International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

An Efficient Technique for Scene Text Extraction from Videos

C. Geetha, A. Thilagavathy, Sudesh Nimmagadda, Ravilla Madhusudhan, Saivignesh Chiruvella

Abstract: ***Text Detection from natural scene images and videos is imperative for applications in real world domain analysis. However the text detection process isperplexingbecause of exigent scenarios that the text exhibit. The information present in the video is either perceptual or it is either in semantic form. Amongst the different content that exists in the video, the text information is a major important content that describes more about the nature of the video. The text present in the video can be categorized into Caption text and Scene text. The caption text is the artificial text that is easy to detect while scene text are natural text which is difficult to identify. In this paper text extraction in natural images by edge based method is implemented. The algorithms are estimated with a set of images of natural scenes that differ alongside the scope of font size, illumination, scale and text direction. Precision, accuracy and recall rates are determined to evaluate the performance. The proposed system worked for all difficult scenarios of varied text and gave better results than the existing methods.***

Keywords: caption text, scene text, text detection, edge based method.

I. INTRODUCTION

Image formats such as TIFF and PDF have seen increased use throughout the years as electronic documents have become more prevalent. In general electronic documents are easier and quicker to use than paper documents. However to make files like PDF fully efficient there is a need to extract text from images using OCR software. As computer technology grows rapidly with advanced features; digital video has become one of the most important in many applications such as education, news and games. Multimedia data usage is also increasing day by day. The text can be posted in an image or video in different kinds of formats. Due to these factors, the predicament of detecting the presence of text develop into a demanding and difficult task. These texts have high level features of the image. Much useful information can be retrieved from this extraction of text like the sequence of video can be understood; searching process can be made easy by using the extracted text as keywords, for classification. The text present in the video can be categorized into Caption text and Scene text. The caption text is the artificial text that is easy to detect while scene text are natural text which is difficult to identify. Detecting the scene text becomes very intricate than the caption text owing to the different formats of the text. In [1],image pyramid is created and text is localized using binarization technique. The method works for multilingual text as well. Jayshree Ghorpade et al [2] implemented a text extraction strategy using java. The system used only static type of text. Michael R.Lyu et al [3] used a robust text extraction method that works for multiple languages. It also dealt with complex background scenarios and various text fonts and styles. A multilayer feed forward network was proposed in [4]. Chucai Yi et al [5] proposed a method for detecting text that involves image partition and candidate grouping. Hybrid approaches in text detection is dealt in [6, 7]. Text detection using soft computing methods are discussed in [8, 9]. Palaihnakote Shivakumara et al [10] proposed a novel approach that mixed the laplacian and sobel edge detection to trace out the text. In [11] text blocks were detected using ANN method.

A. Connected Component method

This method can be used in digital images to detect the connected regions. The text in images is considered as homogeneous regions. For the division of images the splitting and merging technique can be used. If a non-homogeneous region is present it is split into four parts. If two adjacent parts are homogeneous they can be combined. Then some unwanted regions can be deleted after checking for some character rules. Then it is applied to all frames and then redundancies in frames are removed and the text is extracted.

B. Edge based method

Text is having high contrast with the background of an image. So the edges in an image can be detected and the x direction and y direction edges are computed and a threshold value is set. If the computed value is larger compared to the threshold value the region is regarded as text. Then binarization process is done and text is represented in white color against a black background. Then redundancies are removed then the common regions in all frames are chosen and text is extracted.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
Dr. C.Geetha*, Associate Professor, Department of CSE, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India. Email: cga.cse@rmkec.ac.in
A.Thilagavathy, Associate Professor, Department of CSE, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India. Email: atv.cse@rmkec.ac.in
Sudesh Nimmagadda, Student, IV Year CSE, Department of CSE, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India. Email: sudeshnimmagadda@gmail.com
Ravilla Madhusudhan , Student, IV Year CSE, Department of CSE, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India. Email:madhusudhanravilla@gmail.com
Saivignesh Chiruvella, Student, IV Year CSE, Department of CSE, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India. Email:csaivignesh@gmail.com

II. PROPOSED METHOD

In this proposed method the text can be extracted from the image in an efficient way by using the edge detection algorithm. This gives clear extraction of text alone from a complex background. Figure 1 represents the overall workflow of the system.

A. Splitting of Frames

The input to the proposed system is video. As a preprocessing step, the video is first converted to frames. For the conversion of frames from video, the Matlab tool is used. The video is run and from the running video, the frames are retrieved. The contents in the successive frames will be of constant information. Hence it gives the illusion of same picture in all frames.

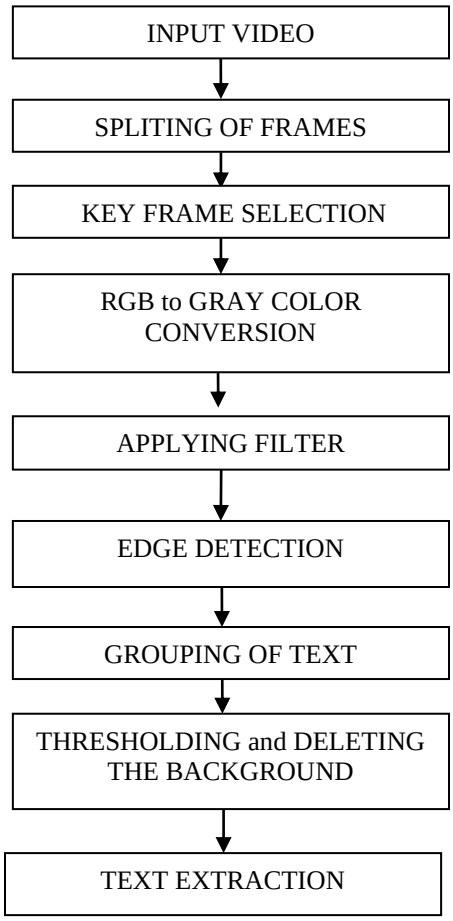


Figure 1. Overall Work Flow

B. Key Frame Selection

Locating a useful frame amongst a series of frames leads to the key frame detection. A mean color histogram of all frames is calculated. The nearby frames with histogram equableness are considered the key-frame. Hence the key-frame is chosen from the split frames of input video. This image is used for further processing. Figure 2 represents the Key frame selected from input video

Figure 2. Input RGB Image

C. Conversion of RGB to Gray Scale Image

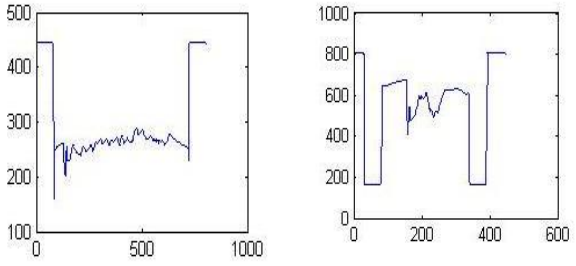
The pixel of a gray scale image lies between 0 to 255 bytes. So the image is sized between 0-255.The RGB image is transformed to gray scale image by the process of plane conversion. Now the image is in one plane. Figure 3 represents the gray scale image after plane conversion.

Figure 3. Gray Scale Image

D. Applying Filters

In this phase, the gray scale image is undergoing filtering and noise reduction process. The noise reduction in images is explored using linear and nonlinear techniques. In this paper an ordinary median filter is used. After the filtration process histogram of horizontal and vertical projections are plotted. Figure 4 represents the filtered image. Figure 5a and b represents the histogram plot of horizontal and vertical projection of filtered image.

Figure 4 Filtered Image



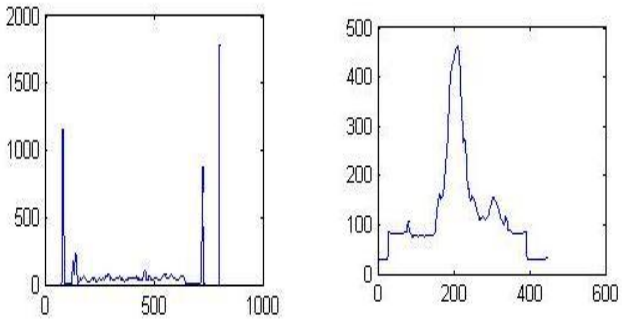
(a) (b)

Figure 5.(a) Horizontal and (b) Vertical projection of Filtered Image

E. Edge Detection

Edge detection remains the basis for the processing of images. An Edge is a sudden change in the pixel intensity of image. It has important features of an image. This sets a separation between the object and background. Edge detection aims at recognizing pointed abstractions of an image. Here Sobel Edge Detector is used to detect the edges of the image and also filtration is done to smoothen the image.The Sobel operator carry out a 2-D spatial slope dimension of an image. Then finds the approximate absolute gradient magnitude at each point in an input grayscale image using the formula **|G| = |Gx| + |Gy|.** Figure 6 represents the edge detected image. Figure 7a and b represents histogram plot of horizontal and vertical projection of Edge image.

Figure 6. Edge Detected Image



(a) **(b)**

Figure 7(a) Horizontal and (b) Vertical projection of Edge Image

F. Grouping of Text and Thresholding

Text is grouped and then Thresholding is done. So based on the histogram plot of the edge detected image the threshold value is chosen. The region that has greater value than the threshold value is considered as text region. Thus text is extracted from the image and displayed in white color with black background. Figure 8 represents the Segmentation of text area and Figure 9 represents the area containing text in the image.

Figure 8. Segmentation of Text

Figure 9 Threshold Image

III. EXPERIMENTS AND RESULTS

The proposed methodology is carried out for various videos and the text extraction results are analysed. The results after each step are shown below. The images represent: (a) Original Frame (b) Gray Image (c) Image after filtering (d)Filtered image horizontal projection (e) filtered image vertical projection (f) Image after edge detection (g) Edge image horizontal projection (h) Edge image vertical projection (i) Segmentation of text (j) Text Extracted Image.

IV. PERFORMANCE ANALYSIS

The performance of the proposed method is evaluated by using precision and recall rates. Recall rates will detect false negatives i.e. the text regions which are not detected. Precision rate will detect the false positives

Image Name	Precision Rate (%)	Recall Rate (%)
Colgate	87.5	63.63
KitKat	75	66.66
Number Plate	40	100
Book	87.09	67.5
T.Shirt	75	69.23

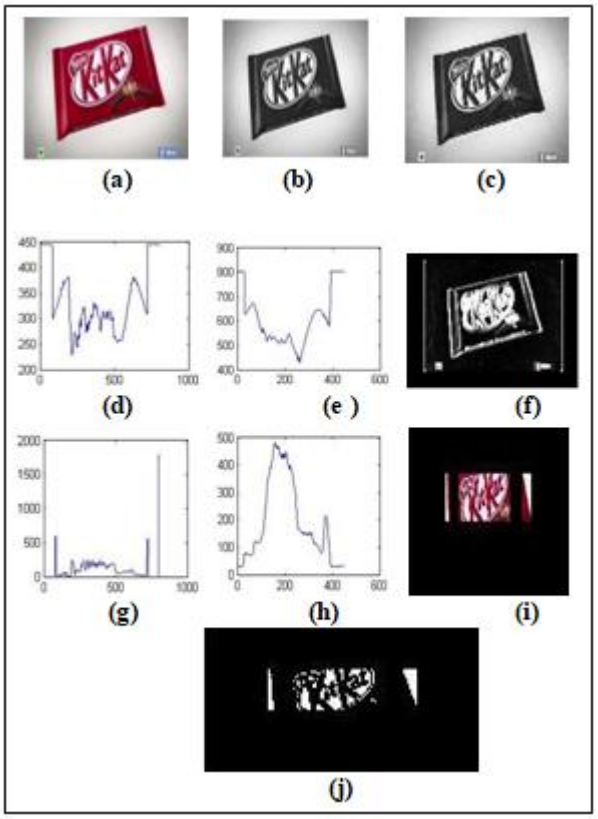


Figure 10 : Output screenshot

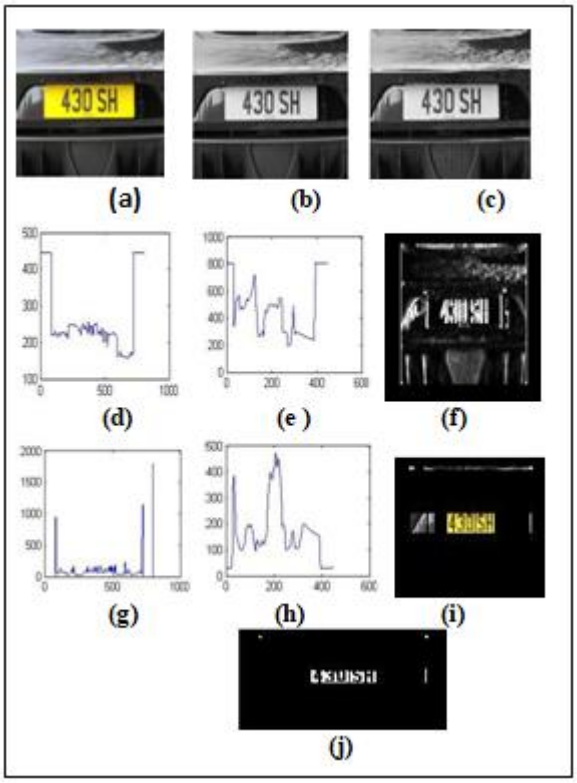


Figure 11 : Output screenshot

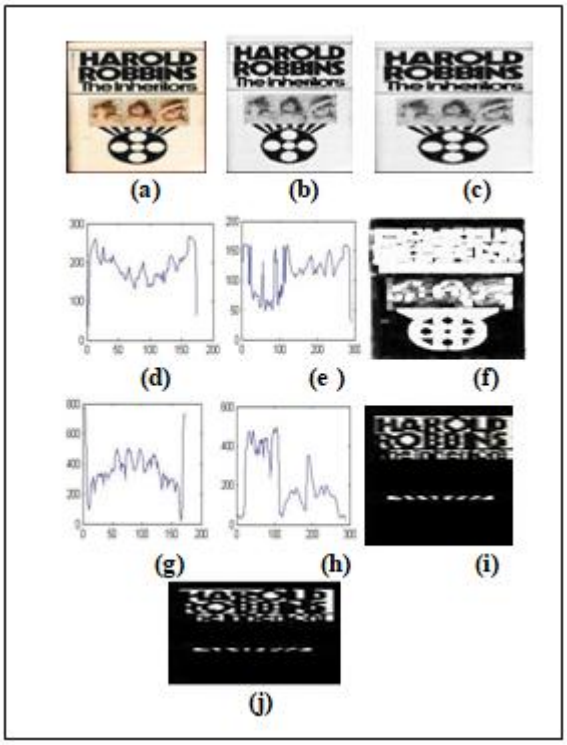


Figure 12 : Output screenshot

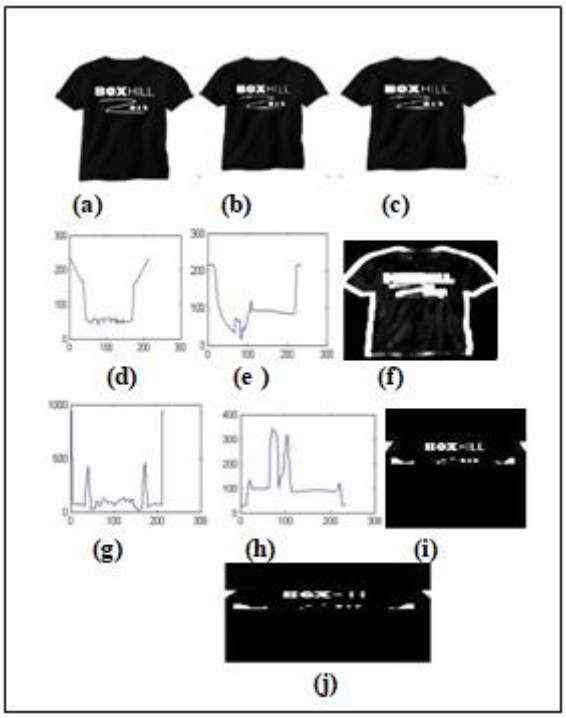


Figure 13: Output Screenshot

Figure 10 – 13 (a) Original Frame (b) Gray Image (c) Image after filtering (d)Filtered image horizontal projection (e) filtered image vertical projection (f) Image after edge detection (g) Edge image horizontal projection (h) Edge image vertical projection (i) Segmentation of text (j) Text Extracted Image.

V. CONCLUSION AND FUTURE WORK

Thus the proposed method is with varied variety of images and also performance is analyzed with precision and recall rates. This gives a clear detection of text from an image by ignoring the unwanted non-text area and also background. This work can be extended by using different edge detection and noise removal algorithms.

REFERENCES

1. Yi-Feng Pan, XinwenHou, and Cheng-Lin proposed, "A Hybrid Approach to Detect and Localize texts in Natural Scene Images", IEEE Transaction on Image Processing, vol. 20, no. 3, March 2011.
2. Jayshree Ghorpade, Raviraj Palvankar, Ajinkya Patankar and Snehal Rathi, "Extracting text from video", Signal & Image Processing: An International Journal (SIPIJ) vol.2, no.2, June 2011.
3. Michael R. Lyu, JiqiangSong and Min 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, 2005.
4. Rainer Lienhart and Axel Wernicke, "Localizing and Segmenting Text in Images and Videos", IEEE transactions on circuits and systems for video technology, vol. 12, no. 4, April 2002.
5. Chucai Yi and YingLi Tian, "Text String Detection from Natural Scenes by Structure-based Partition and Grouping", IEEE Transactions on Image Processing , vol.20 ,Sept. 2011
6. A Thilagavathy, K Aarthi, A Chilambuchelvan, "A hybrid approach to extract scene text from videos" , "International Conference on Computing, Electronics and Electrical Technologies (ICCEET), 2012, pp:1017-1022.

7. Thilagavathy.A, Aarthi.K, Chilambuchelvan. A, "Text Extraction from videos using a hybrid approach", "International Conference on Advances in Computing, Communications and Informatics (ICACCI-2012) " , pp: 193 – 199.
8. Thilagavathy.A , Aarthi.K, Chilambuchelvan. A, "Text Detection and Extraction from videos using ANN Based Network", "International Journal on Soft Computing, Artificial Intelligence and Applications(IJSCAI)" , Vol 1, Aug 2012, pp: 19-28.
9. Thilagavathy, A. & Chilambuchelvan, A.. (2017). Fuzzy based edge enhanced text detection algorithm using MSER. Cluster Computing. 10.1007/s10586-017-1448-5.
10. Palaiahnakote Shivakumara, Rushi Padhuman Sreedhar, TrungQuy Phan, Shijian Lu and Chew Lim Tan, "Multi-Oriented Video SceText Detection through Bayesian Classification and Boundary Growing", IEEE Transactions on Circuits and Systems for Video Technology,vol. 22, Aug 2012.
11. Huiping Li, David Doermann, and Omid Kia , "Automatic Text Detection and Tracking in Digital Video", IEEE transactions on image processing, vol. 9, no. 1, January 2000.