International Journal for Research in Applied Science & Engineering Technology (IJRASET)

# Fusion based text detection and localization for video image

Amit Panwar[1], Himanshu Suyal[2]
*Department of Computer Science and Engineering, GB pant Engineering College, Ghurdhuri, Pauri Garhwal, INDIA*

***Abstract—Text detection is much easier in images but it becomes complex with dealing with video images. Text is important for video indexing summarization and information retrieval text detection involves localization, extraction and recognition from the video images.Different size, style, alignment also low contrast and complex background poses difficulty in automatic detection of the text. In this paper a novel approach is purposed for text detection. The proposed algorithm uses various existing edge based text detection algorithm. Finally fusion process is applied to extract the text from the input image. Proposed method was tested on ICDAR-2003 dataset.***
***Experiment result proves the effectiveness of proposed algorithm. The algorithm produces more significant result as compared to previous algorithms. Detection rate, midsection rate, false positive rate is used for accuracy assessment.***
***Index Terms—text detection, video, image, fusion***

## I. INTRODUCTION

Digital video now a day plays vital role in entertainment, multimedia, education and other applications. For content base multimedia database retrieval and indexing low level features such as color shape or texture is used. These features are easy to extract but they does not provide clear idea about content so more features such as text can be used for this purpose. Text detection is being investigated by researchers for content based indexing. Text in image or video provide important and brief content information such as name, location, date, title, etc. text detection can be broadly classified as connected component based (CC) and texture based.
Connected component based: this method uses one of the following feature:

*A.* Geometry : geometry include the font size, text alignment, shape
*B.* Color: color can vary in a single frame
*C.* Edge: edges are the high frequency component in an image.

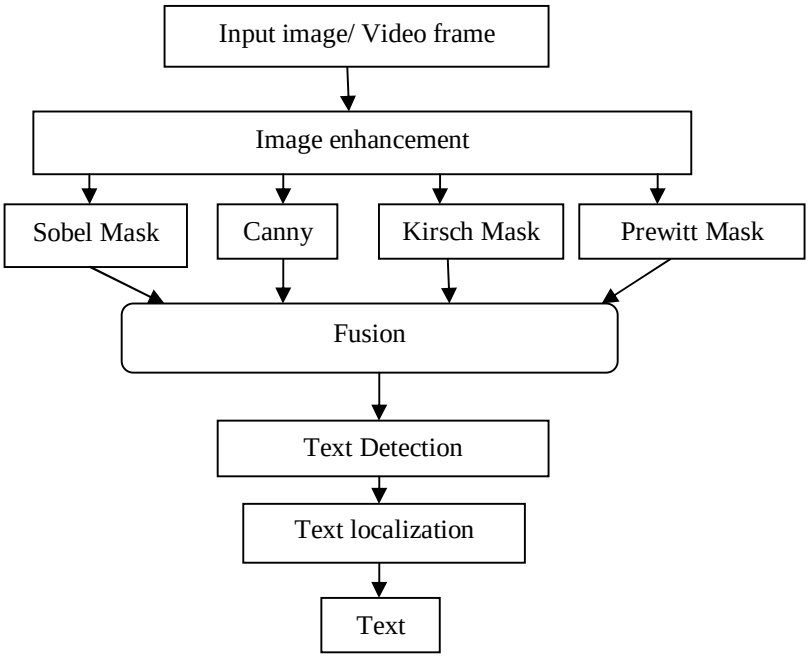
Texture based: these method uses certain texture based filters for text extraction. In this type of method, Text is treated as texture and whole image is divided in sub images.
Several methods has been purposed for text detection. These methods include edges, stroke, corner, texture distribution.The size of character can vary in single frame from small to large also the font can be different. This makes a text detection process difficult. In video the text can either be moving in a direction or still. Presence of special effect and complex makes it even more difficult. Thus existing methods do not fully exploit the presence of text.
In this paper a CC based technique is proposed. Followed by text localization. Text is extracted using fusion based algorithm. The process involves fusion of various edge based method.
Rest of the paper is organized as follows Section II describe the related work, Section III describe the detail discussion on the proposed methodology, Section IV disuses on the experimental results and Section V describe the conclusion and future work

## II. RELATED WORK

Liu, et.al [1] proposed an algorithm for extracting the text from the image and video frames which contain three steps, in first step contain the edge detection, second step contain text candidate detection and at the third step refinement text detection was done. Rong, Li, et.al[2] proposed a two level approach for text detection in natural scene images. Firstly connected component are extracted from input image after that connected text line are extracted to the horizontal or vertical position. Support vector machine (SVM) is used for classification. In the second stage conditional random field model is used for making the final decision. Anthimopoulos, et.al [3] proposed a two stage system for text detection in video images. In first stage text line is detect which is based on the edge map, and used high recall rate and low computation time and the second stage the results is refined using SVM and sliding window. Shivakumara, et.al [4] proposed a technique for extracting graphics text and scene text in video image. For

extracting the text author find the segments containing text in input image and then used statistical features. Shi, et.al [5] proposed a novel approach for scene text detection which used the graph based approach by using Maximally Stable Extremal Regions (MSERs) to incorporating various information sources within one system.Yang, et.al [6] proposed workflow which used fast localization verification schema is used, in which potential text candidate with high recall firstly identify by the multi-scale text detector then image entropy based filter refined detected text candidate line at the end Stroke Width Transfer (SWT) and Support Vector Machine (SVM) applied to eliminate the false alarms. Optical Character Recognition (OCR) is used to separate the text from the complex background.

## III. PROPOSED METHODOLOGY




Flowchart of the proposed method

### *A. Image enhancement*

Image enhancement is used as a preprocessing step. Contrast enhancement is performed to increase quality of the image for further processing. Contrast enhancement increases the perceptibility of the input image. Contrast enhancement enhances the brightness gap between objects and their background. Stretching is performed on the input image in the proposed algorithm. Stretching improves the brightness difference uniformly across the image.

### *B. Edge Detection Algorithms*

Edges are the high frequency component in image. Various edge detection algorithms has been proposed among them. In this paper Robert, Prewitt, canny, kirsch and Sobel edge detection algorithms are used. Diagram 1 shows the mask of the used technique. Canny edge detector is best among the all discussed here. Edges are created by objects which have different local intensity profile due to texture shadow and occlusion.

### *C. Text localization*

Text localization is the process of locating text in the image. This process gives the location of the text from the input image. Text localization process can be classified as region based and texture based. Region based method can be further subdivided in to edge based and connected component based technique. In this paper fusion based edge localization method is proposed which work in bottom up fashion. First identify sub-structures (edges) secondly, merging these sub structure to mark boundaries boxes for text.

| | | |
|---|---|---|
| -1 | 0 | 1 |
| -1 | 0 | 1 |
| -1 | 0 | 1 |

Figure 1Sobel Mask

| | | |
|---|---|---|
| -1 | 0 | 1 |
| -2 | 0 | 2 |
| -1 | 0 | 1 |

Figure 2Prewitt Mask

| | | |
|---|---|---|
| -3 | -3 | 5 |
| -3 | 0 | 5 |
| -3 | -3 | 5 |

Figure 3 Kirsch Mask

## IV. PROPOSED ALGORITHM

*A.* Input image I, convert it into greyscale G.
*B.* Apply contrast stretching to the image G.
*C.* Fuse individual results from the above step using PCA fusion.
*D.* Mark boundaries boxes for text.
*E.* Evaluate efficiency using
*F.* Apply Robert, Prewitt, canny, kirsch and Sobel mask to image G resulting in image F.

## V. EXPERIMENTAL RESULT

ICDAR 2003 test set contains 258 real scene picture. Also some random images from web are used as dataset so that dataset cover's a variety of text in different size, style, font, light text on complex background and text of poor qualities, etc. following categories has been defined for each detected block.


Figure 4 four image with detection result

Truly Detected Block (TDB) – A block that contains a text line, fully or partially.
False detected block (FDB) – A block that does not contain text.
Text block with missing data (MDB) – A detected block that misses some character of a text line.
Ground truth data is collected manually i.e. the actual text blocks (ATB). Performance of the proposed method is evaluated on the following rates. Table 1 shows the three detection rate of the proposed and existing methodology when applied on dataset.
Detection rate (DR) – The ratio of truly detected text blocks and the actual text blocks.
False positive rate (FPR) – The ratio of false detected block and total number of blocks.
Midsection rate (MDR) – the ratio of text block with missing text and truly detected blocks.

| Method | ATB | TDB | FDB | MDB |
|---|---|---|---|---|
| Edge based [7] | 491 | 393 | 86 | 79 |
| Gradient based [8] | 491 | 349 | 48 | 35 |
| Uniform colored [9] | 491 | 252 | 95 | 94 |
| Proposed | 491 | 458 | 39 | 55 |

Table 2 shows the performance of the existing algorithms and the proposed method on the dataset. The proposed method works better as compared to traditional edge based method. It has been observed the proposed methodology does not fit for images with complex background. Presence of many features in the background decreases the efficiency of the proposed algorithm.

| Method | DR | FPR | MDR |
|---|---|---|---|
| Edge based [7] | 80.0 | 18.0 | 20.1 |
| Gradient based [8] | 71.1 | 12.1 | 10.0 |
| Uniform-colored [9] | 51.3 | 27.4 | 37.3 |
| Proposed | 93.3 | 7.9 | 12.0 |

## VI. CONCLUSION AND FUTURE WORK

An efficient method for text detection based on fusion of edge detection mask has been proposed. Presence of text in image is marked by sharp transition as compared to background. Also fusion process displays better result as compared to simple approach.
In the future we plan to improve this method for complex background images using thresholding, it is the simplest method of segmentation. We believe that thresholding can improve results. This method is particularly used for English language. A multilingual method for text detection can be future aspect of this proposed approach.

## REFERENCES

[1] Liu, Chunmei, Chunheng Wang, and Ruwei Dai. "Text detection in images based on unsupervised classification of edge-based features." Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on. IEEE, 2005.
[2] Rong, Li; Suyu, Wang; Shi, Zhixin, "A Two Level Algorithm for Text Detection in Natural Scene Images," Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on , vol., no., pp.329,333, 7-10 April 2014
[3] Anthimopoulos, Marios, Basilis Gatos, and IoannisPratikakis. "A two-stage scheme for text detection in video images." Image and Vision Computing 28.9 (2010): 1413-1426.
[4] Shivakumara, Palaiahnakote, TrungQuyPhan, and Chew Lim Tan. "A laplacian approach to multi-oriented text detection in video." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33.2 (2011): 412-419.
[5] Shi, Cunzhao, et al. "Scene text detection using graph model built upon maximally stable extremal regions." Pattern recognition letters 34.2 (2013): 107-116.
[6] Yang, Haojin, Bernhard Quehl, and Harald Sack. "A framework for improved video text detection and recognition." Multimedia Tools and Applications 69.1 (2014): 217-245.
[7] C. Liu, C. Wang and R. Dai, “Text Detection in ImagesBased on Unsupervised Classification of Edge-basedFeatures”, ICDAR 2005, pp. 610-614.
[8] E. K. Wong and M. Chen, “A new robust algorithm forvideo text extraction”, Pattern Recognition 36, 2003, pp.1397-1406.
[9] Y. Mariano and R. Kasturi, “Locating Uniform-Colored Text in Video Frames”, 15th ICPR, Volume 4, 2000,pp 539-542.