International Journal of Computer Trends and Technology
ISSN: 2231-2803 / https://doi.org/10.14445/22312803/IJCTT-V68I4P117

Volume 68 Issue 4, 96-99, April 2020

Original Article

Text Detection in Video for Video Indexing

Youssef Rachidi

Laboratory IRF-SIC, faculty of sciences Ibn Zohr University, Agadir, Morocco.

Received Date: 09 March 2020
Revised Date: 26 April 2020
Accepted Date: 28 April 2020

Abstract *- Automatic text detection in the video is an important task and a prerequisite for video retrieval, annotation, Recognition, indexing and content analysis. Video OCR is a technique that can greatly help to locate the topics of interest in video via the automatic extraction and reading of captions and annotations. Text in the video can provide key indexing information. Recognizing such text for search applications is critical. A major difficult problem for character recognition for videos is degraded and deformed characters, low-resolution characters or very complex backgrounds. To tackle the problem preprocessing on text images plays a vital role. Most of the OCR engines are working on the binary image, so finding a better binarization procedure for an image to get the desired result is important. The accurate binarization process minimizes the error rate of video OCR.*

Keywords *- Text Detection, OCR, Binarization, video Text, video indexing.*

I. INTRODUCTION

The detection task is one of the most important research areas that gained increasing attention in image processing and machine vision. This task aims to find automatically the location and scales of the regions of interest presented in an image or in the video. These regions could be traffic signs, license plates, cars, animals, people, text, human faces, etc. Video OCR is a technique that can greatly help to locate topics of interest in a large digital video archive via the automatic extraction and reading of captions and annotations [A]. Captions generally provide vital search information about the video being presented - the names of people and places or descriptions of objects. Understanding the content of videos requires the intelligent combination of many technologies: speech recognition, natural language processing, search strategies, image understanding, etc. Extracting and reading captions provides additional information for video understanding. Performing Video OCR on video and combining its results with other video understanding techniques will improve the overall understanding of the video content. Although there is a great need for integrated character recognition in text-based video libraries. Automatic character segmentation was performed for titles and credits in motion picture videos; however, papers have insufficient consideration of character recognition. There are similar research fields that concern character recognition of videos. In character extraction from the car license plate using video images is presented, and characters in scene images are segmented and recognized based on adaptive thresholding [E]. Therefore, automatic detection has been applied in many applications. Among these applications, we mention: Such as biometrics, medical, driver assistance, visual surveillance, security, human-machine interface and robotics and so on. In recent years, a large number of videos Are uploaded and shared every day on YouTube, social networks and TV Channels. Consequently, the automatic indexing and extracting of information from these videos is an issue of great importance and an indispensable process. This is why many companies and research labs contribute and invest in developing efficient algorithms and systems to ensure the segmentation and detection task [H]. These results are related. Character recognition for the video presents its own difficulties because of different conditions of title character size and complex backgrounds. In video caption resolution of the character is lower; also, the background complexity is more severe than in other research. The first problem is the low resolution of the characters. The size of an image is limited by the title number of scan lines defined in the NTSC standard; the character of the video caption is small to avoid the occlusion of interesting objects such as people's faces. Therefore, the resolution of characters in the video caption is insufficient to implement stable and robust Video OCR systems. Another problem is the existence of complex backgrounds. Characters superimposed on videos often have hues and brightness similar to the background, making extraction extremely difficult. These problems in video OCR have opened an area for research [F].

Text recognition in video is a challenging task that has a significant impact on various multimedia applications. And Video OCR is a technique that can greatly help to locate topics of interest in a large digital video via the automatic extraction and reading of captions and annotations. In section 3, we describe the Video Optical Character Recognition process and all the process modules required in video OCR. Applications of video Optical Character Recognition are explained in section 4. Finally, a conclusion is drawn in Section 5.

II. RESEARCH QUESTIONS

Text recognition, even from the detected text lines, remains a challenging problem due to the variety of fonts, colours, the presence of complex backgrounds and the short length of the text strings as well as the Recognition of videotext is also a challenging problem due to various factors such as the presence of rich, dynamic backgrounds,

low resolution, colour, etc. A strategy is required to process the video images to produce high-resolution binarized text images that resemble printed text and minimize the error rate during Recognition of degraded character.

III. VIDEO OPTICAL CHARACTER RECOGNITION

In this section, a similar research topic to camera-based OCR is OCR for texts in videos. This research topic can be further divided into two types. The first type is to recognize scene texts, which are the texts in the scene captured in video frames. The second type is to recognize caption texts, which are the texts superimposed on video frames. Since caption texts are attached intentionally to explain or summarize the contents of the video frames, they are also useful as an accurate index of the video. Caption texts are also captured in multiple video frames like scene texts and have their own characteristics [C]. Fig 1 contains a block diagram that shows the processing sequence in a typical video text processing system

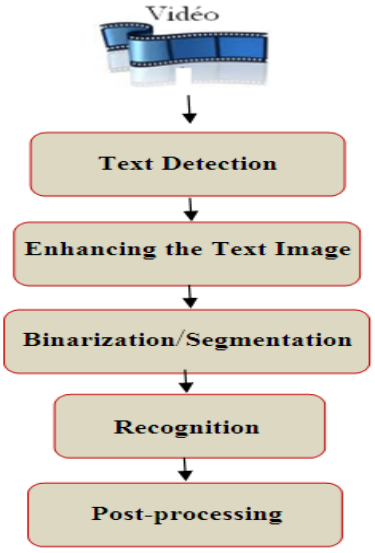

Fig 1. Steps in video OCR

There are two problems in obtaining efficient and robust text detection using machine learning tools. One is how to avoid performing computational intensive classi3cation on the whole image, and the other is how to reduce the variance of character size and greyscale in the feature space before training [D]. The primary challenge is that we need to first track the sequence of instances of a video text object over successive frames of video, even in the presence of factors such as the irregular, spasmodic motion of a handheld or vehicle-mounted camera, the nonlinearity of text plane motion, occlusions, shadows and other illumination changes, and so on. So, following the initial detection step, the next step is enhancement. In addition, as part of the image enhancement step, video text images that exhibit perspective distortion (typically of text that is within the scene and not directly facing the camera); need to be rectified so that their appearance is similar to text that directly faces the camera. Text rectification involves the estimation of the position and orientation of the text plane relative to the camera. As most of the OCR's are working on the binarized image, binarization of enhanced image is required and segmenting text region in video frames acquired by a smartphone has received relatively low attention. However, segmenting the document region from this kind of video clip obtained via mobile camera presents a challenge because of many difficulties: a variety of text formats, document classes, complex background, perspective distortion, illumination variation, poor focusing and motion blurs [E]. Finally, the binarized text is recognized by standard OCR and post-processing is applied to the recognized text.

A. Text Detection

Most of the existing video text detection methods have been proposed on the basis of colour, edge, and texture-based features; we have Color-based approaches that assume that the video text is composed of uniform colour, and Edge-based approaches are also considered useful for overlay text detection since text regions contain rich edge information, and Texture-based approaches, such as the salient point detection and the wavelet transform, have also been used to detect the text regions [F].

B. Enhancing the Text Image

An approach based on multiple-frame processing utilizes the fact that the same text appears in multiple frames in a video. Simply speaking, low-resolution images of a text in multiple frames are a redundant but erroneous representation of the original text. Thus, by applying some error correction technique (like error-correcting code), the original high-resolution image can be recovered [B] [C].

. The enhanced image is computed by aligning the different instances of a particular text region across frames and, for each pixel, choosing the colour that corresponds to the minimum intensity value across frames. We can try other order statistics such as the mean, median, and the maximum, but the minimum order statistic yielded the best image in terms of visual perception [F].

C. Binarization

Our initial aim is to detect the upper and lower baselines of the text image. The area which is limited by the lower and upper baseline defines the main body of the text, and hence, we can perform binarization with a different valuation in parameters for the inside and outside area of the main body of the text. The final result is achieved by extending the text region properly using the convex hulls of the connected components to perform once again the same binarization scheme [G]. We have Baseline and Stroke Width Detection, Binarization using the Baselines Information and Binarization using the Convex Hulls Information. Fig 2, fig 3 present RGB and Greyscale Image and fig 4 present the binarization result of the modified ALLT along with the baselines in grey

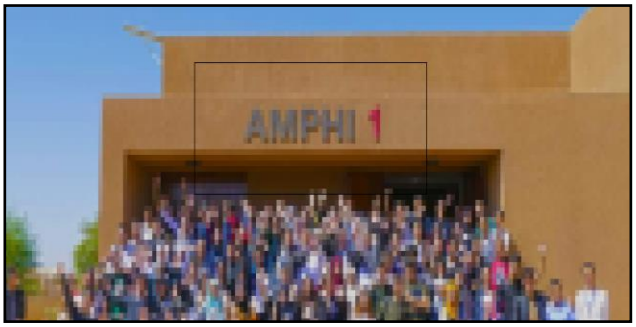

Fig. 2 RGB Image

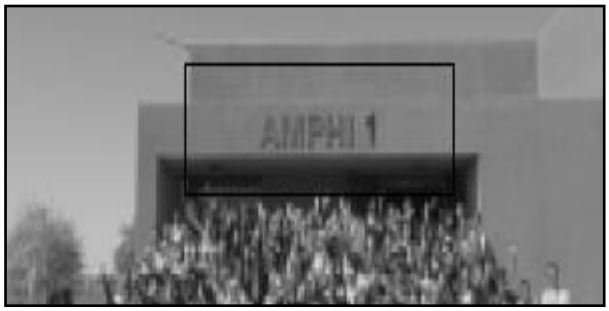

Fig. 3 Greyscale Image

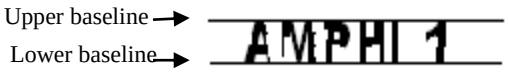

Fig. 4 the binarization result of the modified ALLT along with the baselines in grey

D. Recognition

Most of the previous methods that addressed text recognition in complex images or video worked on improving the binarization method before applying an OCR module. However, an optimal binarization might be difficult to achieve when the background is complex, and the greyscale distribution exhibits several modes. Moreover, the greyscale value of text may not be known in advance

E. Post-processing

In recent years, a large number of videos have been uploaded and shared every day on YouTube, social networks and TV Channels. Consequently, the automatic indexing and extracting of information from these videos is an issue of great importance and an indispensable process. This is why many companies and research labs contribute and invest in developing efficient algorithms and systems to ensure the segmentation and detection task [H]. To acquire text information for content-based access of video databases, high word recognition rates for video OCR are required. Once the text for a frame has been recognized, it is stored to be compared to the text extracted from neighbouring frames for indexing. We apply post-processing, which evaluates differences between recognition results with words in the dictionary and selects a word having the least differences. Different post-processing techniques are used for indexing. Video indexing can be used in various applications like digital libraries digital News.

IV. APPLICATIONS OF VIDEO OPTICAL CHARACTER RECOGNITION

In this work, we focus only on the operation of detecting the text outlines within frames. This operation aims to separate the quadrilateral shape of the text from the background of the image. The relevant text and information can be presented in different types of an image in the video. However, the large changes in text fonts, colours, styles, scales, and the poor contrast between the text region and the background of the image, in addition to the fact that the document text may comprise several types of text information (word, sentence, paragraph, title, subtitle, list, etc.), frequently make text extraction operation more challenging. Performing Video OCR on video and combining its results with other video understanding techniques will improve the overall understanding of the video content. Text in the video provides rich information for content-based search applications. There are various applications in which video OCR is used. Fig.5 shows an application of video OCR where text in the video is detected.

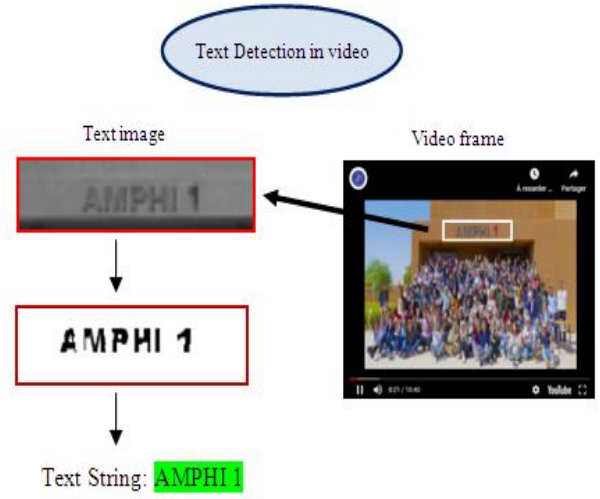

Fig 5. Application of video OCR.

V. CONCLUSION

- Recognizing artificial text embedded in images provides
- high-level semantic clues which enhance au tremendously-
- tomatic image and video indexing. While for printed doc-
- ument, optical character recognition (OCR) systems have
- already reached high recognition rates, and are widely com-
- mercialized, Recognition of superimposed text in images
- and videos are still the subject of active research.
- Current character recognition systems require a binariza-
- tion step that aims at separating the text pixels from the
- background pixels of the processed image. Most of the text
- image binarization methods rely on global or local discrim-
- inating thresholds

Recognizing artificial text embedded in images provides high-level semantic clues which enhance the tremendously automatic image and video indexing. While for a printed document, optical character recognition (OCR) systems have already reached high recognition rates and are widely commercialized, Recognition of superimposed text in images and videos is still the subject of active research. Current character recognition systems require a binarization step that aims at separating the text pixels from the background pixels of the processed image. In fact, experimental results indicate that a single binarization approach may not be adequate for dealing with different kinds of text in the video, and a hybrid technique that combines multiple approaches offers the most promise.

REFERENCES

[1] Sato T., T. Kanade, E. K. Hughes, and M. A. Smith, Video OCR for Digital News Archive, In Proc. IEEE Workshop Content-Based Access Image Video Database, (1998) 52-60.

[2] Tsai, Y. C. Chen, C. L. Fang, A Comprehensive Motion Videotext Detection Localization and Extraction Method, in Proc. CCSP, (2006) 515-519.

[3] Seiichi Uchida, Text Localization and Recognition In Images and Video , (2014) 843-883.

[4] Datong Chen, Jean-Marc Odobez, Herv/E Bourlard, Text Detection And Recognition in Images and Video Frames Pattern Recognition, 37 (2004) 595 – 608.

[5] Liang, J.; Doermann, D, and Li, H. (2005). Camera-Based Analysis of Text and Documents: A Survey. International Journal of Document Analysis And Recognition (IJDAR), 7(2-3) (2005) 84-104.

[6] Sankirti S. And P. M. Kamade, Video OCR for Video Indexing, International Journal of Engineering and Technology, 3(3) (2011).

[7] Konstantinos Ntirogiannis, Basilis Gatos, Ioannis Pratikakis, Binarization of Textual Content in Video Frames, International Conference on Document Analysis and Recognition, (2011).

[8] HASSAN EL BAHI*, ABDELKARIM ZATNI, Document Text Detection In Video Frames Acquired By A Smartphone-Based Online Segment Detector and Dbscan Clustering, Journal of Engineering Science and Technology , 13(2) 2018 540 – 557.