Advanced Research in Electrical and Electronic Engineering
Print ISSN: 2349-5804; Online ISSN: 2349-5812 Volume 2, Number 10 April-June (2015) pp. 7-9

http://www.krishisanskriti.org/areee.html

Multioriented Text Detection in the Video

Sayali S. Kuchekar[1], Sanjay L. Nalbalwar[2] and Swarali Sheth[3]

[1,2,3]*Electronics and Telecommunication Department Dr. Babasaheb Ambedkar Technological University Lonere, Raigad, M.H., India*
E-mail: [1]sayalikuchekar10@gmail.com, [2]nalbalwar_sanjayan@yahoo.com, [3]swarali24185@gmail.com

Abstract—*Text detection in the video is very important for information retrieval. Font sizes and color, contrast, complex background, movement of text or background make the detection process a challenging one. This paper implements an efficient method to detect the multi-oriented text in video. Laplacian Sobel Product is a new concept used here. In order to classify between text and nontext pixels, firstly priory and conditional probabilities are found out. Then by using these two, posterior probability is calculated. By intersecting Bayesian classifier's results with canny edge map text candidates are decided. Boundary Growing Algorithm is applied on these results which fixes the bounding box around the text candidates. For false positive elimination, characteristics of text such as aspect ratio, edge density are considered. After false positive elimination we get the true text as the output. The method is tested and it gives more improved results.*

Keywords*: Laplacian Sobel Product(LSP), Bayesian Classifier, Boundary Growing Algorithm.*

1. INTRODUCTION

Due to increased number of digital image capturing devices, craze of capturing images and videos is increasing now days. Hence to analyze such a huge database which is available for any event or topic, content based analysis is required. Text is the main feature that is useful for this. Most of the video databases are tagged with certain information in the form of text. Also the video itself contains the information in text form. Thus text detection plays an important role in the information retrieval.

The texts in video or image are basically of two types, scene text and graphics text. The text that appears naturally in the image i.e. the text which gets captured while recording the video is called scene text. The examples of scene texts are signs on street, recording of a classroom lecture. On the other hand, the text which is manually added in the video is known as graphics text. The subtitles of movie, tagged information of the video comes under graphics text. Graphics text is comparatively easy to recognize as its structural characteristics such as color, font do not vary.

The main difficulties in the text detection are font size, font color, contrast, complex background, movement of text or background. This makes the detection process more challenging.

Previously a lot of work has been done for the text detection. In these methods, the text having uniform color is detected and method deals with horizontal text only [1, 2]. The method in [3] is based upon foreground and background information of the text. For extracting background information water reservoir concept is used. The method works if the condition that the spacing between the text lines is greater than inter character spacing is true. In [4], the text detection is done using Fuzzy-Clustering Neural Network (FCNN) classifier. C. Liu, C. Wang, R. Dai [5] have proposed an algorithm based on the edge maps in four directions (horizontal, vertical, up-right slanting, and up-left slanting) and on the parameters calculated k-means algorithm is applied.

Here in this paper we have implemented a new method which detects the multi-oriented text in video efficiently. The method is based upon Bayesian classifier. The paper is organized as follows. In section 2 the steps in the algorithm that are followed in the method are discussed. Section 3 gives the experimental results. Conclusion and future scope is given in section 4.

2. IMPLEMENTED METHOD

Here the goal of text detection is achieved using the algorithm shown in fig.1. The main steps of algorithm are text enhancement, parameter extraction, applying Bayesian classifier, false positive elimination. In order to detect the text from the video, we have selected the frames at certain rate. The whole algorithm works on the same frame.

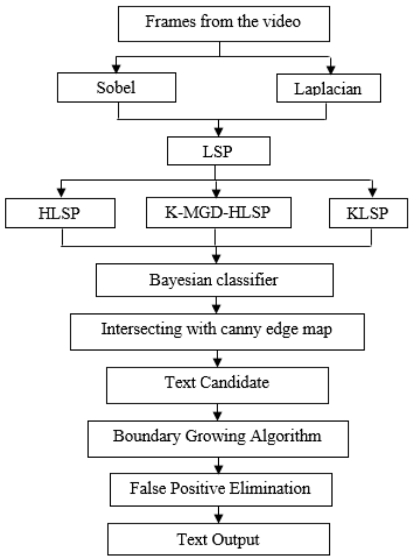


Fig. 1: Algorithm of implemented method

2.1 Text Enhancement

Text regions have large number of discontinuities between text and background. Edge detection plays a very important role in detecting the text regions. For this purpose, laplacian and sobel algorithms are used. Sobel is first order derivative. By using this, text enhancement of high contrast pixels is possible. But it fails for low contrast region. To overcome this laplacian is a more useful tool. Laplacian is a second order derivative. It enhances the pixels in both high contrast and low contrast regions. Here we have used a new concept Laplacian-Sobel product (LSP). The advantages of both methods are combined to get improved results. LSP has more positive and negative peaks as compared to laplacian alone. Hence it is better to use LSP.

2.2 Parameter Extraction

From the LSP, useful information is extracted to detect the text regions. Here to differentiate between text and nontext region we have taken three parameters, viz. HLSP, K-MGD-HLSP and KLSP. HLSP is the classification of high pixel values. The pixel values greater than 0.5 in the normalized LSP are included in the HLSP matrix. K-MGD-HLSP is the classification based upon K-means algorithm. MGD is Maximum Gradient Difference. It is the difference between maximum and minimum gradient values in local window of size 1x n operated on HLSP. Here n is selected based upon the size of text in the image. On this MGD matrix K-means algorithm is applied with K=2. By using K-MGD-HLSP, there are less chances of losing the text pixels. KLSP is the classification done by operating K-Means on LSP with K=2. As a result we get three probable text matrices. In the same way, three nontext matrices are also found out. And both are used for Bayesian classification.

2.3 Bayesian Classifier

The text matrices obtained are named as T1, T2, T3 and nontext matrices as N1, N2 and N3. Each of these represents the binary decision on each pixel. These are collectively used to obtain conditional text probability matrix and conditional nontext probability matrix. They are defined as:

$$P(f(x,y)|T) = \frac{P1(f(x,y)) + P2(f(x,y)) + P3(f(x,y))}{3}$$

And

$$P(f(x,y)|NT) = \frac{N1(f(x,y)) + N2(f(x,y)) + N3(f(x,y))}{3}$$

Where T represents the text class and NT represents nontext class.
These conditional probabilities are used to find the posterior probability.

It is given by:

$$P(T|f(x,y)) = \frac{P(f(x,y)|T)P(T)}{P(f(x,y)|T)P(T) + P(f(x,y)|NT)P(NT)}$$

Where, P (T |f (x, y)) is the probability that a pixel f (x, y) in the resultant output frame of the Bayesian classifier is a text pixel. P (T) and P (NT) are the a priori probabilities.

Then the decision rule is applied as:

$$P(T|f(x,y)) \geq 0.5 \rightarrow BC(x,y) = 1$$

Where BC is resultant matrix obtained after Bayesian classification. Then these results are intersected with canny edge map obtained from original image.

2.4 Boundary Growing Method (BGM)

The text and nontext pixels are detected using Bayesian classifier. Now we need to traverse though these pixels to detect the text. This is done by Boundary Growing method. The method is based upon the nearest neighbour concept. In this method the text image is scanned line by line. Whenever we come across the text candidate a bounding box is created around that. The procedure of fixing the bounding box continues till the end of line. End of the line can be detected by considering inter character spacing. This method can sometime give the false positives, means wrong detection of the text.

2.5 False positive Elimination

The false positives produced due Boundary growing method should be eliminated. For this purpose many parameters such as height, width, aspect ratio, area, etc. are considered. Aspect ratio means height to width ratio. Area means width*height. On these parameters thresholds are applied. And the decision is made based upon the results, whether the text candidate is true or not. Finally false positives are eliminated and we get the text output.

3. EXPERIMENTAL RESULTS

(a)

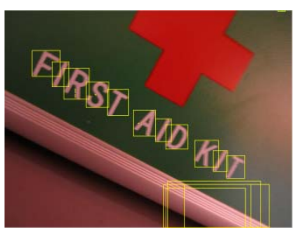
(b)

(c)

(d)

(e)

Fig. 2: (a) Test image (b) Sobel output (c) LSP output (d) Bayesian classifier output (e) Extracted text from the test image



(a) **(b)**

(c) **(d)**

Fig. 3: (a) Test image 1 (b) Text extraction from test image 1 (c) Test image 2 (d) Text extraction from test image 2

4. CONCLUSION AND FUTURE SCOPE

We have implemented a method which detects the text from the video. The method deals with the multioriented text. The method is able to detect both scene and graphics texts.

In future, we will try to eliminate remaining false positives. Also we will be focusing on the curved text in the images. We would also like to detect multilingual text.

REFERENCES

[1] V. Y. Mariano and R. Kasturi, "Locating uniform-colored text in video frames," in *Proc. ICPR*, 2000, pp. 539–542.

[2] Y. Zhong, K. Karu, and A.K. Jain. "Locating text in complex color images", *Pattern Recognition*, 28(10): 1523-1535, 1995.

[3] P. P. Roy, U. Pal, J. Liados, and F. Kimura, "Multi-oriented English text line extraction using background and foreground information," in *Proc.DAS*, 2008, pp. 315–322.

[4] X. Tang, X. Gao, J. Liu, and H. Zhang, "A spatial-temporal approach for video caption detection and recognition," *IEEE Trans. Neural Netw.*,vol. 13, no. 4, pp. 961–971, Jul. 2002.

[5] C. Liu, C. Wang, and R. Dai, "Text detection in images based on unsupervised classification of edge-based features," in *Proc. ICDAR*,2005, pp. 610–614.

[6] Cong Yao, Xiang Bai ,Wenyu Liu, Yi Ma, ZhuowenTu "Detecting Texts of Arbitrary Orientations in Natural Images"

[7] PalaiahnakoteShivakumara, TrungQuyPhan, and Chew Lim Tan,"A Laplacian Approach to Multi-Oriented Text Detection in Video",*IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 2, february 2011.

[8] R.C. Gonzalez and R.E. Woods, Digital Image Processing, pp. 167-180. Addison Wesley and Company, 2002.

[9] R. Lienhart and A. Wernickle, "Localizing and Segmenting Text in Images and Videos," IEEE Trans. Circuits and Systems for Video Technology, vol. 12,no. 4, pp. 256-268, Apr. 2002.

[10] P. Shivakumara, R. P. Sreedhar, T. Q. Phan, S. Lu, C. L. Tan, "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.

[11] Xu Zhao, Kai-Hsiang Lin, Yun Fu, Yuxiao Hu, Yuncai Liu and Thomas S. Huang, "Text From Corners: A Novel Approach to Detect Text and Caption in Videos" *IEEE Transactions on Image Processing*, VOL. 20, NO. 3, March 2011.

[12] Xiaoqian Liu and Weiqiang Wang, "Robustly Extracting Captions in Videos Based on Stroke-Like Edges and Spatio-Temporal Analysis", *IEEE transactions on multimedia*, vol. 14, no. 2, april 2012.