Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 11, November 2014, pg.275 – 282

SURVEY ARTICLE

A SURVEY ON CONTENT BASED LECTURING VIDEO RETRIEVAL

Vigneshwari.G
vigneshwari2411@gmail.com
M E. Scholar, Department of Computer Science and Engineering, NPR College of Engineering Natham, Anna University, TamilNadu, India

Abstract: ***Digital video has become a popular storage and exchange medium due to the rapid development in recording technology. E-lecturing has become more efficient popular. The quantity of lecture video data on the World Wide Web is growing quickly. Therefore, a more proficient method for Video Retrieval within huge lecture video archives is immediately desired. This paper comes close to routine video indexing and video hunt in large lecture videos collection. We apply Automatic video segmentation and key-frame detection to present a image instruction for the video content navigation. In this process, video will be retrieved based on content based video search for the input of speech and video text content. Adaptive Thresholding algorithm is used for detecting text from the video. Histogram of gradient is used for the feature extraction from the video. Classification is done in order to get the accurate result. This will collect required features from lecture video database as labeled with speech recognition and also retrieve Lecture video from database as better than existing algorithm. This better results than existing algorithms. Hence the proposed system will be more efficient for retrieving the images and also improves the recognition rate.***

Keywords: Video Segmentation, Key-frame Selection, Feature Extraction, classification etc.

----------***----------

1. INTRODUCTION

There is an astonishing growth in the amount of digital video data in recent years, but requirement of tool to classify and retrieve the video is not in current. A numeral of colleges and explore institutions are enchanting the chance to record their lectures and publish them online for students to access free of time and location. As a result, there has been an enormous increase in the sum of multimedia data on the Web. An e-lecture can be defined as a media based lecture including an audio or video recording, synchronized slides, table of contents and optional complementary information. Most of all, an e-lecture consists of slides with relevant points mentioned by the lecturer. In only few e-lectures, one can find a written transcript of the oral presentation. An e-lecture can offer additional links with additional information and learning materials for learners. But additional information like external links is not always helpful for learners. Text is a high-level semantic feature which has often been used for content-based information retrieval. In lecture videos, texts from lecture slides serve as an outline for the lecture and are very important for understanding. In order to evaluate the usability, the implementation is done in a large lecture video portal. The developed video analysis methods have been evaluated by using compiled test data sets as well as opened benchmarks. All compiled test sets are publicly available from our website for the further research use. The extraction of

metadata from visual as well as audio resources of lecture videos automatically is done by applying appropriate analysis techniques. For evaluation purposes, several automatic indexing functionalities is developed in a large lecture video portal, which can guide both visually and text oriented users to navigate within lecture video. A user study that intended to verify their search hypothesis and to investigate the usability and the effectiveness of proposed video indexing feature. In this paper, we are presenting Content Based Video Retrieval (CBVR) System it includes various steps: Video Segmentation: Adaptive Thresholding algorithm is used for image segmentation, Feature Extraction: Features are extracted for the key frame and stored into feature vector. Histogram Of Gradient is an algorithm which is used for the feature extraction. Classification: It is a learning algorithm which is used for detection and finally performance analysis is done based on the processing speed of feature extraction of videos and frames, and their effectiveness evaluation of videos.

2. RELATED WORK

In this approach, desktop of the speaker's computer is used which captures his presentations. This method performs by during the lecture through a frame grabber tool. In this method the detection of video is flexibility. This type of method not required extra synchronization between video and slide files and not need to take care of the slide format.

Another research work has been established which can be performed by the screen grabbing method. In this method the temporal scope of a complete unique slide can be considered as a lecture segment. In this segmentation, segmenting two scenes lecture videos can be achieved by only processing slide video streams, which contain most of the visual text metadata. The extracted slide frames can provide a visual guideline for video content navigation. After segmenting a video file into a set of key frames, the text detection procedure will be executed on each key frame, and the extracted text objects will be further used in text recognition and slide structure analysis processes. This information can be applied for automatic lecture video indexing.

A lecture video segmentation method using Scale Invariant Feature Transform feature (SIFT) and the adaptive threshold. In this work SIFT feature is applied to measure slides with similar content. An adaptive threshold selection algorithm is used to detect slide transitions. Another method proposed to annotate lecture video resources by using Linked Data. Their framework enables users to semantically annotate videos using vocabularies defined in the Linked Data cloud. Most of the existing lecture speech recognition systems in the reviewed work cannot achieve a sufficient recognition result which calculated based on Word Error Rates(WERs).The poor recognitionresults not only limit the usability of speech transcript, butalso affect the efficiency of the further indexing process.The video analysis methods may introduce errors due to not take care of the slide format.

A number of content-based video search engines encompass proposed recently. Adcock et al. proposed a lecture webcast search system in which they applied a slide frame segmented to extract lecture slide images. The system retrieved additional lecture videos from different resources such as YouTube, Berkeley Webcast, etc. The search indices are created based on the total metadata obtained from the video hosting website and texts are extracted from slide videos by using a standard Optical Character Recognition [OCR] engine. Since they do not apply text detection and text segmentation process, the OCR recognition accuracy of their approach is therefore lower than our systems. Furthermore, by applying the text detection process we are able to extract the structured text line such as title, subtitle, key-point, etc., that enables a more flexible search function.

3. ARCHITECTURAL DIAGRAM

The following figure shows the architectural block diagram;

In this, video will be converted to frame.. Then preprocessing is done, and noise has been removed from each frame. After that Edge detection process is done to predict the edge. Next feature has been extracted from each frame. Finally classification is used to classify the text frame from the all detected frames.

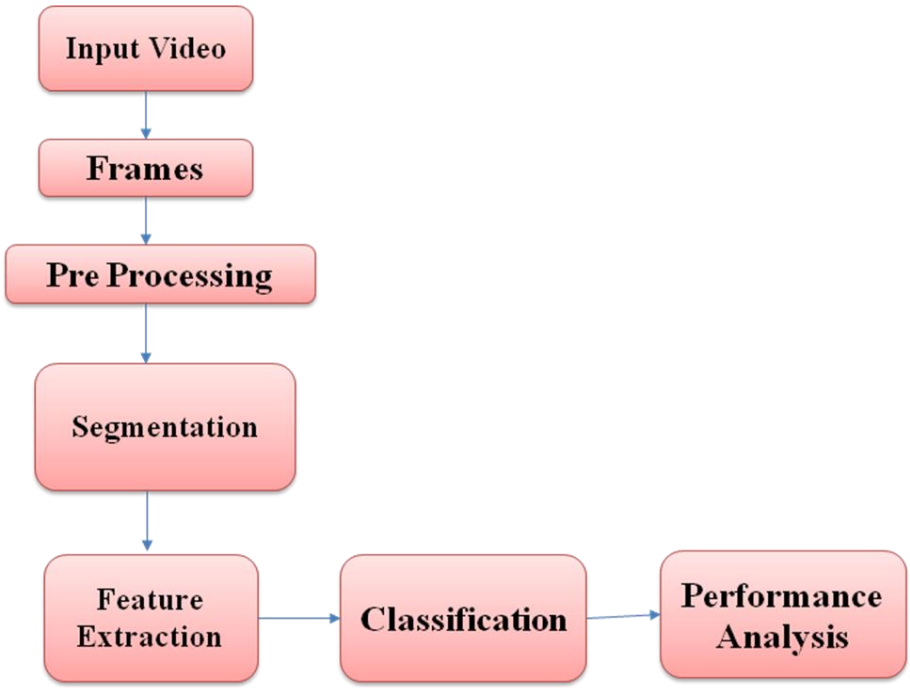


Fig -1: Architectural diagram

4. SYSTEM FLOW

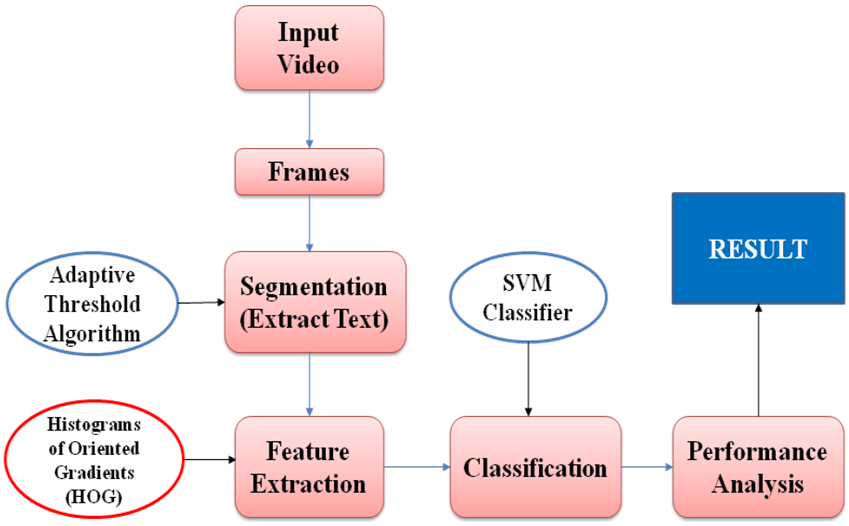


Fig -2: Execution flow of system

The above figure shows the flow of system; During frame conversion videos are separated into image format. Then, Preprocessing is done to remove noise from each frame. After that canny edge detector has been applied to predict the edge. This process is repeated until all frames edge will be detected. Next, Histogram Of Gradient feature extraction method is used for extracting the feature. Finally SVM classifier is used to classify the text frame from the all detected frames. The Canny edge detector is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images. This edge detector

method detects the edge in frames. Histogram of Oriented Gradients (HOG) are feature descriptors which counts occurrences of gradient orientation in localized portions of an image. HOG feature extraction method extracts gradient values of all frames. Finally, Support vector machines (SVM) classifier is used for classification. SVM are supervised learning models with associated learning algorithms, given a set of training examples, each marked as belonging to one of two categories, SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier.

4.1 Video Segmentation

Video segmentation is first step towards the content based video search aiming to segment moving objects in video sequences. Segmentation of Video is done with the help of step by step process of video segmentation.

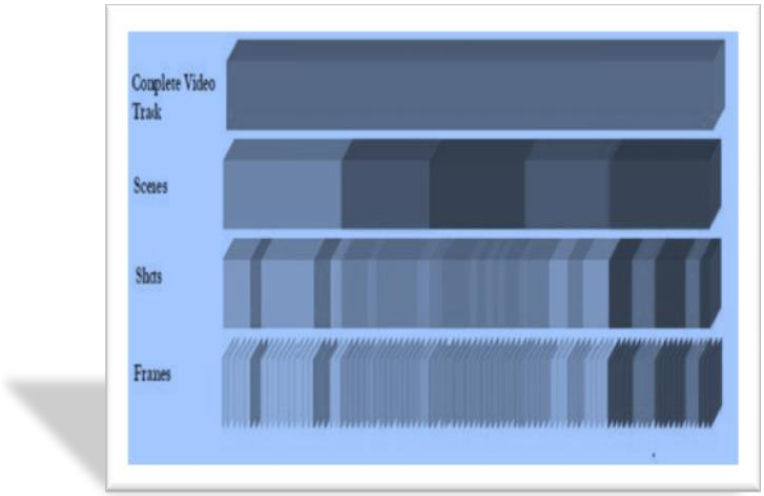


Fig -3:Video Segmentation

The complete video is first converted into scenes, then scenes are converted into shots and finally shots are converted into various frames.

4.2 Feature Extraction

Histogram Of Gradients is an algorithm which is used for the feature extraction. It is extracted based on the histogram of the feature. HOG are feature descriptors used in image processing and computer vision for the purpose of object detection. The technique counts occurrence of gradient orientation in localized portions of an image.

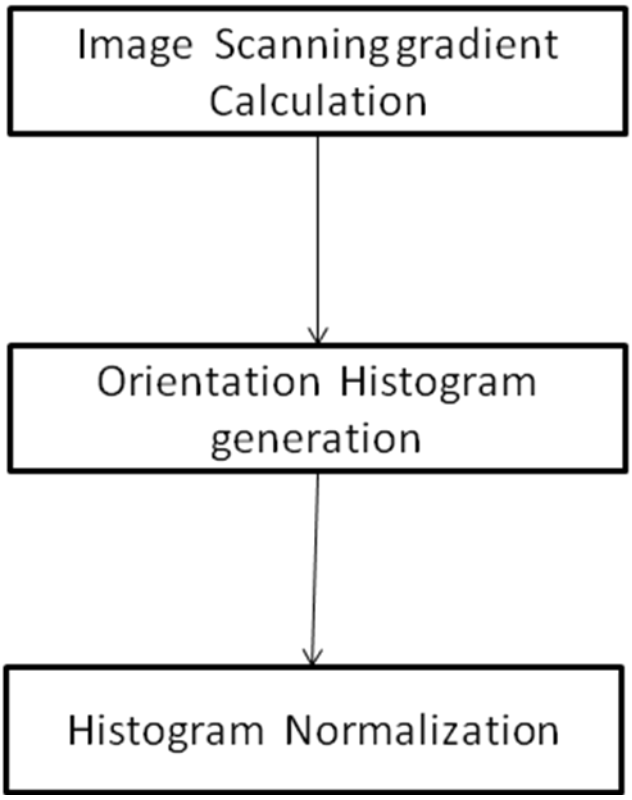


Fig -4:Gradient Calculation

4.3 Classification

It is a learning algorithm which is used for detection. The videos are retrieved using this algorithm. A SVM model is a representation of the model (examples) as points in space mapped. The examples of the separate categories are divided by a clear gap that is as wide as possible New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

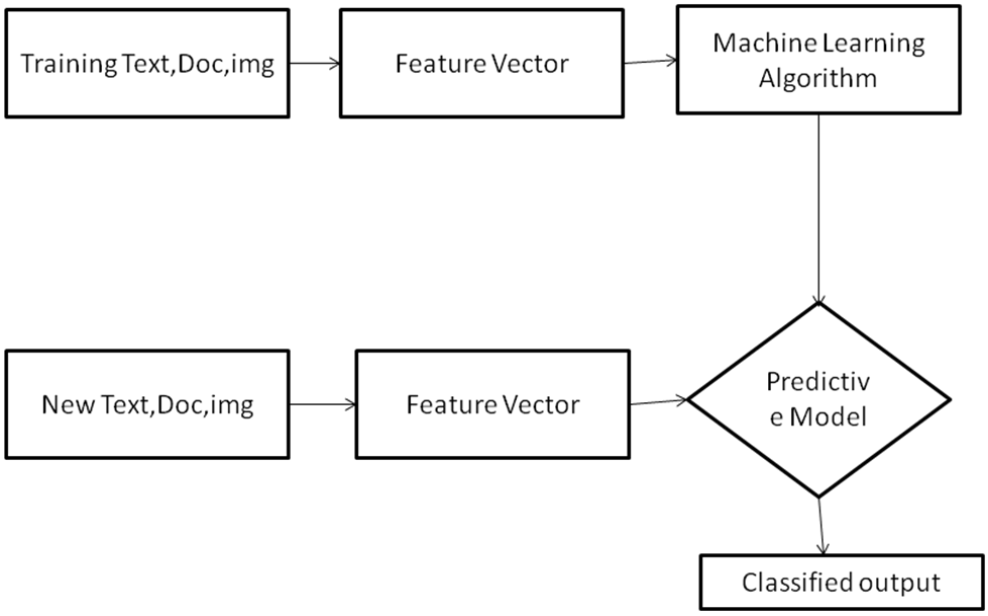


Fig -5:SVM classification

5. PERFORMANCE EVALUATION

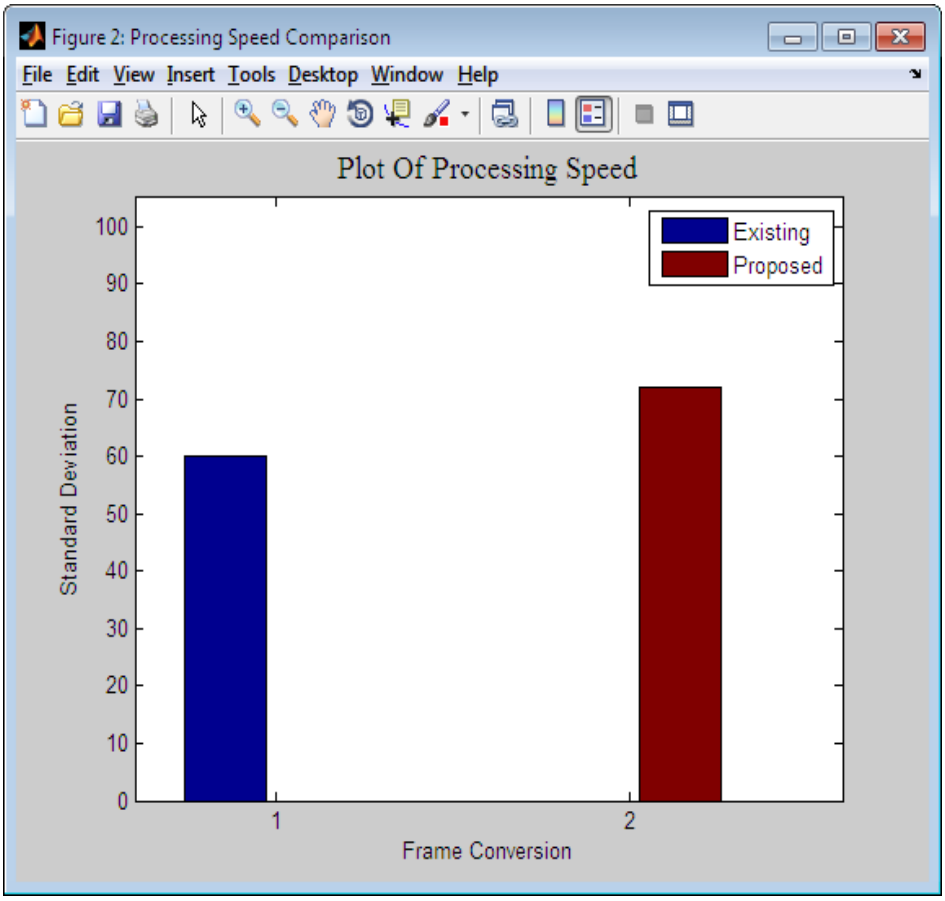


Fig -5:Procesing Speed Comparison

The analysis of the retrieval accuracy of the method. The analysis processing of speed in terms of mean processing time and standard deviation is determined.

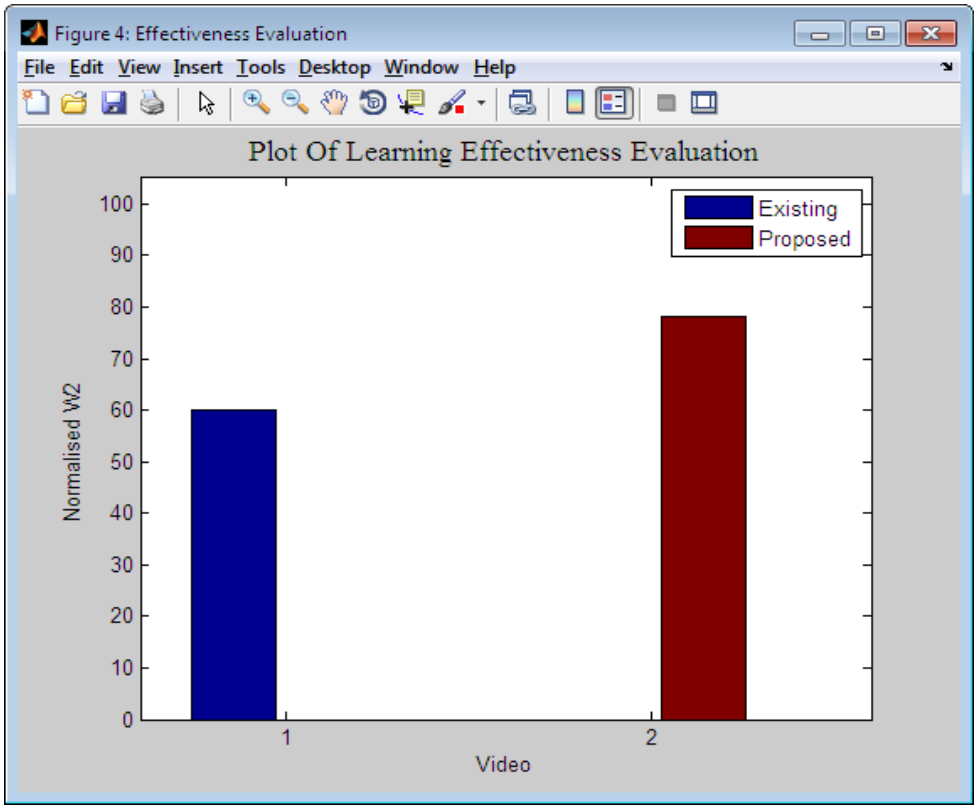
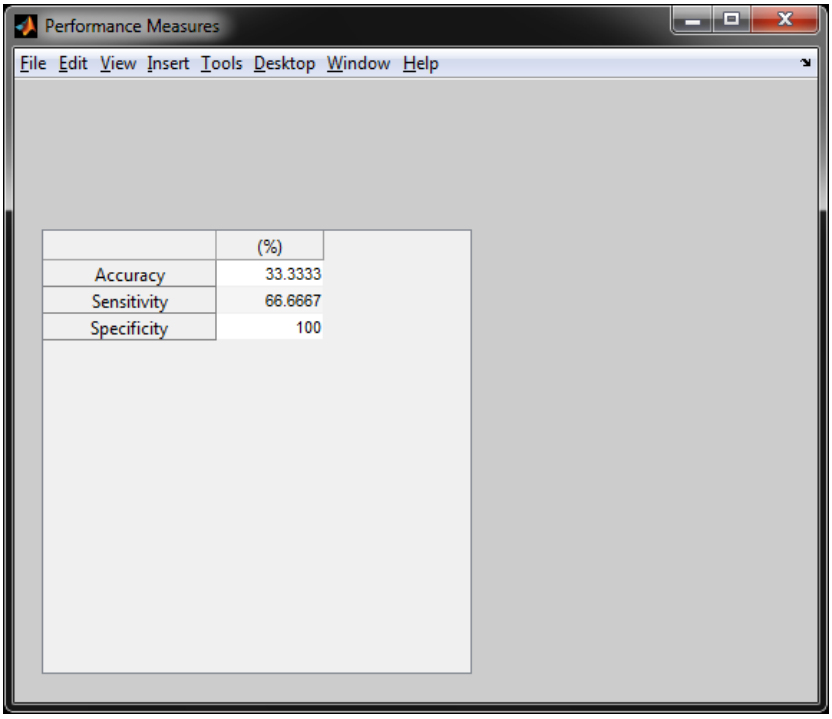


Fig -5:Effective Evaluation Comparison

5.1 Performance Measures

1. Accuracy
2. Sensitivity
3. Specificity



After final retrieval of the video from the database, the text is retrieved from the video.

6. CONCLUSION

This paper presented an approach for content-based lecture video indexing and retrieval in large lecture video archives. The text is retrieved with the usage of the SVM classification and the HOG feature extraction method. The main process of processing speed of both videos and the feature extraction is analyzed and the effectiveness evaluation is higher compared with the presented one.

REFERENCES

[1] E. Leeuwis, M. Federico, and M. Cettolo, "Language modeling and transcription of the ted corpus lectures," in Proc. IEEE International Conference., Speech Signal Process., 2003.

[2] D. Lee and G. G. Lee, "A korean spoken document retrieval system for lecture search," in Proc.ACM Special Interest Group Inf. Retrieval Searching Spontaneous Conversational Speech Workshop, 2008.

[3] J. Glass, T. J. Hazen, L. Hetherington, and C. Wang, "Analysis and processing of lecture audio data: Preliminary investigations," in Proc. HLT-NAACL Workshop Interdisciplinary Approaches Speech Indexing Retrieval, 2004, pp. 9–12.

[4] A. Haubold and J. R. Kender, "Augmented segmentation and visualization for presentation videos," in Proc. 13th Annul. ACMInt. Conf. Multimedia, 2005, pp. 51–60.

[5] W. H€urst, T. Kreuzer, and M. Wiesenh€utter, "A qualitative study towards using large vocabulary automatic speech recognition to
index recorded presentations for search and access over the web,"in Proc. IADIS Int. Conf. WWW/Internet, 2002, pp. 135–143.

[6] C. Munteanu, G. Penn, R. Baecker, and Y. C. Zhang, "Automatic speech recognition for webcasts: How good is good enough and what to do when it isn't," in Proc. 8th Int. Conf. Multimodal Interfaces, 2006.

[7] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Inf. Process. Manage., vol. 24, no. 5, pp. 513–523, 1988.

[8] G. Salton, A. Wong, and C. S. Yang. (Nov. 1975). A vectorspace model for automatic indexing, Commun. ACM, 18(11),pp. 613–620, [Online]. Available: http://doi.acm.org/10.1145/
361219.361220

[9] T.-C. Pong, F. Wang, and C.-W. Ngo, "Structuring low-qualityvideotaped lectures for cross-reference browsing by video text analysis," J. Pattern Recog., vol. 41, no. 10, pp. 3257–3269, 2008.

[10] 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.

[11] T. Tuna, J. Subhlok, L. Barker, V. Varghese, O. Johnson, and S.Shah. (2012), "Development and evaluation of indexed captioned Searchable videos for stem coursework," in Proc. 43rd ACM Tech.Symp. Computer. Sci. Educ.
[12] H.J.Jeong, T.-E. Kim, and M. H. Kim.(2012), "An accurate lecture video segmentation method by using sift and adaptive threshold,"in Proc. 10th Int. Conf. Advances MobileComput., pp.285–288.

[13] H. Sack and J. Waitelonis, "Integrating social tagging and document annotation for content-based search in multimedia data," inProc.1st Semantic Authoring Annotation Workshop, 2006.

[14] C. Meinel, F. Moritz, and M.Siebert, "Community tagging in tele-teaching environments," in Proc. 2nd Int. Conf. eEduc., Bus.,e-Manage. and E-Learn., 2011.

[15] S. Repp, A. Gross, and C. Meinel, "Browsing within lecture videosbased on the chain index of speech transcription," IEEE Trans.Learn. Technol., vol. 1, no. 3, pp. 145–156, Jul. 2008.

