International Journal of Advanced Computer and Mathematical Sciences
ISSN 2230-9624. Vol 5, Issue2, 2014, pp14-18
http://bipublication.com

NOVEL METHOD FOR CHARACTER RECOGNITION IN NATURAL SCENE IMAGES

Pradip P. Shewale and Mayura Kinikar

Department of Computer Engineering, MIT Academy of Engineering Alandi, India
E-mail: pradipshewale28@gmail.com, mukinikar@comp.maepune.ac.in

[Received-27/03/2014, Accepted-07/04/2014]

ABSTRACT:

Now a days due to wide increasing use of digital media text localization and text recognition is one area which is useful for recognizing text. Recognized text can be useful for many purposes like license plate reading, sign detection and to visually impair to get text for accessing text to speech algorithm. This paper provides an improved method to recognize text extracted from video frames and images provided by user. The main objective of this system is to implement template matching and thinning algorithm which is used to develop the system. Template match and thin algorithm are much sensitive about font and size which is easily extracted in the frames or image hence all the text is scaled in one size and then input given to optical character recognition (OCR). Experimental results show improvement in the proposed method.

Keywords: Localization, Segmentation, Template matching, Edge based text.

[I] INTRODUCTION:

The rapid growth of digital media contents like camera images, video which leads to content retrieval for the efficient purpose and useful for describing the actual content in that specific video or image, text can easily retrieved and compared to another semantic contents, For some event occurs we can synchronize the extracted text to the image data. Extracted text can widely use as main component for indexing, automatic annotation, and parsing of video or images. While during the processing some text may not be recognized or localized properly due to some constraints like contrast, color and stationary location[1]. Contrast does not guarantee you due to complex background, because in blurred background low contract text can be recognized easily than in complex background. As per as color issue is concern it is believe that most text have uniform color but bleeding effect at the text edge caused by lossy compression. Stationary location has some issue that scrolling type effect are used only in special effects hence to keep the maximum

efficiency of stationary location it can be handle by system[1].

Image retrieval involves several processes as image preprocessing, edge detection, clustering, localization, blurring and recognition of text from the video or image provided by user[3]. Text in media file consists of useful information for getting information about visual content. In the pattern recognition optical character recognition is important potential area which is used for recognizing the text localized in the image. Some approaches of the OCR like segmentation, thinning and template matching and training. Segmentation takes the binary input image scan through image starts horizontally from the first black pixel to the end while vertical line segmentation carried out some rectangular boxes to the character[4]. Thinning algorithm is applied to the segmentation before applying to the template matching for getting better efficiency of the text. Template matching approach recognition is based on the text provided by user with the text character, number and special character stored in database[1,6]. If some character detected is not present in the stored database then it has to be train o recognize it. Retrieval of text from video or file involves in steps as:

Frame Extraction: Extract the frames from the video provided by user.

Text localization: Find out the area where text is localized.

Text segmentation: Segment the vertical line and horizontal character on binary input image.

Text recognition: Text recognition is carried out by thinning, training the text.

[II] LITERATURE SURVEY

1) Jie Xi worked on detection of text and recognition for retrieving clue text in several superimposed text appear in news videos. In is proposed algorithm there is opening procedure of morphological on the blurred edge area map. Recognition rate for localization in their method is 94.7% while precision rate of recognizing is 67.5%.

2) Rainer Lien hart research on the text localization and Text segmentation in videos and images proposed hat every line text with sub pixel had tracked with the efficient rate and recognition at 69%.

3) Bharatratna P, Gaikwad Ramesh R. Manza Ganesh R. Manza worked on template matching, tracking and feature classification algorithm implements to optical character recognition stated that they got text recognition in uppercase letter at 92%.

4) Palaiahnakote Shivakumara had taken segmented text and worked on non significant elimination of edged to get text line boundary at 93%.

[III] PROPOSED METHODOLOGY

Proposed methodology of text recognition works in the fashion that it involves several processes from extracting frames to the recognized output in the binary image.

1.Video frame extraction:

As video contains combination of images so we have to get these images for the further processing on the frames or images which we are extracting from the video. Only we have to give text containing video the frame extractor. Output of frame extractor is frames in the form of .png. We probably take 10 frames for 30 second video file.

2. Preprocessing:

In the preprocessing phase several contents are there for processing on images is grayscale conversion, thresholding, sobel edge detection, clustering. If extracted frame from video is color image because as described in the introduction it has some drawback about complex background hence for easy process it has to convert in grayscale. Grayscale conversion carried out by retrieving the RGB contents from the image taking average of it and assign to the new pixel. Thresholding is completely black and white image can be get by putting any value between 0-255 e.g. 126 to compare it with value of grayscale pixel. Once we get threshold image sobel edge detection is to be apply on it to detect the edges in the

images as text is combination of strokes of edges. Sobel edge detection algorithm finds the magnitude of edges in x and y direction. Magnitude of x and y direction is combined with the direction of pixel because most of the character has opposite edge pairs going in opposite direction. Sobel edge filter 3x3 matrix template for finding the gradient magnitude[11].

Gradient magnitude in x direction is termed as Gx and Gy fo y direction.

$$Gx = X1 * X2$$

$$X1 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

X1 is the values of default sobel filer and X2 are the values of the edge pixel.

$$Gy = Y1 * Y2$$

$$Y1 = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Sobel gradient= sqrt (Gx^2+Gy^2)

Edge direction $\theta = \tan^{-1}(Gy/Gx)$

In this way we can get the values of edge gradient magnitude and direction.

Clustering:

As in the sobel edge detection we have found the edge values in this section we are clustering the edge pixels got by sobel operator having same properties this can be achieved by finding the frequency counts the x direction in one time to the edge pixels. Those areas having approximate frequency counts are grouped by two lines i.e. red and green lines. The edges found in between these two lines are referred as localized text which is used for further processing for recognition.

3. Text recognition:

There are several processing in text recognition are as follows:

Vertical segmentation, Horizontal character segmentation, Thinning, Scaling, Template matching, Training.

Horizontal character and vertical line segmentation:

Horizontal character and vertical line segment finds the black pixel in image provided by user to start the scan algorithm and by combining it with horizontal a separate rectangular box is assign to each character or text in the image for the thinning purpose[5].

Thinning: As in the above segment we get the segment of character but if it is variable in size it can not recognized easily by OCR hence for getting accuracy it has to thin so in this segment we are thinning the character to recognize it well.

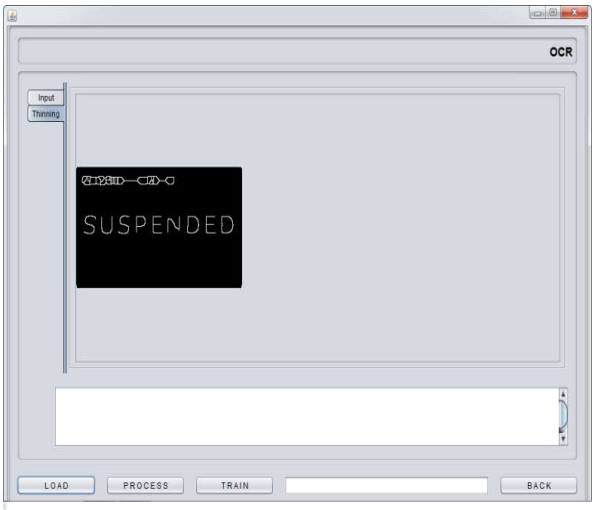

Figure 1. Thinning Character

In above fig we can see we have thin the character thinning works in three stages as pixel which we are deleting should not be connected to more than two pixels. Several template standards should be containing in it.

Scaling:

Scaling is needed to be all character will be in one size for better result.

Template Matching:

Whatever the characters which we have thinned and scaled has to be match in the stored database to recognize what kind of text is present in the image provided by user. Template matching takes the binary image of text and searches it with the stored database if not present in database then it

has to trains in training phase by considering shape of text.

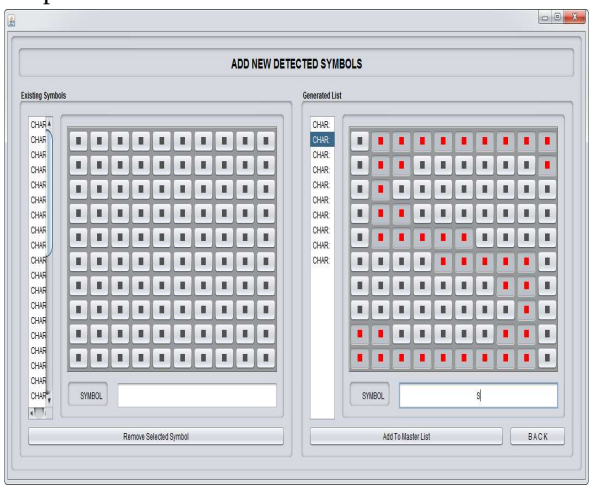

Figure 2. Training the character.

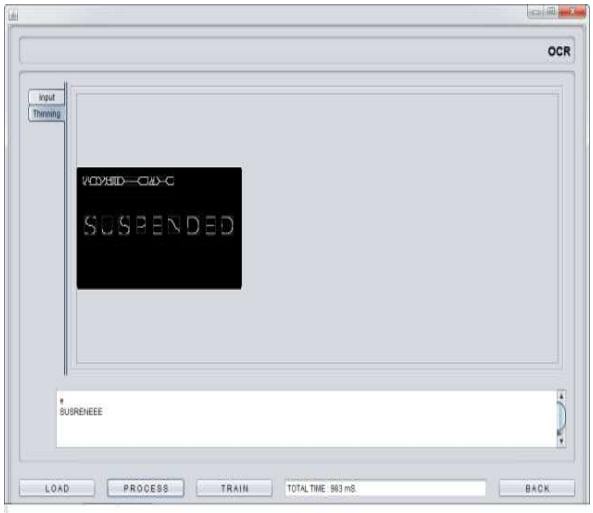

Figure 3. Final output.

[IV] EXPERIMENTAL RESULTS:

Experimental carried out on random images and videos captured by camera in natural scene image to recognize the text.

$$\text{Precision rate} = \frac{\text{No. of accurate detected text}}{\text{No. of accurate detected text} + \text{False positive}}$$

$$\text{False positive rate} = \frac{\text{No. of inaccurate text}}{\text{No. of accurate detected text}}$$

$$\text{Recall rate} = \frac{\text{No. of accurate detected text}}{\text{No. of accurate detected text} + \text{No. of missed text}}$$

Table 1. Recognition Table

Text	Precision Rate	Recall Rate
Uppercase	92	93
Lowercase	92.5	92.6
Numbers	91	93

[V] CONCLUSION:

The proposed system is implemented for extraction of text from video and natural scene images. In this paper we have approached a new novel method for recognizing text in images. The system is based on segmentation, Template matching and training the text. The characters and numbers are detected and recognized well very rare cases shows inaccuracy. The overall results shows there is improvement in the proposed system with precision rate at 92% and recall rate at 93%.

[VI] REFERENCES:

1. Bharatratna P. Gaikwad ,Ramesh R. Manza, Ganesh R. Manza “Video Scene Segmentation to Separate Script” 3rd IEEE International Advance Computing Conference (IACC).March 2013,p.p.1269-1275.
2. Xian-Sheng Hua, Liu Wenyin,” Automatic Performance Evaluation for Video Text Detection,” 6th International Conference on Document Analysis and Recognition (ICDAR2001), September 10-13, (2001), pp545-550.
3. Xi Jie, Xian-Sheng Hua, Xiang-Rong Chen, Liu Wenyin, HongJiang Zhang “A Video Text Detection and Recognition System”, IEEE International conference 2001,
4. P. Shivakumara, W. Huang and C. L. Tan” Efficient Video Text Detection Using Edge Features”, 8th Workshop on Document Analysis Systems(IAPR) , pp 307-314(2008).
5. Chih-Chang Yu, and Hsu-Yung Cheng, Yen-Wen Chung “Detecting and retrieving texts from electronic marquees in natural scenes”, IEEE 17th International Symposium on Consumer Electronics (ISCE), pp.203-204.

6. Huang Xian-Xiang,Zhang Jin-Yu, Chen Yan“Edge Detection of Images Based on Improved Sobel Operator and Genetic Algorithms”□,IEEE international conference 2009.
7. N. A. Otsu, “A threshold selection method from gray-level histograms,” IEEE Transactions on Systems,, vol. 9, no. 1. pp. 62-66.
8. B. Zafarifar, J. Cao “Instantaneously responsive subtitle localization and classification for TV applications,” IEEE Transaction on Consumer Electronics, vol.57, pp.274-282, Feb. 2011.
9. N.Nathiya, K.Pradeepa “Optical Character Recognition for Scene Text Detection, Mining and Recognition”, IEEE International Conference on Computational Intelligence and Computing Research,2013.
10. Y.F.Pan,X.Hou and C.L.Liu,“A hybrid approach to detect and localize texts in natural scene images” IEEE Trans. on Image Proc, vol. 2, pp. 801–811.