Scientific Journal Impact Factor (SJIF) : 3.518
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745 ISSN (PRINT) : 2393 - 8161

CONTENT BASED LECTURE VIDEO RETRIEVAL SYSTEM

Markand Prajakta[1], Pandharkar Vaishnavi[2], Shete Pooja[3] and Sonawane Vaishnavi[4]
[1,2,3,4]*Department of Computer Engineering, K. K. Wagh Institute of Engineering Education and Research, Nashik-03*

Abstract- Video is becoming a prevalent medium for e-learning. Lecture videos contain text information in both the visual and aural channels: the presentation slides and lecturer's speech. To extract the visual information, we apply video content analysis to detect slides and optical character recognition to obtain their text. Automatic speech recognition is used similarly to extract spoken text from the recorded audio. We perform controlled experiments with manually created ground truth for both the slide and spoken text from more than 60 hours of lecture video. We compare the automatically extracted slide and spoken text in terms of accuracy relative to ground truth, overlap with one another, and utility for video retrieval. Results reveal that automatically recovered slide text and spoken text contain different content with varying error profiles. Experiments demonstrate that automatically extracted slide text enables higher precision video retrieval than automatically recovered spoken text. In the last decade e-lecturing has become more and more popular. The amount of lecture video data on the World Wide Web (WWW) is growing rapidly. Therefore, a more efficient method for video retrieval in WWW or within large lecture video archives is urgently needed. This paper presents an approach for automated video indexing and video search in large lecture video archives. First of all, we apply automatic video segmentation and key-frame detection to over a visual guideline for the video content navigation. Subsequently, we extract textual metadata by applying video Optical Character Recognition (OCR) technology on key-frames and Automatic Speech Recognition (ASR) on lecture audio tracks. The OCR and ASR transcript as well as detected slide text line types are adopted for keyword extraction, by which both video and segment-level keywords are extracted for content-based video browsing and search. The performance and the effectiveness of proposed indexing functionalities is proven by evaluation.

Keywords- Lecture videos, automatic video indexing, content-based video search, lecture video archives

I. INTRODUCTION

Computerized video has turned into a well known stockpiling and trade medium because of the fast advancement in recording innovation, enhanced video pressure systems and rapid systems in the most recent couple of years. In this way varying media recordings are utilized increasingly much of the time as a part of e-addressing frameworks. Various colleges and exploration foundations are taking the chance to record their addresses and distribute them online for understudies to get to free of time and area. Accordingly, there has been a tremendous increment in the measure of mixed media information on the Web [4]. In this manner, for a client it is almost difficult to discover wanted recordings without a hunt capacity inside of a video document. Notwithstanding when the client has discovered related video information, it is still troublesome more often than not for him to judge whether a video is helpful by just looking at the title and other worldwide metadata which are regularly short and abnormal state. Also, the asked for data might be secured in just a couple of minutes, the client may in this way need to discover the bit of data he requires without survey the complete video [6,7]. The issue turns out to be the means by which to recover the suitable data in a substantial address video file all the more proficiently.

The majority of the video recovery and video look frameworks, for example, YouTube, Bing and Vimeo answer in light of accessible literary metadata, for example, title, sort, individual, and brief portrayal, and so on. By and large, this sort of metadata must be made by a human to guarantee a high caliber, however the creation step is somewhat time and cost devouring. Moreover, the physically gave metadata is regularly short, abnormal state and subjective. In this manner, past the current methodologies, the up and coming era of video recovery frameworks applies consequently produced metadata by utilizing video examination advances. Customary video recovery in light of visual element extraction can't be just connected to address recordings in view of the homogeneous scene organization of address recordings [1, 5]. A demonstrates an excellent address video recorded utilizing an obsolete arrangement created by a solitary camcorder. Shifting components may bring down the nature of this organization [3].

Presentation video is a quickly developing kind of Internet conveyed content because of its expanding use in training. Proficiently guiding customers to video address substance of interest remains a testing issue [11]. Current video recovery frameworks depend intensely on physically made content metadata because of the semantic hole" between content based components and content based substance portrayals. Presentation video is exceptionally suited to program indexing for recovery. Regularly, presentations are conveyed with the guide of slides that express the creator's topical organizing of the substance [11, 13]. Shots in which an individual slide shows up or is talked about relate to regular units for fleeting video division. Slides contain content depicting the video content that is not accessible in different kinds. The talked content of presentations regularly supplements the slide content, yet is the result of a mix of deliberately created scripts and unconstrained ad lib [14]. Talked content is more bottomless, however can be less particular and unmistakable in contrast with slide content [12].

II. RELATED WORK

Early work on video recovery utilizing talked content ordered news telecasts utilizing ASR. News video has been a relentless center of related work, to some extent since it has more prominent theme structure than non specific video. It additionally will probably contain machine decipherable content from design and tickers regular to that classification. All through the TRECVID evaluations [4] news and documented video has been listed utilizing ASR, now and again in conjunction with machine interpretation. Now and again, the subsequent transcripts displayed high word mistake rates. Intelligent recovery utilizing visual components alone as a part of an adequately effective interface accomplished execution equivalent to conventional recovery utilizing ASR [1]. The Informedia extend additionally connected OCR to video retrieval.6 Compared to presentation video, the representation with content utilized as a part of news telecasts ordinarily possess a littler segment of the casing [9]. Sight and sound recovery research has analyzed projector-based slide catch frameworks that deliver a flood of slide casings as opposed to presentation video [8]. The subsequent stream is lower in unpredictability and can give high determination slide pictures. In this setting created procedures to enhance OCR's recovery power, without utilizing ASR [2].

Along comparable lines, Jones and Edens portrayed techniques for adjusting the slides from a presentation with a sound transcript utilizing a hunt record developed from ASR. Slide content is utilized to inquiry the quest file for arrangement. They amplify this work utilizing a corpus of meeting recordings. ASR is utilized to make a point division of the meeting, and the slide title questions are utilized as a part of investigations for investigation of the corpus. Reliably, the precision of ASR straightforwardly affects the downstream recovery execution [10]. This work demonstrated great results in blend with speaker adjustment and different improvements. They report that high ASR word blunder rates can be passable, however that video-particular catchphrases should be perceived precisely for compelling recovery. Comparable tests were accounted for somewhere else to rectify erroneously perceived terms in ASR transcripts. They report a few diminishments in these blunders [8]. Other picture investigation techniques match slides from a presentation with video outlines in which they show

up. Most related work concentrates on either talked archive recovery or bland video recovery. In spite of years of enduring advancement on execution, precision keeps on posturing difficulties to the consolidation of ASR in sight and sound recovery frameworks [7].

While ASR and shut subtitle (CC) transcripts still by and large beat indexing by substance based visual investigation, slide content recuperated by OCR is profitable for indexing presentation recordings [5]. We concentrate on presentation video as a one of a kind class in which naturally recuperated talked content and slide content can both be separately abused and joined to enhance recovery. In the indication of this paper, we first evaluate the precision of naturally recuperated talked content utilizing ASR and the exactness of consequently recouped slide content utilizing programmed slide discovery and OCR [1]. Also, we think about the qualities of blunders in these two modalities. Next, we lead tries that analyze the effect of translation mistakes on video recovery. At last, we consolidate these modalities for address video recovery.

III. PROPOSED SYSTEM

In the proposed approach, we are actualizing framework power to transfer video is limited to the gateway client it gives affirmation, that the database will just contain address recordings. Gateway client jars embed/erase/overhaul/seek/recover recordings. The easygoing client of framework can hunt down the required video by giving inquiry, and can recover video which is more applicable to his question.

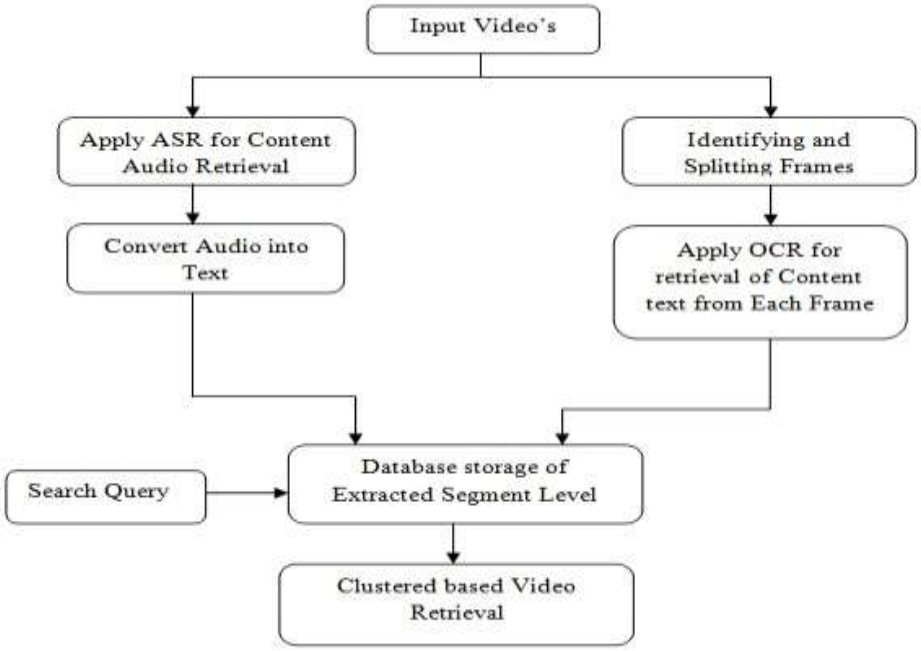

Fig: System Architecture

A lot of literary metadata will be made by utilizing OCR and ASR strategy, which opens up the substance of address recordings. We extricate metadata from visual and in addition sound assets of address video naturally by applying fitting investigation systems. For visual examination, we proposed a strategy for slide video division and connected OCR to assemble content metadata. For sound investigation, we connected an ASR calculation and accumulated content metadata. Concentrated information stockpiling is utilized to assemble the information from ASR and OCR calculations [1, 2]. Utilizing look question the video can be recovered precisely taking into account substance of video. In this framework diverse administrator can transfer recordings at once. Thus the framework can be said to as parallel working framework.

An alternate, printed, representation is determined by perusing the content that present in the video pictures utilizing optical character acknowledgment (OCR). OCR innovation has been industrially accessible for a long time. Be that as it may, perusing the content present in the video stream requires various preparing ventures notwithstanding the genuine character acknowledgment. Our video optical character acknowledgment framework [5] utilizes the accompanying way to deal with distinguish and perceive subtitled content that shows up on the video. Given the quantity of casings contained in common show news, it is not computationally plausible to prepare every single video edge for content [9]. Thus unpleasant or speedy content area discovery is performed first. At that point the content must be separated from the picture, and changed over into a paired high contrast representation, since the monetarily accessible OCR motors don't perceive hued content on a variably shaded foundation [11, 14].

Writings in address slide are firmly identified with the address content and utilized for their recovery assignment. In our methodology we built up a novel video OCR framework for social occasion video content. In the recognition organize; an edge-based multi-scale content identifier is utilized to rapidly restrict hopeful content locales with a low dismissal rate. For the resulting content region check, a picture entropy-based versatile refinement calculation not just serves to dismiss false positives that uncover low edge thickness, additionally assist parts the most content and non-content locales into independent squares [3, 12]. At that point we apply Stroke Width Transform (SWT) based confirmation methodology to evacuate the non-content pieces. Be that as it may, the SWT verifier is not ready to effectively distinguish exceptional non-content examples, for example, circle, window pieces, and garden fence so we embraced an extra SVM classifier to deal with these non-content examples keeping in mind the end goal to assist enhance the location exactness. For content division and acknowledgment, we built up a novel Binarization approach, in which we utilize picture skeleton and edge maps to distinguish content pixels. The proposed strategy incorporates three fundamental steps: content angle heading examination, seed pixel determination, and seed-locale developing. After the seed-locale developing process, the video content pictures are changed over into a suitable arrangement for standard OCR motors [6]

Talked archives are created by extricating sound information from address video records. At that point, we deciphered sound recordings utilizing a programmed discourse acknowledgment (ASR) motor. To start with, the recorded sound document is sectioned into little pieces and uncalled for portions are sorted out. For each remaining section the talked content is translated physically, and added to the transcript document naturally [9]. As a moderate step, a rundown of every single utilized word as a part of the transcript record is made. Keeping in mind the end goal to get the phonetic lexicon, the elocution of every word must be spoken to phonetically.

IV. RESULTS

Utilization of gap and overcome methodologies to abuse appropriated/ parallel/ simultaneous handling of the above to distinguish objects, morphemes, over burdening in capacities, and useful relations and some other conditions. Recovery of video utilizing discourse and content data which is gotten from the different recordings and bringing about the production of metadata naturally without human obstruction. We need to actualize a model which catches the different edges from a video. All the caught casings are then arranged and chose. At that point we get all the content from every one of the edges utilizing OCR for further video recovery framework. Likewise we bring all the voice coming about into content utilizing ASR strategy for the procedure of video recovery framework System provides option for upload video and search the video and also system includes the option for review of the search result.

Fig: System Output

The system include to the videos based on category of the video. User can search the video based on the category of the video and also including the type of video like video with audio/ video with image. The system is develop in simplify and user friendly manner to understand the user quickly. User will get the result in simplify and easy. Also system provides the facility to write the review about their views about the system and the videos.

Fig: System Output

System provides the ability to search video, upload video also user can scan the video with OCR and ASR system. Video can be search in the entire category to simply understand. User will easily get the extracted result from the OCR and ASR scan. System will read the extracted text from the OCR and ASR scan and will display the extracted text.

Fig: System Output

V. CONCLUSION

The proposed approach displays a computerized video look in vast address video files. At first programmed video division and key-outline recognition is done to offer a visual rule for the video content route. Different advancements are utilized for extractions of metadata from key-edges of video address by which both video and fragment level watchwords are extricated for substance based video seek. In address recordings, writings from address slides serve as a blueprint for the address and are critical for comprehension. In this manner in the wake of sectioning a video _le into an arrangement of key edges (all the exceptional slides with complete substance), the content location strategy will be executed on every key casing, and the separated content articles will be further utilized as a part of content acknowledgment and slide structure investigation forms. Discourse is a standout amongst the most essential transporters of data in video addresses. A lot of printed metadata will be made by utilizing OCR und ASR technique, which opens up the substance of address recordings.

VI. ACKNOWLEDGMENT

We are thankful to Prof. Mrs. Priti Vaidya and HOD of computer department Prof. Mr. Shirish Sane for their valuable guidance and encouragement. We would also like to thank the K. K. Wagh Institute of Engineering Education and Research, Nashik-03 for providing the required facilities, Internet access and important books. At last we must express our sincere heartfelt gratitude to all the Teaching and Non-teaching Staff members of Computer Engineering Department who helped us for their valuable time, support, comments, suggestions and persuasion

REFERENCES

[1] 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. 912
[2] 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. 32573269, 2008
[3] 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. 730733
[4] 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. 129134. [Online]. Available: http://doi.acm.org/10.1145/2157136.2157177
[5] 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 Comput., pp. 285288. [Online]. Available: http://doi.acm.org/10.1145/2428955.2429011
[6] C. Meinel, F. Moritz, and M. Siebert, Community tagging in tele-teaching environments, in Proc. 2nd Int. Conf. e-Educ., e-Bus., e-Manage. and E-Learn., 2011
[7] S. Repp, A. Gross, and C. Meinel, Browsing within lecture videos based on the chain index of speech transcription, IEEE Trans. Learn. Technol., vol. 1, no. 3, pp. 145156, Jul. 2008
[8] J. Adcock, M. Cooper, L. Denoue, and H. Pirsiavash, Talkminer: A lecture webcast search engine, in Proc. ACM Int. Conf. Multimedia, 2010, pp. 241250
[9] F. Chang, C.-J. Chen, and C.-J. Lu, A linear-time componentlabeling algorithm using contour tracing technique, Comput. Vis. Image Understanding, vol. 93, no. 2, pp. 206220, Jan. 2004
[10] Bhagyashri Babhale,Kaumudinee Ibitkar,Priyanka Sonawane,Renuka Puntambekar, Content Based Lecture Video Retrieval using Video Text Information. International Journal of Advance Foundation and Research in Computer (IJAFRC) Volume 2, Special Issue (NCRTIT 2015), January 2015. ISSN 2348 – 4853 474.
[11] Alexander G. Hauptmann, Rong Jin, and Tobun D. Ng, Video Retrieval using Speech and Image Information
[12] Deshmukh Bhagyashri, REVIEW ON CONTENT BASED VIDEO LECTURE RETRIEVAL, IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308
[13] Matthew Cooper, "Presentation Video Retrieval using Automatically Recovered Slide and Spoken Text".
[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