A Novel Approach of in halt based Video Recuperation System using Optimal Character Recognition Technologies

Apurva Kandelkar[1], Rohini Deshmane[2], Monali Bachhav[3]
[1,2,3]*Faculty-IT, iNurture Education Solutions, Bengaluru, India.*

Abstract-The online video lecture is now an existing method of the e-learning process. The lectures and videos access from web sites are growing very quickly. So, the best appropriate technique for retrieving videos in a lecture library is needed. This technology will be very beneficial for new users and existing users to search for relevant videos in small amount of time. This article shows a different method to receiving correct results based on inhalt video search. The main purpose of the future system is to restore the video based on its material, rather than retrieving the video to its title and metadata interpretation to provide the correct query check. To mine the text data printed on the slide, we used the OCR and the ASR algorithms to convert the speaker's voice into text.

Keywords **-** Automatic Speech Recognition Algorithm, ASR, e-learning video, Optical Character Recognition, OCR.

I. INTRODUCTION

In today's case, due to the reliable scene properties of the design video, the appropriate results are not valid for video depends on graphical feature concept. For flexible communication, create asimple video in a paired scene format while displaying the speaker and the slides it provides. The exhibition method common toraise students' understanding. These videos were quickly used by scholars for e-learning. At the end, the agency's records upload their lecture videos on the Internet. The Internet is making a lot of video. Finding the right video based on a query is a complicated task. Because when the user searches for the lecture video, the result is displayed depends on video title stand on content. Otherwise, sometimes the data may be found in for minute. Therefore, the user wants to display the video datain a strapped for time without going through the entire video. The difficulty is that it is not possible to efficiently retrieve the correct information in a large speech video archive. Each video recovery search engine like You-Tube and other responses is based on existing text-related information like video titles and their interpretation. Often, such metadata must be formed by people to check for better quality, but the formation phase is tedious and costly. The main purpose of the structure is to restore the video depends on its material, rather than retrieving the video to its title and metadata descriptions so, provide the right outcome for checking the query. At the end, we implemented a model that captured many frames in a video lecture. The captured frames are then distinguished depends on repeating features. Video fragmentation is completed after a specific time interval within two consecutive frames. This is a video lecture that holds a slide presentation for a while. Therefore, good to resolve this problem, key frame segmentation uses the maximum time interval in seconds. All text for all structure advanced videofetchingprocess is abstracted using the OCR algorithm. All speech is also converted to text data using ASR algorithm technology. Therefore, this is used in the process of the video access or retrieval system. Relevant data (text, speech and images from video) are used in content-basedaccess or retrieval systems and collect video based on their text, image and speech parameters. These OCR algorithms are responsible for extracting characters from text data, and the ASR algorithm is useful for retrieving voice data from video speech. OCR and ASR records and identified slide text line types are intended for keyword extraction, where which video keywords are utilized for surfing and searching for content-based video. The future structure is evaluated based on performance and usability. The OCR algorithm will play an important characteras it will bring us with textual information from the keyframes provided by the video segmentation method, which are likely to be used askey or access for the video. Like the automatic speech recognition method, it will provide the final key audio signal for the video. The video indexing and retrieval system will also play a role in responding to users using matching data and user issues. The main issues that scholars must focus on are the keywords of the video, including text and audio, just as we did for text data.

II. RELATED WORK

Yang Haojin [1] proposed a method "content-based speech video retrieval using voice and video text information" for automatic video indexing and video retrieval. The method they use is OCR, which is used to perform slide video segmentation for video segmentation. The video is converted into a frame for collecting text data for each frame and an ASR for translating speech to text data from the lecture video. The main disadvantage of this approach is that it cannot be used to open data resources to add all feature abstractions and appropriate results. John Adcock [2] proposed a method "Talk Miner: A Lecture Webcast Search Engine" for ad-hoc video capture scenes, frame difference and slide detection. You can see some difficulties when trying to identify different slides in a video stream.

For example, a speaker and an image showing a slide show are synthesized in an image, and the swap camera and slide configuration complicates a simple frame difference algorithm for extracting key frame slide images. They use OCR for techniques for slide discovery and frame difference and lexical processes. V. Patel [3] introduced a method for content-based video retrieval through entropy, edge detection, and black-and-white color features for video recovery. They proposed a method for retrieving criminal information e-learning, news video browsing, digital multimedia library retrieval and defense applications. For the purpose of content-based video recovery, they are implemented using the formation of feature databases and video recovery algorithms. Boris Epshtein [4] developed a method of "detecting text in natural scenes with stroke width transform", looking for the stroke width value of each image pixel and showing its use in text detection tasks in conventional images. The technique they use is a text discovery algorithm that provides features that have been verified to be reliable and flexible for text discovery and stroke width transformations for quick answers. Stephan Repp [5] proposed a method of "browsing in a lecture video based on a voice transcription chain index" that allows browsing in units of video in a multimedia knowledge base. The result can be improved data to the result set, with a useful data search to focus on learners. They use the OCR method to automatically index multimedia video. Arpit Jain [6] suggests "text detection and recognition in natural scenes and consumer videos." They have arranged end-to-end text discovery and recovery solutions for multiple areas, such as content-based retrieval systems, video event discovery, human-machine communication, independent robot or vehicle navigation, and vehicle certificate card recycling. Text discovery in natural scenes is a challenging problem and has recently added a lot of consideration. Therefore, there is a need for a robust and fast identification system. By using OCR-based models for recovery, they showed significant improvements in text discovery and recovery tasks on early approaches to large consumer video information. B. Jyothi [7] proposed "using multiple features for Relvance based feedback. Image retrieval of content." In this paper, they propose precise correlation feedback (RFB) content-based image retrieval (CBIR) based on multiple features of the communication recovery method, which will broadly reduce the semantic gap between low-level features and high-level semantics. Content-based image retrieval (CBIR) technology undermines the shortcomings of traditional text-based image restoration methods. In order to improve the addition of all features and recovery demos, they used the Relevant Feed Back method. Yan Yang [8] proposed the "CBVR monitoring video content-based video retrieval (CBVR) scheme." In this paper, they introduce the planning and execution of the framework and data model of CCTV surveillance video on RDBMS. It provides the task of monitoring and monitoring solutions, with a packet structure of event discovery. They proposed a framework to solve the problem of extracting large amounts of data from CCTV surveillance scenarios by classifying video frames and defining a storage model for storing relevant metadata from the surveillance video stream. Their primary method is a content-based video recovery method, which is usually based on video sorting and matching. Mr. PradeepChivadshetti [9] proposed a "content-based video retrieval integrated feature extraction" method for automatic video indexing and video search in large video libraries. In this paper, they recommend end-to-end text discovery and recognition organizations as OCR and ASR for large data sets in pixel-level text discovery and word recognition responsibilities. Due to the extraction and mining script based on the composite background, H.Yan proves the method based on binarization by performing the difference relationship between text and background color. Automatic video segmentation is implemented by Wang et al.

III. PROPOSED SYSTEM

In multimedia-based instructional programs, segmentation of voice and video and its enthusiasm for themes and subtopics are essential. The simple question in any lecture video is to provide semantic requirements and successfully retrieve relevant content from a wide range of videos. If you provide the right browsing capabilities, you can provide effective and efficient student search. The purpose of the proposed system architecture is to plan to use its content to recover video. The proposed system consists of four modules, as shown in the following figure. 1 System architecture.

It usually contains the following components:

1. Get frames from the input video
2. Frame classification
3. Using the OCR algorithm to perform optical character recombination from each frame of the input video
4. Automatic speech recognition (ASR) of the entire audio result of the input video
5. Video plus segment level keywords are extracted with the output of steps 3 and 4 for content-based video surfing and searching.

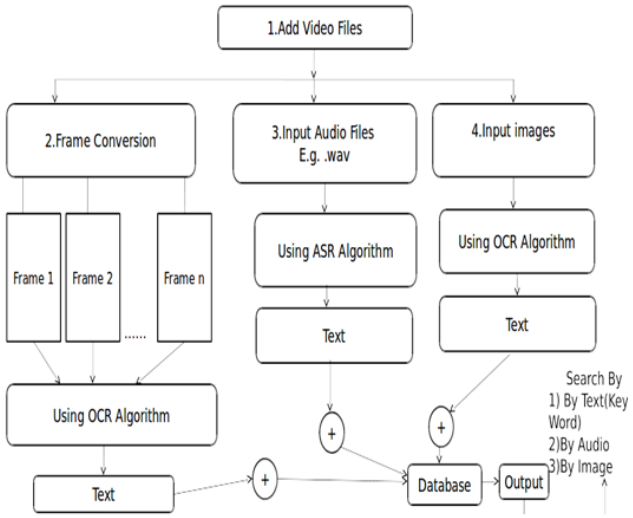

Figure 1: System Architecture

In the acumulation, the proposed solution also provides users with many additional skills, the user can give the video through the specified search request in three steps. The first method used to recover video is a regular search, a text query. Another method is the image format search request, and then the third method representing the recovery of the video is to search for the request as a small video shot and search for the audio search by request format. To implement a model that captures many frames from the video, the Total captured frames are then sorted according to the rendering properties, and then the entire text is fetched from all frames for further video retrieval schemes. Extracting all speech results into text using the ASR method is also a program for video recovery scenarios. The above data (text and speech from video) is used for content-based video recovery schemes and video convergence based on their text and voice constraints. The proposed solution goal is to capture frames from video and frame groupings. The characters are then separated from each video frame used by the OCR algorithm (optical character recombination). Automatic Speech Recombination (ASR) is also implemented for all audio results of the input video. Finally, the results of steps 2 and 3 are used to extract video and segment-level keywords for content-based video browsing and searching.

IV. METHODOLOGY

A. Segmentation Algorithm (OCR) - Video surfing can be used to segment video into descriptive keyframes. The selected keyframes can be navigated through the lecture video portal using graphical suggestions. Video separation and keyframe selection are more frequently accepted as pre-processing for additional inspection jobs for video OCR and visual concept discovery.

The subparts while building the whole OCRapplication are given below:

1. Formulating Training dataset.
2. Pre-processing 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 thanother ones. Parts 2 to 4 are in sequence relying on theresults of the earlier step.

i). Segmentation method consists of two type:

1)First, analyze the entire slideshow video. Capturing the corresponding knowledge adaptation between the approximate structures, by considering the correctness and ability together, establish a three-second analysis break. In the proposed system, it may be possible to refuse to fund a segment with a duration reduced by more than three seconds. Then the same small topic segment is less than three moments, so this setting is not risky. Creates a reliable edge idea for adjacent frames and also forms a pixel difference image since edge mapping.

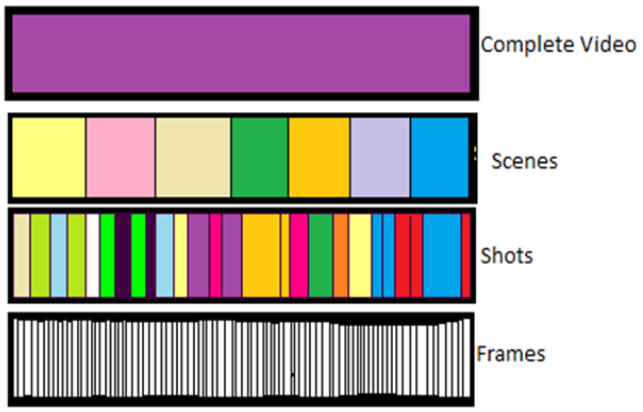

Figure 2: Video Segmentation

2) The second step is to build on the framework. In the second segmentation phase, the actual slides will be captured. The title and content area of the sliding frame are clearly defined first. To build content delivery, use slide styles to check many lecture videos in the database.

V. AUTOMATIC SOUND RECOGNITION ALGORITHM (ASR)

The main purpose of the ASR scheme is to correctly and efficiently convert speech signals into text message records of spoken words without relying on the speaker, the situation or the device used to record the speech (ie the microphone). When a speaker accepts a speech and actually speaks a sentence, the program begins. The software previously produced speech waveforms representing the words being judged and the extraneous noise and interruptions in the speech data input. Next, the software attempts to state the speech as the best estimate of the sentence. Figure 3 shows the audio conversion using the ACR algorithm. First, it converts the speech signal into a vector chain that is measured during the speech signal. A valid representation sequence is then generated using a syntax decoder. The ultimate goal of the ASR study is to allow computers to recognize all words that anyone can understand in real time with 100% accuracy, without being affected by vocabulary, noise, speaker characteristics or accents.

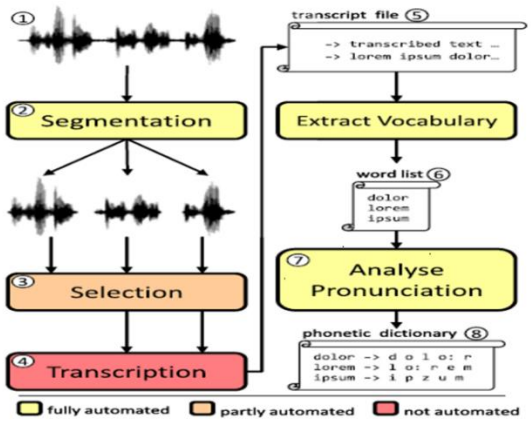

Figure3: Audio Conversion

First, it converts the speech signal into a vector chain that is measured during the speech signal. A valid representation sequence is then generated using a syntax decoder. The ultimate goal of the ASR study is to allow computers to recognize all words that anyone can understand in real time with 100% accuracy, without being affected by vocabulary, noise, speaker characteristics or accents.

VI. FUTURE SCOPE

A usability and usability study of the video search capabilities in the existing lecture video portal will be conducted. Automapping annotations of OCR and ASR results using Linked Open Data resources can significantly increase the number of linked educational resources. Therefore, more efficient search and recommendation methods can be developed in the lecture video archive.

VII. CONCLUSION

In this article, we implemented a content-based video lecture indexing and retrieval method in a large speech video library. It is a simple, scalable, and flexible method that has been widely used to combine visual and textual information in video search processes. It is used to improve the accuracy of search results to help solve the semantic gap problem, and it is difficult to understand the information that users perceive from the low-level features of multimedia data.

VIII. REFERENCES

[1]. Haojin Yang and ChristophMeinel, 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 ,EYALOFEK ,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 ChristophMeinel, 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, XujunPeng, XiaodanZhuang, PradeepNatarajan, 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.MadhaveeLatha, 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 ,FarhadDadgostar, "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. PradeepChivadshetti, Mr. KishorSadafale,Mrs. KalpanaThakare, "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 &# 34; Computer as a Tool&# 34;. 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]. DipaliPatil, Mrs. M. A. Potey "Survey of Content Based Lecture Video Retrieval"International Journal of Computer Trends and Technology (IJCTT) – Volume 19 Number 1 – Jan 2015, ISSN: 2231-2803.

[14]. RupaliKhollam, 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. 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, Ii52-1154.

[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. Rajarathinamand 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-ISSN 1990-9233© IDOSI Publications,2015DOI:10.5829/idosi.mejsr.2015.23.ssps.206

[17]. MadhavGitte, HarshalBawaskar , SourabhSethi, AjinkyaShinde, "content based video retrivalsystem",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 ChristophMeinel. "Lecture video segmentation by automatically analysing the synchronized slides." roceedings 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 &# 34; Computer as a Tool&# 34;. IEEE, 2007.

[22]. Ankush Mittal, SumitGupta(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, EyalOfek, 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.