



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

An Efficient Video Lecture Retrieval Using OCR And Recommendation

Bhagyashri D.Deshmukh, Prof.Y.B.Gurav

M.E., Department of Computer Engineering, Padmabhushan Vasantdada Patil Institute Technology, Bavdhan, Pune, India

Department of Computer Engineering, Padmabhushan Vasantdada Patil Institute Technology, Bavdhan, Pune, India

ABSTRACT: In today's advanced development in multimedia technology allows the capture and storage of video data with relatively very expensive computers. There are new possibilities offered by the information technology have made a large amount of video data publicly available. Without appropriate search techniques to use all these data are very hard. Users are not satisfied with the video retrieval systems that provide analogue VCR functionality. For example, a user analyses a football video will ask for specific events such as goals. That is challenge for video lecture retrieval system of video an important problem. So, there is need for tools that can be manipulate the video content in the same way as traditional databases manage numeric and textual data is significant. Therefore, a more efficient method for video retrieval in World Wide Web or within large lecture video archives is urgently needed. This project presents an approach for automated video indexing and video search in large lecture video archives. In First step, we apply automatic video segmentation and key-frame detection to offer a visual guideline for the video content navigation. Subsequently, we extract textual metadata by applying video Optical Character Recognition (OCR) technology using OCR algorithm on key-frames and (ASR) Automatic Speech Recognition on lecture audio tracks of the video using Google's API. We provide new recommendation method here which is similar to item to item and Pearson recommendation method .

KEYWORDS: video segmentation, video browsing, video retrieval, Optical Character Recognition, Automatic Speech Recognition, Application Programming Interface.

I. INTRODUCTION

Over the last few years, e-lecturing has become very popular. The most of college students are trying to interact with and trying to learn from this e-lecture. Due to very fast development in the recording technology, digital videos are becoming a popular storage and exchange medium. Example of E-lecturing can be that number of universities and research institutions are using it to record their lectures and publish them online for students to access them independent of time and location. As a result, there has been a huge amount of multimedia data on the Web which is very difficult for user to judge whether a video is useful by glancing at the title only and search desired videos without searching within the video archives. So, user can find the piece of information he requires without viewing the complete video. The problem is that how to retrieve the appropriate information in a large lecture video archive more efficiently. The manually provided metadata means data related to video is typically brief, high level and subjective. The next generations of video retrieval systems apply automatically generated metadata by using video analysis technologies.

Hypothesis is of this content based lecture video system is that the relevant metadata can be automatically gathered from lecture videos by using appropriate analysis techniques. These information can help a user to find and to understand lecture contents more efficiently, and the learning effectiveness can thus be improved.

This video lecture retrieval system works with main focus on the lecture videos produced by using the screen grabbing method. Segmenting two-screens lecture videos can be achieved by only processing slide video streams, which contain most of the visual text metadata.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Key frame recognition by OCR technology:-

The OCR technique is Optical Character Recognition only. The OCR technique is used for extracting textual metadata. By using OCR we can convert different type of documents such as paper document and image capture by digital camera into editable and searchable data. With OCR the recognized document looks like the original. The OCR algorithm allows you to save a lot of time and efforts when creating and processing and repurchasing various documents. The search indices are created based on the global metadata obtained from the video hosting website and texts extracted from slide videos by using a standard Optical Character Recognition engine.

Speech recognition by AUTOMATIC SPEECH RECOGNITION technique:

The AUTOMATIC SPEECH RECOGNITION technique is nothing but the Automatic Speech Recognition. It can provide speech to text information from lecture video. In computer science speech recognition is the translation of spoken word into text. Speech is one of the most important carriers of information in video lectures. The poor recognition results not only limit the usability of speech transcript, but also affect the efficiency of the further indexing process. In this paper, we intend to continuously improve the AUTOMATIC SPEECH RECOGNITION result based on the open-source AUTOMATIC SPEECH RECOGNITION Tool.

A large amount of textual metadata will be created using OCR and AUTOMATIC SPEECH RECOGNITION method, which opens up the content of lecture videos. In this project, text metadata is extracted from visual as well as audio resources of lecture videos automatically by applying appropriate analysis techniques. For evaluation purpose we developed several automatic indexing functionalities in a large lecture video portal, which can guide both visually- and text-oriented users to navigate within lecture video. Following fig 1 shows the architecture of the system.

In summary, the major contributions of this paper are the following:

- a) For visual analysis, we propose a new method for slide video segmentation and apply video OCR to gather text metadata. For OCR algorithm is used for extracting text from slides.
- b) In order to remove garbage value from extracted text metadata Stopword removal method is used. After that using xugger API, which is Google's API automatic speech recognition process is done.
- c) To improve the video retrieval result used the Recommendation of videos using user ratings. For recommendation purpose Item-to-item collaborative filtering algorithm and Pearson recommendation algorithm is used.

The rest of the paper is organized as follows: Section 2 reviews of related work and Section 3 is related to automated lecture video indexing; Next section describes Data Preprocessing. After that next section gives details about recommendation process and last result and conclusion.

II.RELATED WORK

H.Sack and J. Waitlonis apply tagging data for lecture video retrieval and video searching[3].The authors T. C.

Pong, F. Wang et al. proposed a new approach for lecture video indexing based on video segmentation and OCR analysis [7]. Grcar et al. introduced videoLectures.net in [1] which is a digital archive for multimedia presentations. Similar to [7], the authors also apply a synchronization process between the recorded lecture video and the slide file, which has to be provided by presenters. Our system contrasts to these two approaches since it directly analyzes the video, which is thus independent of any hardware or presentation technology. The constrained slide format and the synchronization with an external document are not required.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

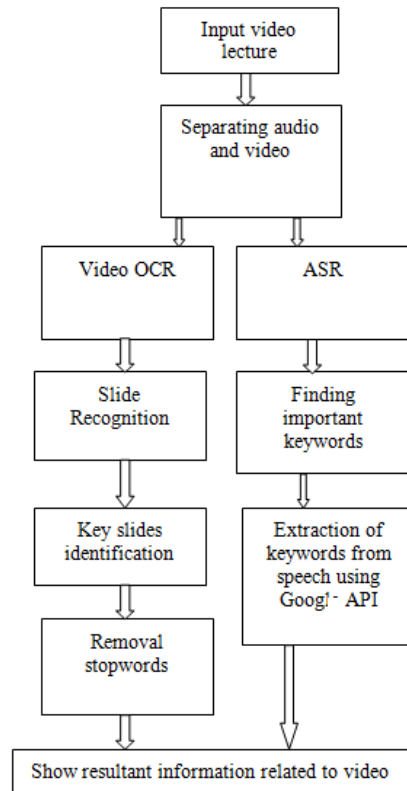


Fig 1:Architecture of Video Retrieval system

H. J. Jeong, T.E. Kim, and M. H. Kim, Proposed a lecture video segmentation method using Scale Invariant Feature Transform (SIFT) feature and the adaptive threshold in [2]. In their work SIFT feature is applied to measure slides with similar content. An adaptive threshold selection algorithm is used to detect slide transitions. The authors T. Tuna, J. Subhlok, L. et al., proposed an approach for Lecture video indexing and search [8].It might have limited use in large lecture video archives with massive amounts of users.

III.MATHEMATICAL MODEL

An efficient Video Lecture Retrieval using OCR system S is defined as:

$$S = \{V, K, I, T\}$$

Where,

$$K = \{K1, K2, K3, \dots, Kn\}$$

$$V = \{V1, V2, V3, \dots, Vn\}$$

$$I = \{I1, I2, I3, I4, \dots, In\}$$

$$T = \{T1, T2, T3, T4, \dots, Tn\}$$

Function $f1(V)$ captures the images from video lecture.

$$f1(V) \rightarrow \{V\} \in \{I1, I2, I3, \dots, In\}$$

Function $f2(I)$ stores the captured the images from video lecture using function $f1(V)$. Initially this set is empty.

$$f2(I) \rightarrow \{\}$$

Function $f3$ extract the text metadata and converted it into important keywords.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

$f3(T) \rightarrow \{T1, T2, T3, \dots, Tn\} \in \{K1, K2, K3, \dots, Kn\}$

Where,

K: is the Set of Keywords

V: set of Videos

I: set of images captured from slides

T: set of Texts

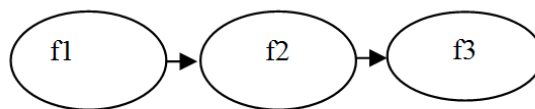


Fig 2: Mathematical Model

IV. DATA PREPROCESSING

In this project, for removing garbage value from extracted text data obtained in OCR processing following stemming method is used for removal of stopword.

What is stemming Algorithm?

A stemming algorithm is nothing but a process of linguistic normalization, in which the variant forms of a word are reduced to a common form.

It is important to appreciate that we use stemming with the intention of improving the performance of content based lecture video retrieval systems.

It is not an exercise in etymology or grammar. It is much easier to write a stemming algorithm for a language when you are familiar with it. Building the stemming algorithm is a process of trying out a small number of ending removals at a time.

For each new ending plus rule added in stemming, decide whether, on average, the stemming process is improved or degraded. If it is degraded the rule is not helpful and can be discarded. What you find eventually is that you can be improving performance in one area of the vocabulary, while causing a similar degradation of performance in another area. When this happens consistently it is time to call a halt to development and to regard the stemming algorithm as finished. It is important to realise that the stemming process cannot be made perfect.

Stopwords:

In this system, stopwords removal method is used. It has been traditional in setting up IR systems to discard the very commonest words of a language the stopwords during indexing. A more modern approach is to index everything, which greatly assists searching for phrases for example. Stopwords can then still be eliminated from the query is as an optional style of retrieval. In either case, a list of stopwords for a language is useful. Getting a list of stopwords can be done by sorting a vocabulary of a text corpus for a language by frequency, and going down the list picking off words to be discarded. The stopword list connects in various ways with the stemming algorithm:

The stemming algorithm can itself be used to detect and remove stopwords. One would add into the irregular-forms table something like this,

```
"" /* null string */  
  
"am/is/are/be/being/been/" /* BE */  
"have/has/having/had/" /* HAD */  
"do/does/doing/did/" /* DID */  
... /* multi-line string */
```



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

so that the words 'am', 'is' etc. map to the null string (or some other easily recognized value). In this video lecture retrieval system, first array of stopwords is created and after extracting text data from video that stopwords are cleared from that text metadata.

V.AUTOMATED LECTURE VIDEO INDEXING

In this section, present analysis processes for retrieving relevant metadata from the two main parts of lecture video, namely the visual screen and audio tracks. From the visual screen of video we firstly detect the slide transitions and extract each unique slide frame with its temporal scope considered as the video segment. Then the video OCR analysis is performed for retrieving textual metadata from slide frames of video. Based on OCR results, we propose a novel solution for lecture outline extraction. In speech-to-text analysis we applied the open-source xuggler Google's API for ASR technique.

Video OCR:

In this video OCR process, following two steps are done for Optical Character Recognition.

First step in this project is segmenting video into representative key frames. The selected key frames can provide a visual guideline for navigation in the lecture video portal. Moreover, video segmentation and key-frame selection is also often adopted as a pre-processing for other analysis tasks such as video OCR, visual concept detection, etc. often considered as a video segment.

We set the time interval for short time period such as 50 seconds for capturing images of slides from Lecture video. Second step of Video OCR is extracting text from that images using OCR Algorithm which is Kohonen Neural Algorithm. Similar text will be deleted in this step and then apply stemmer algorithm for stop word removal. Here we watch every word with our technical dictionary. So finally all unique words are maintained. The Kohonen Neural Algorithm is used for Optical Character Recognition technique. Now see the working of Kohonen Algorithm:

Kohonen Neural Network Algorithm

A Kohonen neural network contains only two levels. The network is presented with an input pattern that is given to the input layer. In this algorithm, this input pattern must be normalized to numbers in the range between -1 and 1. The output from this neural network will be one single winning output neuron. The output neurons can be thought of as groups that the Kohonen neural network has classified the input as part of. To train the Kohonen neural network we present it with the training elements and see which output neuron "wins". This winning neuron's weights are then modified so that it will activate higher on the pattern that caused it to win. There is also a case where there may be one or more neurons that fail to ever win. Such neurons are dead-weight to the neural network. We must identify such neurons and cause them to recognize patterns that are already recognized by other more "overworked" neurons. This causes the burden of recognition to fall more evenly over the output neurons. Self-organization is an unsupervised learning algorithm used by the Kohonen Feature Map neural net.

A neural net tries to simulate the biological human brain and self-organization is probably the best way to realize this. It is commonly known that the context of the human brain is subdivided in different regions, each responsible for certain functions. The neural cells are organizing themselves in groups, according to those incoming information are not only received by a single neural cell, but also influence other cells in its neighborhood. This organization results in some kind of a map, where neural cells with similar functions are arranged close together. This self-organization can also be performed by a neural network. Those neural nets are mostly used for classification purposes, because similar input values are represented in certain areas of the net's map

Steps of Kohonen Neural Algorithm:

A sample structure of a Kohonen Feature Map that uses the self-organization algorithm is as follow:

The algorithm works as follows:

1. Define the range of the input values



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

2. Set all weights to random values taken out of the input value range
3. Define the initial activation area.
4. Take a random input value and pass it to the input layer neuron.
5. Determine the most activated neuron on the map:
Multiply input layer's output with weight values
The map neuron with the greatest resulting value is said to be "most activated"
Compute the feedback value of each of the map neuron using the Gauss function.
6. Change the values using the formulae:
Weight(old)+feedback value*(input value – weight(old))*learning rate
7. Decrease the activation area
8. Go to step 4
9. End algorithm if the activation area is smaller than a specified value

VI. RECOMMENDATION FOR USER RATING

The Recommender systems or recommendation systems are a subclass of information filtering system that seek to predict the 'rating' or 'preference' that user would give to an item. Recommender systems have become extremely common in recent years, and are applied in a variety of applications like for music, movies. The most popular ones are probably movie, music, news, books, research articles, search queries, social tags, and products in general. However, there are also recommender systems for experts, movies, restaurants, financial services, life insurance, persons (online dating), and Twitter followers. Recommender systems typically produce a list of recommendations in one of two ways through collaborative or content-based filtering. Collaborative filtering approaches building a model from a user's past behavior (items previously purchased or selected and/or numerical ratings given to those items) as well as similar decisions made by other users; then use that model to predict items (or ratings for items) that the user may have an interest in. Content-based filtering approaches utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties. In this project we used Collaborative filtering methods which are based on collecting and analyzing a large amount of information on users' behaviours, activities or preferences and predicting what users will like based on their similarity to other users. A key advantage of the collaborative filtering approach is that it does not rely on machine analyzable content and therefore it is capable of accurately recommending complex items such as movies without requiring an "understanding" of the item itself. Many algorithms have been used in measuring user similarity or item similarity in recommender systems. For example, the k-nearest neighbor (k-NN) approach and the Pearson Correlation. One of the most famous examples of collaborative filtering is item-to-item collaborative filtering, an algorithm popularized by Amazon.com's recommender system. In this project we proposed new recommendation method which handles both cases of recommendation one is that it matches each of the user's purchased and rated items to similar items, then combines those similar items into a recommendation list and second is to check similarity between active user and other users who rated video. This proposed method is a combination of item-to-item collaborative filtering algorithm and Pearson Correlation.

We used Hybrid Recommendation approach for this system.

Algorithm Steps are as follow:

1. Start
 2. Get user rating information.
Eg. User item rating (0-5)
- | | | |
|---|---|---|
| 1 | 2 | 4 |
|---|---|---|
3. From current user get rating information for current video.
Rci=N
where N is 0 -5 Rating
and Rci is the Rating of current user to item
 4. Apply the item to item Recommendation Algorithm
 - I. For all item in dataset do following.
 - II. for(for item I1 to In in dataset)
 - i. For(item Ii-1 to In in dataset)
 - ii. Get rating of all user of Ii item

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

- iii. Get rating of all user of I_{i-1} item
- iv. Now find cosine angle between them. Eg. $\cos \Theta = (I1.I2)/(|I1|.|I2|)$
- v. If value is above T then both item have similarity
Else
Both item don't have any similarity.
Where T is Threshold value $T=0.65$
- vi. Go to step i. Do this for all items.

5. Pearson Correlation score calculating for all user in the database.

For(all user $V_i \sum V_i, V_{i+1} \dots V_n$ in database)
For(all user V_{i+1} to $\sum V_{i+2}, V_{i+3} \dots V_{i+n}$)

- I. Get all rating of V_i to all other item.
- II. Get all Rating of V_{i+1} to all other item .
- III. Apply following formulae

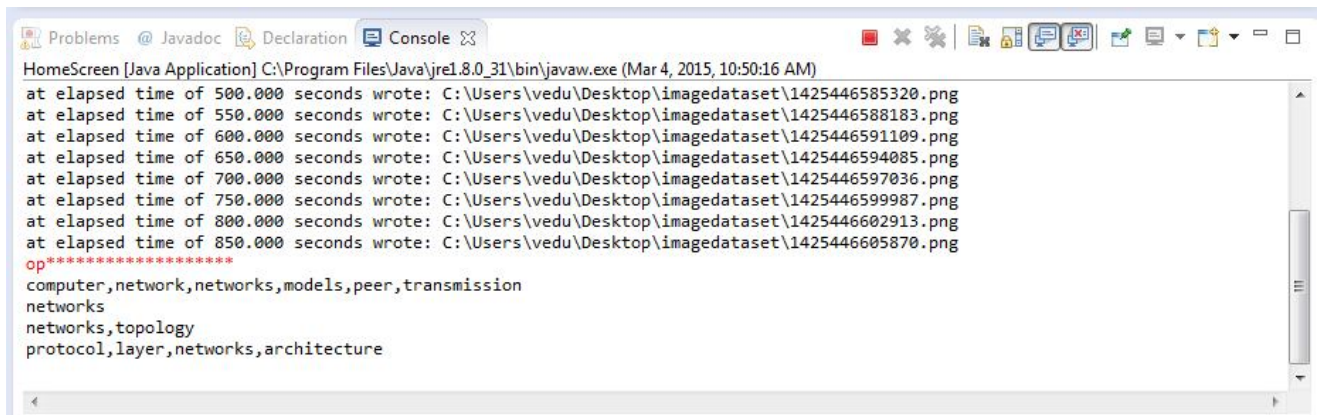
$$W_{xy} = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{(n \sum X_i^2 - (\sum X_i)^2)} \sqrt{(n \sum Y_i^2 - (\sum Y_i)^2)}}$$

Where n =number of user
 X_i = X user of particular i^{th} item
 Y_i = Y user rating for i^{th} item

6. At the end we combine both Recommendation result of these two algorithm and display to the user.

VII.RESULT

The implementation of this video lecture retrieval system is basically on the text information. When we select networking video as a input to the system ,then first data preprocessing is done. After that text information is taken from slides and then used for searching desired video. This is very efficient method for video lecture retrieving called as Optical Character Recognition. This OCR process result is shown in following fig 3 and from that we can calculate precision and recall value of system to measure system performance.



```
HomeScreen [Java Application] C:\Program Files\Java\jre1.8.0_31\bin\javaw.exe (Mar 4, 2015, 10:50:16 AM)
at elapsed time of 500.000 seconds wrote: C:\Users\vedu\Desktop\imagedataset\1425446585320.png
at elapsed time of 550.000 seconds wrote: C:\Users\vedu\Desktop\imagedataset\1425446588183.png
at elapsed time of 600.000 seconds wrote: C:\Users\vedu\Desktop\imagedataset\1425446591109.png
at elapsed time of 650.000 seconds wrote: C:\Users\vedu\Desktop\imagedataset\1425446594085.png
at elapsed time of 700.000 seconds wrote: C:\Users\vedu\Desktop\imagedataset\1425446597036.png
at elapsed time of 750.000 seconds wrote: C:\Users\vedu\Desktop\imagedataset\1425446599987.png
at elapsed time of 800.000 seconds wrote: C:\Users\vedu\Desktop\imagedataset\1425446602913.png
at elapsed time of 850.000 seconds wrote: C:\Users\vedu\Desktop\imagedataset\1425446605870.png
op*****
computer, network, networks, models, peer, transmission
networks
networks, topology
protocol, layer, networks, architecture
```

Fig 3: Output of OCR process

Precision can be seen as a measure of exactness or *quality*, whereas recall is a measure of completeness or *quantity*. From above image we can say that this system gives good Precision and Recall value.

VIII.CONCLUSION

This project, provide very efficient novel video retrieval system for user. Users can judge the video using this lecture video retrieval system using text. To improve the video retrieval result used the Recommendation of videos using user ratings.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

IX.FUTURE SCOPE

In future we can do the Automatic Speech Recognition work using Google API. It helps to find nearly exact video which user wants from video lecture retrieval using the text information which we extract from audio.

REFERENCES

- [1]Greg Linden, Brent Smith, and Jeremy York, " Amazon.com Recommendations Item-to-Item Collaborative Filtering", Industry Report
- [2] 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. 285–288.[Online]. Available: <http://doi.acm.org/10.1145/2428955.2429011>.
- [3] H. Sack and J. Waitelonis, "Integrating social tagging and document annotation for content-based search in multimedia data," in Proc. 1st Semantic Authoring Annotation Workshop, 2006.
- [4] Haojin Yang and Christoph Meinel, Member, "Content Based Lecture Video Retrieval Using Speech and Video Text Information ", IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 7, NO. 2, APRIL-JUNE 2014
- [5] 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. 730–733.
- [6] Priyanka Shenoy, Manoj Jain, Abhishek Shetty, Deepali Vora ,," Web Usage Mining Using Pearson's Correlation Coefficient",International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 3, Issue 2, March -April 2013, pp.676-679 676
- [7] 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. 3257–3269, 2008.
- [8] 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. [Online]. Available: <http://doi.acm.org/10.1145/2157136.2157177>