Asian Journal of Convergence in Technology Volume IV Issue I
ISSN NO: 2350-1146 I.F-5.11

A Novel Approach of Inhalt Based Video Recuperation System Using OCR and ASR Technologies

Swati Khokale,Assistant Professor, Sandip Institute of Technology and Research Centre,Nashik

Apurva Kandelkar,Assistant Professor, Sandip University,Nashik

***Abstract*—Now a days Lecture videos are present methods for E-learning process. The degree of lecture video data on the (WWW)World Wide Web is growing very quickly. Therefore maximum appropriate technique for retrieving video within vast lecture video library is required. This technique will be very beneficial for the new and existing users to search related video within a short period of time. This paper shows a different method for inhalt based video searching for receiving correct outcomes. The main aim of the proposed system is to recover a video on the basis of its substances rather than retrieving video consulting to its title and meta data explanation in order to provide an correct for the examine query. For mining text data printed on slides we put on optical character recognition algorithm(OCR) and automatic speech recognition algorithm(ASR) to convert speaker's speech into text.**

Keywords—Automatic Speech Recognition Algorithm.(ASR),E-learning Videos, Optical Character Recognition (OCR).

I. INTRODUCTION (*HEADING 1*)

In today's situation, because of the reliable scene property of the designed video, appropriate outcomes cannot be useful to videos based on graphical feature concept. For the flexible communication, abilities to makes videos in paired scene format that shows talker and his offering slides simultaneously. Exhibition method is used for greater level of understand ability of students. These videos are quickly used by the scholars for e-learning. For this purpose, records of Institutions upload their lecture videos on internet. Internet is making a great amount of videos .It's a complex job to search a suitable video according to finding query. Because at the time users search a lecture video, Outcomes displayed according to title of video not on the base of contents. Otherwise, sometimes it will be possible like finding data may be covered within few minutes. Thus, user wants to show this data within a small period of time without going through a whole video. The difficulty is that one can't retrieve the correct data in a huge lecture video archive extra efficiently. Each video recovery search engine like You-Tube and other response is based on existing textual relative data like video title and its explanation, etc. Normally, this kind of meta data has to be formed by a human to checked better-quality, but the stage of formation is time and cost consuming. The main aim of the system is to recover a video on the basis of its substances rather than retrieving video consulting to its title and metadata explanation in order to provide an correct result for the examine query. For that purpose, we implement a model which captures the numerous frames from a video lecture. Caused captured frames are then distinguished according to the repetition features. Video fragmentation is completed after specific time interval within two sequential frames. It is a chance of that the video lecture holds one slide presentation for a insufficient period of time. So to solve this difficulty, maximum time interval is used in seconds for key frames segmentation. All the text from all the frames for advance video retrieval system is abstracted using OCR algorithm. Likewise converting all the voice resulting into text using ASR algorithm technique. So, this is used in the process of video retrieval system. The associated data (Text, Voice and Images from Video) is used for content based video retrieval system and gathering of video according to their text, image and voice parameters. This OCR algorithm is answerable for extracting characters from the textual data as well as ASR algorithm is useful to retrieve the speaking data from the video speech. The OCR and ASR record as well as identified slide text line types are expected for keyword extraction, with the benefit of which video keywords are log on for surfing and searching content-based video. The suggested system is evaluated on the base of performance and the practicality. The OCR algorithm will play a important role as it will deliver us the textual information from the key frames provided by Video Segmentation methods, which in chance will be used as key-words for the video. Similar with Automatic Speech Recognition method, it will provide resulting key audio signals for the video. The Video Indexing and Retrieval systems will too play their part in responding the user with the matching data with the user questions. The main concern that scholars have to focus on is the generation of key words for the video, both textual and audio, as we do for the textual data.

II. RELATED WORK

Haojin Yang[1] proposed an approach "Content Based Lecture Video Retrieval Using Speech and Video Text Information" for the purpose of automatic video indexing and video retrieval. The method used by them is OCR used for slide video segmentation that perform video segmentation. The video is transformed into its frames for gathering text data to each frame and ASR used for translation of speech-to-text data from lecture videos. The main disadvantage of this method is that it does not work on Open Data resources for addition of all the feature abstraction and appropriate outcomes. John Adcock [2] proposed an approach "Talk Miner: A Lecture Webcast Search Engine" for ad-hoc video capturing scenarios, Frame Differencing and Slide Detection. Some difficulties were visible when trying to recognize different slides in the video stream. Example , image-in- image compositing of a speaker and a showing slide, swapping cameras, and slide constructs complicate simple frame differencing algorithms for extracting key frame slide images. The technique they used OCR for slide discovery and frame differencing and lexical procedures'. V. Patel[3] introduce an approach "Content Based Video Retrieval by Entropy, Edge Detection, Black and White Colour Features" for video recovery. They proposed an approach for retrieval of criminal information e-learning, news video browsing, digital multimedia library retrieval and defence applications. For the purpose of Content Based Video Recovery they achieved using formation of feature database and video recovery algorithms. Boris Epshtein[4] developed an approach "Detecting Text in Natural Scenes with Stroke Width Transform" that look for finding the value of stroke width for each image pixel, and show its use on the task of text detection in regular images. The technique used by them is the text discovery algorithm for providing a feature that has verified to be dependable and flexible for text discovery and The Stroke Width Transform for fast answers. Stephan Repp[5] proposed an approach "Browsing within Lecture Videos Based on the Chain Index of Speech Transcription" that permits browsing in units of videos from a multimedia knowledge base. The outcomes are possible to improve data to the result set that cares the learners with the helpful data searching. They used OCR method for automatic indexing of multimedia videos.Arpit Jain[6] suggested "Text Detection and Recognition in Natural Scenes and Consumer Videos". They scheduled an end-to-end scheme for text discovery and recovery is essential in multiple fields such as content based retrieval systems, video event discovery, human computer communication, independent robot or vehicle navigation and vehicle certificate plate recovery. Text discovery in natural scenes is a challenging difficulty and has increased a lot of consideration recently. Therefore a robust and fast recognition system is desirable. They showed significant improvement in text discovery and recovery tasks over earlier methods on a large consumer video information by using a model based on OCR system for recovery. B.Jyothi[7] proposed "Relevance Feed Back Content Based Image Retrieval Using Multiple Features". In this paper, they suggested exact Relevance Feed Back (RFB) Content Based Image Retrieval (CBIR) by multiple features based on communicating recovery method which will widely decrease the semantic gap among low-level features and high-level semantics. Content-based Image Retrieval (CBIR) technology destroys the faults of traditional text-based image recovery method. To improve the addition of all features and recovery presentation, they used Relevance Feed Back method. Yan Yang[8] suggested "Content-Based Video Retrieval (CBVR) Scheme for CCTV Surveillance Videos". In this paper, they presented the plan and execution of a framework and a data model for CCTV surveillance videos on RDBMS which offers the tasks of a surveillance monitoring scheme, with a grouping structure for event discovery. They presented a framework to resolve the problem of extracting the great volume of data from CCTV surveillance schemes by classifying video frames and define the storing model for storing associated metadata from surveillance video streams. Their key approach was Content-Based Video Recovery methods which are typically based on video sorting and matching. Mr Pradeep Chivadshetti[9] suggested "Content Based Video Retrieval By Integrated Feature Extraction" method for automatic video indexing and video search in large video libraries. In this paper, they suggested an end-to-end text discovery and recognition organization as OCR and ASR applying for huge dataset in both pixel-level text discovery and word recognition responsibilities. For the reason of extraction and mining script from composite background skeleton based binarization method is proven by H.Yang by performing difference relation of text and background color. Automatic video segmentation is implemented by Wang et al. They suggested a method based on threshold slide transformations. Grcar et al. put on technique of synchronization approaches from recorded videos and files those delivered by presenters [10].Tuna et al. show that the study of lecture video indexing and search. This can be complete with the benefit of using global frame differencing metrics [11].T. Tuna proposed an approach for positive OCR outcomes; they have suggested one image conversion technique as global differencing metrics appropriate for segmentation outcomes when the slides are considered by animations. On behalf of lecture video panel Jeong et al. suggested scale invariant feature convert and adaptive threshold procedure for slides containing corresponding substances [12].Dipali Patil[13]suggested an approach for Survey of Content Based Lecture Video Retrieval for lecture video segmentation methods and requirements using approaches like OCR,ASR, video segmentation classification but the trouble is to discovery of specific portions of their immediate interest. Rupali Khollam[14] proposed an approach for A Survey on Content Based Lecture Video Retrieval by Speech and Video Text data for Automatic Lecture Video Indexing, Slide Video Division, Video Content Surfing and Video Search. Laxmikant S. Kate[15] studied an approach a Survey on Content based Video Retrieval Using Speech and Text information for the purpose of retrieved data, both textual and audio, content-based video indexing and retrieval, using text and audio. The method they used OCR,ASR.R.

Rajarathinam[16] suggested an approach for the reason of analyzing Video Retrieval Using Speech and Text for Content Based data using OCR,ASR. They also studied automated video indexing and video browsing. Madhav Gitte et.[17] studies an approach for content based video retrieval system. They show the methods of video recovery using video segmentation, key frame selection and indexing and clustering using Euclidean Distance Algorithm.

III. Proposed System

In multimedia-based teaching schemes, there is essential of segmentation of speech videos and making them keen on topic and subtopics. Simple trouble in any lecture video is to deliver semantic demand and successfully retrieving related contents from extensive video. Effective and efficient search capability for the students can be providing if correct browsing ability is provided. The purpose of suggested system architecture is to plan for recovery of video using its contents. The proposed system contains of four modules as shown in the following figure.1 System Architecture.

It generally holds following components:
1. Taking Frames from input video
2. Frame Classification
3.Use OCR Algorithm for Optical Character Reorganization from Each Frame of Input Video
4. Automatic Speech Recognition (ASR) for entire Audio outcome by Input Video
5. Video plus Segment Level Keywords are extracted with Output of step 3 & 4 for content based video surfing and search.

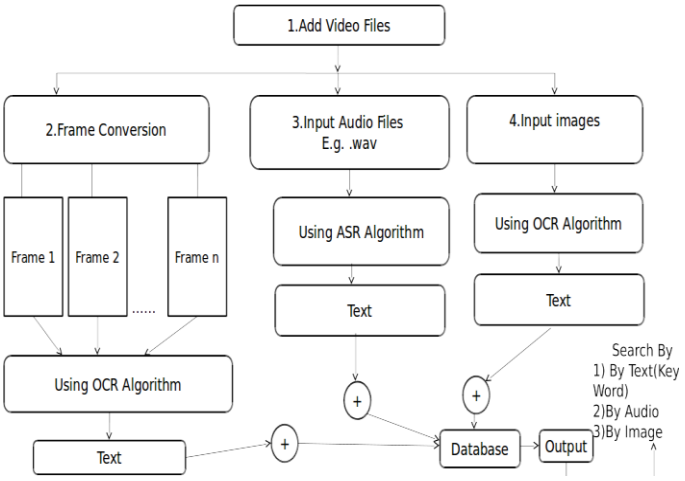

Fig. 1. System Architecture

In accumulation, the suggested scheme also offer numerous extra skill to user that user can giving videos through specified search request in three steps. The first method used for recovering video is regular search i.e. textual query. Another method is search request in image format, and then third method on behalf of recovering video is search request as small video shot and search via audio in search request format. To implementing a model which captures the numerous frames from a video, total captures frames are then categorized according to the reappearance property then fetching whole text from all the frames for further video retrieval scheme. Fetch all the voice resultant into text using ASR method is also used in the procedure of video recovery scheme. The above data (Text and Voice from Video) is used for content based video recovery scheme and bunching of video according to their text and voice constraints. The suggested scheme aim is to Capture the Frames from Video and Frame Grouping. Then separate character from every video frame used by OCR Algorithm (Optical Character Reorganization). Also implementing Automatic Speech Reorganization (ASR) for all Audio outcome from Input Video. Lastly Video and Segment Level keywords are extracted using Outcomes of step 2 & 3 for content based Video browsing and search.

IV. Methodology

A. Segmentation Algorithm(OCR)

Video surfing can be achieve for segmenting of video into descriptive key frames. The selected key frames can provide using graphical advice for navigation in lecture video portal .The video separation and key-frame choice is more frequently accepted as a preprocessed for extra checking jobs for video OCR and visual concept discovery, etc.
The subparts while building the whole OCR application are given below:
1. Formulating Training dataset.
2. Preprocessing file image.
3. Formulate the Tesseract supported image.
4. Accomplish Recognition by the Tesseract engine.
5. Post-processing the produced text outcomes.
Among the sub parts number 1 is self-reliant than other ones. Parts 2 to 4 are in sequence relying on the results of the earlier step. Segmentation method consists of two type:
1) In the first, the whole slide video is analyzed. To capture respective knowledge adaption between approximate structures, for which founding an analysis break of three seconds through taking together correctness and competence into account. That funds that segments by duration reduced than three seconds might be rejected in proposed system. Subsequently there are same little topic segments smaller than three instants, this setting is therefore not risky. To create suitable edge idea for adjacent frames and also form the pixel differencing image since the edge

maps.

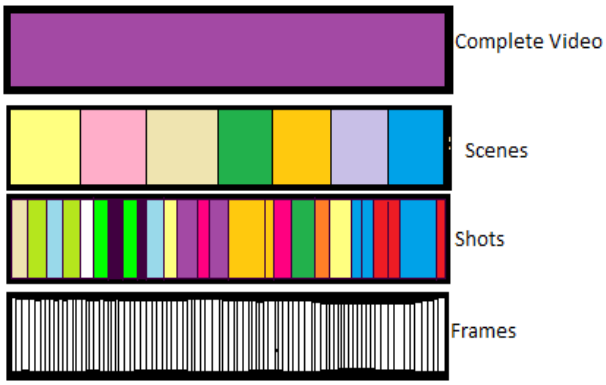

Fig. 2. Video Segmentation

2) Second step constructed at the frames. In the second segmentation stage the actual slide adaption will be captured. The heading as well as content area of a slide frame is first clearly defined. For the purpose of building the content delivery used slide styling for examining a great amount of lecture videos in the database.

B. Automatic Sound Recognition Algorithm(ASR):

The main ambition of an ASR scheme is to correctly and efficiently translate a speech signal into a text message record of the oral words independent on presenter, situation or the device used for record the speech (i.e. the microphone). That procedure starts when a speaker accepts what to say and really speaks a sentence. The software previously make a speech wave form, that represents the words of judgment as well as the extraneous noises and breaks in the spoken data input. Next, software attempts to elucidate speech into best estimate of the sentence. Fig.3 shows audio conversion using ACR algorithm. Firstly it converts the speech signal into a chain of vectors which are measured during the speech signal. Then, using a syntactic decoder it generates a valid sequence of representations. The ultimate goal of ASR research is to allow a computer to recognize in real-time, with 100% accuracy, all words that are intelligibly spoken by any person, independent of vocabulary size, noise, speaker characteristics or accent.

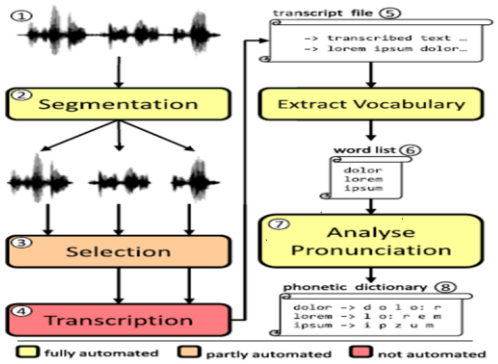

Fig. 3. Audio Conversion

V. Result Analysis

A. Comparison of Normal & OCR Search

TABLE I. Comparison of Normal and OCR Search.

Video	Video Size	Normal Search	OCR Search
1	4.13904mb	7018-Sec	1005-Sec
2	4.13904mb	9018-Sec	1005-Sec
3	4.13904mb	8021-Sec	1005-Sec

The objective of an OCR system is to recognize alphabetic letters, numbers, or other characters, which are in the form of digital images, without any human intervention. This is accomplished by searching a match between the features extracted from the given characters image and the library of image models. Fig 4. shows the time taken by video searching using OCR algorithm and normal searching. The time taken by OCR algorithm is less as compare to normal search and giver very appropriate result on the basis of image as well as text.

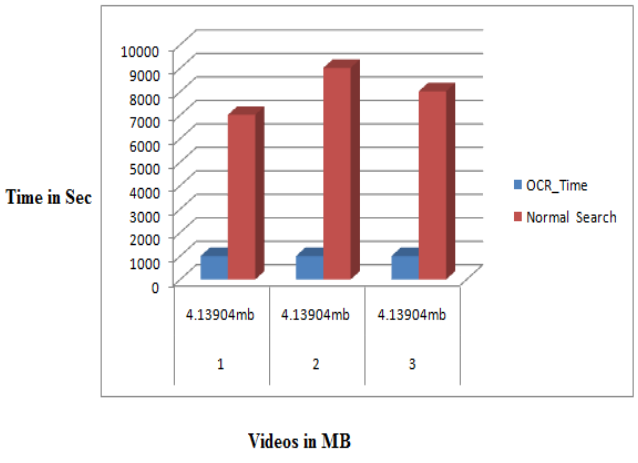

Fig. 4. Comparison of Normal Search & OCR Search

B. Comparison of Normal Search & OCR Search

TABLE II. Comparison of Normal and ASR Search.

Video	Video Size	Normal Search	ASR Search
1	4.13904mb	7018-Sec	1005-Sec
2	4.13904mb	9018-Sec	1005-Sec
3	4.13904mb	8021-Sec	1005-Sec

The ultimate goal of ASR research is to allow a computer to recognize in real-time, with 100 accuracy, all words that are intelligibly spoken by any person, independent of vocabulary size, noise, speaker characteristics or accent. Fig 5 shows the time taken by video searching using ASR algorithm and normal searching. The time taken by ASR algorithm is less as compare to normal search and giver very appropriate result on the basis of Recorded Audio as well as sound.

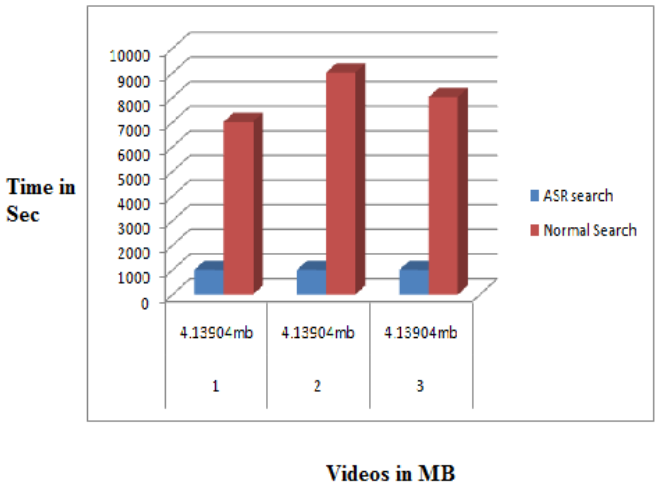

Fig. 5. Comparison of Normal Search & ASR Search

VI. Future Scope

The usability and utility study for the video search function in existing lecture video portal will be conducted. Automated annotation for OCR and ASR results using Linked Open Data resources offers the opportunity to enhance the amount of linked educational resources significantly. Therefore more efficient search and recommendation method could be developed in lecture video archives.

VII. Conclusion

In this paper we implement an approach for content based video lecture indexing and retrieval in large lecture video library. It is a simple, scalable and flexible approach which has been widely used for combining visual and textual information for video search processes. It is used to increase the accuracy of the retrieved results, in order to help in solving the semantic gap problem, referred to the difficulty in understanding the information that the user perceives from the low level characteristics of the multimedia data.

References

[1] Haojin Yang and Christoph Meinel, Member, IEEE "Content Based Lecture Video Retrieval Using Speech and Video Text Information'' in Proc. IEEE transaction on learning technologies, vol.7, no.2,April-June 2014,pp.144-154.

[2] John Adcock, Matthew Cooper, Laurent Denoue, Lawrence A. Rowe, "TalkMiner: A Lecture Webcast Search Engine" MM'10, October 25–29, 2010, Firenze, Italy. Copyright 2010 ACM 978-1-60558-933-6/10/10

[3] B. V. Patel ,A. V. Deorankar, B. B. Meshram, "Content Based Video Retrieval using Entropy, Edge Detection, Black and White Color Features" 978-1-4244-6349-7/10/$26.00 c 2010 IEEE.

[4] Boris Epshtein ,Eyal Ofek,Yonatan Wexler,"Detecting Text in Natural Scenes with Stroke Width Transform"2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP).

[5] Stephan Repp, Andreas GroB, and Christoph Meinel, Member, IEEE "Browsing within Lecture Videos Based on the Chain Index of Speech Transcription" IEEE Transactions on learning technologies ,VOL. 1, NO. 3, JULY-SEPTEMBER 2008,145-156.

[6] Arpit Jain, Xujun Peng, Xiaodan Zhuang, Pradeep Natarajan, Huaigu Cao "Text Detection and Recognition in Natural Scenes and Consumer Videos",2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP),1245-1249.

[7] B.Jyothi, Y.Madhavee Latha, V.S.K.Reddy, "Relvance Feed Back Content Based Image Retrieval Using Multiple Features",978-1-4244-5967-4/10/$26.00 ©2010 IEEE.

[8] Yan Yang Brian, C Lovell ,Farhad Dadgostar, "Content-Based Video Retrieval (CBVR) System for CCTV Surveillance Videos",978-0-7695-3866-2/09 $26.00 © 2009 IEEE/DOI 10.1109/DICTA.2009.36.

[9] Mr. Pradeep Chivadshetti, Mr. Kishor Sadafale, Mrs. Kalpana Thakare, "Content Based Video Retrieval Using Integrated Feature Extraction" ,24th & 25th March 2015 Fourth Post Graduate Conference,IEEE.

[10] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc.Int. Conf. Comput. Vis. Pattern Recog., 2010, pp. 2963–2970.

[11] Dimitrovski, Ivica, et al. "Video Content-Based Retrieval System." EUROCON,2007. The International Conference on &Computer as a Tool , IEEE, 2007.

[12] 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. Comput.Sci. Educ., pp. 129–134.

[13] Dipali Patil, Mrs. M. A. Potey "Survey of Content Based Lecture Video Retrieval" International Journalof Computer Trends and Technology (IJCTT) – Volume 19 Number 1 – Jan 2015, ISSN: 2231-2803.

[14] Rupali Khollam, S. Pratap Singh, "A Survey on Content Based Lecture Video Retrieval Using Speech and Video Text information", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013)Volume 4 Issue 1,January 2015.

[15] Laxmikant S. Kate, M. M. Waghmare, "A Survey on Content based Video Retrieval Using Speech and Text information", International Journal of Science and Research (IJSR)ISSN (Online): 2319-7064,Volume 3 Issue 11, November 2014,1152-1154.

[16] R. Rajarathinam and R. Latha, "Analysis on Video Retrieval Using Speech and Text for Content-Based Information",Middle-East Journal of Scientific Research 23 (Sensing, Signal Processing and Security): 370-376, 2015-ISSN1990-9233©IDOSI Publications, 2015 DOI:10.5829/ idosi.mejsr. 2015.23.ssps.206.

[17] Madhav Gitte, Harshal Bawaskar , Sourabh Sethi, Ajinkya Shinde, "content based video retrival system",IJRET: International Journal of Research in Engineering and Technology Volume: 03 Issue: 06 | Jun-2014, 430-435.eISSN: 2319-1163 | pISSN: 2321-7308.

[18] Che, Xiaoyin, Haojin Yang, and Christoph Meinel. "Lecture video segmentation by automatically analysing the synchronized slides."Proceedings of the 21st ACM international conference on Multimedia. ACM, 2013.

[19] Ashok Ghatol "Implementation of Parallel Image Processing Using NVIDIA GPU framework." Advances in Computing Communication and Control. Springer Berlin Heidelberg, 2011. 457-464.

[20] Journal article – NianhuaXie, Li Li, XianglinZeng, and Stephen Maybank A Survey on Visual Content- Based Video Indexing and Retrieval‖ IEEE Transactions On Systems, Man, And Cybernetics—Part C:Applications And Reviews, Vol. 41, No. 6, November 2011.

[21] Dimitrovski, Ivica, et al. "Video Content-Based Retrieval System." EUROCON, 2007. The International Conference on &; Computer as a Tool ;. IEEE, 2007.

[22] Ankush Mittal, Sumit Gupta(2006), ―Automatic content-based retrieval and semantic classification of video content‖, Int. J. on Digital Libraries 6(1): pp. 30- 38.

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

[24] 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 Mobile Compute., pp. 285–288.

[25] Epshtein, Boris, Eyal Ofek, and Yonatan Wexler. "Detecting text in natural scenes with stroke width transform." Computer Vision and Pattern Recognition(CVPR), 2010 IEEE Conference on. IEEE,2010.international conference on Multimedia. ACM, 2013.