International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Textual Metadata Retrieval of Lecture Videos Using Optical Character Recognition

Aditi V. Lawate[1], M. M. Wankhade[2]

[1]Sinhgad College of Engineering, Savitribai Phule Pune University, Vadgaon (Bk), Pune-041, India

[2]Professor, Sinhgad College of Engineering, Savitribai Phule Pune University, Vadgaon (Bk), Pune-041, India

Abstract: ***The increase of video lecture data on World Wide Web is rapid therefore an efficient method of data retrieval is needed. So the system providing a method for data retrieval from the lecture video is implemented which will extract the text data. Automatic video segmentation and then key frame detection is applied first. Optical Character Recognition (OCR) Technology is applied to extract the textual metadata. The extracted text data is saved in the form of templates for future reference. The 80% of accuracy in recognising the letters from lecture videos is achieved and the extracted information is saved so that the quality of learning is improved.***

Keywords: OCR, slide detection, character segmentation, feature extraction.

1. Introduction

In the age of e-learning, the amount of lecture data or any video data providing an approach for e-learning is increasing daily. So, it is required to design a system which will retrieve the textual data from the lecture video. In order to fulfill the need of data retrieval from the lecture videos, a system which provides an approach for data retrieval from the lecture video can be implemented.

The main aim of the system is to design is to provide an efficient way of data retrieval from a lecture video. First of all apply the automatic video segmentation. Subsequently extract the textual metadata by applying video Optical Character Recognition (OCR) technology on the key frames. This extracted information can be used to improve the quality of learning. So one can also store it in the text lines in the form of templates and can be used whenever it is needed without going through the lecture video again.

2. Literature Review

In the last decade e-lecturing has become more and more popular. Now a days the amount of video lecture data on world wide web is growing rapidly. Author Haojin Yang and Christoph Meinel [1], have presented an idea for content based lecture video retrieval using speech and video text information. Their idea is a more efficient method for video retrieval in World Wide Web or within large lecture video archives is urgently needed. They presented an approach for automated video indexing and video search in large lecture video archives.

For the same purpose, they presented an idea which includes automatic video segmentation and key-frame detection to offer a visual guideline for the video content navigation. Also author E. Leewis, M. Federico [2] stated that the word error rate (WER) for the character recognition from any video is in between 40-60%. Wang et al. proposed an approach for lecture video indexing based on automated video segmentation and OCR analysis [3]. The proposed segmentation algorithm in their work is based on the differential ratio of text and background regions. Using thresholds they attempt to capture the slide transition. The final segmentation results are determined by synchronizing detected slide key-frames and related text books, where the text similarity between them was calculated as indicator. They also apply the synchronization process between the recorded lecture video and the slide file which has to be provided by the presenters.

Using this approach one can develop a system in which we can extract the text from the lecture video and also can save the information in the form of templates which is used to generate the relevant documents and to improve the quality of learning.

3. Proposed Approach

This system presents an approach for getting the textual information from a given video lecture. The video lecture here must be taken in .avi format. If the video is in any other unspecified format then it must be converted into .avi format. Frame rate of the video is 30 frames/second. So, every fifth frame is taken into the consideration to increase the processing speed and there is no loss of information.

3.1 Slide Detection

For detection of new slide in the video, the motion detection method is used. Difference between the pixels of two successive frames is calculated and the difference value is more than 10000 then that slide is taken as new slide. Such detected slides are taken for the further processing.

3.2 Line Separation

After the slide detection, lines must be separated in order to extract the words and letters. There is always noise around the extracted lines in the frames, therefore we need to remove or reduce the noise before segment the characters. We can use horizontal projection to see the distribution of the pixels (Figure 1). Horizontal summation method is used to segment

the character line of the frame. The summation of horizontal elements of image (Row wise) & then finding the nonzero indexes of summation, gets the character line segmentation.

Figure 1: Horizontal Summation with Histogram

3.3 Character Segmentation

There is always noise around the characters, so we need to distinguish the characters from the noise. We can use vertical projection to see the distribution of the pixels (Figure 2). Vertical summation method is used to segment separate the characters of the extracted lines. The summation of vertical elements of image (Column wise) & then finding the nonzero indexes of summation, separates the characters.

Figure 2: Vertical Summation with Histogram

3.4 Character Weights

3.4.1. Size Normalization

The size of the character images is an important factor for the accuracy of character recognition. All the characters images are normalized to predefined height (Vertical Length) in pixel (Figure 3). The characters shall have variable width (Horizontal Length). The scaling shall depend on the calculated vertical length of the characters.

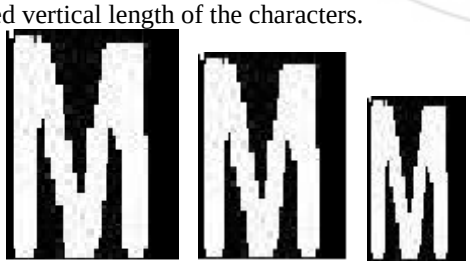

Figure 3: Size Normalized Image

3.4.2. Thinning Process

After size normalization each letter or number, the character skeleton can be found by using the thinning process. The width is defined pixels and the structure is then used to recognize the letters and numbers. Thinning shall improve the recognition rate, by avoiding the variations in depth (Thickness of the characters) of the characters & numbers.

3.4.3. Image Mapping

After thinning each letter is mapped in the left top area of the predefined map of 64*64. The single block is represented by 8*8 pixels. The complete map is formed with 64 blocks (Figure 4).

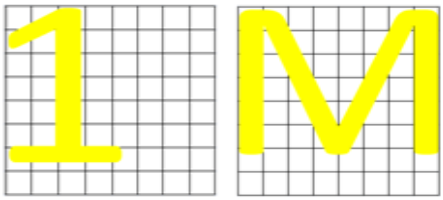

Figure 4: Image Mapping

3.4.4. Horizontal and Vertical Weights

The horizontal weight shall be calculated using horizontal line scanner, the horizontal line map shall slide across the character & the weights are added where the union of the character the line is matched. As shown in the figure 5 the horizontal weight of the Character is 1264.

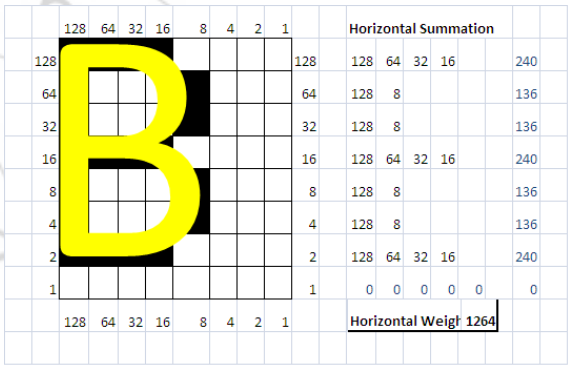

Figure 5: Horizontal Weight

The vertical weight shall be calculated using vertical line scanner, the vertical line map shall slide across the character & the weights are added where the union of the character the line is matched. As shown in the figure 6 the vertical weight of the Character is 808.

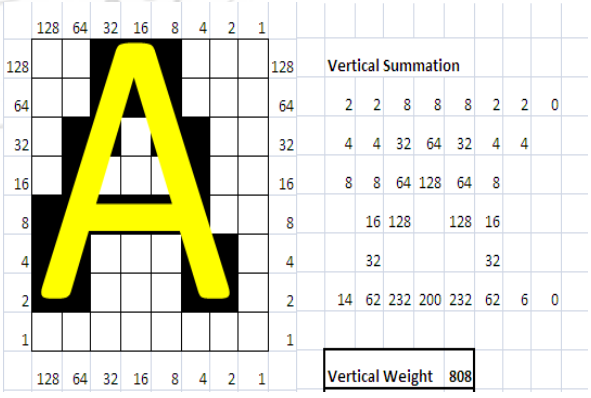

Figure 6: Vertical Weights

The horizontal and vertical weights of all the characters and numbers are calculated and stored in the database, which shall be used to compare with the characters to be recognized.

3.5 Optical Character Recognition (OCR)

The weights of all segmented characters are calculated and then compared with the weights of the stored database. The minimum distance between the vertical and horizontal weights identifies the character (Figure 7). For minimum distance identification, the threshold value can be set in order to recognize the letter.

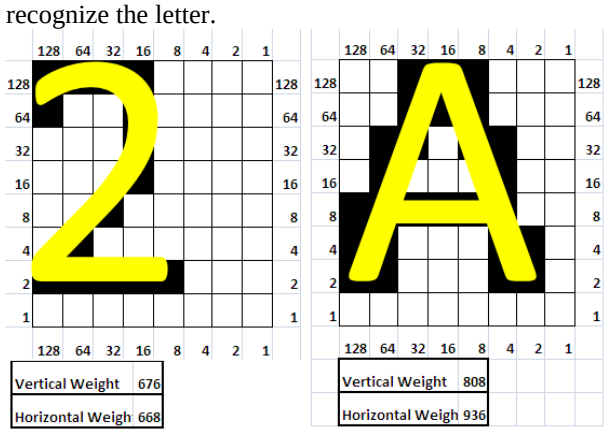

Figure 8: Weights for comparison

4. System Flowchart

The system flowchart shows the algorithm. Using this algorithm the textual metadata is extracted and is saved in the form of text which can be used for future reference (Figure 9).

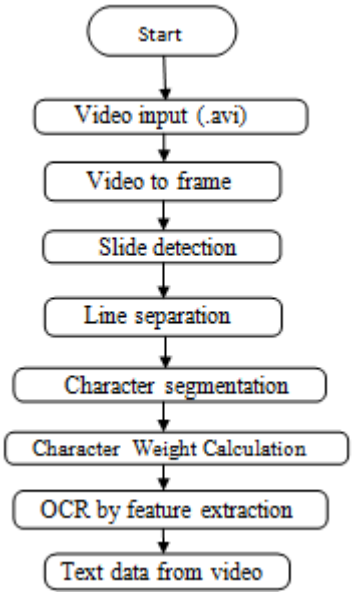

Figure 9: System Flowchart

4. Results

The system mentioned above can extract the textual metadata and the accuracy for character recognition is about 80%. Hence one can use the system for text extraction. The following images shows the captured input frame of the video (Figure 10) and its corresponding text output by applying OCR algorithm (Figure 11).

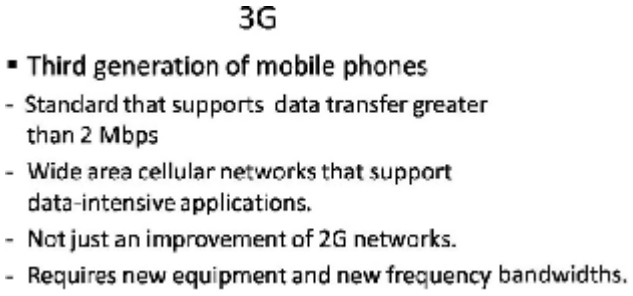

Figure 10: Input Frame

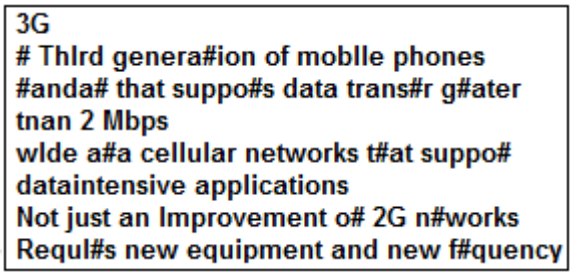

Figure 11: OCR output

References

[1] Haojin Yang and Christoph Meinel, "Content Based Lecture Video Retrieval Using Speech And Video Text Information" in IEEE Transactions On Learning Technologies, VOL. 7, NO. 2, April-June 2014.
[2] E. Leeuwis, M. Federico, and M. Cettolo, "Language modeling and transcription of the ted corpus lectures," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 2003, pp. 232–235.
[3] T.-C. Pong, F. Wang, and C.-W. Ngo, "Structuring low-quality videotaped lectures for cross-reference browsing by video text analysis," J. Pattern Recog. , vol. 41, no. 10, pp. 3257–3269, 2008.
[4] H. Yang, B. Quehl, and H. Sack. (2012), "A framework for improved video text detection and recognition," Multimedia Tools Appl., pp. 1–29, [Online].
[5] M. Grcar, D. Mladenic, and P. Kese, "Semi-automatic categorization of videos on videolectures.net," in Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases, 2009, pp. 730–733.
[6] www.computer.org/cdsl/trans/lt/2014/02/06750040-abs.html

Author Profile

Aditi Lawate received the B.E. degree in Electronics from Solapur University in 2013. Now is pursuing M.E. degree in Signal Processing (E&TC) from Savitribai Phule Pune University. Her area of interest is signal processing and image processing.

Prof. M. M. Wankhade is assistant professor in Electronics and Telecommunication department at Sinhgad College of Engineering, Pune. She has completed her B.E. degree in Electronics and Telecommunication from Amravati University and M.E. in Electronics and Telecommunications from Savitribai Phule Pune University. She has published around 20 papers in national and international journals and conferences.