining the results as follows. We use a 3×3 mask to obtain fine enhanced details of text pixels. The Laplacian operation is a second-order derivative and is used to detect discontinuities in four directions: horizontal, vertical, up-left, and up-right. As a result, it enhances details of both low and high contrast pixels in the image. However, this operation produces more noisy pixels than the Sobel operation. This noise may be the cause for poor performance of the text detection. On the other hand, it is known that Sobel mask operation is a first-order derivative and hence it produces fine details at discontinuities in horizontal and vertical directions. This results in an enhancement at high contrast text pixels, but no enhancement at low contrast text pixels. LSP is used to preserve details at both high and low contrast text pixels while reducing noise in relatively flat areas.

B. Intersecting with canny edge map

The Canny edge detection algorithm is known to many as the optimal edge detector.

There are many advantages of canny edge detection compared to others such as using probability for finding error rate, Localisation and response ,improving signal-to noise ratio.

C. Boundary Growing Method (BGM)

Boundary growing method is based on the nearest-neighbor concept. The basis for this method is that text lines in the image always appear with characters and words in regular spacing in one direction.

D. False Positive Elimination

Since the canny edge map is used for obtaining text candidates, the proposed method may produce a larger number of false positives. It is noted that false positive elimination is challenging and difficult . In this paper, we use geometrical properties of text blocks for the purpose of false positiveelimination as these properties are quite common in the literature to use for false positive elimination. Let W, H, A, AR, and EA be the width, height, area, aspect ratio, and edge of text block B, respectively

$$AR=W/H$$

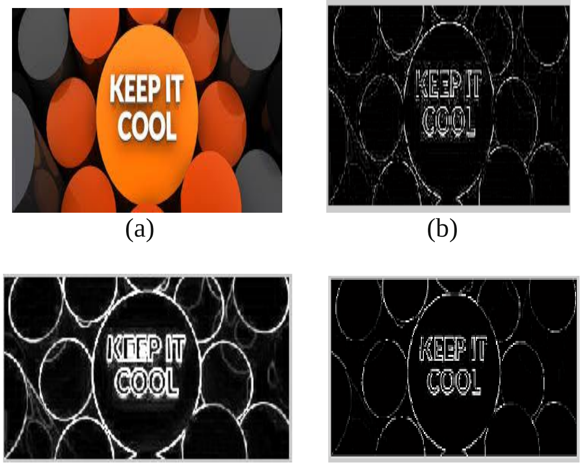
Where $A=W * H$

$$EA=\sum BC(i,j)$$

If $AR < T1$ and $EA/A < T2$, the text block is considered as a false positive; otherwise, it is accepted as a text block. The first rule checks whether the aspect ratio is below a certain threshold. The second rule assumes that a text block has a high edge density due to the transitions between text and background. Here, T1 and T2 are determined based on the experimental study given in our earlier work , and the same dataset is used for both T1 and T2.

III. Experimental Results

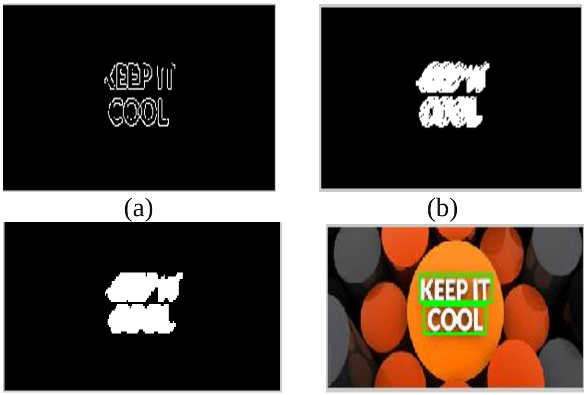
LSP Process



(a) (b)

(c) (d)

Fig.2 a)Input image b)Laplacian Image c) Sobel Image d) LSP Image



(a) (b)

(c) (d)

Fig.3 a)Canny Edge Map b)Line Process c) Fill holes d) Text Detection Image

IV. Conclusion And Future Work

In this paper, we proposed a new video scene text detection method that made use of a new enhancement method using Laplacian and Sobel operation of input images to enhance low contrast text pixels. We proposed a boundary growing method. based on the nearest neighbour concept. Experimental result shows that proposed method works on horizontal text with uniform background. There are few problems in handling false positives. We planned to extend this method to detection of multi-oriented and curve-shaped text lines with good recall, precision, F-measure, and low computational times

References

[1] Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu" A Hybrid Approach to Detect and Localize Texts in Natural Scene Images" IEEE transactions on image processing, vol. 20, no. 3, march 2011

[2] Palaiahnakote Shivakumara, Trung Quy Phan and Chew Lim Tan" A Robust Wavelet Transform Based Technique for Video Text Detection" IEEE 2009

[3] NabinSharma,Palaiahnakote,Shivakumara,Umapada Pal" A New Method for Arbitrarily-Oriented Text Detection in Video" 2012 10th IAPR International Workshop on Document Analysis Systems

[4] Palaiahnakote Shivakumara, Trung Quy Phan, and Chew Lim Tan," A Laplacian Approach to Multi-Oriented Text Detection in Video" IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 2, february 2011

[5] Keechul Jung, Kwang In Kim, Anil K. Jain,"Text information extraction in images andvid eo: a survey" Pattern Recognition, vol. 37, no. 5, pp. 977–997, 2004.

[6] J. Zang and R. Kasturi, "Extraction of text objects in video documents:Recent progress," in Proc. DAS, 2008, pp. 5–17.

[7] X. R. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'04), Washington, DC, 2004, pp. 366–373.

[8] Tianyi Gui, Jun Sun, Satoshi Naoi" A Fast Caption Detection Method for Low Quality Video Images"in 2012 10th IAPR International Workshop on Document Analysis Systems