Video Scene Segmentation to Separate Script

Bharatratna P. Gaikwad
Department of CS and IT
Dr. B. A.M. University,
Aurangabad (MS), India
bharat.gaikwad08@gmail.com

Ramesh R. Manza
Department of CS and IT
Dr. B. A.M. University
Aurangabad (MS), India
manzaramesh @gmail.com

Ganesh R. Manza
Department of CS and IT
Dr. B. A.M. University
Aurangabad (MS), India
ganesh.manza@gmail.com

***Abstract*—This paper proposes improved Template Matching algorithm that applied for the automatic extraction of text from image and video frames. Optical Character Recognition using Template Matching is a system model that is useful to recognize the character, alphabet & special character by comparing two images of the alphabet. The objectives of this system model are to develop a model for the Optical Character Recognition (OCR) system and to implement the Template Matching algorithm in developing the system model .The template matching techniques are more sensitive to font and size variations of the characters than the feature classification methods. This system tested the 25 videos with 370 video frames for each video. In this system 91% of the Character gets recognized successfully using Texture-based approaches to automatic detection, segmentation and recognition of visual text occurrences in images and video frames.**

***Keywords*— *video processing, character detection, localization, tracking, segmentation, OCR*.**

I. INTRODUCTION

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 are using with one or a combination of image, audio, and textual information in the video [1]. A variety of approaches to text information extraction from images and video have been proposed for specific applications including page segmentation , address block location , license plate location , and content-based image/video indexing [2].In the extraction of this information involves the detection, localization, tracking, extraction, enhancement and recognition of text from the images and video frames are provided. Text in images and video frames carries important information for visual content understanding and retrieval. Optical character recognition (OCR) is one of the most popular areas of research in pattern recognition because of its immense application potential. The two fundamental approaches to OCR are template matching and feature classification. In the template matching approach, recognition is based on the correlation of a test character with a set of stored templates. In the feature classification method, features are extracted from a standard character image to generate a feature vector. A decision tree is formed based on the presence or absence of some of the elements in the feature vector. When an unknown character pattern is encountered, this tree is traversed from node to node till a unique decision is reached. The template matching techniques are more sensitive to font and size variations of the characters than the feature classification methods. However, selection and extraction of useful features is not always straight forward [5].
Extracting text information from videos generally involves three major steps:

• Text detection: Find the regions that contain text.
• Text segmentation: Segment text in the detected text regions. The result is usually a binary image for text recognition.
• Text recognition: Convert the text in the video frames into ASCII characters.

II. TECHNICAL IMPLIMENTATION AND ANALYSIS

A. VidoPocessing

Shot: Frames recorded in one camera operation form a shot.
Scene: One or several related shots are combined in a scene.
Sequence: A series of related scenes forms a sequence.
Video: A video is composed of different story units such as shots, scenes, and sequences arranged according to some logical structure (defined by the screen play). These concepts can be used to organize video data. The video consists of sequence of images (video frames). In the first step, we convert video into all frames and saved as JPEG images.

B. Pre Processing

A scaled image was the input which was then converted into a gray scaled image. This image formed the first stage of the pre-processing part. This was carried out by considering the RGB color contents of each pixel of the image and converting them to grayscale. The conversion of a colored image to a gray scaled image was done for easier recognition of the text appearing in the images as after grayscale conversion, the image was converted to a black and white image containing black text with a higher contrast on white background [12].

C. Detection and Localization

978-1-4673-4529-3/12/$31.00 ©2012 IEEE

In the text detection stage, since there was no prior information on whether or not the input image contains any text, the existence or nonexistence of text in the image must be determined. However, in the case of video, the number of frames containing text is much smaller than the number of frames without text. The text detection stage seeks to detect the presence of text in a given image. Text localization methods can be categorized into two types: region-based and texture-based. Select a frame containing text from shots elected by video framing, this stage used region Based Methods for text tracking. Region based methods use the properties of the color or gray scale in a text region [1].

D. Ttracking and Segmentation

When text was tack, the text segmentation step deals with the separation of the text pixels from the background pixels indirectly separate single character from whole text. The output of this step is a binary image where black text characters appear on a white background. This stage included extraction of actual text regions by dividing pixels with similar properties into contours or segments [2][9].

E. Recognition

This stage included actual recognition of extracted characters , The result of recognition was a ratio between the number of correctly extracted characters and that of total characters and evaluates what percentage of a character were extracted correctly from its background. For each extraction result of correct character [4].

F. Survey of Literature:

1) Jie Xi and et.al. has work on Text detection, tracking and recognition to extract the text information in news and commercial videos. He has used Techniques morphological opening procedure on the smoothed edge map. They got the text detection rate is 94.7% and the recognition rate is 67.5% [7].

2) Palaiahnakote Shivakumara and et.al. has work on elimination of non-significant edges from the segmented text portion of a video frame to detect accurate boundary of the text lines in video images. They got percentage 93% [8].

3) Rainer Lienhart and et.al. has worked on the text localizing and segmenting text in complex images and videos, It is able to track each text line with sub-pixel accuracy over the entire occurrence in a video. They got percentage text recognition 69.9% [9].

4) Qixiang Ye and et.al.has worked on the detection and verification of English Text and Chinese text from images and video fames, He has used Techniques for detection is based on Sobel edges feature and the verification uses the wavelet-based features and svm classifer. They got percentage detection rate English 93.9% [10].

III. EXPERIMENTAL RESULTS

A. Canny Edge detector

Among the several textual properties in an image, edge-based methods focus on the 'high contrast between the text and the background'. The edges of the text boundary are identified and merged, and then several heuristics are used to filter out the non-text regions. Usually, an edge filters (e.g. canny operator) is used for the edge detection, and a smoothing operation. The Canny method finds edges by looking for local maxima of the gradient of I. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges [13][15]. This method is therefore less likely than the others to be fooled by noise, and more likely to detect true weak edges [3][16].

$G_X =$

-1	-2	-1
0	0	0
1	2	1

$G_Y =$

-1	0	+1
-2	0	+2
-1	0	+1

Fig.1 Canny Edge detection operator (a) x direction (b) y direction

1. Compute f_x and f_y

$$f_x = \frac{\partial}{\partial x}(f * G) = f * \frac{\partial}{\partial x}G = f * G_x \quad (1)$$

$$f_y = \frac{\partial}{\partial y}(f * G) = f * \frac{\partial}{\partial y}G = f * G_y \quad (2)$$

$G(x, y)$ is the Gaussian function $G_x(x, y)$ is the derivate of $G(x, y)$ with respect to x:

$$G_x(x, y) = \frac{-x}{\sigma^2}G(x, y) \quad (3)$$

$G_y(x, y)$ is the derivate of $G(x, y)$ with respect to y:

$$G_y(x, y) \frac{-y}{\sigma^2}G(x, y) \quad (4)$$

2. Compute the gradient magnitude $magn(i,j)$

$$= \sqrt{f_x^2 + f_y^2} \quad (5)$$

3. Apply non − maxima suppression.

4. Apply hysteresis thresholding / edge linking .

The canny edge detection algorithm is easy to implement, and more efficient than other algorithms. From this edge detected images, text region is identified [3].Text in images and video frames can exhibit many variations with respect to the following properties:

TABLE I. DIFFERENT PROPERTIES OF TEXT IN IMAGES AND VIDEO FRAMES

Category	**Detail Properties**	**sub-classes**
Geometry	Size Regularity in size of text	Font Size, width, height, Bold, Italic
	Alignment	Horizontal/vertical/center
		Straight line with skew (implies vertical direction)
		Curves
		3D perspective distortion
	Inter-Character distance	Aggregation of characters with uniform distance
	Strokes	Different stroke density and statistics
Color		Gray
		Color (monochrome, polychrome)
		Text color is dark or light
Motion		Static
		Linear movement
		Static
		Free movement
Edge		Strong contrast (edges) at text boundaries
Compression		Un-compressed image
		JPEG, MPEG-compressed image [2].

B. Design a System and Implementing Algorithm

The template matching worked on this following Algorithm

a) Load the video (E.g. Avi, Mpeg etc.).

b) Then video is converted into frames with frames name from"img-1 to img-N "till the video will be come to an end.

c) Template is made of Upper case, Lower case, Special character & digit with size 24x42 size.

d) Applying OCR techniques, select the frame among one of them (E.g.img-50).

e) Image is Converted to gray scale and then converted to binary.

f) Then top-down: extracting texture features of the image and then locating text regions.

g) Bottom-up: separating the image into small regions and then grouping character regions into text regions.

h) The character image from the detected string is selected.

i) Segmentation: Each character was automatically selected and thresholding using methods.

j) After that, the image to the size of the first template is rescaled.

k) After rescale the image to the size of the first original image then comprising letters with template and the matching metric is computed.

l) Then the highest match found is stored. If the template image is not match, it might be getting recognized as some other character.

m) The index of the best match is stored as the recognized character.

n) All recognized character showing on Word file.

Architecture of Video Scene Segmentation and Recognition system

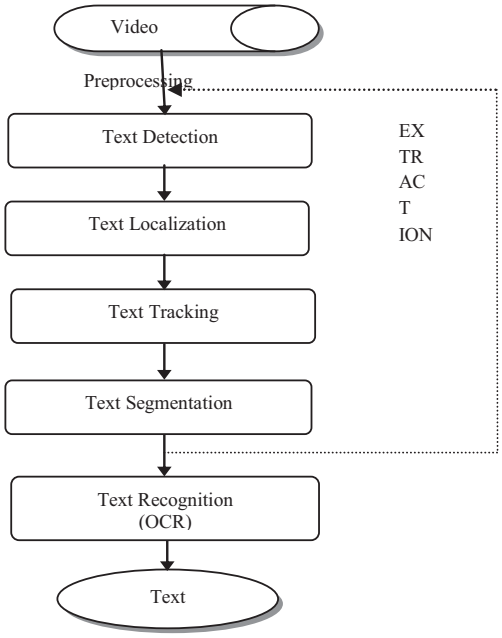
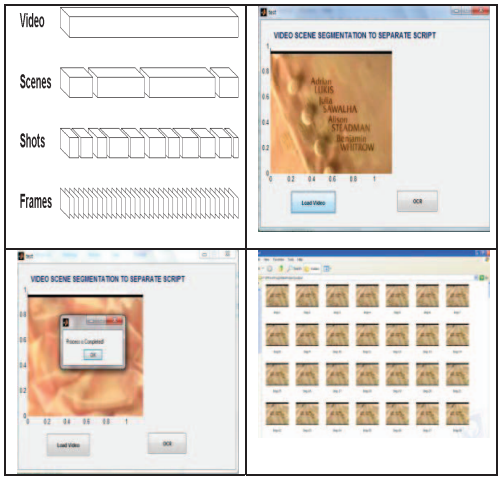


Figure 2.System for character detection and recognition from video/image



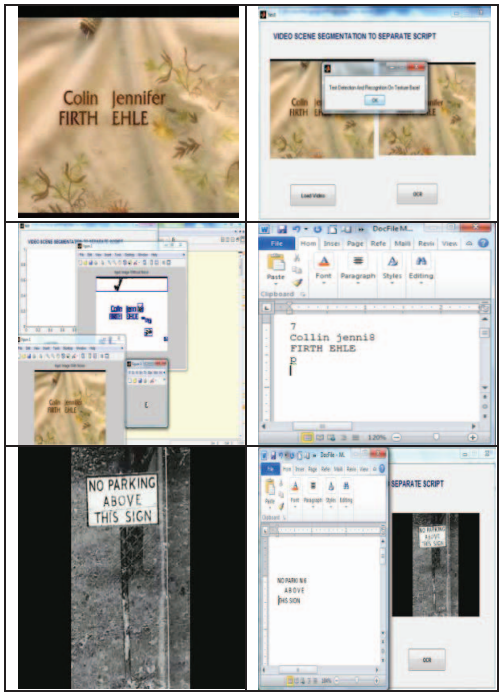


Figure 3. (a) Units for Video data (b) Load video system (c) Video process completed (d) Video into frames (eg.Img-1 to N) (e)Select Video frame (eg.img-50) (f) Detect text (g) Text tracking & Segment texte(h)Character recognizes in word file (i) Select other video frame (eg.Img-150) (j)Text or Character get recognizes.

Description of figure 3: Load Video:-This Button is used for Loading the Video of (*.MPEG,*.AVI) and Converted into frames at destination folder, every frame save with extension JPG.
OCR: This Button is used for optical character recognition; open Frame files then detect the character, Tracking, segmentation and lastly recognition.

C. Character Recognition

Among the 256 ASCII characters, only 94 are used in document images or frame and among these 94 characters, only 80 are frequently used. In the present scope of experiment, we have considered 80 classes recognition problem. These 80 characters are listed in Table 1. These include 26 capital letters, 26 small letters, 10 numeric digits and 18 special characters table 2. These include 26 capital letters, 26 small letters, 10 numeric digits and 18 special characters.

TABLE II. VIDEOS FRAME TEMPLATE CLASSES

1	2	3	4	5	6	7	8	9	0
A	B	C	D	E	F	G	H	I	J
K	L	M	N	O	P	Q	R	S	T
U	V	W	X	Y	Z	a	b	c	d
e	f	g	h	i	j	k	l	m	n
o	p	q	r	s	t	u	v	w	x
y	z	“	;	,	.	#	&	@	(
)	-	%	!	:	‘	$	?	+	/

$$\text{Recall}=\frac{\text{Correct Detected}}{(\text{Correct Detected} + \text{Missed Text Lines})} \quad (6)$$

Whereas precision is defined as:

$$\text{False alarm rate}=\frac{\text{Number of falsely detected text}}{\text{Number of detected text}} \quad (7)$$

$$\text{Precision}=\frac{\text{Correct Detected}}{(\text{Correct Detected} + \text{False Positives})} \quad (8)$$

D. Test result and analysis

TABLE III. EXPERIMENTAL RESULT FOR THE PROPOSED ALGORITHM

#images/frames	370
# textlines	610
#correct detected	583
#false positives	27
Recall (%)	91%
Precision (%)	92%

The following table compares recognition result of improved template matching method and traditional template matching method, the test result is shown as table IV.

Table IV. CHARACTERS RECOGNITION TEST TABLE

Test group	**Recognition Result**
Uppercase	92.05%
Lowercase	91.05%
Digits	93.10%
Special Character	90.05%

Every character set the text box as per the character ,digit, special character size is detected correctly, all character is completely surrounded by a box, some character is not match with template data set then showing other character ,so a detected text box is considered as a false alarm, if no text appears in that box. The text localization algorithm achieved a recall of 91% and a precision of 92%.As seen from the table III&IV, using the improved template matching method, the average recognition rate and Recognition speeds of upper, lower letters, numeric and special characters have been enhanced.

CONCLUSIONS

There are many cases this system are useful for video text information extraction system, vehicle license plate extraction, text based video indexing, video content analysis and video event identification .In this work, we have new approach for character recognition system based on template matching. This system tested the 25 videos with 370 video frames for each video. The system is texture-based approaches to automatic detection, segmentation and recognition of visual text occurrences in images and video frames. The characters are recognized automatically on run-time basis, In a few cases in which 9% characters could not get detected but some other character get recognized . The overall empirical performance of this system recognizing rate is 91% successfully.

REFERENCES

[1] Xian-Sheng Hua, Liu Wenyin, Hong-Jiang Zhang,” Automatic Performance Evaluation for Video Text Detection,” Sixth International Conference on Document Analysis and Recognition (ICDAR2001), pp 545-550, Seattle, Washington, U.S.A., September 10-13, (2001).

[2] Keechul Junga, Kwang In Kimb, Anil K. Jain “Text information extraction in images and video: a survey,” Published by Elsevier Ltd.(2003).

[3] Canny, J., “A Computational Approach to Edge Detection,” IEEE Trans. Pattern Analysis and Machine Intelligence, 8:679-714, November (1986).

[4] H.K. Kim, ECcien ,” automatic text location method and content-based indexing and structuring of video database,”
J. Visual Commun. Image Representation 7 (4) 336–344(1996).

[5] Y. Zhong, A.K. Jain,” Object localization using color, texture, and shape,” Pattern Recognition 33 671–684(2000).

[6] S. Antani, R. Kasturi, R. Jain” A survey on the use of pattern recognition methods for abstraction, indexing, and retrieval of
Images and video, “Pattern Recognition 35 945–965(2002).

[7] Xi Jie, Xian-Sheng Hua, Xiang-Rong Chen, Liu Wenyin, HongJiang Zhang" A Video Text Detection and Recognition System, "IEEE International(2009).

[8] P. Shivakumara, W. Huang and C. L. Tan: Efficient Video Text Detection Using Edge Features, The Eighth IAPR Workshop on Document Analysis Systems (DAS2008), Nara, Japan, pp 307-314(2008).

[9] Rainer Lienhart and Frank Stuber," Automatic text recognition in digital videos," University of Mannheim, Praktische Informatik IV, 68131 Mannheim, Germany.

[10] Qixiang Ye, W. Gao, W. Wang and W. Zeng," A Robust Text Detection Algorithm in Images and Video Frames," IEEE ICICS-PCM, pp. 802-806, (2003).

[11] G. Aghajari, J. Shanbehzadeh, and A. Sarrafzadeh," A Text Localization Algorithm in Color Image via New Projection Profile,"IMECS Hong Kong (2010).

[12] Jayshree Ghorpade, Raviraj Palvankar," Extracting Text From Video," Signal & Image Processing , An International Journal (SIPIJ) Vol.2, No.2, (2011).

[13] Bharatratna Gaikwad Ramesh R. Manza etc," Critical review on video scene segmentation and Recognition ," International Journal of Computer Information Systems (IJCIS), Vol 3, and Number 3, (2011).

[14] Ramesh R. Manza and Bharatratna P. Gaikwad,"A Video Edge Detection Using Adaptive Edge Detection Operator," Issue: January 2012, DOI: DIP012012006, CiiT International Journal of Digital Image Processing: ISSN: 0974–9691 & Online: ISSN: 0974-9586.

[15] Manza R.R., GaikwadB.P., Manza G.R.,"Use Of Edge Detection Operators For Agriculture Video Scene Feature Ex-Traction From Mango Fruits," Advances in Computational Research, ISSN: 0975-3273 & E-ISSN: 0975-9085, Vol 4, Issue 1, 2012, pp.-50-53.

[16] Manza Ramesh R., Bharatratna P. Gaikwad, Manza Ganesh R.," Used of Various Edge Detection Operators for Feature Extraction in Video Scene," ICACEEE-Jan-2012Proc. of the Intl. Conf. on Advances in Computer, Electronics and Electrical Engineering ,ISBN: 978-981-07-1847-3(2012).