International Journal of Engineering Research in Electronic and Communication Engineering (IJERECE) Vol 3, Issue 5, May 2016

# Character Segmentation of Video Subtitles

[1] Satish S Hiremath, [2] K V Suresh
[1][2]Department of ECE, Siddaganga Institute of Technology, Tumakuru, Karnataka, India
[1] satishsh36@gmail.com, [2] sureshkvsit@yahoo.com

***Abstract*— Subtitle extraction from video sequences finds numerous useful applications in video classification. The complex background, illumination and font size of text present in the video makes subtitles extraction extremely challenging. Text in the video are classified as scene text and graphics text. In this paper, character segmentation of graphics text is implemented using morphological operations. Compared to word segmentation in video subtitles, single character segmentation can achieve higher precision because of the simple background. In this paper, the English text present in subtitle of the video is considered for processing. All the character segmentation steps are implemented in Microsoft Visual Studio 2010.**

***Index Terms*—Character segmentation, Complex background, Font size, Graphics text, Illumination, Microsoft Visual Studio 2010, Morphological operations, Subtitle extraction..**

## I. INTRODUCTION

This paper mainly highlights on character segmentation of subtitles present in the videos. The topic of image to text processing has attracted many image processing experts. In modern technology, similar to image, videos are also the main source of information media. As multimedia database is increasing, video indexing has become a challenging and gained lot of interest. There are several methods involved in video indexing, among which one of the method is by using text present in the video. The videos with subtitles carry semantic information, which could provide valuable cues about content of the videos and thus it is very important to analyze. Video index can be put up by detecting, extracting and recognizing text from the video. In text processing, one of the important step is segmentation and in which character segmentation is formidable, as it involves separation of the touching characters.

As proved by Judd et al. [1], when an image containing of text, object, human or any scene is viewed then viewers tends to focus on the text first. Hence, the text present in videos can play a vital part in video indexing. The video text has been categorized into two groups [2] namely 'Scene text' (e.g. text on vehicle, building and sign boards on the road etc...) and 'Graphics text' (e.g. subtitle video, news video). To extract graphics text the complex background, illumination and low resolution creates major problem. Even after text extraction from each frame there may be still considerable amount of non-text region in it. The non-text region in the video frame
may hamper video indexing. Hence, to minimize the non-text region we aim for word and character segmentation. Word segmentation splits the text region into small portions, which decreases the interaction with background noise. Whereas, character segmentation further divides the words into small portions comprising of single character which in turn further reduces the non-text region in video frames. The video index enables the user to enjoy the videos with specific topics. For instance, one can search for the term "Sports News" to get the Sports news of the day. Some examples of video frame with subtitles are shown in Figure 1 (a) and (b).

In this paper, automatic character segmentation method is presented. In section 2, the methods used in literature for character segmentation are discussed. The methodology used to segment the characters is given in section 3. Section 4 provides the results of the presented method followed by conclusion in section 5.

## II. LITERATURE REVIEW

In [3], the number plate (NP) extraction and each character segmentation of NP are discussed using different methods. As segmentation is one of the steps in optical character recognition (OCR), the efficiency of OCR increases with increase in accuracy of character segmentation. Hence in this paper different methods are explained to obtain accurate NP character segmentation.
Disadvantage: As the precision of OCR rest on the character segment method; the wrongly segmented character can result in misidentification of the character.

The author in [4] considers some key features of the character like Characters are monochrome; Characters are rigid; Characters have size restrictions; the same characters appear in multiple successive frames and so on for character segmentation. Text segmentation extracts all pixels out of the video that are part of text characters and

rejects all pixels which do not belong to characters. The Split-and-Merge algorithm is used to achieve segmentation and it is built on a hierarchical decomposition of a frame. The accuracy of segmentation is in the range from 96% to 99%.
Disadvantage: Segmentation performance is higher for video samples with moving text and moving background than for those where the text and background are static. The algorithm cannot profit from multiple samples of same text in consecutive frames.

Basavaraj A and Nagaraj Patil in [5] have explained about video text extraction. Authors use morphological edge detection, sobel filter and dilation to extract the text from the gray scale image. The main objective of the text extraction is to decrease the number of false text region to be served to the OCR.

Disadvantage: The proposed process is insensitive to skew and text orientation.


**(A)**


**(B)**

***Figure 1: Examples Of Video Frames With Subtitles***

In [6] X Huang, Huadong Ma and He Zhang presents the method for text extraction in video where the edge map is obtained by the color gradient method and it helps to process the text rows. The adaptive Thresholding and inward filling is used to process the edge map and get contour of the entire text row. Using the proposed method character segmentation is achieved easily.
Disadvantage: The proposed method assumes that the text character has unchanging color. Hence, this method is not applicable to the text whose character has no even color.

The efficient character segmentation of handwritten devnagari text is proposed in [7]. The authors use the morphological operation dilation and erosion to obtain high accuracy of segmentation of devnagari handwritten characters.
Disadvantage: The accuracy of segmentation of devnagari handwritten characters depends on the preprocessing steps which include removal of noise and distortions of an input image.

## III. METHODOLOGY

The video frames are extracted from the video [8] and the video frame with subtitle is considered for character segmentation. The flow chart of the method is shown in Figure 2. All the steps are implemented in Microsoft Visual Studio 2010 (Open CV).

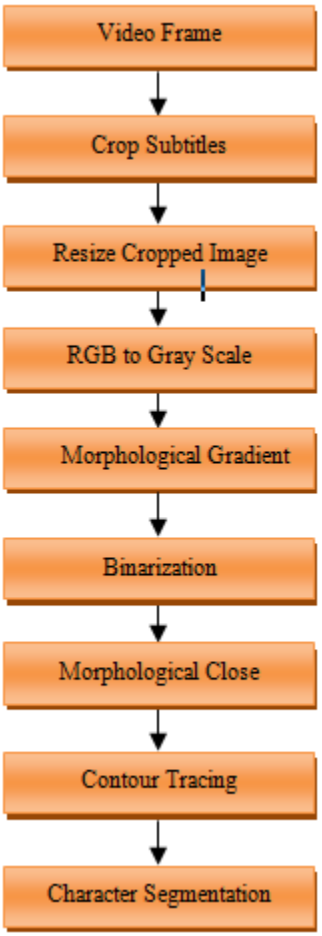



***Figure 2: Flow Chart For Character Segmentation***

### A. Video Frame

The video frames are extracted from the video [8] and one of a frame with text is considered for processing. The original video specifications are Bits per Pixel: 24, Frame Rate: 23.9760, Height: 720, Width: 1280 and Video Format: 'RGB24'.The video frame consisting of text is shown in Figure 3.

### B. Cropping of Subtitles

In videos, subtitles appear at the bottom of the video frame. As subtitles are region of interest (ROI) for us we crop the bottom portion of the video frame so subtitles can be used for the character segmentation. The result of the cropped subtitle is shown in Figure 4.

### C. Resizing

The cropped image is resized to twice its resolution, so that the chances of detecting character increases. The resized image is shown in Figure 5.


*Figure 3: Video Frame with text*

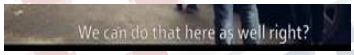



*Figure 4: Cropped Subtitles of the video*

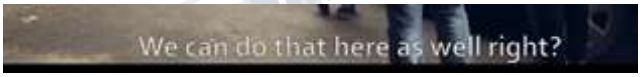



*Figure 5: Resized Image Of The Cropped Subtitles*

### D. RGB to Gray Conversion

As we are working on pixels, the conversion of video frames from RGB to gray scale makes the operation simple. The gray scale outcome is shown in the Figure 6.

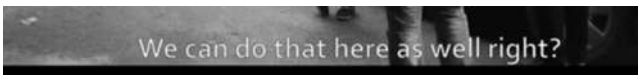



*Figure 6: Gray Scale Of Resized Image*

### E. Morphological Gradient

Morphological gradient is the difference between the dilation and the erosion of a given image. It is an image where each pixel value specifies the contrast intensity in the nearby neighborhood of that pixel. It is most preferred for the segmentation application.

Let $f(x)$ be the gray scale image [9]. The gray scale gradient can be given as

$$\nabla(f) = \max_{x \in g}\{f(x)\} - \min_{x \in g}\{f(x)\} \quad (1)$$

For a structuring element g, the morphological gradient result of the image with subtitle is shown in Figure 7.


*Figure 7: Morphological gradient of subtitle image*

### F. Binarization

After applying morphological gradient to the cropped subtitle image, Otsu Thresholding is applied to convert gray scale image into binary image.

The Otsu Thresholding algorithm assumes that the image comprises two classes of pixels. It calculates the optimum threshold separating the two classes so that their collective spread is minimal and which in turn maximizes inter-class variance. The binarized image is shown in Figure 8 and the weighted inter-class variance [10] given by the relation:

$$\sigma_w^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t) \quad (2)$$

Where,

$$q_1(t) = \sum_{i=1}^{t} P(i) \;\&\; q_2(t) = \sum_{i=t+1}^{I} P(i) \quad (3)$$

$$\mu_1(t) = \sum_{i=1}^{t} \frac{iP(i)}{q_1(t)} \;\&\; \mu_1(t) = \sum_{i=t+1}^{I} \frac{iP(i)}{q_2(t)} \quad (4)$$

$$\sigma_1^2(t) = \sum_{i=1}^{t} [i - \mu_1(t)]^2 \frac{P(i)}{q_1(t)} \;\&\; \sigma_2^2(t) = \sum_{i=t+1}^{I} [i - \mu_2(t)]^2 \frac{P(i)}{q_2(t)}$$

(5)


*Figure 8: Binarized Image by Otsu Method*

### G. Morphological Close

The morphological close operation is a dilation followed by erosion, using the same structuring element for both the operations.

Dilation [11] adds pixels to the borders of objects in an image by finding the local maxima and creates the output matrix from these extreme values.

$$A \oplus B = \left\{ z \mid (\hat{B})_z \cap A \neq \phi \right\} \qquad (6)$$

Erosion removes pixels on object boundaries in an image by finding the local minima and generates the output matrix from these minimum values.

$$A \Theta B = \left\{ z \mid (\hat{B})_z \subseteq A \right\} \qquad (7)$$

Closing of set A by structuring element B, denoted $A \bullet B$, and it is defined as

$$A \bullet B = (A \oplus B) \Theta B \qquad (8)$$

The morphological close operation is applied on the binarized subtitle image. The result of morphological close operation is shown in the Figure 9.


*Figure 9: Morphological close Operation result*.

### H. Contour Tracing

Contour tracing is a method that is applied to digital images in order to extract their border. After applying the morphological operation, the boundary of the characters is traced.

### I. Character Segmentation

The contour tracing helps to distinct each character, using rectangular box the segmented character can be visualized as shown in the Figure 10.

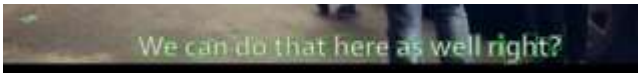


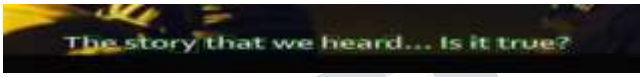

*Figure 10: Final result with segmented characters*

## IV. RESULTS

The character segmentation results for different videos are shown in Figure 11.


**(A)**


**(B)**




**(C)**

***Figure 11: Character Segmentation Results For Different Videos***

One Of The Ways To Calculate Accuracy Of Segmentation Can Be Defined As

$$Accuracy = \frac{No.\text{ of Characters Segmented}}{\text{Total No. of Characters}} \qquad (9)$$

In the video frame with text (Figure 3), the total number of characters is 27 and the number of characters segmented is 25 (Figure 10 and Figure 11 (a)) and in case of Figure 11 (b) and (c) 56 and 19 characters are segmented respectively. Therefore, the accuracy of character segmentation is approximately above 90%. Using morphological operation, characters of video subtitle can be segmented effectively.

## V. CONCLUSION

In this paper, the character segmentation using morphological operation is discussed in detail. The morphological operations involved in character segmentation are gradient and close. The proposed technique effectively segments the characters of video subtitle text. All the steps are implemented in Microsoft Visual Studio 2010 (Open CV).

## REFERENCES

[1] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," IEEE 12th International Conference on Computer Vision (ICCV), pp. 2106-2113, October 2009.

[2] Jiamin Xu, Palaiahnakote Shivakumara, Tong Lu, Trung Quy Phan and Chew Lim Tan, "Graphics and Scene Text Classification in Video," 22nd International Conference on Pattern Recognition (ICPR), pp. 4714 - 4719, August 2014.

[3] Chirag Patel, Atul Patel and Dipti Shah, "A Review of Character Segmentation Methods," Internationl Journal of Current Engineering and Technolgy, Vol. 3, No. 5, pp. 2075 - 2078, December 2013.

[4] Rainer Lienhart, "Indexing and retrieval of digital video sequences based on automatic text recognition," Proc. 4th Association for Computing Machinery (ACM) International Multimedia Conference, pp. 11-20, November 1996.

[5] Basavaraj Amarapur and Nagaraj Patil, "Video Text Extraction from Images for Character Recognition," Canadian Conference on Electrical and Computer Engineering, pp. 198-201, May 2006.

[6] Xiaodong Huang, Huadong Ma and He Zhang, "A New Video Text Extraction Approach," IEEE International Conference on Multimedia and Expo, pp. 650-653, July 2009.

[7] Saniya M.Ansari and Dr. Udaysingh Sutar, "An efficient method of segmentation for handwritten devnagari word recognition," International Journal of Scientific & Engineering Research, Vol. 6, No. 5, pp. 1126-1131, May 2015.

[8] "Goa alla Gokarna New Short Film Kannada," https://www.youtube.com/watch?v=bvEKJHqVkXY, 2015.

[9] Evans.A.N, "Morphological gradient operators for colour images," International Conference on Image Processing (ICIP), Vol. 5, pp. 3089 - 3092, October 2004.

[10] "OpenCV: Image Thresholding (Otsu's Binarization)," http://docs.opencv.org/master/d7/d4d/tutorial_py_thresholding.html#gsc.tab=0.

[11] Rafel C. Gonzalez and Richard E. Woods, "Digital Image Processing," Pearson Prentice Hall, 2013.

[12] "Naa kandanthe," https://youtu.be/04yjh3f_jlw, 2014.