© 2016 IJSRSET | Volume 2 | Issue 5 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Survey on Content-Based Lecture Video Retrieval By Text

Prashant Zanjugade, Neha Shinde, Omkar Devle, Darshana Gandre

Department of Computer Engineering, Mumbai University, Mumbai, Maharashtra, India

ABSTRACT

E-Teaching is becoming popular nowadays, as a result of this, there is a huge increase in the amount of lecture video data on the World Wide Web. Lecture videos contain text information in the visual channels: the presentation slides and lecturer's speech. Therefore, a more effective method for retrieval of video within large lecture video archives is needed. So, in this paper, we present an approach for automated video indexing and searching of video in archives. Firstly, we apply automatic video segmentation which will fragment the video into a number of frames and then keyframe detection is applied to offer a visual guideline for the video content navigation. Subsequently, by applying video Optical Character Recognition (OCR) technology on key-frames we extract textual metadata. The OCR detects as well as transcripts slide text for keyword extraction, by which all keywords are extracted from the video for content-based video browsing and search.

Keywords: Lecture Videos, Video Segmentation, Optical Character Recognition (OCR), Keyframe detection, Indexing, Ranking.

I. INTRODUCTION

DIGITAL video has become a popular storage and exchange medium due to the rapid development in recording technology, improved video compression techniques and high-speed networks in the last few years. Therefore audiovisual recordings are used more and more frequently in e-lecturing systems.A number of Universities and Institutions are taking opportunities to record their lectures and publish them online for students to access.It is impossible to find desired videos without search function within video archive. Even when the user has found related video it is still difficult for him to judge whether the video is useful or not only glancing at the title. Therefore beyond current approaches method for retrieval urgently needed.Moreover, the requested information may be covered in only a few minutes, the user might thus want to find the piece of information he requires without viewing the complete video. The problem becomes how to retrieve the appropriate information in a large lecture video archive more efficiently.

Therefore, beyond the current approaches, the next generation of video retrieval systems apply automatically generated metadata by using video analysis technologies. The relevant metadata can be automatically gathered from lecture videos by using appropriate analysis techniques. They can help a user to find and to understand lecture contents more efficiently, and the learning effectiveness can be improved.We present an approach for automated video indexing and video search in large lecture video archive. We extract the outline of a lecture from visual resources and organize videos based on contents.

Proposed System provides different functionality for two main clients which are Administrator and user. The administrator is responsible for controlling the entire database including security and adding, updating and deleting videos to and from the database. The administrator also can add or remove the user as well as it can search the specific video by using text query. The user can only retrieve videos based on submitted query based on the content i.e. text.

II. METHODS AND MATERIAL

1. Literature review

A. Content-based Video Retrieval Systems - Methods, Techniques, Trends and Challenges

Content-Based Video Retrieval (CBVR) has been increasingly used to describe the process of retrieving desired videos from a large collection on the basis of features that are extracted from the videos. The extracted features are used to index, classify and retrieve desired and relevant videos while filtering out undesired ones. Videos can be represented by their audio, texts, faces and objects in their frames. An individual video possesses unique motion features, color histograms, motion histograms, text features, audio features, features extracted from faces and objects existing in its frames. Videos containing useful information and occupying significant space in the databases are under-utilized unless CBVR systems capable of retrieving desired videos by sharply selecting relevant while filtering out undesired videos exist. Results have shown performance improvement (higher precision and recall values) when features suitable to particular types of videos are utilized wisely. Various combinations of these features can also be used to achieve desired performance. In this paper, a complex and wide area of CBVR and CBVR systems have been presented in a comprehensive and simple way. Processes at different stages in CBVR systems are described in a systematic way. Types of features, their combinations and their utilization methods, techniques and algorithms are also shown. Various querying methods, some of the features like GLCM, Gabor Magnitude, an algorithm to obtain similarities like Kullback-Leibler distance method and Relevance Feedback Method are discussed.

Shot matching is done and hence video retrieval is achieved by comparing their features. The drawback of techniques employing key frames matching is that temporal information and the related information between the key frames in a shot is lost.

B. Efficient Video Feature Extraction and Retrieval on Multimodal Search

The core research in Content-Based Video Retrieval (CBVR) is to automatically parse video and text to identify meaningful composition structure. To facilitate fast and accurate content access to video data, a video document should be segmented into shots and scenes. Recognizing objects sequence from videos is an important Problem of computer vision applications such as web searching, the target recognition, surveillance, crime detection etc. To build an Efficient video retrieval system that focuses on features such as color, texture, shape, motion, Visual text embedded in an image .Multimodal is the capacity of the system to communicate with one or more input given in search process, It can text, Image, audio embedded in it with signals based, Text embedded in an image. Thus it provides potentially accurate results. These results are used in video searching, video surveillance, text embedded in the image. In CBVR, a Video is segmented for its preprocessing Key Frames are used for Feature Extraction. Then clustering and indexing is done with k-means clustering HCT(Hierarchical Clustering Tree),Then Similarity matching is done.

C. Multimedia Indexing and Retrieval Techniques

Retrieval of multimedia has become a requirement for many contemporary information systems. These systems need to provide browsing, querying, navigation, and, sometimes, composition capabilities involving various forms of media. In this survey, we review techniques for text, image, audio and video retrieval. We first look at indexing and retrieval techniques for text, audio, image and video. We also discuss features visual features for video retrieval such as color, texture, shape. The indexing techniques are discussed for these features. They also compare most popular techniques used for indexing and retrieval.

D. Content-Based Video Retrieval System

Content-based video retrieval is an approach for facilitating the searching and browsing of large image collections over World Wide Web. In this approach, video analysis is conducted on low-level visual properties extracted from the video frame. We believed that in order to create an effective video retrieval system, visual perception must be taken into account. We conjectured that a technique which employs multiple features for indexing and retrieval would be more effective in the discrimination and search tasks of videos. In order to validate this claim,

content-based indexing and retrieval systems were implemented
using the color histogram, various texture features, and other approaches. Videos were stored in Oracle 9i Database and a user study measured the correctness of the response.

2. Flow of System

We present an approach for automated video indexing & video search engine to find particular video from large lecture video archives. It will extract the outline of the lectures from visual resources and organize the videos based on the contents. This method will reduce the time required for searching the desired video. The process flow for intended retrieval system is described below.

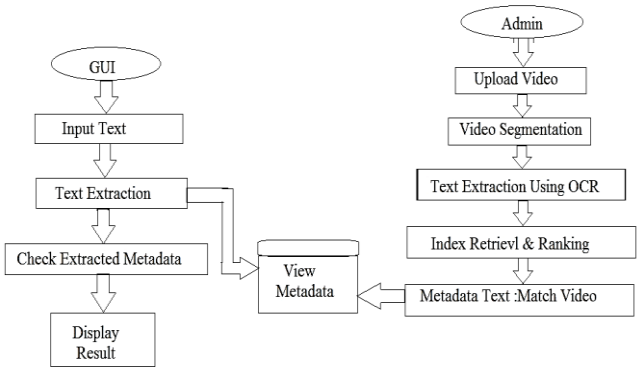

Figure 1. Flow of System

We propose to develop and implement a system for searching the specific video from large video database. The content based lecture video retrieval system is to search lecture video efficiently within lecture video database using text occurs in the lecture videos. In the system's flow process the entire video is analyzed first, then the process of segmentation is carried out on that video. We try to capture frames from video by providing appropriate time interval, this provides us multiple redundant frames. For removing these redundant frames we use sequential comparison-based approach. After gathered these keyframes we apply OCR(Optical character recognition) technique, which will extract textual metadata. This data is then index by using inverted indexing. For retrieval, the specific video system should know the weight of every word which occurs in the document. TF-IDF(Term Frequency-Inverse Document Frequency) is a weighting factor used for getting the how many time the word occurs in the document? Then videos are arranged in decreasing order of the calculated frequency.

Ranking of videos is done by Relevance Feedback method, this method will improve the retrieval performance. We will use vector space model which contain three module TF-IDF weighting factor, inverted indexing and relevance feedback method.

3. Positioning Techniques

This section presents the various techniques that are appropriate for Content-Based Lecture Video Retrieval By Text.

1. Video Segmentation

In the proposed algorithm, a large video is being taken as an input and multiple frames are created as an output. Steps for the algorithm are as follows:

i. Input Video.

ii. Create frames from the input video after every constant interval of the time till video ends.

2. Sequential Comparison-Based Approach for key frame selection

In these algorithms, frames subsequent to a previously extracted key frame are sequentially compared with the key frame until a frame which is very different from the key frame is obtained. This frame is selected as the next key frame. The merits of the sequential comparison based algorithms include their simplicity, intuitiveness, low computational complexity, and adaptation of the number of keyframes to the length of the shot.

However, these algorithms suffer from problems such as the keyframes represent local properties of the shot rather than the global properties, the irregular distribution and uncontrolled number of key frames make these algorithms unsuitable for applications that need an even distribution or a fixed number of keyframes and also redundancy can occur when there are contents appearing repeatedly in the same shot.

3. Optical Character Recognition For Text Extraction

OCR is the electronic conversion of scanned images of printed text into computer-readable text. OCR is used for text extraction of a video. It is a common method of digitizing printed texts so that they can be easily edited, searched, stored more compactly, displayed on-line, and used in machine processes such as machine translation, text-to-speech, key data extraction and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision .

3.Vector Space Model

A. Inverted Indexing

i. Fetching the Document

The job is really simple if I get a text file (.txt). But if it was a doc or pdf, I'll need to parse them using some libraries to retrieve their text. Let's say I'm successful in reading the text. What next?

ii. Removing the Stop Words

Consider the last paragraph. What were the important words we may be looking for? "text", "libraries", "doc", "pdf", "retrieve", "successful". But most of the other words are just a waste. We denote the most occurring words as "stop words" and remove them so that I don't get indexes for words like "I", "the", "we", "is", "an". In regular use, we have a list of 500-1000 words. But it may differ depending on use.

iii. Stem to the Root Word

Then comes Stemming. Now whenever I want to search for "retrieval", I want to see a document that has information about it. But the word present in the document is called "retrieve" instead of "retrieval". To relate the both words, I'll chop some part of each and every word I read so that I could get the "root word". Retrieve may become "retriev". So will "retrieval". We have to be sure about the rules we use to chop the words. There are standard tools for performing this like "Porter's Stemmer". You can play around with a porter stemmer here : Porter Stemmer Online

iv. Record Document IDs

Now get ready for the main task - Indexing.
Every document I have has got a unique document id. As I encounter a non-stop word that is stemmed now, I save it in my memory in the form :
retriev ==> docID104007
If I get same word in some other document, I may write
retriev ==> docID104007
retriev ==> docID154033
But very soon I've to combine them in a single list
retriev ==> docID104007&docID154033
I can further improve by writing how many timed did the word occur in the document so that we can rank the most important documents while retrieving.
retriev ==> docID104007|5|&docID154033|2|

v. Merge and Store the Terms

Finally, we save all of them in disk files. It's great if we sort the index based on the words for quick and easy retrieval.

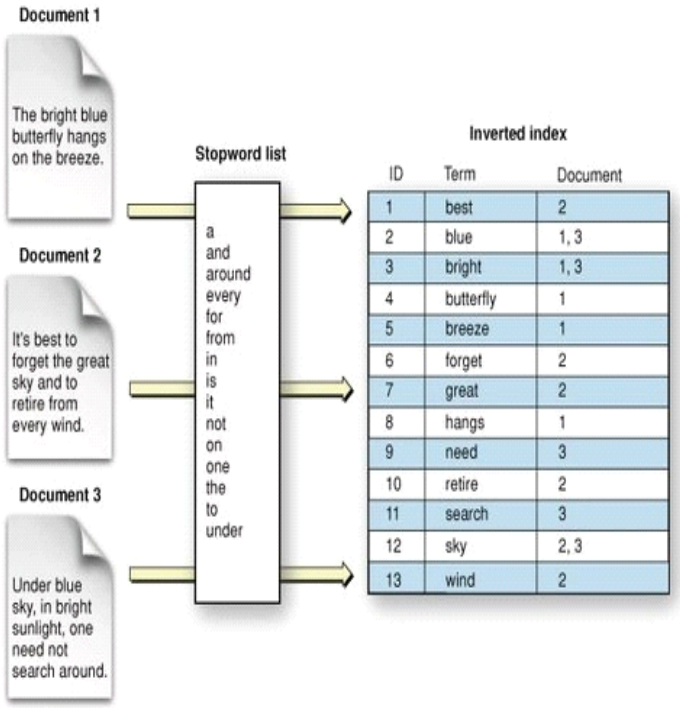

Figure 2. Inverted Indexing

4. TF-IDF weighting factor

TF-IDF: It is used as a weighting factor in information retrieval and text mining. The tf-idf value increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus, which helps to control for the fact that some words are generally more common than others. Variations of the tfidf[2] weighting scheme are often used by search engines as a central tool in scoring and ranking a document's relevance given a user query.

TF: term frequency tf(t,d), the simplest choice is to use the raw frequency of a term in a document, i.e. the number of times that term t occurs in document d. If we denote the raw frequency of t by f(t,d), then the simple tf scheme is

$$tf(t,d) = f(t,d)$$

IDF: The inverse document frequency is the logarithmically scaled fraction of the documents that contain the word, obtained by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient.

$$idf(t,d) = log\frac{N}{|(d\epsilon D{:}t\epsilon d)|}.$$

Then tfidf is calculated as –

$$tfidf(t,d,D) = tf(t,d) * idf(t,D)$$

5. Relevance feedback method for ranking

To improve the retrieval performance, relevance feedback technique can be used . Systems using relevance feedback are effective in ranking and retrieving similar videos.Relevance feedback is an important strength of the vector-space model. The idea is that queries specified by end users typically fail to describe completely what those users want. Thus, typical queries miss many relevant documents. However, if the user can identify some retrieved documents as relevant, then the system can use this information to reformulate the original query into a new one that may capture some of the concepts not explicitly specified in the original query.

The main problem with relevance feedback is that including a large piece of text indiscriminately in the original query will adversely affect effectiveness. Doing so also incurs the typically high cost of processing long queries. Continuing with our example, if the relevant document contains “gold price in the London market,” then “London” and “market” may also be included in the reformulated query, resulting in retrieval of irrelevant articles mentioning “London.” Therefore, criteria must be carefully set up to select terms from the relevant texts for query expansion. We define three methods— High, Mid, and Low—that select high-, medium-, and low-frequency terms, respectively, from the relevant text for query expansion.

III. RESULTS AND DISCUSSION

A performance of video retrieval is evaluated with recall-precision values. Recall and precision are the two parameters as given in equations .

Recall: Percentage of the Relavant items that are found by the query,the extent to which the query found all the items that satisfies the requirment.

$$\text{Recall} = DC/\,DB$$

Precision: Percentage of the hit that are relavant to extent to which the set of hits retrieved by a querysatisfies the requirment that generated the qyery.

$$\text{Precision} = C\,/DT$$

DC = number of similar clips detected correctly

DB = number of similar clip in the database

DT = total number of detected clips

IV. CONCLUSION

The existing search engines results are not that much content based, one cannot judge whether it contains required data thoroughly, required information may be covered within few minutes. So content based lecture video retrieval approach presents an appropriate method to retrieve the desired video more efficiently which is more content specific, which in turns increases learning effectiveness? This will also reduce the required time for searching.

V. REFERENCES

[1] Haojin Yang and Cristoph Meinel, “Content Based Lecture Video Retrieval Using Speech And Video Text Information”, IEEE transactions on learning technologies, vol. 7,no. 2,April June 2014.

[2] Pratyush Pranjal, Rohit Lakhdive, Aditya Vasave, AtulkumarVarma and AmitBarve ,"An Effective Method of Video Segmentation and Summarization for Surveillance ",International

[3] Journal of Computer Applications (0975 – 8887) Volume 95 – No 2, June 2014 Aasif Ansari and Muzammil H Mohammed,"Content based Video Retrieval Systems - Methods, Techniques, Trends and Challenges ",International Journal of Computer Applications (0975 – 8887) Volume 112 – No. 7, February 2015

[4] 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