MANTECH Publications

Journal of Embedded Systems & Its Applications
Volume 2, Issue 1, January-June 2017

Embedded Technology Based Effective and Efficient Video Text Extraction

Sangita Wamane
Assistant Professor
Department of E &T.C,
P.D.V.V.P.COE, Ahmednagar, Maharashtra, India
Corresponding Author: *wamanesangita6@gmail.com*

Abstract

The rapid growth of video data leads to an urgent demand for efficient and true content-based browsing and retrieving systems. In response to such needs, various video content analysis schemes using one or a combination of image, audio, and text information in videos have been proposed to parse, index, or abstract massive amount of data text in video is a very compact and accurate clue for video indexing and summarization. Most video text detection and extraction methods hold assumptions on text color, background contrast, and font style. Moreover, few methods can handle multilingual text well since different languages may have quite different appearances.

In this paper, an efficient overlay text detection and extraction method is implemented which deals with complex backgrounds. Based on our observation that there exist transient colors between inserted text and its adjacent background. It is robust with respect to font size, style text, color, orientation and noise and can be used in a large variety of application fields such as mobile robot navigation vehicle license detection and recognition, object identification , document retrieving, etc. System is implemented using ARM 7 microcontroller. Extracted text is finally displayed on LCD.

Keywords: *Optical character recognition (OCR), overlay text, transition map, video summarization, video information retrieval.*

1. INTRODUCTION

Video editing technology is more developed due to increasing uses of superimposed text inserted into video contents which provides better visual understanding for the viewers. Most propagation of videos tend to increase the use of overlay text to convey more direct summary of semantics and deliver better viewing experience such as headlines summarize the reports in news videos and subtitles in the documentary drama help viewers understand the content. Videos of sports contain text describing the scores, team, player names or speakers, location, date of an event, etc [1].In general, Video text can be classified into scene text and overlay text [2].

Scene text naturally occurs in the background as a part of the scene, such as the advertising banners, boards and so on; whereas overlay text is superimposed on the video scene and used to help viewers' understanding. Since the overlay text is highly structured and compact, it can be used for video indexing and retrieval [3]. However, for video optical character reorganization, overlay text extraction becomes more challenging, compared to the extraction of text for OCR tasks of document images, due to the difficulties resulting from complex background, size, unknown text, and color and so on. Two steps are mainly involved before the overlay text recognition is carried out, which include detection and extraction of overlay text. First, superimposed text regions are differentiated from background.

To determine the accurate boundaries of overlay text strings, the detected overlay text regions are refined. Background pixels are removed from the overlay text strings in the extraction step, to generate a binary text image for video OCR [4]. Although many methods have been proposed to detect and extract the video text, small number of method can effectively deal with different shape, color and multilingual text.

To address the problem of unknown text, complex background, size, color and so on, we propose a new method for overlay text detection and extraction using the transition region between the overlay text and background. First, the transition map is generated based on our observation that there exist transient colors between overlay text and its adjacent background. After that overlay text regions are detected roughly by computing the density of transition pixels and the consistency of texture around the

transition pixels. The detected overlay text regions are localized accurately using the projection of transition map with an improved color-based thresholding method to extract text strings correctly.

2. SYSTEM OVERVIEW

The main objective of the proposed system is to extract the text from complex videos scene and images. The extracting text from videos comprises many stages namely text detection, text localization and text extraction. The text detection is used to identify the presence of text in the video frame whereas text localization is used to determine the location of the text in the video frame and generate the bounding box in order to indicate the candidate region.

The candidate region is a portion of the frame which contains the text. In text extraction stage the text are extracted from the frame and passed on to the OCR for character verification. In the proposed system video is splitted into frames based on the shots. Redundant frames are discarded by performing frame similarity which results in selection of key frames [8]. In pre-processing stage, the text existing confidence is identified and its scale in the key frames. This stage identifies the region where the text is present i.e. candidate region. The adaptive thresholding (binarization) is applied to identify the presence of text in the key frame. After the detection of text region, the connected component analysis is performed where both horizontal and vertical projection in the key frame is used to detect the text. The extracted text is passed to the OCR (Optical character recognition) for character confirmation.

3. PROPOSED METHOD

3.1. Text detection method

Text detection is mainly used for finding only text in the image region that can be easily highlighted to the user. Text can be detected by manipulating the different properties of text characters such as vertical edge density, edge orientation variance or the texture. Two main problems of text detection are first is, how to avoid performing computational intensive classification on whole image and second is how to reduce difference of character size. To address these problems, we propose new text detection method that successfully detect superimposed text regions regardless of color, position, size, style, and contrast and also exist different size of texts mixed in each image frame.

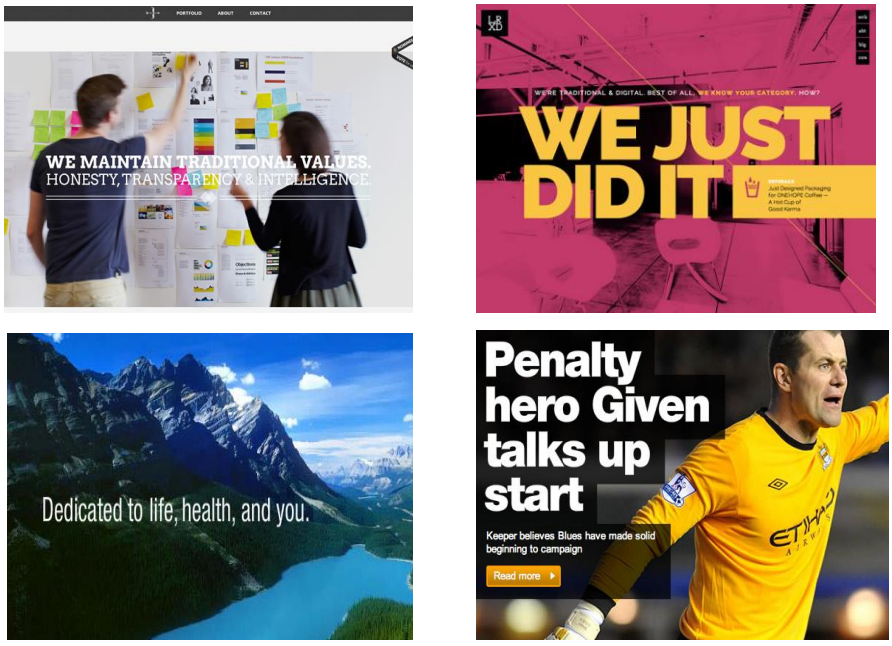


Fig1.Example of Overlay text images

Overall procedure of the proposed detection method is shown in fig.1.1.

Fig.1.1. Procedure of the proposed detection method [1]

1. Transition Map generation

Transient colors between Overlay text and its adjacent background are existing due logarithmic change in intensity at the boundary of superimposed text. The transition map can be used as a useful indicator for the superimposed text region [6].The intensities at the boundary of

overlay text are observed to have the logarithmical change due to color bleeding. Since at the boundary the change of intensity of overlay text may be small in the low contrast image, to effectively determine whether a pixel is within a transition region, the modified saturation is first introduced as a weight value based on the fact that overlay text is in the form of overlay graphics.

The modified saturation is defined as follows:

sxyval = (1.0 - ((3.0 / sum(R+G+B) * minrgb))-------(1)

intensity = sum / 3.0f;

barsxy = sxyval / maxsxy

where

maxsxy = 2.0f * (0.5f - intensity) ; if intensity > 0.5

maxsxy = 2.0f * intensity ; otherwise--------(2)

sxyval and maxsxy denote saturation value and maximum saturation at the corresponding intensity level respectively. maximum value of saturation is normalized in accordance with intensity compared to 0.5.The transition can be define by combination of change of intensity and modified saturation as follows:

Dl = (1.0 + dSl) * (I1 - I2);

Dh = (1.0 + dSh) * (I2 - I3);

Where dSl = ds1 - ds2;

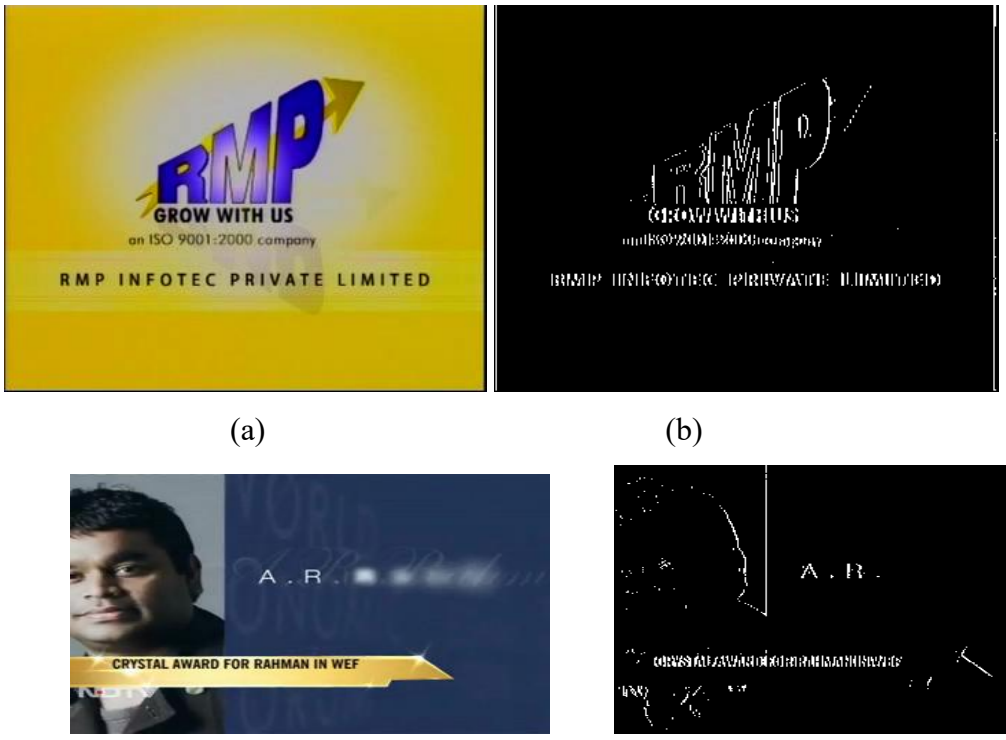
dSh = ds2 - ds3;----------(3)

since,dsl and dsh can be zero by the achromatic overlay text and background ,we add 1 to the weight in ds3.

T(x,y)={1; if(Dh > (Dl + TH)

={0;otherwise----------(4)

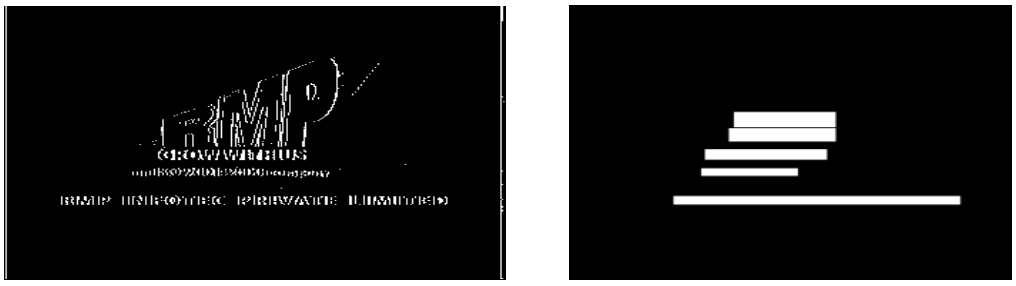
If the pixel satisfies this logarithmical change constraint, three consecutive pixel centered by the current pixel are detected as the transition pixel and transition map is generated. The thresholding value (TH) is empirically set to 80 in consideration of the logarithmical change.



(a) (b)

(c) (d)

Fig 2(a), (c) original image (b),(d)transition Map



(a) (b)

Fig. 3 (a) Transition map. (b) Smoothed candidate regions based on rectangular bounding box fitting.

2. Candidate Region Extraction

Next step is to generate the connected components; we first generate a linked map. If a gap of consecutive pixels between two nonzero points in the same row is shorter than 5% of the image width, they are filled

with 1s. If the connected components are smaller than the threshold value, they are removed [5]. The threshold value is empirically selected by observing the minimum size of overlay text region.

Then each connected component is reshaped to have smooth boundaries. the overlay text regions are generally in rectangular shapes, a rectangular bounding box is generated by linking four points, which correspond to

(min_x,min_y),(max_x,min_y),
(min_x,max_y),(max_x,max_y).

The refined candidate regions are shown in Fig. 3(b).

FivePercent = (bm.Width * 4) / 100;
start = j;

bm.SetPixel= Color.Black; if (j < bm.Width-1);

gap = j - start;
bm.SetPixel=Color.White;if (gap < 5% && gap > 0))

j = (start + gap) - 1; if (start + gap) - 1) > start)
----------(5)

3. Overlay Text Region Determination

The next step is to determine the real overlay text region among the boundary smoothed candidate regions. We use a texture-based approach for overlay text region determination. The local binary pattern (LBP) is employ to describe the texture around the transition pixel. LBP is a very efficient and simple tool to represent the consistency of texture using only the intensity pattern. We also define the probability of overlay text (POT) using the operator as follows:

i. The LBP operator is first applied to every transition pixel in each candidate region.

ii. Then, we compute the number of different LBPs to consider the intensity variation around the transition pixel.

iii. Since we use the 8 neighbor pixels to obtain the LBP value, the total number of potentially different LBPs is 28=256.

Let denote the density of transition pixels in each candidate region and can be easily obtained from dividing the number of transition pixels by the size of each candidate region. POT is defined as follows:

POT = density * NOL;

NOL = (NOL / 256.0) / LBPImg.Count;-----(6)

Where NOL denotes the number of different LBPs, which is normalized by the maximum of the number of different LBPs (i.e., 256) in each candidate region. If POT of the candidate region is larger than a predefined value, the corresponding region is finally determined as the overlay text region. The thresholding (TH) value in POT is empirically set to 0.05.

4. Overlay Text Region Refinement

A modified projection of transition pixels in the transition map is used to perform the overlay text region refinement. First, the horizontal projection is performed to accumulate all the transition pixel counts in each row of the detected overlay text region to form a histogram of the number of transition pixels. Then the null points, which denote the pixel row without transition pixels, are removed and separated regions are re-labeled [7]. The projection is conducted vertically and null points are removed once again.

A) Text Extraction Method

Text extraction is used for converting the grayscale image of a text region into the ready binary image in which all picture elements of characters are in black and others are in white.

Three main problem of text extraction are:

1)The unknown color polarity means whether text is light or dark
2) complex background and
3) various stroke widths.

To address these problems we propose fast and efficient text extraction system consists of color polarity computation, adaptive thresholding, dam point labeling and inward filling. Overall procedure of text extraction system is shown in fig.4.

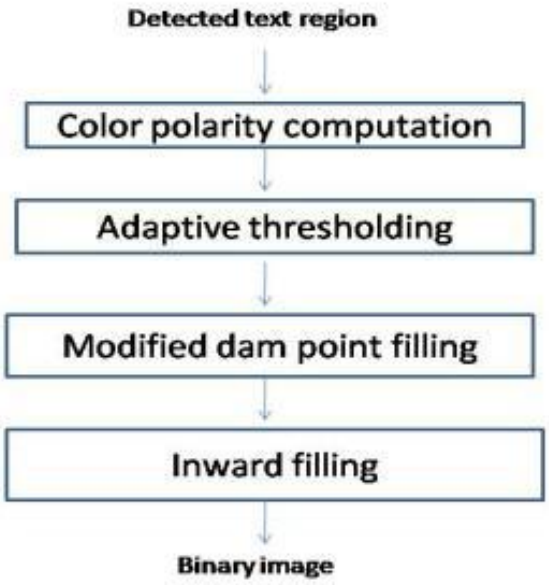


Fig.4. Procedure of proposed extraction method [1]

The text extraction methods fall into two groups namely stroke based and color-based methods. The refined superimposed text region is converted into a binary image in which all picture elements consisting of superimposed text are highlighted and others are inhabited [6]. Since First, each superimposed text region is expanded wider by two pixels to develop the continuity of background. Expanded outer region is denoted by ER.In next step comparison of picture element inside the text region and the pixels in ER are done, so that pixels connected to the expanded region can be eliminated. The text region is denoted as TR and the expanded text region as ETR. Next, adaptive shareholding based on sliding window is performed in the horizontal and the vertical directions with different window sizes, respectively. Inside TR, dam points are defined to prevent filling from flooding into text pixels. Finally, corrected characters are obtained from each superimposed text region by the inward filling [7].

5. BLOCK DIAGRAM

Block diagram of the system is shown in fig.7

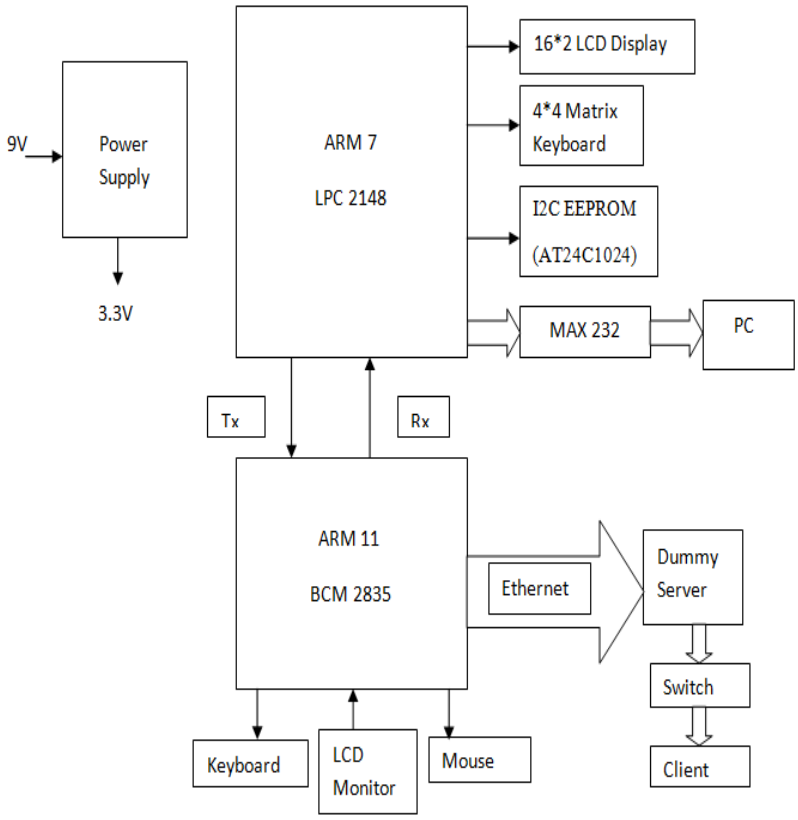


Fig.5.Block diagram for image and video text data extraction

We implement a system to detect and extract overlay text present on images and complex video scene. The system is basically implemented on ARM 7 (LPC 2148)microcontroller. Text data is serially transmitted and received by UART0/UART 1 port of arm 7 controller and finally displayed on LCD,their is serial communication between them. System is operated in two modes namely process mode and play mode.

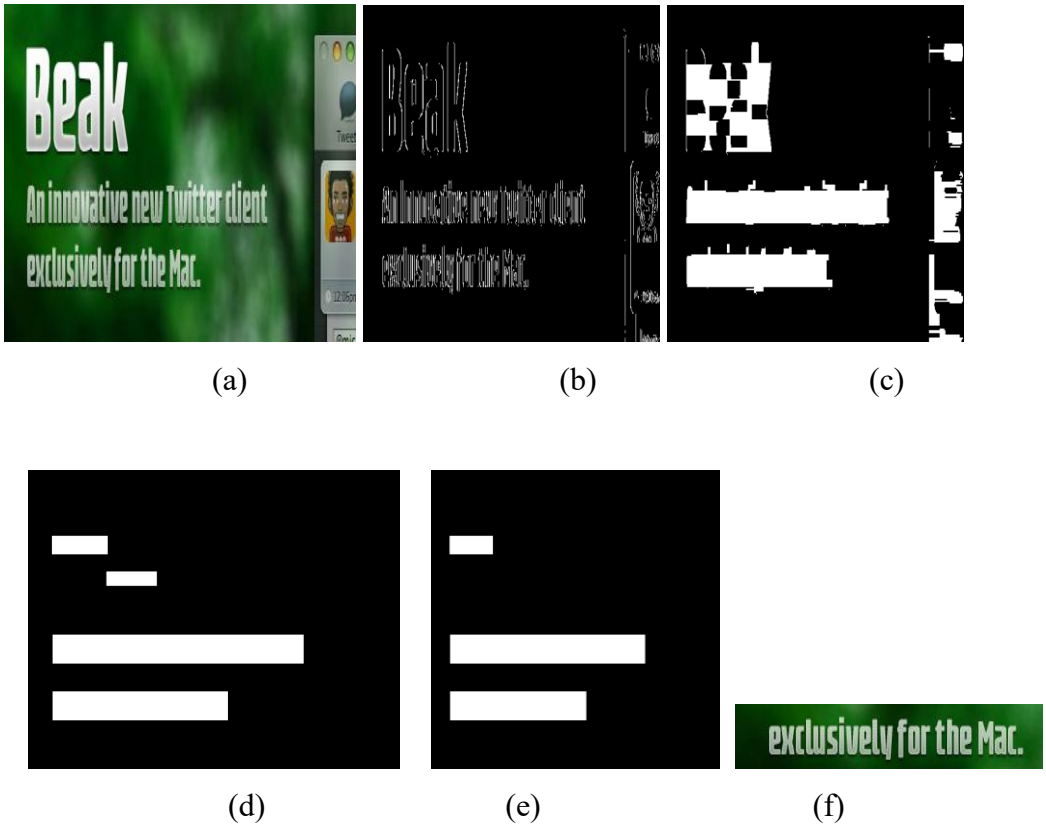
In play mode, we will play the video stored on PC, according to our choice whereas in process mode, it will perform overlay text extraction on any online or stored video on PC then that video is splited into number of key frames(images). Text region indicator is developed to compute the text existing confidence and candidate region by performing binarization.VB.net is used to perform overlay text data extraction. After that Overlay text data is again rollback to the pc and displayed on LCD.External

Memory isused for storing the obtained result.

ARM11 board (Raspberry Pi) is interfacing with ARM7 microcontroller. Due to this data available at the serial port is converted into TCP/IP format and sent it to Ethernet port. Through Ethernet it gets uploaded to the dummy server. Via internet, this data can be monitored by clients located worldwide.

6. EXPERIMENTAL RESULT

In this section, the system is constructed on the PC platform and the development environment is VB.NET 2008 .The video and image sequences are taken from internet and movies .Experiment results are shown in Fig 8 and Fig 9.



(a) (b) (c)

(d) (e) (f)

Fig.6. (a) Original image (b) Transition map (c)Linked map (d)Candidate region extraction (e)text region determination(f) Detected text image

(a) *(b)*

Fig.7. (a) Gray scale image (b) Extracted image

The framework for evaluating performance has been implemented with the image size of 320 X 240. Since the TH value plays an important role to generate a robust transition map, it is carefully set to 80 The minimum size of overlay text region used to remove small components is set to be 300. The parameters, such as window size for adaptive thresholding, the minimum size of overlay text region, and the threshold value for the overlay text region update, can be consistently set according to the image width or height.

between inserted text and its adjacent background. First the transition map is generated based on logarithmical change of intensity and modified saturation, then each connected component is reshaped to have smooth boundaries. We compute the density of transition pixels and the consistency of texture around the transition pixels to distinguish the overlay text regions from other candidate regions. The local binary pattern is used for the intensity variation around the transition pixel in the proposed method.

CONCLUSION

A novel method for overlay text detection and extraction from complex videos scene is proposed in this paper. Detection method is based on that there exist transient colors

The boundaries of the detected overlay text regions are localized accurately using the projection of overlay text pixels in the transition map. System is implemented using Arm 7 microcontroller. Various videos have

been tested for validating the performance of our detection and extraction method. The proposed method is mostly applicable for real time applications. To expand the algorithm for more advanced and intelligent application, our future work is to detect and extract the moving superimposed text.

REFERENCES

1. Wonjun Kim Dept. of Electron. Eng., Inf.& Commun. Univ.Changick Kim Daejeon, "A New Approach for Overlay Text Detection and Extraction from Complex Video Scene" 2009 IEEE.

2. Partha Sarathi Giri ,"Text Information Extraction And Analysis From Images Using Digital Image Processing Techniques," International Journal on Advanced Computer Theory and Engineering (IJACTE) ISSN (Print) : 2319 – 2526, Volume-2, Issue-1, 2013

3. Cong Yao, Xiang Bai, Wenyu Liu, Yi Ma2 Zhuowen Tu Huazhong, University of Science and Technology,, Microsoft Research Asia, Lab of Neuro Imaging and Department of Computer Science, UCLA "Detecting Texts of Arbitrary Orientations in Natural Images" 978-1-4673-1228-8/12/$31.00 ©2012 IEEE.

4. Keechul Jung, Kwang In Kim, Anil K. Jain ," Text Information Extraction in Images and Video: A Survey"

5. C.P. Sumathi, T. Santhanam, N. Priya,"Techniques and chanllnges of automatic Text extraction in complex images:A Survey" Journal of Theoretical and Applied Information Technology (JATIT)31st January 2012. ISSN: 1992-8645 ,Vol. 35 No.2

6. Michael R. Lyu, Fellow, IEEE, Jiqiang Song, Member, IEEE, and Min Cai," A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction"IEEE transaction on circuit and system for video technology,Vol.15,NO.2,FEB 2005

7. Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu, Senior Member, IEEE, "A Hybrid Approach to Detect

and Localize Texts in Natural Scene Images", IEEE transactions on image processing,vol. 20, no. 3, March 2011.

8. Ce'line Mancas-Thillou and Bernard Gosselin "spatial and color spaces combination for natural scene text extraction " 1-4244-0481-9/06/$20.00 c2006 ieee