Implementation of Text Extraction From Video Using Morphology and DWT

Mr. Sumit R. Dhobale
ME (CSE) Scholar, Department of CSE,
P R Patil College of Engg. & Tech
Amravati-444602, India
sumit.dhobale@gmail.com

Prof. Akhilesh A. Tayade
ME (CSE) Scholar, Department of CSE,
P R Patil College of Engg. & Tech
Amravati-444602, India
akhilesh_tayde2006@yahoo.com

Abstract—Video has 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 algorithm performance on the basis of text localization and false positive rate has improved for different types of video. The overall accuracy of this methodology is high than that of any other methods. The advantage of this algorithm is that it minimise processing time.

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

*****

I. INTRODUCTION

The video has become most popular nowadays in the field of internet, wireless, broadcast. That's the number of video consist different types of video retrieval and analysis. Text which has been retrieve from the video is important in the field of news, social media, tutorials, lecture video etc.[7] The video is composed of individual images the frames that are replayed at a certain speed to create the effect of continuous motions.[3] Several frames form shots, which are in turn combined to form sties or, episodes, or scenes that the video is composed of. A video sequence is a succession of 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 intervals (25 images per second).Text data present in the videos and images contains useful information for automatic indexing, annotation and structuring of images. [1]

Several frames form shots, which are in turn combined to form sties or, episodes, or scenes that the video is composed of. A video sequence is a succession of 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 intervals (25 images per second).

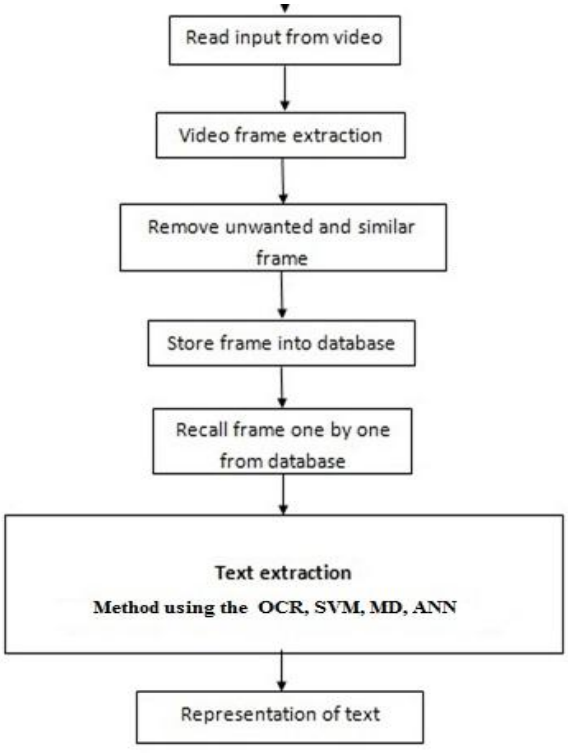


Figure 1: Data flow diagram of extracting text from video

Text data present in the videos and images contains useful information for automatic indexing, annotation and structuring of images.

II. METHODOLOGY IMPLEMENTATION OF EXTRACT TEXT FROM VIDEO USING MD ALGORITHM

2.1 Implementation of extracting frame on video.

This module focuses on the implementation aspects ofextracting frame on video . Following are the steps for it.

Firstly video of different categories are collected. It includes educational video, e-learning video, Multimedia video and Animation Videos. 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.

Step 1: Read the Input Video

The input gives to matlab which read the video that basically consist of various feature out of them only text on the image need. Pick up the video file from system

Step 2: Read video frmae

Video consist of various features color, images, text etc. Read the Video Features and compression format, if it is uncompressed Video, Process on it.

Step 3: Find the Number of Frames from Video called N

This module input will given in the form of video and it will convert them into the frames called them N.

Step 4: Identify the Unique Frame

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. Minimum number of the frequent changes in a video leads to high efficiency. For reducing the similar frames DWT based similarity measure is implemented for identification of frames.

Step 5: Avoid concurrency of frames

The extracting frame are also consist of same features those of same text thus it avoid concurrency of frame. For selecting one frame out of numbers of frame there calculate the difference between two frames. Only if there is no difference then use next frame for processing.

Step 7: Text Frame Detection

After identifying the unique frames, next is to identify text frame by using DWT based classification. DWT will classify the image into text frame and non text frame and identify only the text frame portion. The transform of a signal is just another form of representing the signal. It does not change the information content present in the signal. DWT offers multi-resolution representation of an imageand gives perfect reconstruction of decomposed image. Imageitself is considered as two dimensional signals. When imageis passed through series of low pass and high pass filters, itdecomposes the image into sub bands of distinct resolutions.Decompositions can be done at different DWT levels. DWToffers multi-resolution representation of a signal.

It has been widely accepted that maximum energy of most of natural images is concentrated in ‘approximate (LL) subband’ which is low frequency sub-band. Hence modification to the coefficients of these low frequency sub-bands would cause severe and unacceptable image degradation.

2.2 Implementation of localizing and extract text from frame.

The text detection stage search for to detect the occurrence of content in a camera captured natural scene images. Because of different font, highlights, different cluttered background image alteration and demeaning correct and quick text detection in scene images is still difficult task. The approach uses a character descriptor to fragment text from an image. Initially content is detected in multi size images using edge based system, morphological Function and projection report of the image. These detected text area are then confirmed using descriptor and wavelet features. The algorithm is strong when

difference in style, size of font and color. Vertical edges with a predefined pattern are used to detect the edges, then grouping vertical boundaries into text area using a filtering process.

Step 1: Converted into gray scale

The current frame converted into gray scale image which gives better text observation for further processing which chooses the threshold to minimize the intraclass variance of

the black and white pixels.

Step 2: Text Detection

Once text frames are identified, the next step is to apply segmentation approach so that the text can be extracted easily from the video. To perform this, a combined approach of morphology and DWT is used. DWT is used to perform the decomposition and on the each decomposition morphological operator is used to extract the text in better way.

The dwt performs a single-level two-dimensional wavelet decomposition with respect to either a particular wavelet or particular wavelet decomposition filters (Lo_D and Hi_D) specify.[cA,cH,cV,cD] = dwt2(X,'wname') computes the approximation coefficients matrix cA and details coefficients matrices cH, cV, and cD (horizontal, vertical, and diagonal, respectively), obtained by wavelet decomposition of the input matrix X.

The 'wname' string contains the wavelet name.[cA,cH,cV,cD] = dwt2(X,Lo_D,Hi_D) computes

the two-dimensional wavelet decomposition as above, based on wavelet decomposition filters that you specify.Lo_D is the decomposition low-pass filter.Hi_D is the decomposition high-pass filter.

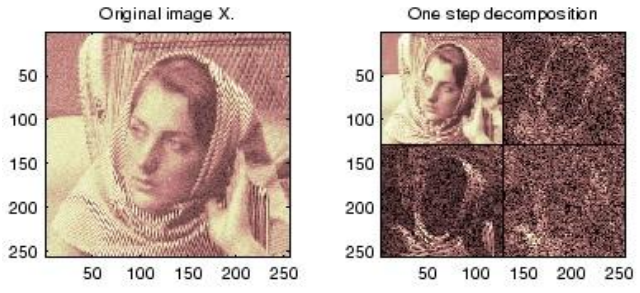


Figure 2 : DWT Synthesis

Step 3: Text localisation

The text localization process that to find out particular region from the whole image. For that localization threshold region MSER detector incrementally steps through the intensity range of the input image to detect stable regions. The Threshold Delta parameter determines the number of increments the detector tests for stability. You can think of the threshold delta value as the size of a cup to fill a bucket with water. The smaller the cup, the more number of increments it takes to fill up the bucket. The bucket can be thought of as the intensity profile of the region. Find the number of component of text and create vector to store it. Then go to each component.

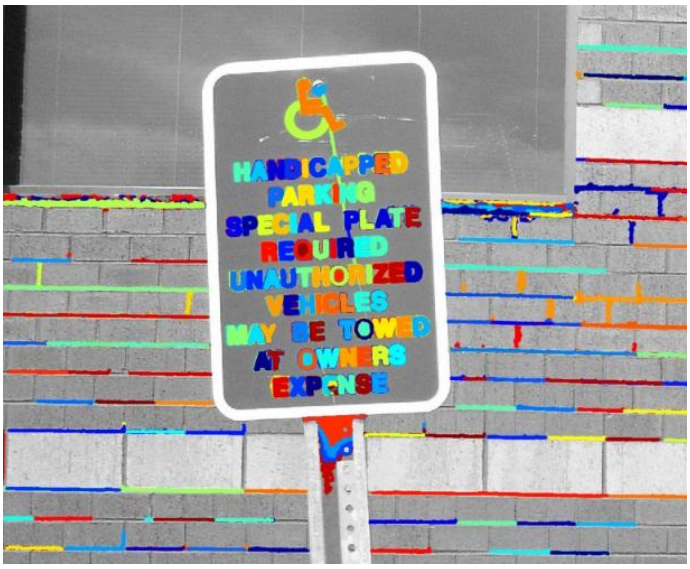


Figure 3 : Text Localisation

A pixel on the edge of the input image might not be considered to be a border pixel if a non-default connectivity is specified the elements on the first and last row are not considered to be border pixels because, according to that connectivity definition, they are not connected to the region outside the image.

Step 4: Remove Non-Text Regions Based On Basic Geometric Properties

Although the MSER algorithm picks out most of the text, it also detects many other stable regions in the image that are not text. You can use a rule-based approach to remove non-text regions. For example, geometric properties of text can be used to filter out non-text regions using simple thresholds.

Alternatively, you can use a machine learning approach to train a text vs. non-text classifier. Typically, a combination of the two approaches produces better results . This example uses a simple rule-based approach to filter non-text regions based on geometric properties threshold the data to determine which regions to remove.

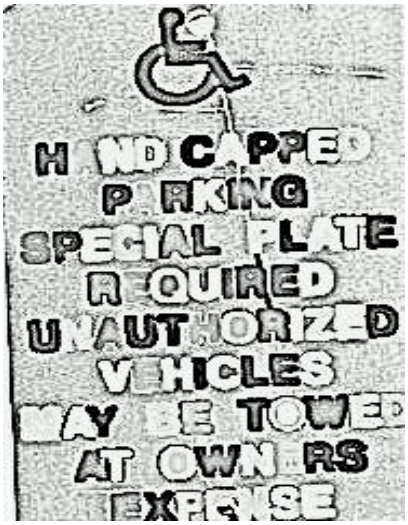


Figure 4: Extracted Text Region

Step 5: Morphological Algorithm

The morphological approach is used for segmentation of the images from grayscale image. Morphology is a theory for the analysis and processing of geometrical structures, based on lattice theory, set theory, random function, and topology. Normally, this is applied to digital images, but this can also be applied on graphs, solids, surface meshes, and many other spatial structures. Geometrical and Topological continuous-space concepts such as size, shape, connectivity, convexity, and geodesic distance, can be characterized in both continuous and discrete spaces. The morphological operation use to thickens the character without changing their parametric value. Frist converted image into binary image then measures a set of properties for each connected component (object) in the binary image.

The logical(A) converts numeric input A into an array of logical values. Any nonzero element of input A is converted to logical 1 (true) and zeros are converted to logical 0 (false). Complex values and NaNs cannot be converted to logical values and result in a conversion error.

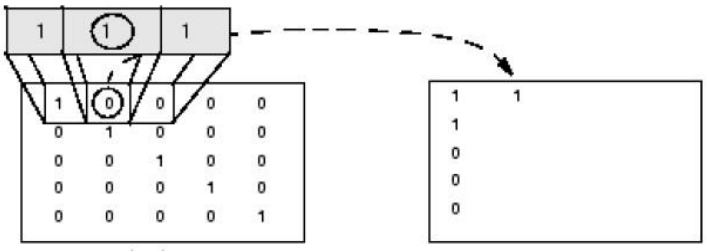


Figure 5 : Dilution Of Region

. This indicates that the region is more likely to be a text region because the lines and curves that make up the region all have similar widths, which is a common characteristic of human readable text.

Now, the overlapping bounding boxes can be merged together to form a single bounding box around individual words or text lines. To do this, compute the overlap ratio between all bounding box pairs. This quantifies the distance between all pairs of text regions so that it is possible to find groups of neighboring text regions by looking for non-zero overlap ratios. Once the pair-wise overlap ratios are computed, use a graph to find all the text regions "connected" by a non-zero overlap ratio.

Step 6: Recognize Detected Text Using OCR

After detecting the text regions, use the OCR function to recognize the text within each bounding box. Note that without first finding the text regions, the output of the OCR function would be considerably more noisy..

Step 7: Stored into text file

Extracted text portion compare with words and matching text store into documents in the form of .doc, .txt or .pdf format.

III. RESULT ANALYSIS

This section show the performance analysis of the system and the result gatherd from the various video can response to its performance and accuracy.

3.1 Localisation Rate

The localisation rate is the fraction of retrieved instances of frame that are relevant. It is defined as the ratio of correctly

detected blocks to the sum of correctly detected blocks plus false positives.

Correctly detected blocks are the reasons which are correctly detected by the presented MD algorithm. False positives are the regions which are actually not characters of a text, but have been detected as text regions.

$$\text{Localization Rate} = \frac{\text{Correctly detected blocks}}{\text{Correct detected blocks} + \text{false positives}} * 100$$

The Recall rate (also known as sensitivity) is the fraction of relevant instances. It is defined as the ratio of correctly detected blocks to the sum of correctly detected blocks plus false negatives. False Negatives are the regions which are actually text characters, but have not been detected as text regions.

$$\text{Recall Rate} = \frac{\text{Correctly detected blocks}}{\text{Correctly detected blocks} + \text{false negativ e}} * 100$$

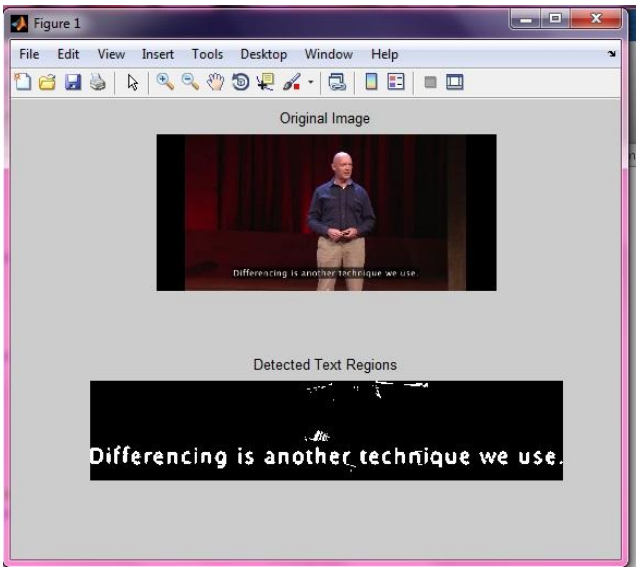
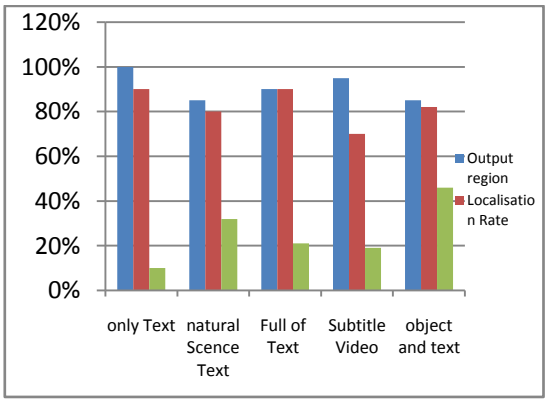


Figure 5 : Display GUI for full subtitle Region.

TABLE 1 : System localization and call result

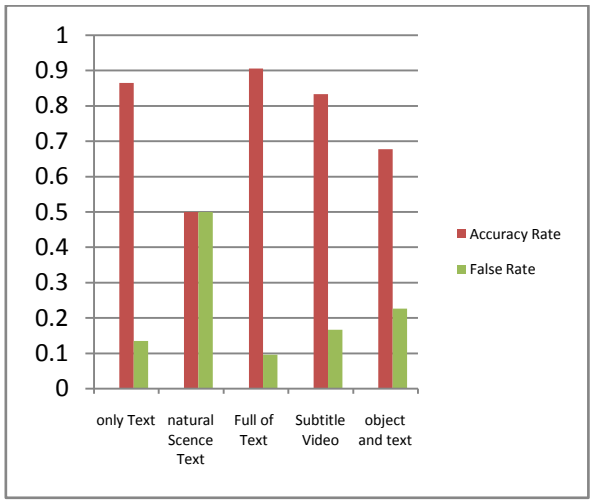
Types of Video	Text containing region	Output region	False Positive	False Negative	Localisation Rate	Call Rate
only Text	80%	100%	90%	10%	90%	10 %
natural Scence	20%	70%	55%	25%	80%	32 %
Full of Text	100%	90%	80%	20%	90%	21 %
Subtitle Video	10%	95%	70%	20%	70%	19 %
object and text	20%	65%	56%	29%	82%	46 %



3.2 Accuracy Rate

The accuracy rate are calculated by the number of character present in current frame and output character accuracy. Every charetor match and synthesize the accuracy

Types of Video	Number of charector on Frame	Number of match charector	Accuracy Rate	False Rate
only Text	37	32	86.46%	13.54%
natural Scence	10	5	50%	50%
Full of Text	814	737	90.54%	9.56%
Subtitle Video	36	30	83.33%	16.66%
object and text	18	11	67.77%	22.66%



The testing is perform manualy one by one frame analysys and match the output with current frame. The following frame cintains total 37 chatector. The frist line cansist of four character out of them the system all acurate output, in the next line six character out of that system cant juge the value of "S" and gives incorrect output 5. After analising the all character 32 are give acurately by the system. The percent of accuracy calculated by the fraction of number of character to the system gives correct value of respective character.

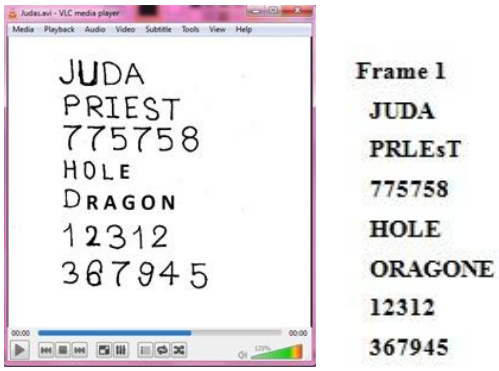


Figure 6 : Output in text format.

The following graph shows the analisys of different charactoristics of video and their respective result.

IV. CONCLUSION AND FUTURE SCOPE

In this methodology for text extraction is implemented with the help of MD algorithm. The approach is used in this algorithm morphology and DWT for low resolution as well as high resolution. Result of all work are shown that the localization rate and false positive rate have been improved for extraction of text from video in different types of video such as Video containing only Text, Video containing natural Sense Text, Video containing Full of Text, Subtitle Video, Video with object and text. The overall accuracy of this methodology is high than that of any other methods. The advantage of this algorithm is that it minimize processing time.

This method will be enhance with real time voice synthesis. So that it can be use as real time system for text and voice matching. Also the extracted text use for metadata using data mining technology to search the video which consist of any type of text present in the video.

REFERENCES

[1] Sonam, Mukesh Kumar "Implementation of MD algorithm for Text Extraction from Video" 2013 Nirma University International Conference on Engineering (NUiCONE) 978-1-4799-0727-4/13/$31.00 ©2013 IEEE

[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 ComputerVision 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 ProcessingSociety 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. IntConf on multimediaand Expo.2009.pp.1-4,IEEE.

[12] J. Serra, "Introduction to mathematical morphology" Computer Vision, Graphics, Image Processing, vol. 35, pp. 283-305, 1986.

[13] 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.

[14] 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

[15] 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.

[16] 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.