ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 1, January 2018

Novel Approach for Caption Localization and Retrieval in Video using Fast Text Localization Method

[1] Renuka Ganeshrao Sitafale, [2] Prof. A. U. Chhajed
[1] Anuradha Engineering College, [2] Anuradha Engineering College

***Abstract*— Now a days caption in digital media images, capture images or natural images as important for text retrieval in images. In this paper we proposed a fast text localization methd in which sobel edge detection and clustering method is used to localize all the possible edges in the image. As text is made up of edge strokes and opposite edge pair lies in the text. Sobel edge detection algorithm finds magnitude of edges in X and Y direction and angle of edge pixels. Clustering method is used to group the text having same properties while Optical character recognition method to recognize is the localized edges are text or not. As these recognized text can be useful for many purposes like licecence plate reading, sign detection and visually impair to get text for accessing text to speech algorithm. Experimental results show the improvance in the proposed algorithm.**

Keywords— Caption, Clustering, Edge detection, Text localization, OCR.

I. INTRODUCTION

As the wide use of internet, digital media, digital media capturing devices e.g. mobile phones, cameras people are capturing images, video and upload it to the internet websites like instagram, mypics, youtube etc. Text retrieval in this medium has become research area now a day's [2]. As it may contain logo, superimposed text, cricket score, information about players, breaking news, Temperature etc. This text is relevant to the video or image it has many applications like content based web search, logo detection in CCTV video feeds, sign detection, license plate reading. As visually impaired person cannot see it will be useful for them to access text and clustered with text to speech algorithm and make them to read cover of book, labels on door, medicine labels etc. hence caption localization and detection become important research now a days. Our main aim of the proposed system is to detect text in internet video, low quality images downloaded from internet or capturing devices[8]. Texts in natural scenes provide rich information for people.

Retrieval of text from video or image involve several steps:

1. Frame Extraction: Extract the frames from video provided by user.
2. Text Localization: Find out area where text is localized.
3. Text Segmentation: Segment vertical line and horizontal character on binary input image.
4. Text Recognition: Text recognition is carried out by thinning, training the text.
5. Optical Character Recognition(OCR): OCR having following main constraints as vertical line segmentation and horizontal character segmentation.

Recognizing video text information directly from video provides unique benefits are as:

- ➢ It is very useful for describing the contents of video sequence.
- ➢ It can be easily extracted and compared to other semantic contents.
- ➢ Extracted text is exactly synchronized with the image data when the event occurs.
- ➢ Manual logging may not be feasible for large collection of archived videos.
- ➢ It enables applications such as keyword-based image search, automatic video logging, and text-based image indexing.
- ➢ It is an important component for the automatic annotation, indexing and parsing of images and videos. For example the sports videos text information displays valuable game information such as scores and players. The valuable text information extracted from sports videos is called key captions text and they can be used for video highlights or content search.

In this paper we proposed a fast text localization method to localize all the possible edges and text in images. Optical character recognition is used to recognize the clustered get from the output of sobel edge detection algorithm [6].

II. PROPOSED METHOD

A. Text Localization

Text localization include several steps.

a) Preprocessing:

- Input image: input to the system is image containing text or .avi uncompressed video file. Java media framework which optional package of java platform takes the frames of playing video. Frames can be taken from 10 to 50 in numbers.
- Gray Scale Conversion: After taking frames or snapshot next is to convert to gray scale if the frames are color in nature for easy processing and saving tie for finding text in after processes. Color image is to convert into grayscale by retrieving each R,G and B contents from color image it can be retrieved by some formulae's like B=(RGB) AND 0XFF, G=(RGB>>8AND 0XFF, R=(RGB>>16) AND 0XFF.Once value of R,G and B is retrieved we can take average of these three constraint as R+G+B/3 and assign it to new pixel in image like these way we can convert color image to grayscale pixel by pixel value[9].

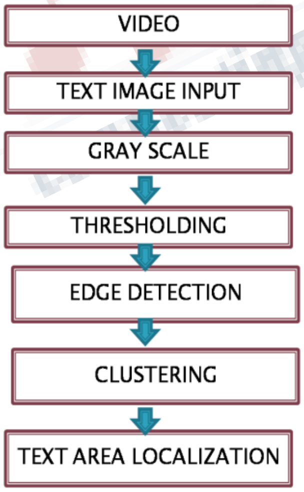


Fig1: Text Localization

•Thresholding: Gray scale image is now can be used for binary thresholding as it is next step in proposed system. Thresholding can be achieved by taking thresholding value user can set value in between range of pixels 0 to 255 and comparing it with the gray scale image. If we get less value of pixel than threshold value then assign it as "BLACK". and if get more value than threshold assign it as "WHITE" or vice versa. In this way we can get complete black and white image than gray scale image because it contains variable value of pixels [2].

b) Edge detection: Next step in pre-processing is edge detection by sobel operator. Input to the sobel edge detection algorithm id threshold image. Sobel operator takes threshold image treates each pixel as constraint of edge sobel operator find out the strength and direction of each pixel. Stregth of pixel is calculated on x and y axis.Magnitude of X axis can be calculate as Gx and Gy for y axis. Total magnitude can be calculated as $G=\sqrt{Gx}+\sqrt{Gy}$. Direction of edge pixel has to be calculating because character strokes lies in pairs and opposite pair as shown in fig.
$\Theta=\tan(Gy/Gx)$ Magnitue and direction are important in getting strong density of edge pixel[1].

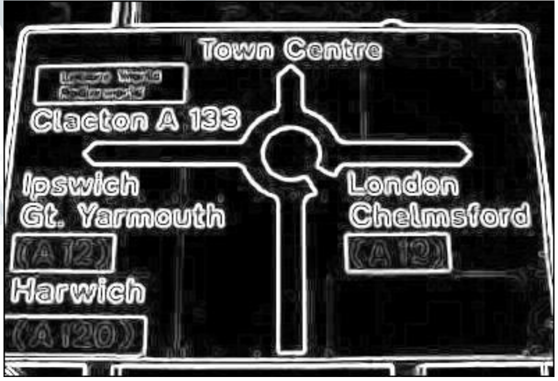


Fig2: Edge detection by using sobel operator

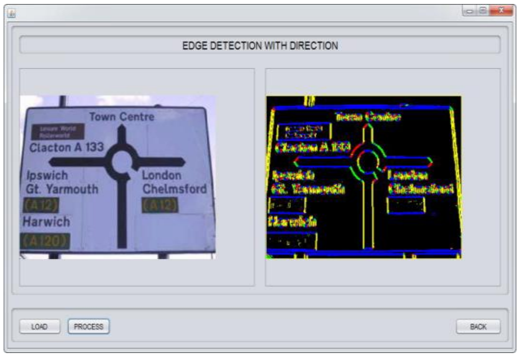


Fig3: Relation of edge direction with color.

c) Clustering: In this step edges from sobel operator is used as we get all the possible edges in the images from sobel edge detection algorithm likewise if there is some noise remain in the image before going it to the clustering then blurring can be done on getting clear edge pixels or some inclear images.Blurring can be on each pixel as matrices took of 3*3 or 5*5 is treated and average of R,G and b is assign to center pixel[1,4]. Frequency counting part of k-means clustering is used to localize the text in the output image of sobel edge detection algorithm. K-meas clustering approach.

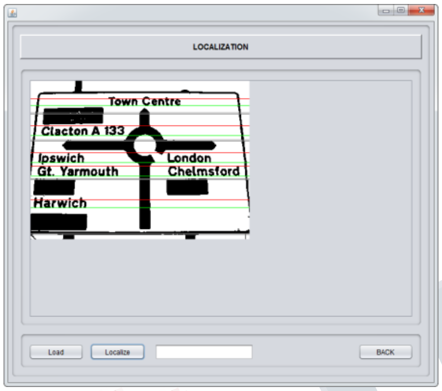


Fig4: Clustering(Text localization)

B. Text Recognition

All the possible edges are localized by using sobel edge detection algorithm and k-means clustering algorithm. Whatever the edged we have localized is that text,number or special character we have to recognized to get the proper output. For that reason we are applying the output of clustering to the optical character recognition (OCR).

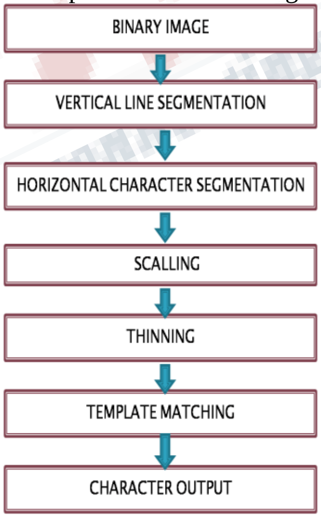


Fig5: Text Recognition by using Optical

OCR has several approaches as follows.

- Binary input image: Input to the optical character recognition is binary input which we get from ouput of hreshold image as in the total black and white form or we can use output image of clustering or sobel edge detection algorithm to recognize text in the caption.

- Thinning : Thinning can be applied to binary input image for easy process and time saving process in the optical character recognition.As thick characters can be complex in process and cannot easily process due to color contrast and background hence we need to thin it.Thinning can be done in three stages first stage is to check whether it satisfies the pixel range like 001,100,011,110.If the text or character provided by user contains the above column and row matix simultaneously then can move to second condition that pixel we are deleting or erasing should not be end point or it should not be connected to more than two pixel.If it satisfies above two condition we mark the pixel as erasable and after checking n point connectivity the pixel is deleted[5].

- Scaling: As all characters should be in random in size,shape and position.All has to be in one size it an be achieved by scalling it.Dirty function is use to scale the characters.

- Template matching: By using technique of template matching we can find small part of template image.Template match process can be used to navigate mobile robot and to detect edges. Template match works in two fashion like template based matching and feature based matching. Feature based matching technique used some of the feature of the images to match like edges and template match. If the source has the features then only it is used for feature based template otherwise template based matching is used. The basic method of template matching works in fashion that it uses gray scale or binary threshold image.All the black pixel which has value 1is consider for template matching. All the 1's compared to the templates in the stored database if the stored database contains same template value of 1's and 1's cane be assign as count+ and then can get same or approximate value which can be consider.

- Training: training is the important constraint in optical character recognition. As we get same behavioural pattern by template match technique we have to train the system what is the character generated by user to the character present in database. We can create same behavioral pastern and save it database or automatically we can give image containing text to retrieve and save text in database [7].

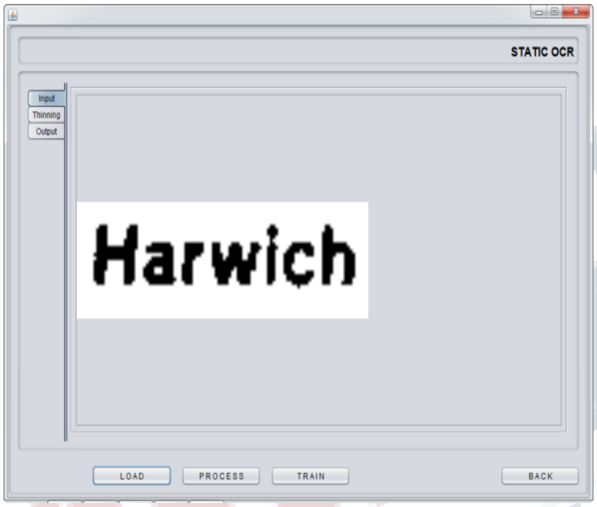


Fig6.1: Binary input image

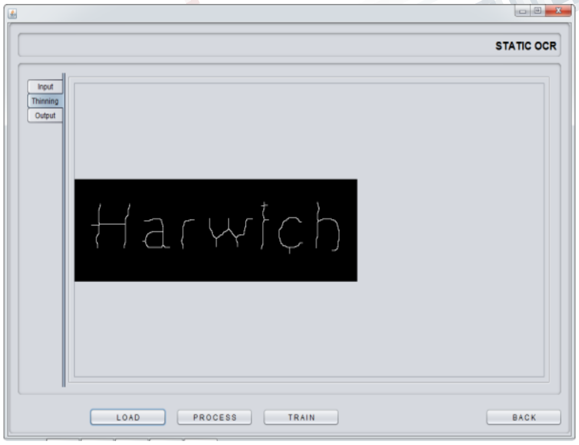
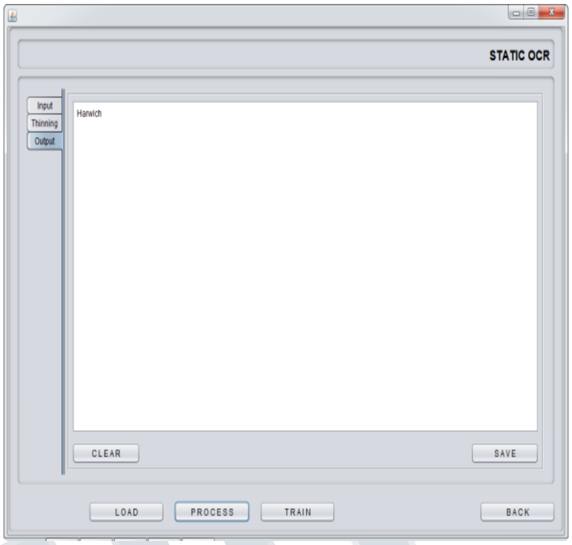


Fig6.2: Thinning and process



```
Harwich
```

Fig6.3: Final output

III. EXPERIMENTAL RESULTS

Experimental results carried out on some random images and one .avi video file is taken for localization of text.

We measure the localization performance on precision rate and recall rate
Precision={relevant document} ∩{retrieved document}/retrieved document
Recall={relevant document} ∩{retrieved document}/relevant document

CONCLUSION

This paper presented a fast text localization method to localize all the text in image and video. Experimental results show efficiency of proposed method to precision rate at 91.2% and recall rate at 94%.In future we will contribute to improve efficiency of our system.

REFERENCES

[1] Bo Bai, Fei Yin, Cheng-Lin Liu "A Fast Stroke-Based Method for Text Detection in Vide" IEEE 10th IAPR International Workshop on Document Analysis Systems,march 2012,pp 69-73. J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," IEEE Trans. Electron Devices, vol. ED-11, pp. 34-39, Jan. 1959.

[2] Khaled Alsabti,Sanjay Ranka,Vineet Singh "An Efficient K-Means Clustering Algorithm" .

[3] LuoB,Tang X, Liu J and Zhang H. "Video caption detection and extraction using temporal information" In ICIP'03 pages 297-300, 2003 Barcelona Spain, September2010.

[4] Cheolkon Jung and JoongkyuKim "A Novel Approach for Key Caption Detection in Golf Videos Using Color Patterns" ETRI Journal, Volume 30, Number 5, pp-750-753, October 2008.

[5] David Crandall, Sameer Antani, RangacharKasturi "Extraction of special effects caption text events from digital video" International journal on document analysis and recognition, Volume: 5, pp:138-157, 2003.

[6] Huiping Li Doermann, D. KiaO.,"Automatic text detection and tracking in digital video" ,IEEE Transactions on Image Processing, Vol.9, p.p. 147-156, January 2000.

[7] Cheolko Jun,Su Young Li, Joongk yu Kim, "Robust Detection Of Key Captions For Sports Video Understanding" IEEE International Conference on Image Processing (ICIP 2008), pp-2520-2524, 2008.

[8] Minhua Li, Meng Bai "Text Segmentation from Image with Complex Background Based on Markov Random Fields" IEEE 2012 International Conference on Computer Science and Electronics Engineering,pp. 486-489.

[9] Tianyi Gui, Jun Sun, Satoshi Naoi, Yutaka Katsuyama, Akihiro Minagawa "A Fast Caption Detection Method for Low Quality Video Images" IEEE 10th IAPR International Workshop on Document Analysis Systems,feb 2012,pp 302-307.

[10] A.Thilagavathy, K.Aarthi, A. Chilambuchelvan "A Hybrid Approach to Extract Scene Text from Video", Int Conference on Computing, Electronics and Electrical Technologies [ICCEET], 2012, pp 1017-1023.