Video-Text Detection by Implementing a Laplacian Operator

[1] Bharath Vyas B [2] Prof. Narendra Kumar
[1]UG Scholar [2] Assistant Professor,
[1]Dayananda Sagar College of Engineering, Bangalore, India [2] Department of Electronics and Communication
R.N.S. Institute of Technology, Bangalore, India
[1]9595vyas@gmail.com [2] nkrnsit@gmail.com

Abstract: **Text is an important form of information. Any information in the form of text present in a document-image or video, is difficult to be modified, if the text is static in it. Hence, modification or analysis of a text is possible only by separating it from a document-image or a video. This project deals with an efficient method of isolating text present in a video, by using a Laplacian operator. The document image is convolved with Laplacian operator to highlight text regions present in the document image. The pixels related only to text (textual pixels) need to be stressed upon, and isolated from other non-textual pixels. This is achieved by computing a gradient-difference amongst the neighborhood pixels. Clusters of texts are required, in order to differentiate between textual and non-textual pixels. To identify text cluster from non-text cluster, the mean of both the cluster is computed, by employing K-means clustering technique. When this method is employed, it results in the mean of the text cluster possessing a higher value than that of a non-text cluster. In order to prune the text data contained in the identified text blocks, for each candidate text block, the corresponding region in the Sobel edge map of the input image undergoes projection profile analysis to determine the boundary of the text blocks. At the end of the process, false positives regarding the textual clusters are removed by employing geometrical properties-based empirical rules. Experimental results on the standard document image database collected from ICDAR-2003 dataset show that the Laplacian operator based text detection method is able to detect text of different fonts, contrast and backgrounds.**

Index Terms**—Laplacian operator, Sobel Edge, Clustering**

I. INTRODUCTION

There is huge quantity of video databases on the Internet, by using the text present in those videos we can extract or retrieve those video. Video text consists of two types: graphic text and scene text. Graphic text is artificially added to the video during the editing process. Scene text appears naturally in the scenes captured by the camera. Although many methods have been proposed over the past years, text detection is still a challenging problem because videos often have low resolution and complex backgrounds and text can be of different sizes, styles and alignments. In addition, scene text is usually affected by lighting conditions and perspective distortions.

Various methods were proposed for the isolation of text data from the documented image. Among which text extraction based on laplacian method have been widely used as effective tool in text segmentation. This project work implements an efficient text isolation algorithm for the extraction of text data from the documented video clips. The proposed task implements neural network for the recognition of text characters from the isolated text image for making it editable. The proposed work is implemented using mat lab tool.
Many efforts have been made earlier to address the problems of text area detection, text segmentation and text recognition. Current text detection approaches can be classified into three categories

Connected component-based [2], edge based [1] and texture-based [3]. The first category is connected component-based method, which can locate text quickly but have difficulties when text is embedded in complex background or touches other graphical objects and this approach does not work well for all video images because it assumes that text pixels in the same region have similar colors or grayscale intensities. The second category is texture-based, which is hard to find accurate boundaries of text areas and usually yields many false alarms in “text-like” background texture areas and this second approach requires text to have a reasonably high contrast to the background in order to detect the edges. The third category is edge-based method. Generally, analyzing the projection profiles of edge intensity maps can decompose text regions and can efficiently predict the text data from a given video image clip.

There is a considerable amount of text occurring in video that is a useful source of information, which can be used to improve the indexing of video. The presence of text in a scene, to some extent, naturally describes its content. If this text information can be harnessed, it can be used along with the temporal segmentation methods to provide a much truer form of content-based access to the video data. text detection and recognition in videos can help a lot in video content analysis and understanding, since text can provide concise and direct description of the stories presented in the videos. In digital news videos, the superimposed captions usually present the involved person's name and the summary of the news event. Hence, the recognized text can become a part of index in a video retrieval system.

Here we consider three existing methods [7, 8] for comparative study. Liu et al. [7] extract edge features by using the sobel operator. This method is able to determine the accurate boundary of each text block. However, it is sensitive to the threshold values for edge detection. Wong et al. [8] compute the maximum gradient difference values to identify candidate text regions. This method has a low false positive rate but uses many threshold values and heuristic rules. Therefore, it may only work well for specific datasets. Finally, mariano et al. [4] perform clustering in the l*a*b* color space to locate uniform-colored text. Although it is good at detecting low contrast text and scene text, this method is extremely slow and produces many false positives.

II. PROPOSED SYSTEM

In this paper, text detection method which consists of three steps: text detection, boundary refinement and false positive elimination. In the first step, by using the laplacan operator text region is identified. The second step uses sobel edge operator to determine the accurate boundary of each text block which we got from the first stage. Finally, false positives are removed based on geometrical properties. Experimental results show that the proposed method outperforms the above three methods in terms of detection and false positive rates.

Following diagram shows flow chart of proposed method

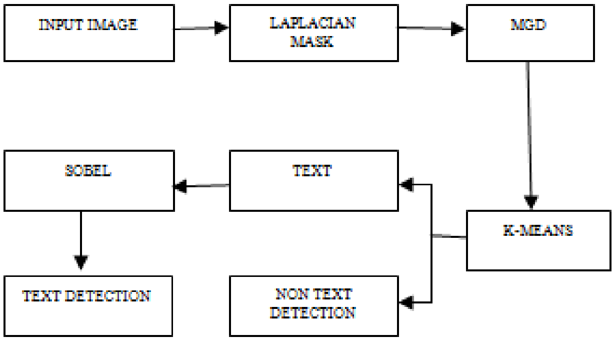

Fig. 1. Flow diagram

A. Text detection

Text regions will be usually having huge quantity of discontinuities compared to image content. In order to make processing easy, input image is converted to grayscale and filtered by a 3 × 3 laplacian mask to detect the discontinuities in four directions: horizontal, vertical, up-left and up-right. The following figure shows 3*3 laplacian mask.

1	1	1
1	-8	1
1	1	1

*Fig. 2. 3*3 laplacian mask*

Because the mask produces two values for every edge, the laplacian-filtered image contains both positive and negative values. The transitions between these values (the zero crossings) correspond to the transitions between text and background. In order to capture the relationship between positive and negative values, we use the maximum gradient difference (MGD), defined as the difference between the maximum and minimum values within a local 1 × n window.
The MGD value at pixel (i, j) is computed from the laplacian-filtered image f as follows.

$$MGD\,(i, j) = \max\,(f\,(i, j + t)) - \min\,(f\,(i, j + t)) \qquad (1)$$

where t € [-(n-1)/2, (n-1)/2]

The MGD map is obtained by moving the window over the image. Text regions typically have larger MGD values compared to non-text regions because they have many positive and negative peaks. Once we obtain MGD map normalization is used so that all pixel values are converted into a value between 0 and 1 where 1 will be maximum and use k-means to classify all the pixels into two clusters, text and non text, based on the Euclidean distance or city block distance between MGD sample profiles of text and non-text regions. Each graph shows the positive and negative values of the middle row of the corresponding laplacian-filtered image (not shown here) values. Let the two clusters returned by k-means be cl1 (cluster mean mean1) and cl2 (cluster mean mean2). Since the cluster order varies for different runs, we have the following rule to identify the text cluster. If mean1 > mean2, cl1 is the text cluster; otherwise, cl2 is the text cluster. This is because it is expected that text regions have larger MGD values than non-text regions. At the end of this step, each connected component in the text cluster is a candidate text region.

B. Boundary refinement

It is bit difficult to locate exact boundary from the text region because of false positive. In order to overcome this BINAY SOBEL edge map is computed for the horizontal text region, it will be denoted as horizontal projection profile analysis.

$$Hp\ (i) = \sum SM\ (i, j) \qquad (2)$$

If hp (i) is greater than a certain threshold (threshold will be decided by going through image proparties) row i(pixel under process and it is horizontal pixel) is part of a text line; otherwise, it is part of the gap between different text lines. From this rule, we can determine the top row i1 and bottom row i2 of each text line. Line. The vertical projection profile is then defined as follows.

$$VP\ (J) = \sum SM\ (I, J) \qquad (3)$$

Here also same rule we will use if vp(j) is greater than some pre defined text it is considered as part of text line else it will be considered as gap between the charater or the word present in the image.

C. False positive elimination

We eliminate false positives based on geometrical properties. Let w, h, ar, a and ea be the width, height, aspect ratio, area and edge area of text block.

$$Ar = w / h \qquad (4)$$

$$A = w \times h \qquad (5)$$

$$Ea = \sum sm\ (i, j) \qquad (6)$$

If ar < t1 or ea / a < t2, the candidate text block is considered as a false positive; otherwise, it is accepted as a text block. The first rule checks whether the aspect ratio is below a certain threshold. The second rule assumes that a text block has a high edge density due to the transitions between text and background.

III. CLUSTERING

Partitioning clustering approach construct. A typical clustering analysis approach via partitioning data set iteratively It construct a partition of a data set to produce several non-empty clusters (usually, the number of clusters given in advance) in principle, partitions achieved via minimizing the sum of squared distance in each cluster. Given a k, find a partition of k clusters to optimize the chosen partitioning criterion. Global optimal: exhaustively enumerate all partitions. K-means algorithm (macqueen'67): each cluster is represented by the centre of the cluster and the algorithm converges to stable centers of clusters. "k" stands for number of clusters; it is a user input to the algorithm. From a set of data points or observations (all numerical), k-means attempts to classify them into k clusters. The algorithm is iterative in nature.

Algorithm k-means
1. Decide on a value for k.
2. Initialize the k cluster centers (randomly, if necessary).
3. Decide the class memberships of the n objects by assigning them to the nearest cluster center.
4.re-estimate the k cluster centers, by assuming the memberships found above are correct.
5. If none of the n objects changed membership in the last iteration, exit. Otherwise go to step 4

IV. EXPERIMENTAL RESULT

As there is no standard dataset available, we have selected 101 video images, extracted from news programmers, sports videos and movie clips, for our own dataset. There are both graphic text and scene text of different languages, e.g. English, chinese and korean, in the dataset.

Following snap shot shows both input and output images.

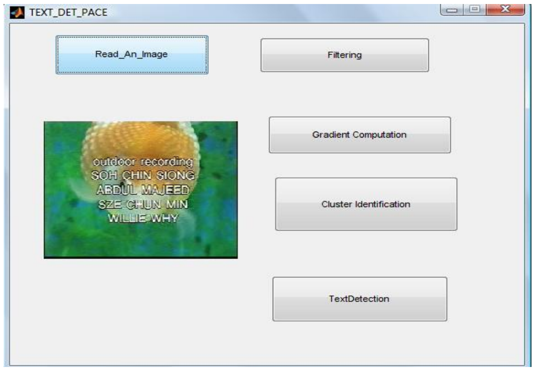

Fig 3. Input image

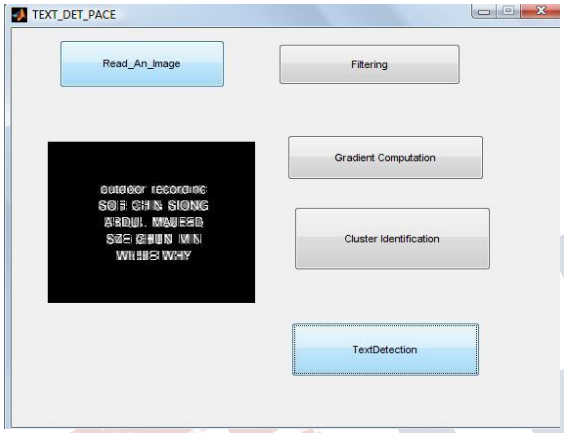

Fig 4. Output image

V. CONCLUSION

Text data present in images and video contain useful information for automatic annotation, indexing, and structuring of images. Extraction of this information involves detection, localization, tracking, extraction, enhancement, and recognition of the text from a given image. However, variations of text due to differences in size, style, orientation, and alignment, as well as low image contrast and complex background make the problem of automatic text extraction extremely challenging. In this context, we have developed an efficient method for text detection based on the Laplacian operator. The gradient information helps to identify the candidate text regions and the edge information serves to determine the accurate boundary of each text block. Experimental results show that the proposed method is capable of detecting text blocks accurately.

There has already been a lot of work done on vehicle license plate and container plate recognition. Although container and vehicle license plates share many characteristics with scene text, many assumptions have been made regarding the image acquisition process (camera and vehicle position and direction, illumination, character types, and color) and geometric attributes of the text. Currently, the text detection step can show white patches even for non-horizontal text however, the refinement step is only able to detect the boundary for horizontal text because of the use of horizontal and vertical projection profiles. Our future work will focus on extending this method to detect text blocks having arbitrary orientation.

REFERENCES

[1] J. Zang and R. Kasturi, "Extraction of Text Objects in Video Documents: Recent Progress" The Eighth IAPR Workshop on Document Analysis Systems (DAS2008), Nara, Japan, September 2008, pp 5-17.

[2] J. Zhang, D. Goldgof and R. Kasturi, "A New Edge-Based Text Verification Approach for Video", ICPR, December 2008, pp 1-4.

[3] K. Jung, K.I. Kim and A.K. Jain, "Text information extraction in images and video: a survey", Pattern Recognition, 37, 2004, pp. 977-997.

[4] A.K. Jain and B. Yu, "Automatic Text Location in Images and Video Frames", Pattern Recognition, Vol. 31(12), 1998, pp. 2055-2076.

[5] M. Anthimopoulos, B. Gatos and I. Pratikakis, "A Hybrid System for Text Detection in Video Frames", The Eighth IAPR Workshop on Document Analysis Systems (DAS2008), Nara, Japan, September 2008, pp 286-293.

[6] M. R. Lyu, J. Song and M. Cai, "A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 2, February 2005, pp 243-255.

[7] C. Liu, C.Wang and R. Dai, "Text Detection in Images Based on Unsupervised Classification of Edge-based Features", ICDAR 2005, pp. 610-614.

[8] E. K. Wong and M. Chen, "A new robust algorithm for video text extraction", Pattern Recognition 36, 2003, pp. 1397-1406.

[9] P. Shivakumara, W. Huang and C. L. Tan, "An Efficient Edge based Technique for Text Detection in Video Frames", The Eighth IAPR Workshop on Document Analysis Systems (DAS2008), Nara, Japan, September 2008, pp 307-314.

[10] Trung Quy Phan, Palaiahnakote Shivakumara, Chew Lim Tan, “A Laplacian method for video text detection” 2009 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, July 2009, pp 66-70.

IFERP
connecting engineers... developing research