Embedded Technology Based Image and Video Data Extraction

Ms.S.J.Wamane, T.A.More

PG Student, Dept. of E &T.C, Pravara Rural Engineering College, Loni, Maharashtra, India.

Assistant Professor, Dept. of E &T.C, Pravara Rural Engineering College, Loni, Maharashtra, India.

Abstract: Text data present in video and images contain useful information for automatic indexing, explanation and structuring of images and is also useful for video information recovery and summarization This information is extracted through detection, localization and segmentation of the text from a given image, but dissimilarity of text due to differences in style, size direction, alignment and complex background make the problem of automatic text extraction extremely difficult. Many techniques have been proposed to address this problem, and the purpose of this paper is to classify and review these methods and performance evaluation. We propose new methods to detect and extract the text from the video scene and also extract the audio. System is also implemented using Arm 11 microcontroller in which auto and manual mode is used for video and audio selection. The proposed method is strong about the character of different size, color and location. To reduce the processing time superimposed text region update between the frames is also employed.

Keywords: text detection, text extraction, video information recovery, video summarization, Arm 11 microcontroller.

I. 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 superimposed 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. 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.

We propose a new superimposed detection and extraction method to detect and extract text using Arm 11 microcontroller. First, we generate the detection algorithm to detect text of any online video or image using auto and manual mode. After the detection, text from image or video is extracted through extraction algorithm and obtained result will be displayed on LCD.

II. SYSTEM OVERVIEW

The main objective of the proposed system is to extract the text from videos 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]. The key frames are the one that contains the scene text. 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.

III. PROPOSED METHODS

Text detection and extraction from video and images is become an raising area in the field of information resources relevant to an information need from a collection of information resources to solve the fundamental problem of content based video information retrival.Detection and extraction of text helps to understand the video content with the help of text recognition using optical character recognition (OCR).

A) 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: 1) How to avoid performing computational intensive classification on whole image.2) 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. Overall procedure of the proposed detection method is shown in fig.1.
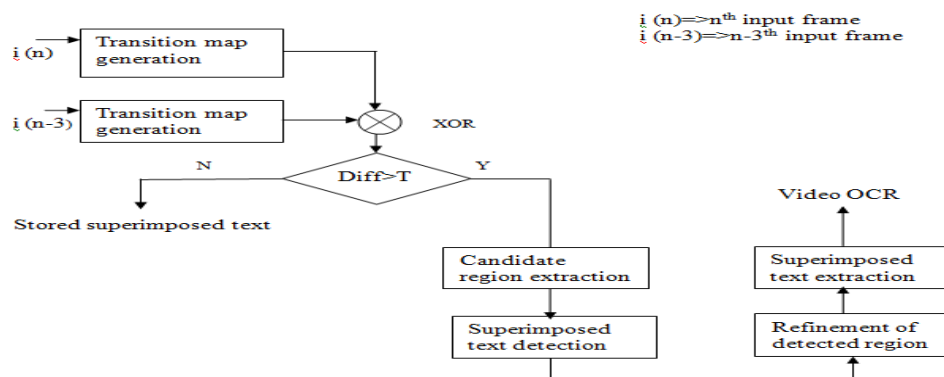

Fig.1. Procedure of the proposed detection method [7]

According to rule of thumb, if the background of superimposed text is dark, then the superimposed text tends to be bright and vice-versa, due to this transient colors between superimposed text and its adjacent background is exist. Therefore, there is logarithmic change in intensity at the boundary of superimposed text [4]. The transition map can be used as a useful indicator for the superimposed text region. Linked map is generated first for the generation of the connected components. If the threshold value is greater than the connected component, they are removed. The TH value is generally selected by observing the minimum size of superimposed text region. Then that connected component is reshaped to have smooth boundaries [7]. The next step to determine the real superimposed text region

depends on the aspect ratio of text region. For better accurate text extraction the overlay text region or the bounding box obtained is needs to be refined. Once the superimposed text regions are detected in the current frame, then to take advantage of continuity of superimposed text between consecutive frames for the text region detection of the next frame [5]. If the difference obtained by XOR of current transition map and previous transition map, is smaller than a predefined value, the superimposed text regions of previous frame are directly applied as the detection result without further refinement.

B) Text Extraction Method

Text extraction is used for converting the greyscale 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 new text extraction system. Overall procedure of text extraction system is shown in fig.2.
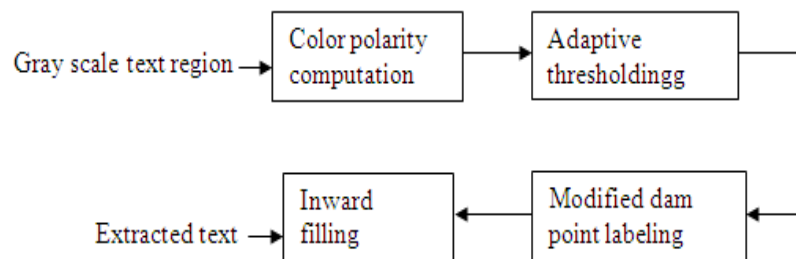

Fig.2. Procedure of proposed extraction method [7]

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. 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 [6]. 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. Finally, corrected characters are obtained from each superimposed text region by the inward filling [7].

IV. BLOCK DIAGRAM OF THE SYSTEM
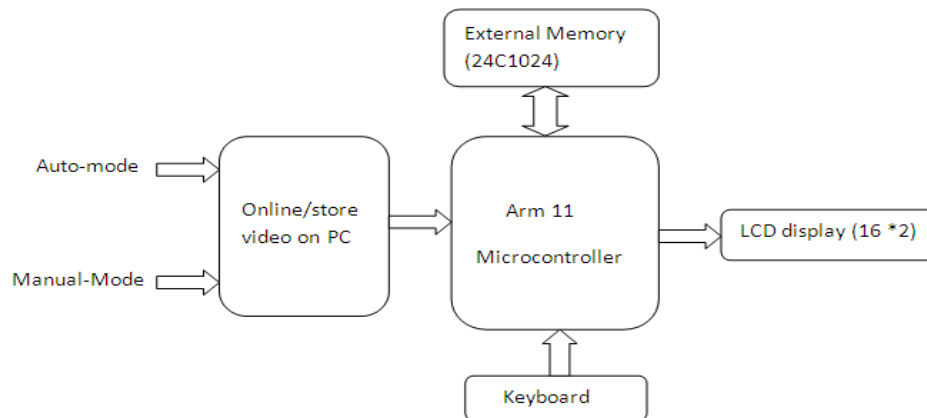
Block diagram of the system is shown in fig.3

Fig .3. Block diagram of the system

We implement hardware to detect and extract data from images and video. The implemented system is operated in two modes namely auto mode and manual mode. These two modes are mainly used for the selection of online or stored video or audio on PC. In auto mode stored or any online video or audio on PC is selected then in order to detect text on that video, it is divided into frames and then detected and extracted by using proposed methods; whereas in manual mode online video or any stored video on pc is manually selected for text detection and extraction.ARM11 microcontroller is interface with PC.Arm 11 is a family of ARM architecture 32-bit RISC microprocessor cores.VB.net is used for coding on Arm 11 microcontroller to detect and extract text on any online or stored video on PC. After that data is again rollback to the pc and displayed on LCD.External memory is used for storing the obtained result.

V. CONCLUSION

A new method for text detection and extraction from videos and images is proposed in this paper. System is implemented using Arm 11 microcontroller. Two modes are mainly used for the images and video selection which includes auto and manual mode. Video is split into frames and key frames obtained. Text region indicator is developed to compute the text existing confidence and candidate region by performing binarization.C or C compiler is used for coding on Arm 11 microcontroller and VB.net is used for data extraction and finally the obtained result will be displayed on LCD. 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.   C. G. M. Snoek and M. Worring, "Time interval maximum entropy based event indexing in soccer video," in *Proc. Int. Conf. Multimedia and Expo*, Jul. 2003, vol. 3, pp. 481–484.
2.   J. Gllavata, R. Ewerth, and B. Freisleben, "Text detection in images based on unsupervised classification of high-frequency wavelet coefficients,"in *Proc. Int. Conf. Pattern Recognition*, Aug. 2004, vol. 1, pp.425–428.
3.   J. Cho, S. Jeong, and B. Choi, "News video retrieval using automatic indexing of korean closed-caption," *Lecture Notes in Computer Science*,vol. 2945, pp. 694–703, Aug. 2004.

4.   L. Agnihotri and N. Dimitrova, "Text detection for video analysis," *Proc. IEEE Int.Workshop on Content-Based Access of Image and Video Libraries*, Jun. 1999, pp. 109–113.
5.   X. S. Hua, P. Yin, and H. J. Zhang, "Efficient video text recognition using multiple frame integration," in *Proc. Int. Conf. Image Processing*,Sep. 2004, vol. 2, pp. 22–25.
6.   K. C. K. Kim *et al.*, "Scene text extraction in natural scene images using hierarchical feature combining and verification," in *Proc. Int. Conf. Pattern Recognition*, Aug. 2004, vol. 2, pp. 679–682.
7.   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.