A survey on Text Retrieval from Video

Mr. Sumit R. Dhobale[1], Prof. Akhilesh A. Tayade[2]

1ME (CSE) Scholar, Department of CSE, P R Patil College of Engg. & Tech., Amravati-444602, India

2Assitantant Professor, Department of CSE, P R Patil College of Engg. & Tech., Amravati -444602, India

ABSTRACT

Nowadays, video has become one of the most popular media for entertainment, study and types delivered through the internet, wireless network, broadcast which deals with the video content analysis and retrieval. The content based retrieval of image and video databases is an important application due to rapid proliferation of digital video data on the Internet and corporate intranets. Text which is extracted from video either embedded or superimposed within video frames is very useful for describing the contents of the frames, it enables both keyword and free-text based search from internet that find out the any contained display in the video.The number of methods are available to retrieve text from the video that's are OCR system, SVM system, DWT system, ANN base on their uses.

Keywords:- OCR(Optical Character Recognition), SVM(Support Vector Machine), DWT , superimposed

1.INTRODUCTION

The video are categorised most of news, social media, tutorials, lectures, e-learning videos consist of text which is very important regarding to the video information. The digital video use increasing day by day. Such digital video are efficient access involves indexing, retrieval, querying and browsing, which will require automated methods to understand its content and information of video. A digital video sequence is a combination of sequential images so to store a video on a computing support means to store a sequence of images which will have to be perfectly presented to the user at sufficient time intervals e.g. In standard 25 images per second. Some of that video consist text data that contains useful information for automatic indexing, annotation and structuring of images. Annotation is most useful contains in video in the form of text. From that video which consist of text are use to processing for Text information extraction system (TIE).That text information extraction system consists of four stages: text detection, text localization, text extraction & enhancement and text recognition. All the txt display in video are not same in visualisation. It consist of much of variation, that variations of the text are due to differences in font, style, size, orientation, alignment, low quality, high quality, complex back-ground, unknown layout makes the text extraction from video a challenging task as compare to high quality video. The MD algorithm is latest algorithm of text retrieval from video, for the text extraction from video algorithm is combination of the DWT (Discrete Wavelet Transform) and Morphology. Discrete Wavelet Transform is mathematical term, mathematically DWT transforms an image into frequency components and performed on the whole images, which differs from the other traditional methods that work on smaller pieces of the desired data. The morphology is a branch of image processing and useful for analyzing the shapes in the image. The morphology tool is useful for extracting image components that are use in the representation and description of region shapes such as boundaries, frames, skeletons and the convex hull from the image. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbours.

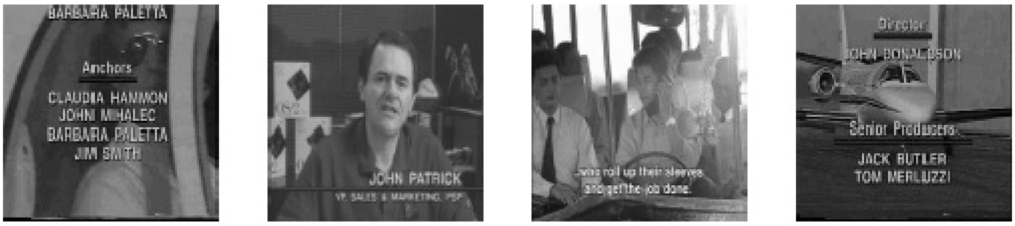


Figure 1:- Text in videos appears in different contexts, backgrounds, and font sizes**.**

2. LITERATURE SURVEY

The text information extraction system proposed by K. Jung, K. I. Kim, and A. K. Jain in 2004 which consists of 4 stages first text detection, second text localization, third text extraction & enhancement and last text recognition. The proposed method that 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 variations of text due to differences in orientation, size, style and alignment as well as low image contrast and complex background make the problem of automatic text extraction. However comprehensive surveys of related problems such as face detection, document analysis, and image & video indexing can be found, the problem of text information extraction is not well surveyed. The large number of techniques has been proposed to address this problem to classify and review these algorithms, discuss benchmark data and performance evaluation, and to point out promising directions for future research.[1] Thilagavathy invent another method of text extraction from video with fast identification of existing multimedia documents and mounting demand for information indexing and retrieval has been done on extracting the text from images and videos based on Artificial Neural network (ANN) in 2012. Extracting the scene text from video is demanding due to complex background, varying font size, different style, lower resolution and blurring, position, viewing angle and so on. In hybrid method where the two most well-liked text extraction techniques i.e. region based method and connected component (CC) based method used. The first step is the video is split into frames and key frames obtained. The obtained are Text region indicator (TRI) is being developed to compute the text prevailing confidence and candidate region by performing binarization to particular text region. Out of two modules Artificial Neural network (ANN) is used as the classifier and Optical Character Recognition (OCR) is used for character verification. All text is grouped by constructing the minimum spanning tree with the use of bounding box distance.[6] In traditional method of handwritten document images incorporate noise as binarization process by product. Because of this it causes some difficulties in text analysis process . To overcome this problem D. Brodic propose method in 2011 that morphological preprocessing is performed noise as binarization process. The main aim to make document image noiseless is evaluated. Results are given and comparative analysis is made. From the obtained results, decision-making is performed and convenient preprocessing morphological operation is proposed.[7] Epshtein B, Ofek E and Wexler Y (2010) present a novel image operator that seeks to find the value of stroke width for each image pixel, and demonstrate its use on the task of text detection in natural images. The suggested operator is local and data dependent that makes it fast and robust enough to eliminate the need for multi-scale computation or scanning windows. Its simplicity allows the algorithm to detect texts in many fonts and languages. HENK *et al* proposed the connected morphological operators and filters for binary images From digital video. [4] Connected operators cannot introduce new is continuities but these are suited for applications where contour information is important. In Huang, proposed a framework to reduce the effects of Ink Bleed in old documents which visualisation is low quality. [5] Thilagavathy proposed a hybrid method where the two text extraction techniques i.e. region based methods and connected component (CC) based method comes together. The region based method is used for segmentation and CC for filtering the text and non-text components. The cconnected component based approaches generally extracts characters based on their color or edge of text information. [6] In Wayne proposed the mathematical morphology as a systematic approach to analyze the geometric characteristics of signals or images, and provides an overview of MM and some Morphological filters which are widely used in image processing. [9] The classifier used in recently machine learning based approaches have been proposed which use popular SVM Classifier and Neural networks to train the features obtained by the Local Binary Pattern (LBP) based operator to detect Video text. [10] The SVM used as a classifier and then perform the CAMSHIFT to identify the text regions from the image frame in text base video. Kim et.al. proposed method that partial considered the case of multi-oriented scene text but caption text to a larger extent which are horizontally aligned. [11] Mathematical morphology is becoming increasingly important in image processing and computer vision applications. Neural network implementation of morphological operations has also been suggested. [12] Sharma, Palaiahnakote Shivakumara, Umapada Pal, Michael Blumenstein and Chew Lim Tan presents a new method for arbitrarily-oriented text detection in video, based on dominant text pixel selection, text representatives and region growing. The method uses gradient pixel direction and magnitude corresponding to Sobel edge pixels of the input frame to obtain dominant text pixels. Edge components in the Sobel edge map corresponding to dominant text pixels are then extracted and we call them text representatives. Then the perimeter of candidate text representatives grows along the text direction in the Sobel edge map to group the neighboring text components which we call word patches. The word patches are used for finding the direction of text lines and then the word patches are expanded in the same direction in the Sobel edge map to group the neighboring word patches and to restore missing text information. This results in extraction of arbitrarilyoriented text from the video frame. To evaluate the method, we considered arbitrarily-oriented data, non-horizontal data, horizontal data, Hua’s data and ICDAR-2003 competition data (Camera images).The experimental results show that the proposed method outperforms the existing method in terms of recall and f-measure.[14] Jia Yu , Yan Wang Video artificial text detection is a challenging problem of pattern recognition. Current methods which are usually based on edge, texture, connected domain, feature orlearning are always limited by size, location, language of artificial text in video. To solve the problems mentioned above, this paper applied SOM (Self-Organizing Map) based on supervised learning to video artificial text detection. First, text features were extracted. And considering the video artificial text's limitations mentioned, artificial text’s location and gradient of each pixel were used as the features which were used to classify. Then three layers supervised SOM was proposed to classify the text and non-text areas in video image. At last, the morphologic operating was used to get a much more accurate result of text area. Experiments showed that this method could locate and detect artificial text area in video efficiently.[15] Mohammad Khodadadi Azadboni , Alireza Behrad propose approach for text detection and localization is proposed. For this purpose, we first localize text location and then determine characters' pixels. The proposed text detection approach is a two-stage algorithm that in first stage, apply low pass filter on image in FFT domain to remove noisy element and then we apply Laplacian operator to the resultant image to highlight high contrast areas in the image. Then the product of corner dilated points and Laplacian enhanced image is calculated and text blocks are extracted using image vertical and horizontal projection. In the second stage of the algorithm, the extracted text blocks are verified using an SVM classifier. Text textures such as text angles and variance, momentum, entropy in co-occurrence matrix of text block are used for SVM training. We assumed that the characters of each text block have the same color. Therefore, we first estimate background color using image pixels in borders of detected text areas. Then the text color is estimated using the color clusters of pixels in text block and background color. We use color segmentation to extract character pixels. Experimental results show the promise of the proposed algorithm.[16]

Table: Different types of method implemented for text extraction.

Sr. No.	Author	Year	Methods	Implementation	Sr. No.	Author
1	K. Jung, K. I. Kim, and A. K. Jain	2004	Pattern Reorganisation	Extraction of this information involves detection, localization, tracking, extraction, enhancement, and recognition of the text variations of text due to differences in orientation, size, style and alignment as well as low image contrast and complex background make the problem of automatic text extraction	1	K. Jung, K. I. Kim, and A. K. Jain
2	Thilagavathy	2012	ANN	fast identification of existing multimedia documents and mounting demand for information indexing and retrieval has been done on extracting the text from images and videos based on Artificial Neural network	2	Thilagavathy
3	Jia Yu , Yan Wang	2011	SVM	The proposed approach is a two-stage algorithm that in first stage, apply low pass filter on image in FFT domain to remove noisy element and then we apply Laplacian operator to the resultant image to highlight high contrast areas in the image. In the second stage of the	3	Jia Yu , Yan Wang

				algorithm, the extracted text blocks are verified using an SVM classifier.		
4	*Jia Yu, Yan Wang*	2009	SOM	Current methods which are usually based on edge, texture, connected domain, feature or learning are always limited by size, location, language of artificial text in video. To solve the problems mentioned above, this paper applied SOM (Self-Organizing Map) based on supervised learning to video artificial text detection.	4	*Jia Yu, Yan Wang*

3. PROPOSED WORK

3.1 Identify the Unique Frame

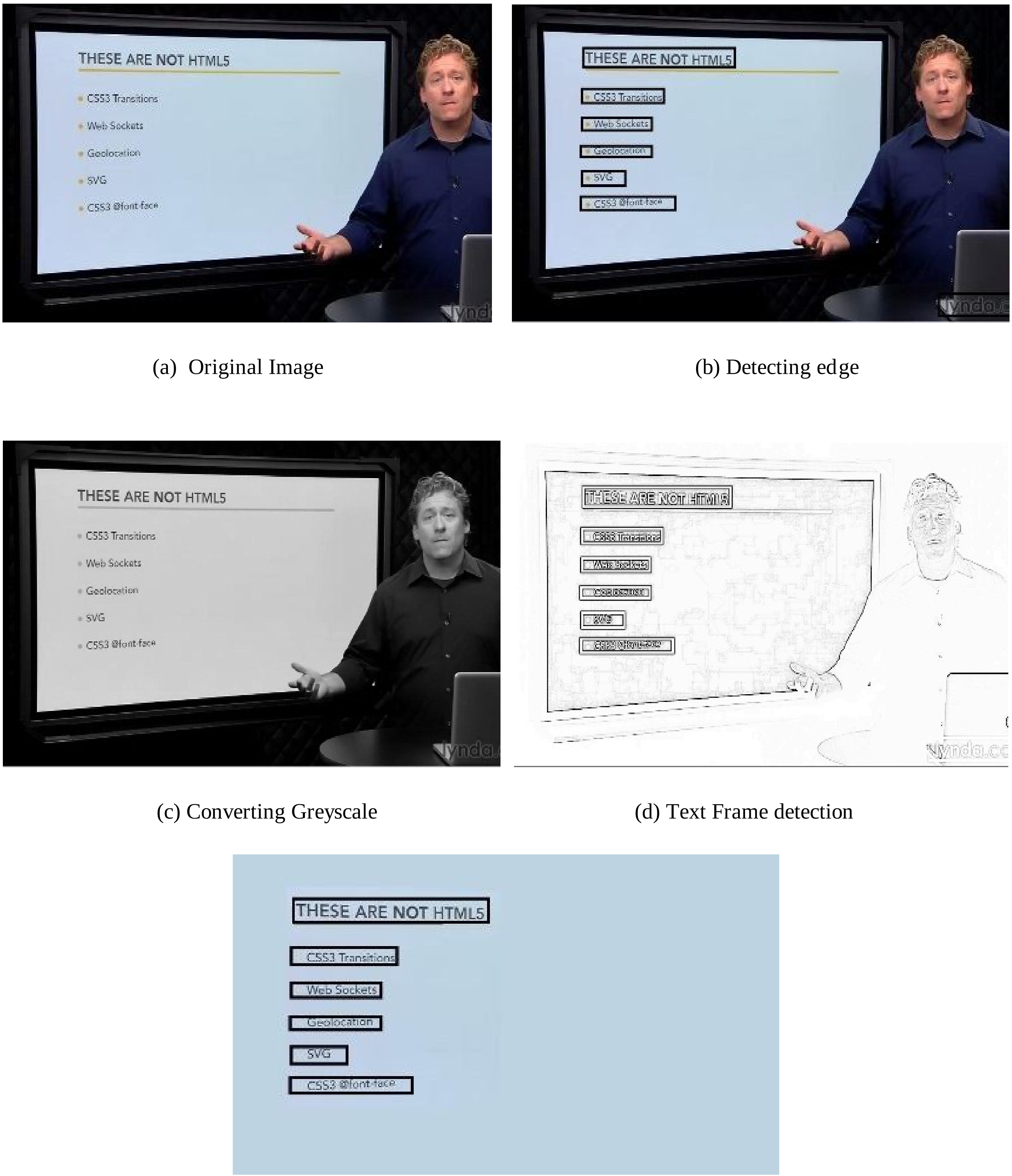
In the first module input will given in the form of video and it will convert them into the frames. The input will be given to the system in the form of a video. But instead of processing all frames, first step is the identification of the unique frames from the video, that process is known as pre-processing stage for reduction of number of processed frames. This process can done in number of method that invented some of them are SVM, OCR , DWT etc which identify only the text appear image. Text in video appears as either *scene text* or as *superimposed text* to extract superimposed text and scene text that possesses typical text attributes. Its do not assume any prior knowledge about frame resolution, text location, and font styles. Remove unwanted frame and the image which consist of homogeneous information.

3.2 Text Frame Detection

After identifying the unique frames, next is to identify text frame based classification. The method need to classify the image into text frame and non text frame and identify only the text frame portion. The different methodology use to classify that is OCR is use for identify the text base and non text based frame text. SVM use to identify the text based region. DWT is capture region which consist of text based on the frequency transform. The objective of this first step is to remove the background from an input gray scale image where the background is interpreted as containing non-text scene contents. Having obtained a number of homogenous regions in the labelled image, the non-text background regions are removed based on their spatial (width and height) proportions.

3.3 Refinement of Text Regions

The OCR systems require text to be printed against a clean background for character recognition, a local thresholding operation based on an interactive selection method is performed in each candidate region to separate the text from its surroundings and from other extraneous background pixels contained within its interior. Once thresholds are determined for all candidate regions, positive and negative images are computed, where the positive image contains region pixels whose gray levels are above their respective local thresholds and the negative image contains region pixels whose gray levels fall below their respective thresholds. A negative image will contain candidate text regions if text appears in inverse video mode in the input. All the remaining processing steps are performed on both positive and negative images and their results are combined at the end of the last stage. We further sharpen and separate the character region boundaries by performing a region boundary analysis based on the gray level contrast between the regions and their boundaries. This is necessary especially when characters within a text string appear connected with each other and need to be separated. For more details, we refer the reader to.



(a) Original Image (b) Detecting edge

(c) Converting Greyscale (d) Text Frame detection

(e) Text extraction

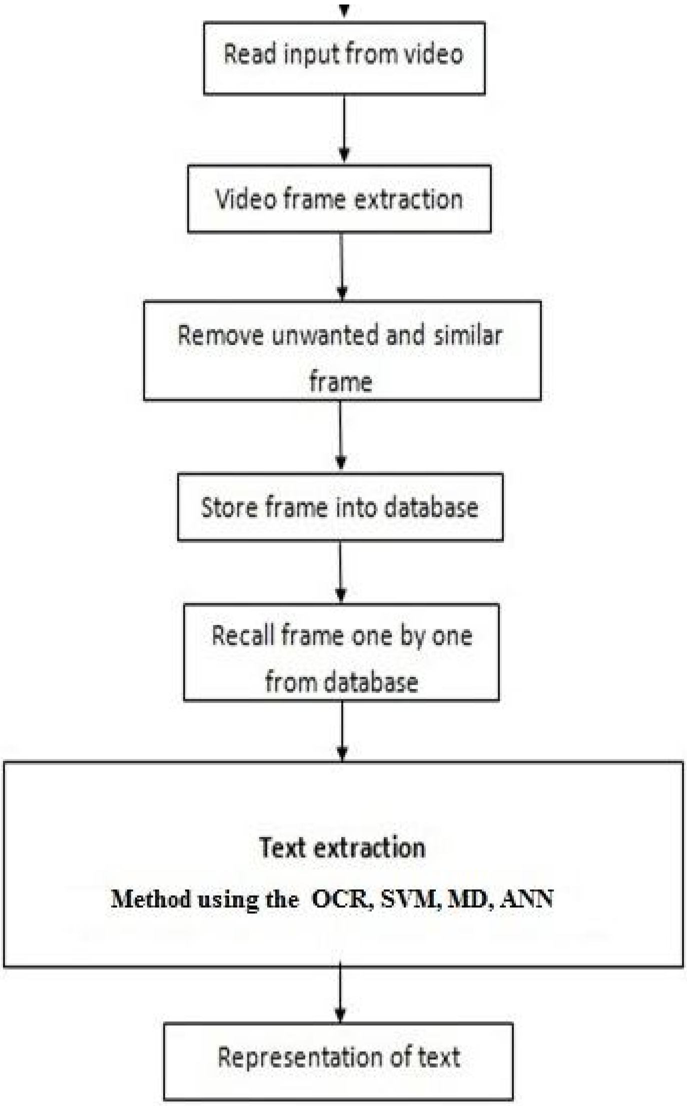
Figure 2:- An example of edge detection and text extraction from text contained video.

3.4 Verification of Text Characteristics

The candidate character regions that remain in the image are then subjected to a verification step where they are examined for typical text font characteristics. A candidate region is removed (i) if its area is less than 12 or its height less than 4 pixels, because OCR systems find it difficult to recognize small fonts; (ii) if the ratio of the area of its MBR to the region area (fill factor) is greater than 4; (iii) if the gray level contrast with the background is low.

3.5 Text Consistency Analysis

Neighboring text regions are examined for consistency to eliminate false positive regions. Unlike many other systems, ours attempts to ensure that regions adjacent in a line in the image exhibit the characteristics of a text string, thus verifying the global structure of a row of text in a local manner. This text consistency test includes (i) position analysism that checks intercharacter spacing, (ii) horizontal alignment verification of characters, and (iii) vertical proportion analysis of adjacent character regions. If all three conditions are satisfied, we retain the candidate word region as a text string. The final outputs are a binary image containing the text characters, which can be directly used as input to an OCR system to generate the text string in ASCII, and a text file containing the feature values of the character regions. Algorithm for text extraction from video has following steps defined in the form of flowchart:


Read input from video
Video frame extraction
Remove unwanted and similar frame
Store frame into database
Recall frame one by one from database
Text extraction
Method using the OCR, SVM, MD, ANN
Representation of text

3.6 Interframe Analysis for Text Refinement

Since text in videos persists over multiple consecutive frames, intraframe processing is followed by interframe verification. The text regions in each set of five consecutive frames are analyzed together to add missing characters and to delete regions incorrectly identified as text. This interframe analysis involves examination of the similarity of text regions in terms of their positions, intensities, and shape features and mitigates false alarm.

4.CONCLUSION

Here we identified a research gap for text extraction from video and it is motivated to propose novel method for extraction from text and characters from video. In addition to this propose method is not applicable for complex non linear motion video or some other effect within video. This work can be implemented as future research or proposed method.

REFERENCES

[1] K. Jung, K. I. Kim, and A. K. Jain, “Text information extraction in images and video: A survey,” Pattern Recogn., vol. 37, no. 5, pp. 977–997, 2004.
[2] NINA S.T. HIRATA, ”Text Segmentation by Automatically Designed Morphological Operators ”, 0-7695-0878-2/2000 IEEE.
[3] X. C. Jin, ” A Domain Operator For Binary Morphological Processing”, IEEE transactions on image processing 1057- 7149/95@1995 IEEE.

[4] Henk J. A. M. Heijmans, ” Connected Morphological Operators and Filters for Binary Images”, 0-8186-8183-7/97@ 1997 IEEE.

[5] Yi Huang, ” A Framework for Reducing Ink-Bleed in Old Documents”, 978-1-4244-2243-2/08©2008 IEEE.

[6] A. Thilagavathy, ”Text Detection And Extraction From Video Using Ann Based Network” ,IJSCAI, vol.1, Aug 2012.

[7] D. Brodic, ”Preprocessing of Binary Document Images by Morphological Operators”, MIPRO 2011.

[8] Neha Gupta and V.K Banga, ”Image Segmentation for Text Extraction”, ICEECE, 2012.

[9] Epshtein B, Ofek E and Wexler Y (2010) “Detecting text in natural scenes with stroke width transform”. IEEE, Conference on Computer Vision and Pattern Recognition (CVPR) CVPR, 2963–2970.

[10] K. I. Kim, C. S. Shin, M. H. Park, and H. J. Kim. “Support Vector Machine based text detection in digital video” .In proc. IEEE signal Processing Society Workshop, pages 634-641, 2000. [11] Wayne, "Mathematical Morphology and Its Application on image segmentation

[11] Shivakumara P, Phan TQ, and Tan CL “Video text detection based on filters and edge features”. In proc. of the 2009. Int Conf on multimedia and Expo.2009.pp.1-4,IEEE.

[12] Y. Pan, X. Hou, and C. Liu, "A hybrid approach to detect and localize texts in natural scene images," IEEE Transactions on Image Processing, vol. 20, pp. 1-1, 2011.

[13] Nabin Sharma, Palaiahnakote Shivakumara, Umapada Pal, Michael Blumenstein and Chew Lim Tan “A New Method for Arbitrarily-Oriented Text Detection in Video” 2012 IEEE DOI 10.1109/DAS.2012.6

[14] Jia Yu , Yan Wang “Apply SOM to Video Artificial Text Area Detection” 2009 Fourth International Conference on Internet Computing for Science and Engineering 978-0-7695-4027-6/10 $26.00 © 2010 IEEE DOI 10.1109/ICICSE.2009.

[15] Mohammad Khodadadi Azadboni , Alireza Behrad “Text Detection and Character Extraction in Colour Images using FFT Domain Filtering and SVM Classification” 6'th International Symposium on Telecommunications (IST'2012) 978-1-4673-2073-3/12/$31.00 ©2012 IEEE.

AUTHOR

Mr. Sumit R. Dhobale Received Bachelor degree in computer science and Engg from Amravati University in 2012 and pursuing master degree in C.S.E from P.R. Patil college of Engg Amravati -444602

Prof. Akhilesh A. Tayade Working as Assistant Professor in department of C.S.E at P.R. Patil College of Engg Amravati -44602