Text Detection and Extraction from Video

Pallavi P. Shinde(ME II Year)
Department of Computer Engg.
VPCOE
Baramati--413133

Prof. P. M. Patil(Assistant Prof.)
Department of Computer Engg.(PG)
VPCOE
Baramati--413133

***Abstract*—** Detecting and extracting text from videos is important task. Also it can be used for video indexing, video retrieval, automatic annotation, and content analysis. Proposed method is a corner based approach to detect text from videos. This approach is introduced because numbers of corner points are present orderly in characters of text. In proposed method text regions formed by the corner points are detected by combining some text features. Finally using OCR, text is extracted from detected text region and also words of extracted text are segmented. Extracted text and words can be used for video retrieval, video matching, annotation, indexing.

Keywords- *Harris corner detector, Text detection, Text Extraction, Word Segmentation.*

*****

I. INTRODUCTION

Text detection from images is an important component for a wide range of applications because text in images and video frames carries important information for automatic annotation, and indexing.

Images and videos on webs and in databases are increasing rapidly. So the important task to develop effective methods to manage and retrieve these multimedia resources by their content. Text can be used for this task because text usually carries high-level semantic information. For example, text in web images can reflect the content of the web pages. Text on journal and book covers can be helpful to retrieve these digital resources. In news videos caption text usually gives the breaking news. Sub-title in sport videos often annotates information of score, athlete and highlight. Therefore, there is a need of system which detects and recognizes text in images and videos. The role of text detection is to find the image regions containing only text.

Detecting text from videos is a challenging task because there exist difficulties in text detection such as:

1. Text may be embedded in complex background;
2. It is difficult to find effective features to discriminate text with other text-like things, such as leaves, window curtains or other general textures;
3. Text pattern varies with different font-size, font-color and languages;
4. Text quality decreases due to noise and image encoding/decoding procedure.

To overcome this challenge proposed system uses corner points to detect text from video.

Proposed system describes the text regions with the discriminative features, from which the nontext regions formed by the corner points appeared in the background can be filtered out efficiently. This paper presents a method to detect the still text and then extract the text. Also this method segments the words of extracted text. Then this extracted text and words can be used for applications like indexing of video, video retrieval, annotation.

II. LITERATURE SURVEY

Here we present some existing techniques which are used for text detection.

1. Texture based method: This method treats the text region as a special type of texture with distinct texture properties because texture may be fine, smooth, regular, irregular, or linear. Texture properties are used to distinguish text from the background. Texture gives us information about the spatial arrangement of the colors or intensities in an image. The techniques that are extract texture features include support vector machine (SVM)[1], wavelet[2], FFT[3].

Advantage: Texture based approaches works well in extracting texts from complex and textured background, insensitive to image resolution.

Disadvantage: Texture based approach is time consuming so it is not suitable for large database, fail to detect characters with low contrast.

2. Connected component based method: This method [4] segments an image into a set of connected components and successively merges the small components into larger ones. By analyzing geometrical characteristics the final connected components are classified as text or background.

Advantages: Low computational cost Robust to font size, High accuracy.

Disadvantages: This approach have difficulties when the text is noisy, multi colored and textured, some characters with small size are missed.

3. Edged based method: Edge based methods [5]-[6] utilize the structural and geometry properties of character and text. An

edge extractor is applied for the edge detection because characters are composed of line segments and text regions contain rich edge information. Then smoothing operation or a morphological operator is used in the merging stage to reduce the noise or numerous imperfections in image.

Advantage: Edge based method is effective in detecting text regions when the other parts in the image do not have too many strong edges, not sensitive to image color/intensity.

Disadvantage: This method is not applicable when background has similar strong edge distributions as the text regions, non horizontally aligned texts cannot be localized.

III. PROPOSED SYSTEM

The system architecture of proposed text detection system is illustrated in Fig. 1. Firstly video is divided into frames. Then Corner point detection algorithm is applied on each frame to detect corner points. Text features of region containing corner points are computed to detect and locate the text regions in static images. Then by using OCR text from text region is extracted. Finally word segmentation is done to separate the words of extracted text.

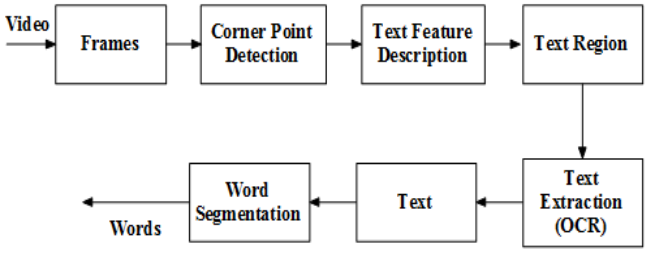

Fig. 1 Architecture of Proposed System

A. Corner point detection

A corner can be defined as the intersection of two edges or a point. It is junctions of contours. Generally corner points are more stable features over changes of viewpoint. Corner detection is widely used in computer vision application such as motion detection, image matching, tracking.

In this project, Harris corner detector [7] is used to extract the corner points. The Harris corner detector is a popular interest point detector. Because there is no effect of rotation, scale, illumination variation, and image noise on the performance of Harris corner detector. It is based upon the local auto-correlation function of a signal, where the local auto-correlation function measures the local changes of the signal with patches shifted by a small amount in different directions. Suppose we have a Gray scale 2-D image. Consider taking an image patch over the window W(u,υ) and shifting it by (∂_x, ∂_y). The change produced by the shift is given by

$$E(\partial_x, \partial_y) = \sum_W [I(u+\delta x, v+\delta y) - I(u,v)]^2 \quad (1)$$

First order Taylor's expansion is given by,

$$f(u+\partial_x, v+\partial_y) \approx f(u,v) + \partial_x f_x(u,v) + \partial_y f_y(u,v)$$

The shifted image is approximated by a first order Taylor expansion,

$$I(u+\delta x, v+\delta y) \approx I(u,v) + \partial_x I_x(u,v) + \partial_y I_y(u,v) \quad (2)$$

where Ix and Iy denote the partial derivatives in x and y directions, respectively. Substituting equation (2) into (1)

$$\approx \sum_w [I(u,v) + \partial_x I_x(u,v) + \partial_y I_y(u,v) - I(u,v)]^2$$

$$= \sum_w \partial_x^2 (I_x(u,v))^2 + 2\partial_x \partial_y I_x(u,v) I_y(u,v) + \partial_y^2 (I_y(u,v))^2$$

By rewriting as matrix equation we get,

$$= \sum_w \begin{bmatrix} \partial_x & \partial_y \end{bmatrix} \begin{bmatrix} (I_x(u,v))^2 & I_x(u,v) I_y(u,v) \\ I_x(u,v) I_y(u,v) & (I_y(u,v))^2 \end{bmatrix} \begin{bmatrix} \partial_x \\ \partial_y \end{bmatrix}$$

$$= [\partial_x \quad \partial_y] \left(\begin{bmatrix} \sum_w (I_x(u,v))^2 & \sum_w I_x(u,v) I_y(u,v) \\ \sum_w I_x(u,v) I_y(u,v) & \sum_w (I_y(u,v))^2 \end{bmatrix} \right) \begin{bmatrix} \partial_x \\ \partial_y \end{bmatrix}$$

$$E(\delta_x, \delta_y) = \begin{bmatrix} \partial_x & \partial_y \end{bmatrix} C \begin{bmatrix} \partial_x & \partial_y \end{bmatrix}^T \quad (3)$$

Where the Hessian matrix C captures the intensity structure of the local neighbourhood.

$$C = \begin{pmatrix} \sum_W (Ix(u,v))^2 & \sum_W Ix(u,v) Iy(u,v) \\ \sum_W Ix(u,v) Iy(u,v) & \sum_W (Iy(u,v))^2 \end{pmatrix}$$

Harris and Stephens design the response function

$$f_R = \lambda_1 \lambda_2 - \kappa(\lambda_1 + \lambda_2)^2 = \det C - \kappa trace^2(C) \quad (4)$$

Where λ_1, λ_2 be the eigenvalues of matrix C and κ is a tunable sensitivity parameter. f_R is positive in the corner region, negative in the edge region, and small in the flat region.

B. Text Feature Extraction

 ISSN: 2321-8169
1746 – 1749

Once corner points are extracted, the shape properties of the regions containing corner points are computed. These properties are used to make decision to accept the regions as text or nontext. Generally characters appear with other characters, numbers of corner points are present in text region. Also corner points are usually placed in a horizontal line. Therefore, the text can be effectively detected by finding the shape properties of the formed regions. The numbers of corners may also appear in the nontext regions but they usually are unordered and can be filtered out using some features. The region properties are used as the features to describe text regions are: area, saturation, aspect ratio. For simplicity these features are represented as Ra, Rs, and Ras, respectively.

These features [8] are combined to detect text region.

1. Area: The area of a region is defined as the number of foreground pixels in the region enclosed by a rectangle bounding box. Area is the basic feature for text detection. The small regions generated by the disorderly corner points can be easily filtered out according to the area measurement. In nontext region corner points are distributed randomly and irregularly in image frames. So for nontext region area is very large as compared to area of text region.
2. Saturation: the saturation specifies the proportion of the foreground pixels in the bounding box that also belong to the region, which can be calculated by

$$Rs=Ra/RB$$

where RB represents the whole region enclosed by the bounding box. For nontext region value of Rs is small and for text region value of Rs is large.
3. Aspect Ratio: Aspect Ratio of a bounding box is defined as the ratio of its width to its height. Usually texts are placed horizontally in videos. Therefore, the width of text region is much greater than the height of text region. This characteristic can be utilized to filter out some of the false alarms. A false alarm occurs where a nontext region has same properties as like text region and nontext region is identified as a text region.

C. Text Extraction and Word Segmentation

After detecting text region in the image, from that text region text is extracted from the image using optical character recognition (OCR) system. OCR is used to convert images with text into editable formats. OCR processes input images with text and get editable documents like TXT file. The process of OCR involves several steps including segmentation, feature extraction, and classification. OCR firstly converts input image into relatively high-resolution image TIFF. Then it recognizes the character one by one and finds the best match for each character. Then all characters are written into the text file and lastly we will get the text file with extracted text. Finally word of extracted text is segmented. Generally words are separated by space so when space character come the word before the space is displayed at new line. In this way we will get the separate word of extracted text.

IV. RESULT ANALYSIS

In this section the results obtained from proposed text detection system is analyzed by giving different videos as a input to the system. This section also deals with performance analysis and comparison of proposed method and existing text detection method. All the classes in proposed system were coded and compiled in the JAVA 1.7. All tests were carried out on an Intel(R) Core(TM)i3-2310M CPU with 2.10 GHz Pentium processor and 4 GB RAM under MS Windows 7 (64 bit) operating system. Different videos are used for testing. Results shows that the proposed corner based approach gives good precision result than texture based approach.

Table I shows comparison between performance of our approach and texture based approach. Table II shows the performance of our text detection system.

Table I: Performance Comparison between Our Approach and Texture Based Approach

Table I: Performance Comparison between Our Approach and Texture Based Approach

No. Of Frames	Our Approach		Texture Based Approach	
	Precision	Recall	Precision	Recall
45	92.10%	83.33%	88.86%	85.5%

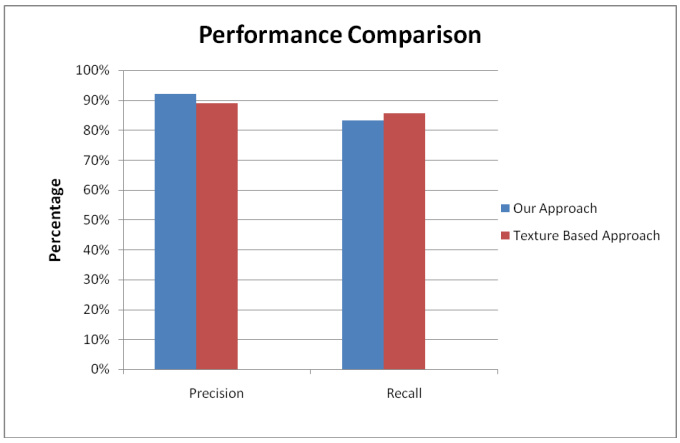

Fig. 2 Performance Comparison Graph

Table II: Performance of Text Detection

No. Of Frames	Hits	Misses	False Alarms	Precision	Recall
45	35	7	3	92.10	83.33

V. CONCLUSION

This proposed method develops a text detection system which detects as well as extracts the text from video. Usually numbers of corner points are present orderly in text region. So by using the corner points as the fundamental feature of character and text, the system detects video text with high precision and efficiency. Based on the results of text detection, actual text is recognized and words of extracted texts are segmented which can be used in future applications

REFERENCES

[1] K. Kim, K. Jung, and J. Kim, Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm, IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 16311639, Dec. 2003.

[2] W. Mao, F. Chung, K. K. M. Lam, and W. Sun, Hybrid chinese/english text detection in images and video frames, in Proc. 16th Int. Conf. Pattern Recognit., 2002, vol. 3, pp. 10151018.

[3] B. K. Sin, S. K. Kim, and B. J. Cho, Locating characters in scene images using frequency features, in Proc. 16th Int. Conf. Pattern Recognit., 2002, vol. 3, pp.489492.

[4] A. K. Jain and B. Yu, Automatic text location in images and video frames, Pattern Recognit., vol. 31, no. 12, pp. 20552076, 1998.

[5] M. R. Lyu, J. Song, and M. Cai, A comprehensive method for multilingual video text detection, localization, and extraction, IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 2, pp. 243255, Feb. 2005.

[6] X. Liu and J. Samarabandu, Multiscale edge-based text extraction from complex images, in Proc. Int. Conf. Multimedia Expo., 2006, pp. 17211724.

[7] C. Harris and M. Stephens, A combined corner and edge detector, in Proc. Alvey Vis. Conf., 1988, pp. 147151.

[8] Xu Zhao, Kai-Hsiang, Yun Fu, Text From Corner: A Novel Approach to Detect Text and Caption in Videos, IEEE Trans. Image Processing, vol. 20, no.3, March 2011