A Review on Arbitrarily and Horizontal Scene Text Extraction from Images or Videos

Palvi Garg

Punjabi University, YCOE, Talwandi Sabo, India

Email: palvigarg234761@gmail.com

ABSTRACT

The extraction of text in an image is classical problem in the computer vision. With the increasing popularity of practical vision system text recognition in videos becomes a critical task. Text data present in images and video contain useful information for automatic annotation, image indexing. But as video frames contain complex background, multiple fonts and orientation, and different alignments make the problem of automatic text extraction extremely challenging. There are two types of text in video frames. One is scene text and other is a graphic text. Scene text is captured in scene while recording while graphic text is more structured and closely related to subject. Scene text is difficult to detect than a graphic text. In proposed technology extraction of scene text is done.

Keywords-Text recognition; Text extraction; image indexing; scene text; graphic text

1. INTRODUCTION

Text detection and extraction from video is an emerging area for research in the field of image processing and multimedia as it is useful in bridging a gap between low level feature and high level features to retrieve video events based on semantic with the help of Optical Character Recognition (OCR). Besides, scene text detection in video is challenging because of low resolution, complex background, different fonts, font size, orientation and color bleeding [2]. Arbitrary orientation of text in video makes the problem even more complex and

challenging. In spite of such Extensive studies, it is still not easy to design a general-purpose text information extraction 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. There are two types of text present in images and videos.

1. Scene text
2. Graphic text
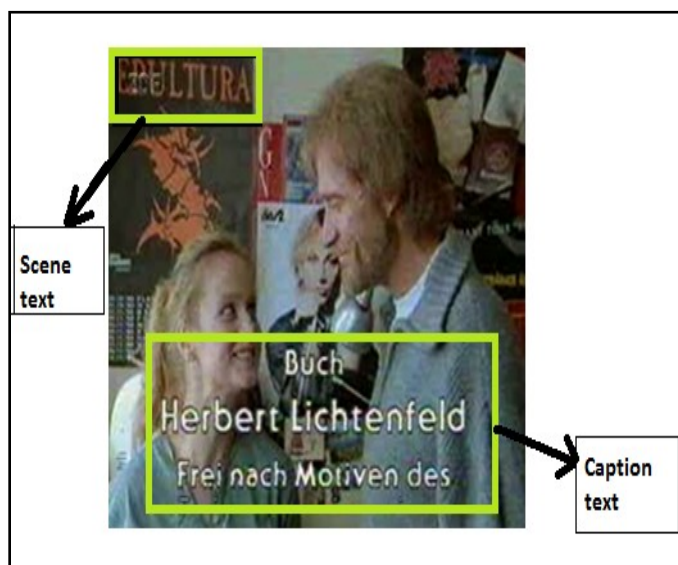
Figure 1: Scene text and caption text images

There are several methods for natural scene text detection in camera based images in document analysis. It is seen that these methods required

high resolution and clear shape of the character to identify the regular pattern of text for text detection in natural scenes. For instance, proposed a hybrid approach for text detection in natural scene in images based on conditional random field. The conditional random field involves connected component analysis to label the text candidates. We can also see method on multi-oriented text extraction from camera images in [9] but this method works well if text with clear character shape is present in the images. These constraints are true for high resolution like scanned and camera images but not necessarily true for video based images due to undesirable properties of video. Thus, document analysis based methods used for text extraction from camera images and natural scene images may not be suitable without modifications for scene text detection or extraction from video frames. Generally, video contains two types of text that are scene text and graphics text. Scene text is captured by the camera. Examples of scene text include street signs, billboards, and text on trucks and writing on shirts. Graphics text is manually added to video frames to supplement the visual and audio content. Since it is manually added, detection of such text is easier than scene text. In case of sports domains, scene text helps in retrieving sports events and it is useful in many applications such as navigation, surveillance, video classification, or analysis of sports events.

The major categories of text detection method are

- Connected component-based.
- Texture-based methods.
- Edge and gradient based methods.

Since connected component based methods expect character shape, the methods may not be suitable for scene text detection in video with complex background. While texture based method are better than connected component as they work well for complex background of video. However, there is a problem in defining texture property for scene text detection as background may give

defined texture property and they are sensitive to fonts and font size. On the other hand, the combination of edge and gradient feature based method are good for text detection in terms of efficiency and some extent to complex background. However, these methods suffer from setting threshold values at several stages of the algorithms. Based on the above discussion, we can conclude that arbitrarily-oriented text detection in video frames is not addressed fully. Multi-oriented text has only been partially addressed in where the algorithm is limited to caption text and a few selected directions.

2. CHALLENGES IN EXTRACTING TEXT

1. PROBLEM IN REMOVE NOISE

Remove noise is challenging task in extracting text from videos. Images in digital cameras and conventional camera pick up noise from a variety of source. So, our aim is to remove noise in background while extracting text from images or videos.

2.TO FIND OUT FALSE POSITIVES IN IMAGES OR VIDEOS

False positives elimination is a problem occur in extracting text from videos or images. False positives are the information which is not necessary for extracting text but there are present with extracting text. Example: logo of any company or college which is present in scene images.

3.TO FIND OUT TEXT IN LOW RESOLUTION IMAGES

Low resolution images means image resolution is not higher than that necessary for the image to convey its information. While extracting text from images or videos, low resolution images create a problem because there is useful information can't be find from low resolution images.

3. TECHNIQUE USED IN EXTRACTING TEXT

To attain the objective, step by step methodology is used to accomplish this work. First of all, existed literature will be collected regarding text detection in video clips or images. Merits and demerits are concluded from that literature. Proposed technique is applied and then results are evaluated.

Step 1: From which video text is extracted, take a video frame.

Step 2: A video frame is taken as input and text is detected.

Step 3: After text is detected, find the location of text in video frame.

Step 4: False positives are removed from that image and text line is segmented.

Step 5: Text is recognized by OCR from video frame .

Step 6: Finally, text is extracted from a video frame.

ACKNOWLEDGEMENT

REFERENCES

[1] Bhavadharani R., Sowmya P. M., Thilagavathy A., "A Dynamic Approach to Extract Texts and Captions from Videos", International journal of computer science and mobile computing, Vol. 3, Issue. 4, April 2014, pg.1047 – 1052.

[2] Choksi A.,Desai N., Chauhan A., Revdiwala V., Patel K., " Text Extraction from Natural Scene Images using Prewitt Edge Detection Method"International Journal of Advanced Research in Computer Science and Software Engineering,Vol.3, Issue 12, December 2013.

[3] Dhobale1 S.R., Tayade A. A., "A survey on Text Retrieval from Video"International Journal of Application or Innovation in Engineering & Management,Vol.3, Issue 11, November 2014, pg. 79-85.

[4] GhorpadeJ., PalvankarR., Patankar A., Rathi S., "Extracting text from video"Signal & Image Processing : An International Journal (SIPIJ), Vol.2, No.2, June 2011, pg. 103-112.

[5] Liu X., Wang W., "Robustly Extracting Captions in Videos Based onStroke-Like Edges and Spatio-Temporal Analysis", IEEE transactions on multimedia, Vol. 14, No. 2, April 2012, pg. 482-489.

[6]Murthy N., Kumaraswamy Y. S., "Robust Model for Text Extraction from Complex Video Inputs Based on SUSAN Contour Detection and Fuzzy C Means Clustering", International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, September 2011, pg.225-234.

[7]Palma D., Ascenso D., Pereira F., "Automatic Text Extraction in Digital Video Based on Motion Analysis", Springer-Verlag Berlin Heidelberg 2004, pg. 588–596.

[8] Shivkumara P., Phan T.Q., Lu S., Tan C.L., "Gradient Vector Flow and Grouping-based Methodfor Arbitrarily Oriented Scene TextDetection in Video Images", IEEE transactions on circuits and systems for video technology, Vol. 23, No. 10, October 2013, pg. 1729-1739.

[9]Shivakumara P., Sreedhar R.P., PhanT.Q., Lu S., Tan C.L., "Multioriented Video Scene Text Detection Through Bayesian Classification and Boundary Growing" IEEE transactions on circuits and systems for video technology, Vol. 22, No. 8, August 2012, pg. 1227-1235.

[10]Sonam., Kumar M., "Implementation of MD algorithm for TextExtraction from Video"Nirma University International Conference on Engineering IEEE 2013, pg. 1-5.

[11] Zhong Y., Zhang H., Jain A.K., "Automatic Caption Localization in Compressed Video", IEEE transactions on pattern analysis and machine intelligence, Vol. 22, No. 4, April 2000, pg. 385-392.