ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

Survey and Comparative Analysis on Video Retrieval Techniques

Dipali Patil[1], Vrushali Uttarwar[2]

Assistant Professor, Department of Computer Engineering, D Y Patil College of Engineering, Pune, Maharashtra, India[1]

Assistant Professor, Department of Computer Engineering, D Y Patil College of Engineering, Pune, Maharashtra, India[2]

ABSTRACT:In this paper we present a deep survey and comparative analysis on different available approaches for the content based video lectures retrieval. The extensive growth of Internet has provided availability of e-learning content and lecture videos which have brought forth an imperative need for developing effective content based retrieval system to speed up the process. The key factors for developing video retrieval system are Comprehensive metadata extraction and support for topic level search within videos. In this paper, we study different lecture video segmentation techniques. As there is firm needs for segmenting lecture videos into topic units in order to organize videos for browsing and to provide search capability. Universal metadata extraction & support for topic level search within videos are key factors in formulating such systems.

KEYWORDS: Content based video Retrieval, E-learning and Video Segmentation

I. INTRODUCTION

Distance Learning is a new concept used to provide the education to learners who are willing to learn but cannot be present at the lecture hall. Hence many universities provide facility to students by providing them with the recordings of the lectures which were conducted; in order to provide this facility it is necessary to maintain hundreds of lecture video recordings in a repository, and make them available to remote students over the Internet. This facility is completely independent of time and location. Digital video has become a popular storage and exchange medium due to the amazing growth in recording technology, improved video compression techniques and high speed networks in the last some years. Hence audio-visual recordings are more often used in e-lecturing systems. As a result, there has been an tremendous gain in the quantity of multimedia data on the Web.
Hence without a search function within a video archive it is impossible for a user to find the video output.

User gets frustrated with duplicate contents. Universally consented video retrieval and indexing methods are not well defined or available. Hence a content based video retrieval system is presented as a motivation from above challenges. Content Based Video Retrieval system it includes various steps like Video Segmentation, Key frame Selection, Feature Extraction, Classification and Clustering, Indexing and Similarity.

There is an exponential increase in the amount of video data, and also it often demands a lot of time and is arduous for students to search through a full videos, in order to find some specific part of their interest. Considering an example where a user needs to find a portion where TCP protocol is discussed in a repository that has a course on Networking containing around 50 lectures of 60 minutes. The user needs to manually check each video tiles and then decide which video might contain the explanation of his interest.

The problem is aggravated if the user is new with area, or the topic is very definite, as it would be difficult for the user to select the videos to be examined. Optical Character Recognition & Automatic Speech Recognition method, which

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

provides the content of lecture videos, will create an enormous amount of textual metadata. The search indices are created from different information resources, covering manual annotations, comments, OCR and ASR keywords, metadata, etc. for content-based video retrieval and search.
The remainder of this paper is organized as follows. We provide an overview of various Video retrieval techniques with their pros and cons in Section II, and shows the comparative analysis in Section III and conclusion is then given in Section IV.

II. RELATED WORK

In recent years the need for content-based retrieval of image and video information from internet archives has gain the attention of nearly all research fellows. Research efforts have caused to the development of methods that provide access to image and video data. The methods are used to determine the similar things in the visual information content educed from low level features. These aspects are clustered for creation of database. Most of the primary work related to Content based video retrieval chiefly emphasis on video segmentation & summarization. To obtain efficient and effective result for video search, very first video has to be segmented and all text information and Meta data need to be extracted for indexing & tagging purpose. Techniques for searching in lecture videos.

1. Existing Lecture Video Repositories:

NPTEL [7], MIT Open Courseware [9] & freevideolectures site [8] are few of the present lecture video repositories. We looked into support for search & browsing features accessible in those repositories explained in table I. Some repositories don't have automatic transcriptions they use manual transcriptions, subtitles for lecture videos but they are not providing use of them to provide search features.

2. Lecture Video Retrieval:

Munteanu et al., [1] concentrate the research on English speech recognition for Technology Entertainment and Design (TED) lecture videos & webcasts. The training corpus is made manually, which is thus hard to be bettered or optimized periodically in this. And hence, OCR & ASR are used to get text & speech transcript respectively.
Automatic Speech Recognition (ASR) allows speech-to-text information on audio lecture videos, which is thus well proper for content based lecture video retrieval is studied in [2]. Authors in [3] proposed a lecture video segmentation method based on Scale Invariant Feature Transform feature & the reconciling threshold. In their work, SIFT feature is mainly applied to measure slides with similar content. Authors in [4] presented method for lecture video indexing and search. First, the lecture videos are divided into representative's key frames by using frame differencing metrics. Then to gather the data from the slide standard Optical character recognition engine is enforced, in which they employ some image transformation methods to improve the OCR result. An adaptive threshold selection algorithm is used to detect slide transitions between video frames. In their rating, this method attained promising results for processing one scene lecture video. Authors in [5] applies tagging data mechanism for lecture video retrieval and search. Recently, collaborative tagging has become a common & standard functionality in lecture video portals.

3. Content based video Retrieval:

The contents in the lecture videos are specially (i) Lecturer Speech: part of video that shows the instructor talking, (ii) Slides: Part of video that shows the current slide of the presentation, and (iii) Lecturer Notes: Part of video that point the board/paper on which instructor is writing. Content based approaches extract metadata from appropriate portions of the video and create an index that can be used for searching within the video. This techniques are difficult to automate and time-consuming to do manually. In multimedia-based learning systems, there is need of segmenting of lecture videos & forming them into topic & subtopics. Basic problem in any lecture video is to give semantic query & efficiently retrieve required contents form long video. Effective and efficient search capability for the students can be given if sound browsing facility is given. A highly accurate method for video segmentation applying SIFT & an Adaptive threshold. Using SIFT, can easily equate two slides, having like Contents but unlike backgrounds. And can calculate Frame transition quite precisely by using Adaptive threshold.[10] OCR was initially developed for high contrast data images,

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

taken from metal & other surfaces with add roughness & reflectivity. Content-based gathering within video data demands textual metadata that has to be given manually by the users or that has to be extracted by automated analysis.

To address this, methods from common OCR focusing on high-resolution scans of printed (text) documents have to be amended & adapted to be also relevant for video OCR. In video OCR, video frames holding visible textual information have to be known first. Then, the text has to be broke from its background, & geometrical transformations have to be employed in front of common OCR algorithms can work the text productively. Texts in the lecture slides are closely related to the lecture content can thus render crucial information for the retrieval task. In the detection stage, an edge-based multi-scale text detector is applied to rapidly localize candidate text regions with a low rejection rate. Then Stroke Width Transform (SWT) based verification methods are applied to take out the non-text blocks. The video text images are converted into a suitable format for standard OCR engines.[11] The synchronization issue is solved by a solution which segments lecture video by analyzing its supplementary synchronized slides. The slides content develops automatically from OCR (Optical Character Recognition). Then partition the slides into various subtopics by analyzing their logical relevance. As the slides are adjusted with the lecture video, the subtopics of the slides point the segments of the video. Then OCR results are the inputs for the entire process [12].

Vendrig &Worring [20] propose a system that allows character identification in movies. In order to achieve this, they associate visual content to names extracted from movie scripts. Denman et al [13] present the tools in a system for creating semantically significant summaries of broadcast Snooker footage. Their system parses the video sequence, identifies proper camera views, &tracks ball movements. A same approach presented by Kim et al [14] extracts semantic information from basketball videos based on audio-visual features. A semantic video retrieval approach using audio analysis is presented by Bakker and Lew [15] in which the audio can be automatically categorized into semantic categories such as explosions, music, speech, etc. A learning method using the AdaBoost algorithm and a k-nearest neighbor approach is proposed by Pickering et al [16] for video retrieval.

A variety of segmentation schemes have been developed. They use rule-based methods based on input from multimedia streams machine learning techniques, or topic detection and tracking (TDT) methods [21]–[23]. For example, authors in [22] arose an automatically induced segmentation system using the Hidden Markov Model (HMM). In this, authors manually divided instructional videos, such as videotaped lectures or interviews, into individual clips based on their content. The segmentation was done at a logical level instead of a physical level by simply identifying time boundaries of each clip within an entire video. We did it by hand because 1) unlike a TV news video consisting of a sequence of reports with frequent scene changes and interruptions for commercials that are commonly used as important clues for automatic video segmentation, most instructional videos have only one speaker and lack cues for segmentation, and 2) video segmentation is not the focus of this research. Therefore, in our study, video segmentation was a manual process. It normally took about 0.5–1 hours to segment a one-hour video using a software package called Final Cut Pro (Apple Computer, Inc.).

A summary of disputes for content-based sailing of digital video is presented by Smeaton and Over [27]. The authors provide a drumhead of the activities in the TREC Video track in 2002 where 17 teams from across the world took part. A quick video retrieval method under sparse training data is proposed by Liu and Kender [28]. Observing that the prompting objects' trajectories play an vital role in content-based retrieval in video databases, Authors in [29] present an effective same trajectory-based retrieval algorithm for moving objects in video databases. A robust content-based video copy identification scheme dedicated to TV broadcast is presented in [30]. The recognition of like videos is depending upon local features extracted at interest points. Similarly, Shao et al [31] extract local constant descriptors for quick picture identification based on local appearance. Video similarity detection issues are also discussed in [32]. Also, new methods are proposed for a wide range of retrieval problems, including object matching [33], shape-based retrieval [34], hierarchical clustering in multidimensional index structures [35], k-d trees for database indexing[36] and trademarks [37]. Authors in [38] investigates the audio-assisted video scene segmentation for semantic story browsing. In [39] authors proposed a novel approach for content-analysis and semantic summarization of instructional videos,

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

while Miura et al [40] studied cooking videos application. Domain-specific information is also in [41] for the retrieval of artworks by color art concepts. In [42] authors shows the documentary video application.

III. COMPARATIVE ANALYSIS

In the above section we have studied the lecture video retrieval mechanism thoroughly. To make the video retrieval process mature researchers will need to understand how to evaluate and benchmark features, methods, and systems. Various papers which focuses these questions are [17], [18], and [19]. While Sebe et al [19] perform an evaluation of different salient point extraction techniques to be used in content-based image retrieval, In [18] authors propose a technique for creation of a reference image set in which the similarity of each image pair is computed applying"visual content words" as a ground vector that allows the multidimensional content of each image to be presented with a content vector. The authors claim that the similarity measure computed with these content vectors correlates with the subjective judgment of human observers and provides a more objective method for evaluating and expressing the image content. Muller et al [17] compare different ways of evaluating the performance of content-based retrieval systems taking a subset of the Corel images. Their aim is to show how easy it is to get differing results, even when the like image group, CBIRS, and performance measuresare used. Lecture video indexing without digital sources (lecture slides) has been already studied in [24], [25], [26]. However, there are still some remaining challenges for the research on lecture videos, including cross-modal and multi-modal retrieval. One of the main issues in existing systems is the lack of use of multi-modal and cross-modal analysis (audio, video, graphics, text...).

TABLE I. SLECTURE VIDEO REPOSITORIES COMPARISION

Repository	Search	Navigation Features
NPTEL	No	No
Freelecturevideos.com	Meta-data	No
Videolectures.com	Meta-data	Slide Synchronization
MIT Open course ware	Content	Speech-transript

IV. CONCLUSION

In this paper we studied the video retrieval techniques; all the studied techniques are having many advantages and disadvantages. Based on this the techniques are enforce for various application. In order to improve the performance of the video retrieval techniques we need to study how to combine these approaches and make them retrieve what the user needs in less time. Also we studied the lecture video segmentation techniques which are used to reduce the effort of the user of watching all video, by providing him a way to jump to his area of interested topic in the video.

REFERENCES

[1] Munteanu, C., Penn, G., Baecker, R., and Zhang, Y. "Automatic speech recognition for webcasts: how good is good enough and what to do when it isn't.", Proceedings of the 8th international conference on Multimodal interfaces. ACM, 2006.
[2] Leeuwis, Erwin, Marcello Federico, and Mauro Cettolo "Language modelling and transcription of the TED corpus lectures", Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03). 2003 IEEE International Conference on. Vol. 1. IEEE, 2003.
[3] Jeong, Hyun Ji, Tak-Eun Kim, and Myoung Ho Kim "An accurate lecture video segmentation method by using sift and adaptive threshold", Proceedings of the 10th International Conference on Advances in Mobile Computing and Multimedia. ACM, 2012
[4] Tuna, T., Subhlok, J., Barker, L., Varghese, V., Johnson, O., and Shah, S. "Development and evaluation of indexed captioned searchable videos for STEM coursework", Proceedings of the 43rd ACM technical symposium on Computer Science Education. ACM, 2012.
[5] Moritz, F., M. Siebert, and C. Meinel "Community tagging in teleteaching environments", 2nd International Conference on eEducation, e-Business, e-Management and E-Learning (to appear). 2011.
[6] Sack, Harald, and Jrg Waitelonis. "Integrating social tagging and document annotation for content-based search in multimedia data.", Semantic Authoring and Annotation Workshop (SAAW). 2006.
[7] http://www.nptel.iitm.ac.in.
[8] http://www.freevideolectures.com.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

[9] http://ocw.mit.edu/courses/audio-video-courses.
[10] H. J. Jeong,, T.-E. Kim, and M. H. Kim, "An accurate lecture video segmentation method by using sift and adaptive threshold," in Proc. 10th Int. Conf. Advances Mobile Comput, 2012.
[11] Smeaton, Alan F. "Techniques used and open challenges to theanalysis, indexing and retrieval of digital video." Information Systems 32.4 (2007): 545-559.
[12] Epshtein, Boris, Eyal Ofek, and Yonatan Wexler, "Detecting text in natural scenes with stroke width transform." Computer Vision and Pattern Recognition (CVPR), IEEE Conference 2010
[13] 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.
[14] Kim, K., Choi, J., Kim, N., Kim, P.: Extracting semantic information from basketball video based on audio-visual features. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, 268–277, Springer (2002).
[15] Bakker, E., Lew, M.: Semantic video retrieval using audio analysis. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, 260–267, Springer (2002)
[16] Pickering, M., R¨uger, S., Sinclair, D.: Video retrieval by feature learning in key frames. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, 298–306, Springer (2002)
[17] M¨uller, H., Marchand-Maillet, S., Pun, T.: The truth about Corel – Evaluation in image retrieval. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, 36–45, Springer (2002)
[18] Black, J., Fahmy, G., Panchanathan, S.: A method for evaluating the performance of content-based image retrieval systems based on subjectively determined similarity between images. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, 343–352, Springer (2002)
[19] Sebe, N., Tian, Q., Loupias, E., Lew, M., Huang, T.: Evaluation of salient point techniques. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, 353–362, , Springer (2002)
[20] Vendrig, J., Worring, M.: Multimodal person identification in movies. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, 166–175, Springer (2002).
[21] O. Javed, S. Khan, Z. Rasheed, and M. Shah, "A framework for segmentation of interview videos," in IASTED Int. Conf. Internet and Multimedia Systems and Applications United States, Las Vegas, NV, 2000.
[22] S. Boykin and A. Merlino, "Machine learning of event segmentation for news on demand," Commun. ACM, vol. 43, pp. 35–41, 2000.
[23] P. Chiu, A. Girgensohn,W. Polak, E. Rieffel, and L.Wilcox, "A genetic algorithm for video segmentation and summarization," in 2000 IEEE Int. Conf. Multimedia and Expo , vol. 3, pp. 1329–1332., , New York, 2000
[24] J. Adcock, M. Cooper, L. Denoue, H. Pirsiavash, and L. A. Rowe.Talkminer: a lecture webcast search engine. In MM '10, pages 241–250, NY, USA, 2010.
[25] T. Martin, A. Boucher, J.-M. Ogier, M. Rossignol, and E. Castelli. Multimedia scenario extraction and content indexing for e-learning. In CBMI '07, pages 204–211, 2207.
[26] M. Merler and J. R. Kender. Semantic keyword extraction via adaptive text binarization of unstructured unsourced video. In ICIP 2010, pages 261–264. IEEE, 2009.
[27] Smeaton, A., Over, P.: TRECVID: Benchmarking the effectiveness of information retrieval tasks in video. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 18–27, , Springer (2003).
[28] Liu, Y., Kender, J.: Fast video retrieval under sparse training data. In: International Conference on Image and Video retrieval, Lecture Notes in Computer Science, vol. 2728, 388–397 , Springer (2003).
[29] Shim, C.B., Chang, J.W.: Efficient similar trajectory-based retrieval for moving objects in video databases. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 158–167, Springer (2003).
[30] Joly, A., Frelicot, C., Buisson, O.: Robust content-based video copy identification in a large reference database. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 398–407, Springer (2003).
[31] Shao, H., Svoboda, T., Tuytelaars, T., van Gool, L.: HPAT indexing for fast object/scene recognition based on local appearance. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 68–77, Springer (2003).
[32] Hoi, C.H., Wang, W., Lyu, M.: A novel scheme for video similarity detection. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 358–367, Springer (2003)
[33] Kim, S., Park, S., Kim, M.: Central object extraction for object-based image retrieval. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 38–47 39., Springer (2003).
[34] Arica, N., Yarman-Vural, F.: A compact shape descriptor based on the beam angle statistics. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 148–157, Springer (2003).
[35] Chen, Z., Ding, J., Zhang, M., Tavanapong, W.: Hierarchical clustering-merging in multidimensional index structures. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 78–87, Springer (2003).
[36] Scott, G., Shyu, C.R.: EBS k-d tree: An entropy balanced statistical k-d tree for image databases with ground-truth labels. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 448–457, Springer (2003)
[37] Eakins, J., Riley, K.J., Edwards, J.: Shape feature matching for trademark image retrieval. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 28–37, Springer (2003).
[38] Cao, Y., Tavanapong, W., Kim, K.: Audio-assisted scene segmentation for story browsing. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 428–437, Springer (2003).
[39] Liu, T., Kender, J.: Spatial-temporal semantic grouping of instructional video content. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 348–357, Springer (2003).
[40] Miura, K., Hamada, R., Ide, I., Sakai, S., Tanaka, H.: Associating cooking video segments with preparation steps. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 168–177, Springer (2003).
[41] Lay, J., Guan, L.: Concept-based retrieval of art documents. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 368–377, Springer (2003).
[42] Velivelli, A., Ngo, C.W., Huang, T.: Detection of documentary scene changes by audio-visual fusion. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, 218–228, Springer (2003).