Content based Video Retrieval using Text Annotation and Low Level Features Technique

Aniket Sugandhi
Research Scholar
Department of Information Technology, SVITS,
Indore, India

Deepshikha Sharma
Assistant Professor
Department of Information Technology, SVITS,
Indore, India

ABSTRACT
The information retrieval processes are playing essential role in the computer based database exploration or finding the essential contents from the databases. Now in these days a number of search techniques and retrieval models are exist by using which the users can find the data. According to the different data formats the information retrieval processes are also varying therefore different data format based retrieval process are works in different manner. In this presented work the content based video retrieval model is presented. In the content based video retrieval model the work is initiated from the segmentation of the videos into the set of frames. The segmentation of video is performed using the FFMPEG media library that works on the basis of two parameters first the video clip and second the time slot duration by which the frames are extracted. After segmentation to enable the text based query the text annotation concept is used and the individual frames of video is tagged with the user defined text. On the other hand to enable the query by example the low level features of individual frames are also computed. To compute the low level descriptors shape, color and texture analysis is performed. Thus the canny edge detection, local binary pattern and the color movement analysis techniques are used. Finally for classifying the relevant videos according to the user query the KNN classifier is implemented. That classifier accepts the user text query or the example query for classifying the similar video objects from the database. The implementation of the presented methodology is performed using JAVA technology. Additionally for finding the performance precision, recall, and f-measures are computed, according to the computed values the example based techniques provides more accurate outcomes as compared to text based query. In addition of that the resource consumption of the proposed technique is also computed in terms of time and space, according to the results the example based query processing consumes additional resources as compared to text based methods. Thus the proposed work is accomplished with the satisfactory performance.

Keywords
Information Retrieval, Video Data Analysis, Search Relevancy, Improvement of Methods

1. INTRODUCTION
After invention of the computers the data is stored in the digital formats. The digital data consumes less space for storing the large data. In addition of that for utilizing the stored data in databases there is need to prepare a retrieval system which can recognize the user query and extract the required knowledge from database. Such kinds of systems are known as the information retrieval systems. The retrieval systems are designed for the data centric approaches. In other words the data processing techniques of different retrieval systems may depends on the kind of data and the formats of data. The information retrieval systems can be developed for web documents, text documents, images, videos and other formats. In this presented system the video retrieval techniques are evaluated and investigated.

The video retrieval systems can be categorized in two broad categories. Text based retrieval systems and the contents based retrieval systems. In text based retrieval systems the videos are retrieved using the associated description, tags, title and other meta-data information. But due to lingual differences in text the performance of text based retrieval systems are performed week as compared to content based methods. In content based techniques the data is retrieved on the basis of the hidden features of the targeted objects therefore the relevancy of the content based techniques are much adoptable as compared to the text based techniques.

In this presented work a content and text annotation based video retrieval data model is demonstrated. The given technique usages both the concepts namely text based retrieval and content based retrieval methods. Thus the proposed working model improvers the performance of the traditionally available techniques of information retrieval.

2. PROPOSED WORK
The various techniques and methods are available for relevant video retrieval. In this presented work the content based video retrieval technique is proposed and their functional aspects are formulated. Therefore this section includes the domain overview, the methodology of the system development and the step process in terms of algorithm.

2.1. Domain overview
Data Retrieval is the process to select or extract user query relevance data from a file or a group of files. The data can be found in structured or in unstructured format. Among them database management systems allows the data in a structured way. Additionally for retrieval of required data searching is performed by processing user input query in terms of Structured Query Language (SQL). But when the data is in unstructured format the retrieval processes becomes complicated. Video has been more widely used data format in computer technology. That is a complex medium of data additionally contains large amount of visual and audio based information and images. On the other hand the search on such complicated data source is a difficult task as compared to other unstructured data sources such as text or web documents.

In this presented work the video retrieval system is investigated and explored. Retrieving of video clips accurately relevant from the user input query from a huge video library is one of the key issues in development of video database. Former video based information retrieval system is finding videos based on keywords, meta-data, and description or by the titles associated with the data. But manual treatment of data can increases time and effort, additionally text

associated with video is not defining the entire video contents more appropriately, or additionally the lingual differences cause a large number of poor results.

The management of multimedia data is one of the crucial tasks in the data mining due to unstructured nature of data. The main challenge is to handle data with complex structure such as video and audio. Now a day's people have access a tremendous amount of video from internet. The video consists of a sequence of images with some temporal information. The video content may be classified into three categories, namely.

1. Low-level feature information that includes features such as color, texture, shape and so on,
2. Syntactic information that describes the contents of video, including salient objects, their spatial-temporal position and spatial temporal relations between them, and
3. Semantic information, which describes what happening in the video along with what is perceived by the users.

The higher-level semantic information of video is extracted by examining the features of the audio, video and text. Additionally different other features are fully exploited and used to capture the semantics for bridging the gap between the high level semantic and the low level features.

Video databases are widespread and video data sets are extremely large. There are tools for managing and searching within such collections, but need some tool to extract hidden knowledge within the video data for applications. Some issues regarding data retrieval:

1. Poor-quality data: noisy data, missing, inadequate size, and poor data sampling.
2. Lack of understanding/lack of diffusion of data retrieval techniques.
3. Data variety - trying to accommodate data that comes from different sources and in a variety of different forms (images, geo data, text, social, numeric, etc.).
4. Dealing with huge datasets, or 'Big Data,' that require distributed approaches.
5. Coming up with the right question or problem - "More data beats the better algorithm, but smarter questions beat more data."

2.2. Methodology

There are a number of issues and challenges are exist for retrieving the user query relevance video search. Therefore the different kinds of techniques are developed for improving the accuracy or relevancy of search systems. In this presented work a new model for video data search is presented using the low level features of segmented video frames. The given method also incorporates the text annotation technique for supporting the text based query search. Therefore the basic conceptual model is presented in two different major modules. In first module the training of the system is performed, the basic concept of training and their participating components are described in figure 1. In addition of that the second module perform the retrieval of relevant videos from the data base the figure 2 provides the utilized components of the information retrieval system.

According to the given diagram (figure 1) the training module of the system is defined, additionally the different components of the system are combined in the following manner.

Input video: Initial input of the system is a video object which is uploaded by the end user. The video is basically the collection of images and the audio but in this proposed technique the visual features are utilized for learning the video objects. In this phase the video which is to be processed for preparing the video data is uploaded as input.

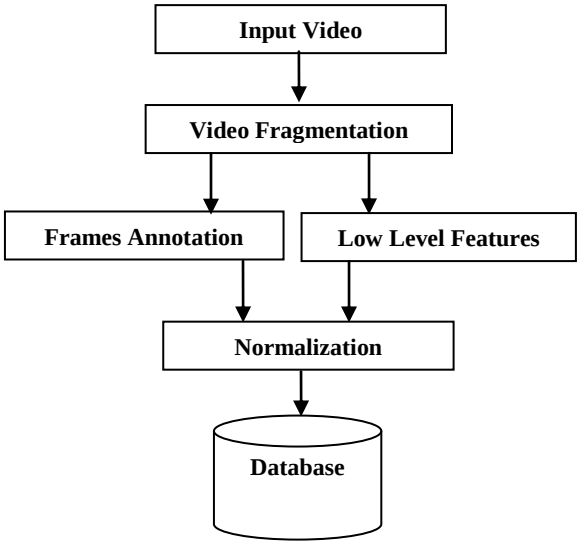

Figure 1 Training Module

Video fragmentation: In order to recover the visual features from the input video the segmentation of video is need to be performing. Therefore in the next step the video is fragmented using the FF-MPEG API. This technique is used to convert entire video into a set of images according to the time parameter or in random manner. This process extracts the images for further utilization in two different processes.

Image annotation: In this step the keyword, description of video frame or a list of key words are combined with the video data by user. This text or tag input are used to recognize the videos or clips during the text based search or query using the text.

Low level features: In this step the low level features are computed from the extracted or tagged images. These features are used to identify the objects in frames or the color patterns available on the extracted frames. To estimate the features from the frames low level features from images are computed.

Thus the texture feature of image is computed using LBP (local binary pattern), edge feature is computed using the canny edge detection technique and the color features are estimated using the grid color movement technique.

Normalization: The computed low level features are different in scale and different in length additionally the associated tags are also available in different formats. Therefore in order to keep organized the normalization process is required to store them into the database. The database contains the normalized low level features with the associated tags and the video name which are identified using this annotation feature.

After successfully development of the training module, there is a need to prepare the search process for retrieving the user query relevant videos from the database. To demonstrate the testing or search process figure 2 is developed. The entire process of the testing or video retrieval is demonstrated in this figure using the different components. In this process both kinds of user query can be used by the end users. Here user can search the video by the example image or by the text query. Therefore the provision for input both kinds of query are performed here.

Query by example: When the query input is provided by example of image then this phase of search is processed. Query by example provides the frame or image as input to recognize the similar frame in a given video or clip.

Low level features: The input example frame or image is processed in this phase for recovering the low level features. These features are similar to the training module thus the local binary pattern, canny edge detection and the grid color movement analysis is performed on the query image. These recovered features are utilized with the classifier for classify the similar video frames or images.

KNN classifier: the KNN (k-nearest neighbor) classifier is a distance based classification schemes. This classifier accepts two major inputs first the user query image features and the database image features. Using the distance of the video features the KNN classifier distinguishes the associated video.

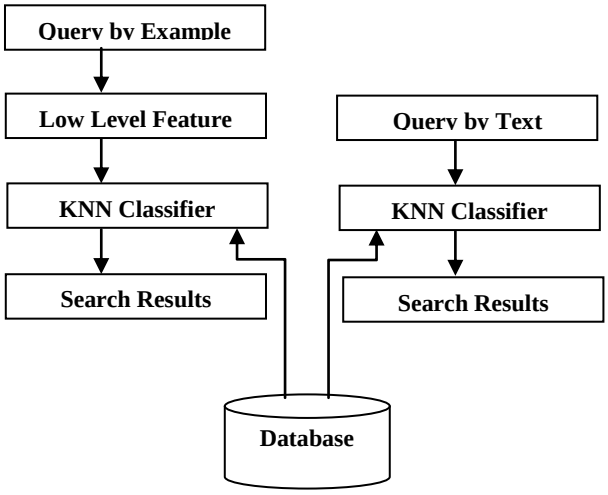

Figure 2 Retrieval Module

Search results: The obtained similar scene containing videos are reported as results in this phase of retrieval.

Query by text: In this kind of search the user provides the text keywords for finding the relevant videos from database.

KNN classifier: In this phase the similar classifier is used where in first parameter the input user query is provided and in second the database keywords are used. Using the input keywords the classifier classifies the similar keyword containing videos.

Search results: The outcome of the text keyword based search outcomes are reported in this phase of video retrieval.

The proposed working model is described in this section for efficient and accurate video retrieval. In next section the proposed technique is explained using step by step algorithm.

2.3. Proposed Algorithm

This section provides the summarized steps of the proposed information retrieval process thus the entire methodology is formulated using the algorithm steps.

Table 1 Training Process

Input: Video clip V, relevant keywords K Output: Trained model database D
Process: 1. $R_v = \text{readVideo}(V)$ 2. $F[M] = \text{FFMPEG.segment}(T_e, R_v)$ 3. $\text{for}(i = 1; i \leq F.\text{length}; i++)$ a. $F_{tag} = \text{frame.tag}(K, F_i)$ b. $LBP = \text{LBP.texture}(F_i, r)$ c. $Color = \text{grid.colormovement}(F_i)$ d. $Canny = \text{canny.edge}(F_i)$ e. $N = \text{normalize}(F_{tag}, LBP, Color, Canny)$ f. $D_i = N$ 4. End for 5. Return D

Table 2 Search by Example

Input: training database D, using example query Q Output: list video L
Process: 1. $R_f = \text{readImage}(Q)$ 2. $LBP = \text{LBP.texture}(R_f, r)$ 3. $Color = \text{grid.colormovement}(R_f)$ 4. $Canny = \text{canny.edge}(R_f)$ 5. $N = \text{normalize}(LBP, Color, Canny)$ 6. $L = \text{KNN.classify}(N, D)$ 7. Return L

Table 3 Search by Keywords

Input: input user query keywords K, training database D Output: list of videos L
Process: 1. $R = \text{readQueryInput}(K)$ 2. $L = \text{KNN.classify}(R, D)$ 3. Return L

3. RESULTS ANALYSIS

This chapter provides the detailed discussion about the evaluation of the proposed system. Therefore the different performance parameters are computed and their observation basis line graphs are represented in this chapter.

3.1. Precision

Precision for an information retrieval system can be defined by the amount of data retrieved that are relevant to the search query. That can be evaluated by the following formula:

$$precision = \frac{relevant\ document\ \cap retrieved\ document}{retrieved\ document}$$

Table 4 Tabular Representation of Precision Rate

Database Size	Query by tags	Query by frames
10	0.6	0.7
30	0.68	0.74
50	0.73	0.78
70	0.77	0.81
100	0.80	0.83
300	0.82	0.84
500	0.85	0.87

The precision for the proposed video retrieval system is evaluated by performing the search using text and frame based technique is computed using figure 3 and table 4. In this diagram the X axis shows the amount of data available in the database for making query and the Y axis shows the amount of correctly retrieved documents from the database. To represent the performance of the proposed system blue line shows the performance of search by text and red line shows the performance of query by the frame. According to the obtained results the performance of the system is increases with the amount of data in database increases. But the results also demonstrate the query by frame is more accurate than the query by tag search systems.

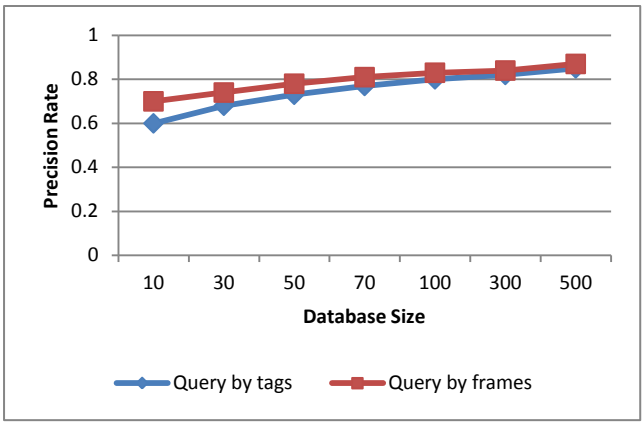

Figure 3 Graphical Comparison of Precision Rate

3.2. Recall

Recall is the amount of data that are extracted during the search is relevant to the user query. That can be estimated using the following formula:

$$Recall = \frac{Relevant\ Documents \cap Retrieved\ Document}{Relevant\ Documents}$$

The figure 4 and the table 5 show the performance of the implemented system in terms of recall rate.

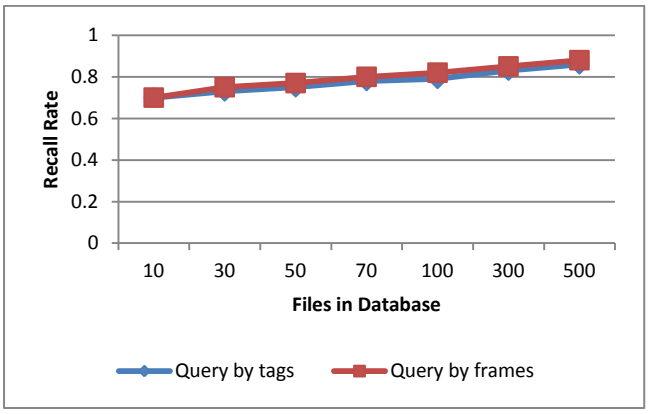

Figure 4 Graphical Comparison of Recall Rate

In this diagram the X axis shows the number of file in the database and the Y axis shows the recall rate of the system. For demonstrating the performance of the techniques the blue line shows the performance of search by text query and the red line shows the performance for the query by frames. According to the computed results the query by frame results more relevant outcomes as compared to the query by tag search process. Thus the proposed technique is efficient and produces accurate results as the amount of data in query database is increases.

Table 5 Tabular Representation of Recall Rate

Database Size	Query by tags	Query by frames
10	0.7	0.7
30	0.73	0.75
50	0.75	0.77
70	0.78	0.80
100	0.79	0.82
300	0.83	0.85
500	0.86	0.88

3.3. F-measure

It measure and combines precision and recall in terms of harmonic mean of precision and recall rate of the obtained results, that can also be termed F-measure or balanced F-score:

$$F - measure = 2.\frac{precision * recall}{precision + recall}$$

Table 6 Tabular Representation of F-Measure

Database Size	Query by tags	Query by frames
10	0.646	0.7
30	0.704	0.744
50	0.739	0.774
70	0.774	0.804
100	0.794	0.824
300	0.824	0.844
500	0.854	0.874

Figure 5 Graphical Comparison of F-Measure

The performance of the proposed technique is reported using the figure 5 and table 6 in terms of f-measures values. According to the diagram the X axis contains the number of files in database and Y axis contains the f-score of the system. The blue line is used to demonstrate the performance of the text based query and red line shows the performance of the frame based query processing outcomes. According to the obtained performance the frame based technique produces more accurate results as compared to the text based technique.

3.4. Memory usages

The amount of main memory required to compute the outcomes using the given algorithm is termed here as the memory utilization of algorithm. That can be computed using the following formula.

$$memory\ used = total\ memory - free\ memory$$

Table 7 Tabular Representation of Memory Consumption

Database Size	Query by tags	Query by frames
10	26839	28847
30	27395	29448
50	28131	30284
70	28957	31834
100	29905	33263
300	30782	33957
500	31993	34823

The memory consumption of both the processes query by text and query by image is reported using figure 6 and table 7. In this diagram the experimental file size is included in X axis and the Y axis contains the amount of main memory consumed in terms of KB (kilobytes). The blue line in the diagram shows the performance of the tag based search process and the red line shows the performance of query by frame. According to the obtained results the query by frame technique consumes more main memory as compared to text based search process.

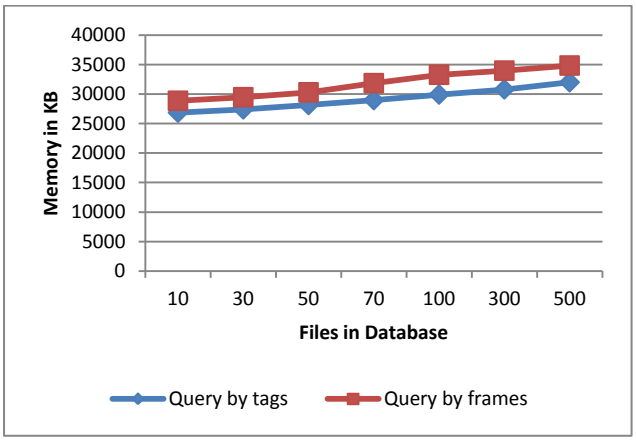

Figure 6 Graphical Comparison of Memory Consumption

3.5. Time consumption

The amount of time required to process the user query request for obtaining the user query relevance data is termed as the time consumption of the algorithm. The time consumption of the algorithm is estimated using the following formula:

$$time\ consumption = finishing\ time - initiation\ time$$

The computed search time of both the techniques are given using figure 7 and table 8. The X axis of the diagram shows the amount of files in the database for processing and the Y axis shows the respective expense of the time for relevant data search. This time is computed in terms of milliseconds (MS). The performance of text based technique is given using blue line and the red line shows the performance of the frame based technique. According to the obtained results the time consumption of the search is increases as the amount of data is increases for making search. But the time is adoptable as compared to other traditional systems. In addition of that text based results are efficiently retrieved as compared to the frame based search technique.

Table 8 Tabular Representation of Time Consumption

Database Size	Query by tags	Query by frames
10	1.38	2.55
30	2.49	4.29
50	6.38	8.91
70	8.43	10.34
100	11.27	15.32
300	25.42	39.48
500	37.15	48.39

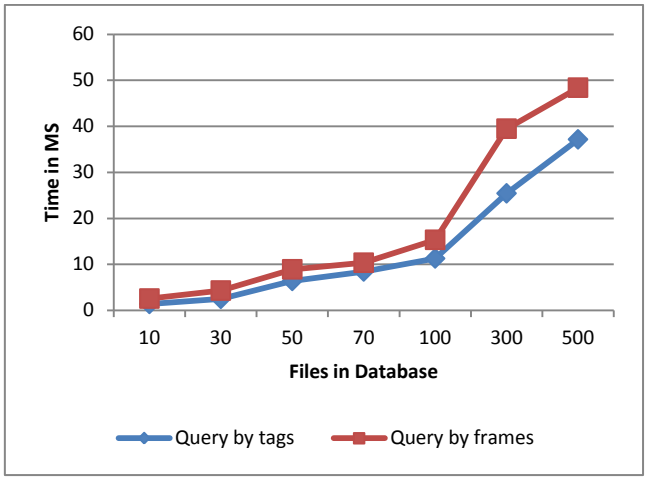

Figure 7 Graphical Comparison of Time Consumption

The obtained performance based summary is reported using the table 9.

Table 9 Comparative Performance Analysis

S. No.	Parameters	Text based query	Example based query
1	Precision	Low	High
2	Recall	Low	High
3	F-measures	Low	High
4	Memory consumption	Low	High
5	Time consumption	Low	High

According to the obtained performance of the proposed system the example based learning provides more accurate results as compared to the text based query outcomes. In addition of that during the process of the example based user query processing needs additional computational resources in terms of time complexity and space complexity. Thus the model is adoptable for both kinds of query processing for the video retrieval.

4. CONCLUSIONS

The proposed work is intended to investigate about the video retrieval systems therefore different video retrieval models are studied and a new model is presented in this work for providing the user query in both the techniques namely query by example and query by keywords. This chapter provides the summary of entire work performed and the possible future extension of the work is also presented.

4.1. Conclusion

Data handing in the computational technology is one of the essential tasks. To proper manage the data in different kinds of databases are prepared and the suitable information retrieval techniques are implemented for identifying the user query relevance data. But the handing of unstructured data is complicated work as compared to the structured formats of data. Among the various unstructured data formats i.e. web documents, text documents, images and others the video data is much complicated format. Therefore the handling and retrieval process of the video contents needs additional efforts.

In this presented work the video retrieval technique is investigated and a new model for text based query and example based query processing technique for the video retrieval technique is demonstrated. The proposed technique incorporates the tag based learning process and the low level feature computation based learning is implemented. In the training module first the video data is segmented into a list of visual objects known as frames additionally the relevant tags are produced with the each frames. The tagged frames are then processed using the three different low level feature computation techniques namely the LBP (local binary pattern) for the texture information, the canny edge detection technique for edge or objects estimation and the color grid movement is applied for computing the color variations in frames. Finally the KNN classifier is implemented for classifying the videos according to user query.

The implementation of the proposed working model is performed using the JAVA technology and their performances in terms of different performance parameters are evaluated.

4.2. Future work

The proposed work's key aim of the efficient and accurate video retrieval system development is completed successfully. The performance evaluation demonstrates the effectiveness of the proposed working model. In near future the following extension is possible for work.

1. Need to improve the computational complexity of the presented working model
2. Need to enhance the feature computation technique for video retrieval process
3. Need to incorporate the audio features also for representing the knowledge in the video frames.

5. REFERENCES

[1] Zhengyu Deng, Jitao Sang, and Changsheng Xu, "Personalized Celebrity Video Search Based on Cross-Space Mining", PCM 2012, LNCS 7674, pp. 455–463, 2012. c Springer-Verlag Berlin Heidelberg 2012.

[2] Norbert Fuhr, "An Information Retrieval View of Environmental Information Systems", Technische Hochschule Darmstadt, Fachbereich Informatik Karolinenplatz 5, W-6100 Darmstadt Germany.

[3] A. Scherp and R. Jain, "Towards an ecosystem for semantics", In Proceedings of Workshop on Many faces of Multimedia Semantics, at ACM Multimedia 2007, pp. 3-12.

[4] B V Patel and B B Meshram, "Content based Video Retrieval Systems", International Journal of UbiComp (IJU), Vol.3, No.2, April 2012.

[5] Yu-Gang Jiang, Jun Yang, Chong-Wah Ngo, Alexander G. Hauptmann, "Representations of Key point-Based Semantic Concept Detection: A Comprehensive Study", IEEE, 2008.

[6] Masoud Nosrati, Ronak Karimi, Mehdi Hariri, "Detecting Circular Shapes From Areal Images Using Median Filter and CHT", World Applied Programming, Vol (2), Issue (1), . 49-54, January 2012.

[7] Zhenhua Guo, Lei Zhang, David Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification", IEEE transaction on image processing, 2010.

[8] Mohammad Jafari, Neda Abdollahi, Ali Amiri, Mahmood Fathy, "Generalization of Determinant Kernels for Non-Square Matrix and its Application in Video Retrieval", International Journal of Scientific Research in Computer Science and Engineering, Volume-3, Issue-4.

[9] Dan Albertson, Melissa P. Johnston, "Connecting with Educators: Science Teachers and Interactive Video Retrieval", iConference 2015 Proceedings.

[10] Haojin Yang and Christoph Meinel, "Content Based Lecture Video Retrieval Using Speech and Video Text Information", IEEE Transactions on Learning Technologies, vol. 7, no. 2, April-June 2014.

[11] Jiajun Liu, Zi Huang, Hongyun Cai, Heng Tao Shen, Chong Wah Ngo, Wei Wang, "Near-Duplicate Video Retrieval: Current Research and Future Trends", ACM Computing Surveys, Vol. 45, No. 4, Article 44, Publication date: August 2013.

[12] Luca Rossetto, Ivan Giangreco, HeikoSchuldt, St´ephaneDupont, Omar Seddati, MetinSezgin, and Yusuf Sahillio˘glu, "IMOTION — A Content-Based Video Retrieval Engine", (Eds.): MMM 2015, Part II, LNCS 8936, pp. 255–260, 2015. c Springer International Publishing Switzerland 2015.

[13] Ling Shao, Simon Jones, and Xuelong Li, "Efficient Search and Localization of Human Actions in Video Databases", IEEE Transactions on Circuits and Systems for Video Technology, Vol 24, No 3, March 2014.