Optical Character Recognition for Alphanumerical Character Verification in Video Frames

Sheshank Shetty, Arun S. Devadiga, S. Sibi Chakkaravarthy, K.A. Varun Kumar, Ethala Kamalanaban and P. Visu

Abstract In real world, optical character recognition (OCR) is one of the key terms challenging image processing stream. Various applications are emerging based on OCR; one fine good example of these terms was recognizing vehicle's number plate in tolls. Since various researchers are under strong discussion in this area, here we proposed a new algorithm for recognizing the characters from the motion pictures. Our proposed model is subjected to two major segregations: One is for character mapping and another one is for character verification. Experimental results have been enclosed with an accuracy level of 97.08 %.

Keywords Optical text extraction · Optical character recognition · Canny filter · RGB image · Gabor filters

S. Shetty (✉) · A.S. Devadiga
Department of Computer Science and Engineering, NMAM Institute of Technology,
Nitte, Karnataka, India
e-mail: sheshankshetty06@gmail.com

A.S. Devadiga
e-mail: arundevadiga1@gmail.com

S.S. Chakkaravarthy · K.A. Varun Kumar · E. Kamalanaban · P. Visu
Department of Computer Science and Engineering, Vel Tech University, Chennai, India
e-mail: sb.sibi@gmail.com

K.A. Varun Kumar
e-mail: varun.kumar300@gmail.com

E. Kamalanaban
e-mail: ethalakamalanaban2009@gmail.com

P. Visu
e-mail: pandu.visu@gmail.com

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324,
DOI 10.1007/978-81-322-2126-5_10

1 Introduction

In the last years, the problem of text detection, recognition, and extraction from a video frame that received a significant attention and lot of work has been proposed. Since this is an era of digitalization, the videos act as an efficient source of information. The one important information hidden inside the video is text, e.g., the title of the movie and scene text hidden in the random video (i.e., shop name and number plate of the vehicle). This text embedded inside the videos may be in small quantity, but these always carry important information of the multimedia content. Hence, retrieving a text from the video makes it more useful and highly desirable.

The proposed method for recognizing and extracting a text from a video frame is optical text extraction (OTE)–optical character recognition (OCR)-based method. This OTE–OCR-based method has three main modules: division of videos into an individual frame, text detection and text recognition and extraction. The major objective of this proposed method OTE–OCR-based text recognition and extraction from the video frames is to efficiently extract a text from video frames. Henceforth, there will be an extension of the application of the OCR systems into the wider area.

Many problems have been faced during the text extraction from the video frames. These problems were due to low resolution of the video, due to complex backgrounds, and also due to the unknown colors. Lot of methods were proposed to overcome these problems by incorporating different architectures and different methods to detect a text out of videos, and these methods have been discussed in the related work section.

The document is organized as follows: In Sect. 2, we discuss various related works on text detection from videos. Section 3 describes the basics of OCR. Section 4 reviews the methodologies involved in the proposed method. In Sect. 5, we present the experimental results of this project compared to the previous work. Finally, we conclude by giving a brief overview of the proposed method and also give a glimpse on future works.

2 Related Work

In this section, we review related works on text retrieval from videos. Lienhart and Stuber [1] briefly discuss automatic text recognition in digital videos. This work comes under the category of bottom-up method where images are segmented to form regions, and later, character regions are grouped to form words. In this work, they deal with text segmentation method, where the aim of segmentation step focuses on dividing frames into regions that contain a text and regions that do not contain a text. Regions that do not contain a text are discarded since these regions are not useful in the recognition process. The regions with text are called as character candidate region, which are passed to recognition process. Motion analysis method is used to detect the character candidate region in consecutive frames. Since this is a bottom-up method, its complexity relies on the image content and also on the segmentation algorithm.

The text localization is one major task performed in this proposed system. In recent years, it has got a significant attention [2, 3]. Chen et al. [2] in their work review about feature extraction using Gabor filters. The text embedded in a video can be a superimposed text and scene text [1]. The motion pictures captured can be from some random scene. Hence, detecting a text from a scene image or video is also an important area to be considered. Ohya et al. [4] in their work briefly discuss about the text detection from the scene image. Scene image is one of the complex images because the images captured occur in a three-dimensional space and might be distorted due to the illumination of light, image might be tilted or slanted, and some part of the image might be partially visible. Hence, these are the major problems faced in the scene images. Since in scene image, the text exists in different orientations, Ohya et al. focused mainly on monochrome images and also on still images rather than on motion pictures or videos. Our main focus will be on detecting a text embedded in a video.

The existing works [1–4] have one or more limitations while retrieving a text from images or videos. There exists sensitivity to different fonts, font sizes, and colors and also restrict to the type of text retrieved (i.e., titles or subtitles only), and not able to handle videos, rather restricting to still images. In our proposed method, we not only restrict to detecting the textual information in the motion picture. But, we present an efficient computational method of extracting the text and passing the segmented characters to the OCR system, and the final outcome will be an editable text in a Word format.

3 Working Methodology

The OCR-based text recognition and extraction is a novel computation scheme which involves three major tasks. The proposed system is divided into three phases: division of videos into an individual frame, text detection phase, and text recognition and extraction phase. As a first phase of our proposed system, we take video as an input. The input video is divided into individual frames, each individual frames are passed through the rest of the two phases, and the individual frame represents the RGB image. The conversion of RGB image to grayscale image is done.

The second phase of the proposed system is text detection; here, we perform two main operations that are text localization and text verification. Text localization is performed on individual frames. Here, feature extraction from an individual frame is done. Henceforth, edge map of the individual frames must be created. There are many methods explained previously for creating the edge map [6]. The edge detector used is a Canny filter [5]. The edge detection (i.e., horizontal and vertical edge detections) is done to the grayscale image using the Sobel and Canny masks. Using edges as the prominent feature of our system gives the opportunity to detect characters with different fonts and colors since every character presents strong edges, despite its font or color, in order to be readable. Canny edge detector is applied to grayscale images. Canny uses Sobel masks in order to find the edge magnitude of the image, in gray scale, and then uses non-maxima suppression and hysteresis threshold. The two operations performed by the Canny edge detector manage to remove non-maxima pixels, without effecting the connectivity of the contour. The resulted edge map will be a binary image with background 0 (black) and connectivity of the contour in 1 (white). Later, dilation on the resulted image is done to find the textlike region. Dilation by a cross-shaped element is performed to connect the character contours of every text line. The dilation is process of increasing the size of pixel and preserving the connectivity of the contour 3 × 13 rectangular structuring element, and octagon structuring element is applied horizontally and vertically, respectively, on the edge map. Different cross-shaped structuring elements (i.e., disk, line) were tried. For more effectiveness of detection, rectangular and octagon were used. The common edge between the vertical and the horizontal edges is extracted, and it is dilated again to get the accurate text like regions.

Morphological binary open image operation is performed to remove the small objects and this operation removes the binary image all connected components that have fewer than 600 pixels. The groove filling mechanism is applied to fill the gap between the non-connected pixels.

3.1 Algorithm—Alphanumeric Character Extraction

```
Begin
Function OCR
      Input:Video file(.avi)
      Output:Image/frames
      Step:Pre-processing
Convert RGB -> grey;
Sobel(horizontal,videoframes);
Canny(vertical,videoframes);
Plot(octagon);
Plot(rectangle);
Dilate(videoframes);
          for i=1:m
               for j=1:n
              FindText(min(n),max(m),min(m),max(n));
             End
         End
Joincharparts(FindText);
End
```

3.2 Algorithm—Alphanumeric Character Verification

```
Begin
Function Text Or Not
      Input: Four coordinates of the dilated regions
(X,X1,Y,Y1) and the edge map image (H)
      Output: Text Or Not
            for i= X:X1
                  for j=Y:Y1
                       hcount=hcount+1
                  end
             end
          tcount=(X1-X) * (Y1-Y)
          Ratio=hcount/tcount
          If (Ratio >=0.065)
          Result= TRUE ( Its Text)
          end
end
```

4 Experiment and Results

The ten sample images were passed to evaluate the performance of the OCR-based system. According to the evaluation, the detection percentage was 99.35 %, and the recognition and extraction percentage was 92.45 %. By using the proposed system, the detection of the text from the image was exceptionally good since there was small miss rate. There was a minor decrease in the recognition and extraction percentage. This was because the template file used for the OCR system was basically trained for Times New Roman and Arial Black fonts. Hence, scene images had a lower recognition rate compared to the normal monochrome images due to the unknown fonts and the colors. This can be overcome by training the template file for more fonts (Fig. 1).

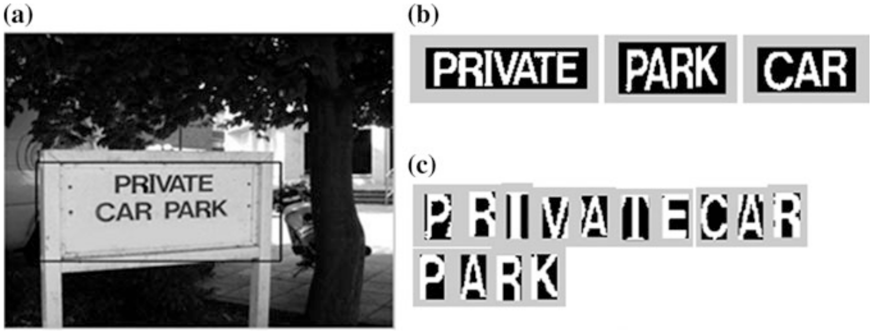

Fig. 1 Results of our proposed methodology. **a** Scene image, **b** representing the extracted region, and **c** segmented text extracted using our proposed methodology

Table 1 Result analysis

Number of video frames	12
Detection percentage	97.08 %
Extracted and recognized percentage	91.60 %

By using the OCR-based text recognition and extraction method, we evaluated the system by passing 12 video frames, the detection percentage was 97.08 %, and the extraction and recognition percentage was 91.60 %. Since large fonts and scene objects (alphanumeric) were used, the processing speed was of average speed (Table 1).

The coordinates of the dilated regions are passed to text verification module. This is performed to check whether the textlike regions extracted are text or not. To verify whether these regions are text or not, the *hcount* of the textlike region of the image is calculated where *hcount* is the total number of white pixels in the detected image. Next, the *tcount* of the textlike region of the image is calculated where *tcount* is the total number of the pixels in the detected image. The ratio of *hcount* to *tcount* is performed, and if the ratio is greater than or equal to a threshold value 0.065, then the coordinates of the dilated regions passed are assumed to be a text, or else the regions extracted are discarded, since it is not useful.

Table 2 Performance analysis of test data

Sample images (with scene text and superimposed text)	Number of characters in an image	Total number of characters detected, recognized, and extracted			
		Text detection phase	Detection percentage (%)	Text recognition and extraction phase	Recognition and extraction percentage (%)
10.avi	14	14	100	12	85.71
5.avi	93	90	96.77	84	90.32
mg1.avi	41	41	100	40	97.56
s2.avi	110	110	100	101	91.81
2.avi	11	11	100	11	100
sample.avi	62	61	98.38	61	98.38
blank.jpg	10	10	100	10	100
txt.jpg	18	18	100	16	88.88
s3.jpg	61	60	98.36	54	88.52
6.avi	24	24	100	20	83.33

The overall detection percentage of OCR-based system for 10 sample images is 99.35 %
The overall recognition and extraction percentage of OCR-based system for 10 sample images is 92.45 %

$$\text{Text ratio} = [(\text{hcount})/(\text{tcount}) \geq 0.065$$

hcount the count of white pixels in an extracted dilated region
tcount the total number of pixels in an extracted dilated region.

5 Conclusion

OCR is one of the emerging areas where most of the researchers are interested in finding the optical version of a character from a digital image to the wrapped version of editable text. Our proposed algorithm is used to perform OCR efficiently with an accuracy of 97.08 %. We have tested nearly about 10 data sets with various video frames/rates/fps. Table 2 denotes clearly the performance analysis of the video frames for various data sets. We have tested our proposed model in images also and achieved the accuracy level better than the accuracy level of video frames. The overall detection rate is very high when compared to recognition rate. Table 2 clearly states the performance metrics of our proposed model.

References

1. R. Lienhart, F. Stuber, Automatic text recognition in digital videos, in *Proceedings SPIE, Image and Video Processing IV.* (1995) pp. 2666–2675
2. X. Chen, J. Yang, J. Zhang, A. Waibel, Automatic detection and recognition of signs from natural scenes. IEEE Trans. Image Process. **13**, 87–99 (2004)
3. V. Wu, R. Manmatha, E.M. Riseman, Text finder: an automatic system to detect and recognize text in images. IEEE Trans. Pattern Anal. Mach. Intell. (1999)
4. J. Ohya, A. Shio, A. Akamatsu, Recognition of characters in scene images. IEEE Trans. Pattern Anal. Mach. Intell. **16**, 214–220 (1994)
5. J. Canny, A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 679–698 (1986)
6. R. Gonzalez, R. Woods, *Digital Image Processing* (Addison Wesley, Boston, 1992), pp. 414–428