

News Video Segmentation And Categorization Using Text Extraction Technique

Harshal Gaikwad, Abhijeet Hapase, Chinmay Kelkar, Nutan Khairnar
Sinhgad College of Engineering, Pune University, India

Abstract

: Video is one of the sources for presenting the valuable information. It contains sequence of video images, audio and text information. Text data present in video contain useful information for automatic annotation, structuring, mining, indexing and retrieval of video. Nowadays superimposed text in video sequences provides useful information about their contents. While video content is often stored in rather large files or broadcasted in continuous streams, users are often interested in retrieving only a particular topic of interest of them. It is, therefore, necessary to split video documents or streams into shorter segments corresponding to appropriate retrieval units. A TV news program comprises continuous video stream containing a number of news headlines. The proposed method consists of division of this video on the basis of news caption, text extraction from partitioned frames, string matching and recombination of frames.

Keywords: Optical character recognition, Pattern Matching, Segmentation, Superimposed text, text extraction, Vision and Scene understanding, Video Analysis

1. Introduction

Today people have access to a tremendous amount of video, both on television and the Internet. The amount of video that a viewer has to choose from is now so large that it is infeasible for a human to go through it all to find video of interest. Traditionally, the images and video sequences have been manually annotated with a small number of key word descriptors after visual inspection by the human reviewer. This process can be very time consuming. One method that viewers use to narrow their choices is to look for video within specific categories or genre. Because of the huge amount of video to categorize, research has begun on automatically classifying video. Text

Information retrieval from video images (Video Annotation) has become an increasingly important research area in recent years for the video

information retrieval and video mining applications. Detection and recognition of text captions embedded in image frames of videos is an important component for video retrieval and indexing. Video text detection and extraction is an important step for information retrieval and indexing in video images. Recognizing video text information directly from videos provides unique benefits. 'NEWS ON DEMAND' provides facilities mentioned above.

Video text may be divided into two types: *Scene text* (which exist in the real-world objects and scenes) and *Superimposed text* or *Graphic text* (which are added during editing processes). Scene text appears within the scene which is then captured by the recording device. It is showing naturally in scenes like text on cloth, street signs, bill boards, and text on vehicle and etc. The appearance of the text is typically incidental to the scene content and only useful in applications such as in navigation, surveillance or reading text appearing on the known objects rather than general indexing and retrieval. It is difficult to detect and extract since it may appear in a virtually unlimited number of poses, size, shapes and colors. Super imposed text is mechanically added into the video frame to supplement the visual and audio content, and is often more structured and closely related to the subject than the scene text is. The superimposed text is one powerful source of high-level semantics. If these text occurrences could be detected, segmented, and recognized automatically, they would be a valuable source of high-level semantics for indexing and retrieval. An example of the superimposed text includes headlines, Keyword summaries, time and locations stamps, names of the people and sports scores. The superimposed text is the most reliable clue for enable users to quickly locate their interested content in an enormous quantity of video data, many research efforts have been put into video indexing and summarization. This superimposed text is used as key feature.

2. Problem Definition

Given a downloaded news video in any format ,the system must be able to segment the required news video into different category for ex political, economics, sports, weather etc and provide the user with necessary news clip as per his/her choice

3. Proposed System

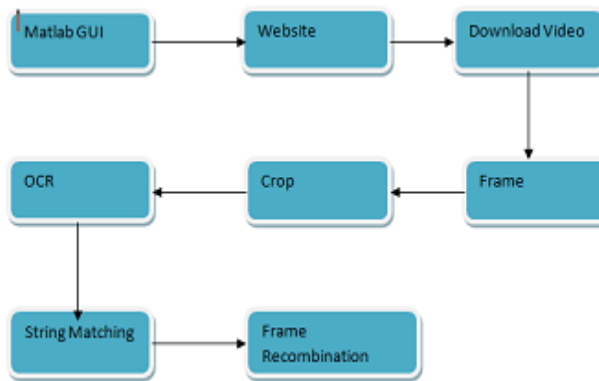


Fig. 2: Preliminary Design

4.Objective

The main objective of the system is to provide the user with the necessary category of the news video from the video of the standard formatted news for ex BBC news, CNN IBN or bollywood news etc

Related Work

I. PROPOSED SOLUTION AND MATHEMATICAL MODEL

The System can be represented as

$$S = \{ I, O, F, D \}$$

Where

I = Input set

O = Output set

F = Set of functions

D = Dead states (Failure states).

Input Set:

$$I = \{ V, F, d \}$$

Where

$V = \{v1, v2, v3, \dots\}$ (Set of videos to be segmented)

$F = \{f1, f2, f3, \dots\}$ (Set of frames)

d = destination of downloaded video.

Output Set:

$$O = \{ V_o \}$$

Where

V_o = Set of partitioned videos.

Set Of Functions:

$$F = \{ F1, F2, F3 \}$$

F1 = Function implemented using Framing tool.

F2 = Function used to extract text from video.

F3 = Function used to club the frames to form video.

Dead states: These are the dead states.

5. METHODOLOGY

A. Matlab framing tool

The Matlab Framing tool is used as a framing tool for the video downloaded by the user. The news video is used as input. It frames the given video into several frames. The frames are also cropped here since the input is needed to be given to the ocr.

B. Optical Character Recognition

Optical character recognition, usually abbreviated to OCR, is the mechanical or electronic conversion of scanned images of handwritten, typewritten or printed text into machine-encoded text. It is widely used as a form of data entry from some sort of original paper data source, whether documents, sales receipts, mail, or any number of printed records. It is a common method of digitizing printed texts so that they can be electronically searched, stored more compactly, displayed on-line, and used in machine processes such as machine translation, text-to-speech and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

What OCR contains?

Since OCR technology has been more and more widely applied to paper-intensive industry, it is facing more complex images environment in the real world. For example: complicated backgrounds, degraded-images, heavy-noise, paper skew, picture distortion, low-resolution, disturbed by grid & lines, text image consisting of special fonts, symbols, glossary words and etc. All the factors affect OCR products' stability in recognition accuracy.

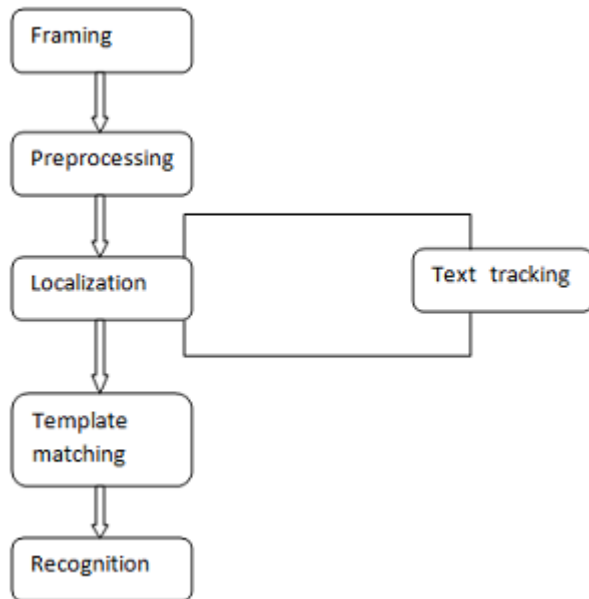


Fig 3.OCR Workflow



Fig 4.Input image required for OCR

The figure shown above shows the intermediate input required for the OCR.

For the OCR, the flow of the system goes like this..

1. Read image.
2. If size=3 goto 3 else goto 4.
3. Convert image to grayscale.
4. Convert image to binary.
5. Update image with opened area.
6. Open text file.
7. Load templates.
8. Calculate template size.
9. Clip image for multiline.
10. Read letter by matching template.
11. Update word.
12. Store word in text file.
13. next image ? goto 9 else goto 14.
14. End

Templates in OCR

Basically the templates are needed to be designed for template matching. Basically these templates are images of letters. The font used for our system. An OCR could be font-free but if you need to be more specific then a font-based OCR can be designed. The font used for our system is rockwell. Initially, the normal letter is created in paint. It is sized to 24*42. Then the image is first converted to binary. Hence all the values in the image are converted to 0 and 1. This basically removes noise from the image as well as the third component of the image is removed which provides ease of reading.

After that, the image is negated i.e if the image contains the black letter on white background, then it converts to the white letter on black background. This process basically provides an ease of matching the template with the letter in the input image. This how the font based template is created for OCR.

The corr2 function in matlab is used for matching of two instance images. $\text{corr2}(A,B)$ computes the correlation coefficient between A and B where A and B are the matrices and vectors of the same size. Since we know that image is represented in the term of matrix, we can easily use this function over here.

C.String matching and recombination of video.

After the text is extracted from the video frame, the text needs to be matched with the text received from the user at the dropdown list at the GUI. The string is matched within the loop. When the strings from both the interfaces are equal, recombination starts from that instance. This recombination continues until the new string is encountered. This recombined video should be shown to the user as the output.

6. CONCLUSION

Nowadays superimposed text in video sequence provides useful information about their contents. Thus our system aims at dividing a particular news into different sections which will provide the ease to our user to see the video section of his own choice. In short our system aims in providing News on Demand feature to our user.

ACKNOWLEDGEMENT

We wish to express true sense of gratitude towards our guides Mrs Nakil nad Mrs G.C Chiddarwar. Both of them contributed their valuable guidance and gave us plenty of their precious time to solve every problem that arose. We are also grateful to our Head of Department for his constant encouragement in the fulfilment of our work.

7. References

- [1] C.Sujatha1 and Dr. D. Selvathi, "An optimal solution for image edge detection problem using simplified gabor wavelet", International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.2, No.3, June 2012.
- [2] Huiping Li, David Doermann, and Omid Kia," Automatic Text Detection and Tracking in Digital Video", IEEE Transaction on image processing, Vol. 9, No. 1, Jan 2000.
- [3] Alfredo Petrosino and Giuseppe Salvi, "A Two-Subcycle Thinning Algorithm and Its Parallel Implementation on SIM Machines", IEEE Transaction on image processing, Vol. 9, NO. 2, Feb 2000.
- [4] Darin Brezeale and Diane J. Cook,"Automatic Video Classification:A Survey of the Literature", IEEE Transactions On Systems, Man And Cybernetics—Part C: Application And Reviews, Vol. 38, No. 3, May 2008.
- [5] Wei Jiang, Kin-Man Lam, and Ting-Zhi Shen,"Efficient Edge Detection Using Simplified Gabor Wavelets", IEEE Transactions On Systems , Man And Cybernetics—Part B: Cybernetics, Vol.39, No. 4, Aug 2009
- [6] Angshumi Sarma and Amrita Ganguly,"An Entropy based Video Watermarking Scheme", International Journal of Computer Applications (0975 – 8887) Volume 50 – No.7, July 2012.
- [7] Alexander G. Hauptmann and Michael J. Witbrock," Story Segmentation and Detection of Commercial In Broadcast News Video, ADL-98 Advances in Digital Libraries Conference, Santa Barbara, CA, April 22-24, 1998.
- [8] Miriam Le´on, Sergio Mallo and Antoni Gasull, "A Tree Structure Based Caption Text Detection Approach", Visualization imaging and image processing.Sept 7-9 2005
- [9] V.Vijayakumar , R.Nedunchezian "A Novel Method for Super Imposed Text Extraction in a Sports Video" International Journal of Computer Applications (0975 8887)Volume15,No.1,February 2011
- [10] Jayshree Ghorpade, Raviraj Palvankar, Ajinkya Patankar and Snehal Rathi , " Extracting text from video". Signal &ImageProcessing : An International Journal (SIPIJ) Vol.2, No.2, June 2011