Detect the Text and Caption in Video

Megha Khokhra
Department of Electronics and Communication
KIRC, Gujarat, India

Abstract- *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. The features to describe text regions are area, saturation, orientation, aspect ratio and position. Then convert into binary image. 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.*

Keywords- Text detection, video converted, image frames, caption video, detect text from video, methodology, edge selection

I. INTRODUCTION

In today's world, video plays an important role as a media delivered over TV broadcasting, internet and wireless network. It is often required to automatically detect and extract the text information from these video frames. 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.[8]

The quality of video degrades when it is transmitted through a medium and this makes the video frames of low intensity and low contrast. These are the main problems of working with video images.[3]

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.[5]

With the expansion of digital media resources, there is an increase in the demand for video semantic analysis, retrieval and annotation. Text contains semantic information and thus can contribute significantly to video retrieval and understanding tasks. Therefore, video text recognition is crucial to research in all video indexing and summarization.

II. BACKGROUND

A. Text detection

Text area detection from video frame is an essential step for Video OCR. Text Extraction and recognition in general have quite a lot of relevant application for automatic indexing or information retrieval such document indexing, content-based image retrieval, and the famous car plate recognition which further opens up the possibility for more improved and advanced systems. [5]

B. Steps involved for detection of text

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 following steps:

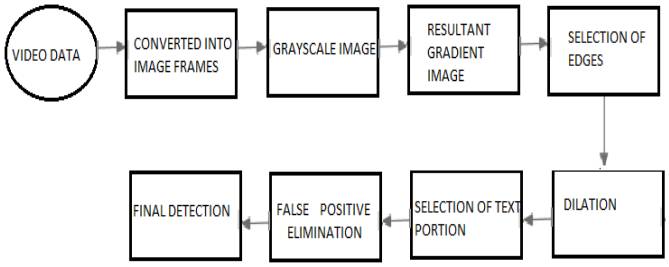

Fig 1 Block Diagram for Text Detection in Video

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]

C. Gradient Based Procedure

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**).

$$HGI(i, j) = GI(i, j) - GI(i+1, j)$$
$$\forall (i, j) \in GI \text{ and } (i+1, j) \in GI$$

Similarly the gradient value of a particular pixel (i, j) of the vertical gradient image (**VGI**) of the image (**I**) is calculated by taking the difference of the immediate right pixel (i, j+1) with the current pixel (i, j) of the gray image in the following way.

$$VGI(i,j) = |GI(i, j) - GI(i, j+1)|$$
$$\forall (i, j) \in GI \text{ and } (i+1, j) \in GI$$

The horizontal and vertical gradient images are then merged to find the Resultant gradient image (**RGI**).

$$RGI(i, j) = HGI(i, j) + VGI(i, j)$$
$$\forall (i, j) \in HGI \cap VGI$$

The resultant gradient images are shown in Figure 2.

(a) (b) (c)

Fig. 2 (a) Horizontal Gradient Image (b) Vertical Gradient Image (c)Resultant Gradient Image

D. Edge Based procedure

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. The original canny edge image and the selected edge image are shown in Figure 3

Fig. 3: (a) Original Canny Image

Fig. 3: (b) Selected Canny Image

E. Uniform Color Based Procedure

In this stage we take the intersection of the images we got after opening morphological operation and the binarized gradient image. 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.

We next applied the false positive removal methodology. In that method they used a scaling factor α for scaling both the internal and external boxes of the detected text block. In that algorithm set the scaling factor α to 0.2. We

did a simple modification in this method. In our case we set two different scaling factors *a* and *b* for scaling internal and external boxes of the original box respectively. We set *a=0.4* and *b=0.2* experimentally on the dataset of images we considered.

IV. EXPERIMENTS AND RESULTS

A. Comparison

To give an objective comparison of all the above methods, we define the following performance metrics, i.e. detection rate, false positive rate and misdetection rate. The metrics are defined as follows. The detected text blocks are represented by their bounding boxes. A text block is considered as truly detected if the bounding box covers all or some characters in the same text block. If a truly text block has some characters not covered by the bounding box, it is considered as a misdetection. To judge the correctness of the text blocks detected, we manually count Actual Text Blocks (ATB) in the images in the dataset.

Also we manually label each of the detected blocks as one of the following categories:

Truly detected text blocks (TDB): a detected block that contains text fully or partially.

Falsely detected text blocks (FDB): a detected block that does not contain text.

Text block with missing data (MDB): a truly detected text block that misses some characters.

Based on the number of blocks in each of the categories mentioned above, the following metrics are calculated to evaluate the performance of the methods:

Detection rate (DR) = Number of TDB /Number of ATB.

False positive rate (FPR) = Number of FDB / Number of detected text blocks (Truly + falsely).

Misdetection rate (MDR) = Number of MDB / Number of TDB.

The results in terms of the above metrics are reported in Table 1. The table shows the performance of the proposed method in comparison with the existing methods where we can see that the detection rate, false positive and misdetection rates of the proposed method are higher than the three existing methods.[1]

Table 1: Performance of the existing methods

Method	DR	FPR	MDR
Edge based	80.0	18.3	20.1
Gradient based	71.0	12.0	10.0
Uniform colour based	51.3	27.3	37.3

B. Software to be used

MATLAB (Matrix Laboratory) is a high-level language and interactive environment for numerical computation, visualization, and programming. Using MATLAB, one can analyze data, develop algorithms, and create models and applications. The language, tools, and built-in math functions enable to explore multiple approaches and reach a solution faster than with spreadsheets or traditional programming languages, such as C/C++ or Java. MATLAB is used for a range of applications, including signal processing and communications, image and video processing, control systems, test and measurement, computational finance, and computational biology. More than a million engineers and scientists in industry and academia use MATLAB, the language of technical computing.

C. Results

The results of our proposed method are shown in below Fingers.

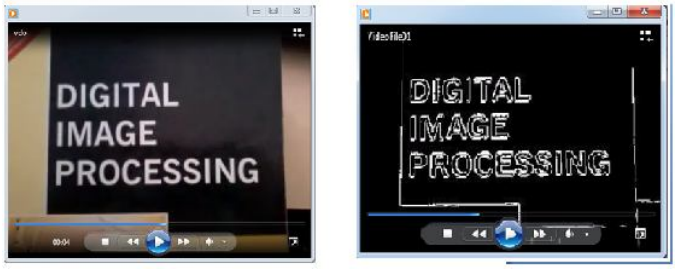

Fig 4 Result 1

Fig 5 Result 2

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] 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, September 2009.

[2] Bindhu. N. and Bala Murugan**.** "An Adaptive Novel Approach for Detection of Text and Caption in Videos", IEEE transactions on Image Processing, September 2012.

[3] Christian Wolf AND Jean-Michel Jolion, "Model based text detection in images and videos," in Proc. IEEE Conf. Comput. Vis. Pattern Rec 2004.

[4] Xu Zhao, Kai-Hsiang Lin, Yun Fu "Text From Corners: A Novel Approach to Detect Text and Caption in Videos", IEEE Transactions on Image Processing, Vol. 20, No. 3, March 2011.

[5] Xiangrong Chen and Hongjiang Zhang "Text Area Detection from Video Frames", Microsoft Research China, 2000.

[6] Datong Chen, Jean-Marc Odobez and Herv/e Bourlard "Text detection and recognition in images and video frames", in Vis. Pattern Rec 2004.

[7] Qixiang Ye, Wen Gao, Weiqiang Wang and Wei Zeng, "A Robust Text Detection Algorithm in Images and Video Frames", IEEE, 2003.

[8] Marios Anthimopoulos "Text Detection in Images and Videos", Department of Informatics and Telecommunications National and Kapodistrian University of Athens, 2007.