# A Text Based Video Retrieval Using Semantic and Visual Approach

**VAIDEHI BANTE, AVINASH BHUTE**

[1] *PG Student, Department of Information Technology, Sinhgad College of Engineering, Maharashtra, India*

[2] *Associate Professor, Department of Information Technology, Sinhgad College of Engineering, Maharashtra, India*

**Abstract -** ***There is interesting growth in the amount of digital video data can be explained by the new models of digital media production, distribution and consumption. But still there is lack of tool to categorize and retrieve the video. Categorizing web based videos is yet more challenging task plus the repeated and duplicate contents in video frustrate the user. Well known accepted video retrieval and indexing techniques are not properly defined and available. Almost all the multimedia search systems rely on the available contextual or metadata information in text form. These all challenges motivate us to present multimedia warehouse using multimodal features. The Content based video retrieval which includes features like visual and concept based video. The content based retrieval includes the query content i.e. semantic content in which the human perceptions can be measured.***

***In this paper the visual and semantic approaches are considered. The visual feature includes color, shape, and texture. The semantic approach focus of work is the notion of a semantic concept: an objective linguistic description of an observable entity. The paper provides an overview of the different existing techniques in text, image and video based video retrieval and different approaches to search with in long videos. The proposed method showed the impressive results on the average with precision and recall in testing on the developed dataset.***

***Keywords:*** *Multimedia Retrieval, Similarity Matching Algorithm, Text frames classification, Video Retrieval System.*

## 1. INTRODUCTION

Video retrieval is one of the most popular and interesting topics in both multimedia research and real life applications. As there is ample number of video archives including documentary videos, broadcast news meeting movies, video etc. On the other side video sharing on the web is growing with a tremendous speed which creates perhaps the most heterogeneous and the largest publicly available video archive. To find the desired videos and accurate one is becoming harder and harder every day for the users. Research focus is now been on video retrieval by aiming at the facilitation of this task.

In video archives, there are two main aspects of information available in video retrieval: visual content, semantic information. Apart from visual approach, the concept based i.e. semantic concepts can be extracted from video with more information like multimedia content e.g. audio, text, visual are available. The semantic concepts meaning an intermediate layer of multimedia descriptors that aims to bridge the gap between user information need and low-level multimedia content. Nowadays most of the active research is based on the utilization of the semantic concepts and visual content.[1,2]

This paper, focus of work is the two main frameworks for video retrieval: text-based visual approach and semantic approach. The text-based methods are originated from the information retrieval [IR] community and can be tracked back to 1970s. In content based videos are utilized through the visual features such as texture, shape, color, motion. In semantic approach query to concept mapping stage finds the concept for query term with built in dictionary and computes the weight factor.

## 2. RELATED WORK

### A. Semiautomatic Method to generate annotation for Cricket Videos

According to Dr. Sunitha Abburu [13] presents a paper which is based on indexing and, semantic video analysis and retrieval are necessary for the adequate usage of video repositories. To extract from the proper source the semantics of the video, the superimposed text this will increase the efficiency of retrieval system. The author proposed a method which is semi-automatic to generate annotation and metadata for cricket videos and to extract the semantics of cricket video an automated tool DLER is used.

The aim of this paper is to propose unique techniques for the text based video extraction includes steps like video text detection, localization, extraction, and recognition. The first step includes in the text extraction is that in video frames, it is difficult to divide the video into video frames because of complex or weird background, unknown text character color, and various other features. Here the author proposed a fully automatic method, which is a simple approach for preprocessing which integrates all the steps involved in text detection, text localization, and then extraction, and recognition as a simple and with the use of single tool.

### B. Semantic Multimedia Retrieval Using Lexical Query Expansion

According to Milind R. Naphade, Alexander Haubold, Apostol (paul) Natsev, [3] presents methods to enhance text based search retrieval of visual multimedia content by developing a set of visual models of semantic concepts from a lexicon of concepts allow uniform for the collection via queries of words or fully qualified sentences text search is performed, and results are returned in the form of ranked video shots.

The proposed approach presents methods which involves a query expansion stage, in which query terms are compared to the visual concepts for which authors independently build classifier models. During expansion, this advantages a synonym dictionary and wordnet similarities. In particular, authors spotlight on lexical query analysis and expansion mapping query words and phrases to concepts and build a ranked list vector of matching shots which is based on concept detection scores and which will automatically computed query-to-concept relevance scores.

### C. Latent Semantic Indexing

According to Roshan Fernandes, Rashmi M, [8] concentrated on the concept of the classifying web based videos is an important yet challenging task. Therefore this paper focused on the accuracy of retrieval system which depends on the method used for detecting shots and scenes, kind of key frames etc. video features used for retrieval. For this kind semantic video indexing is a step towards automatic video indexing and retrieval, therefore a latent semantic indexing (LSI) technique is proposed. The LSI method is based on singular value decomposition and fusion of visual features like color and edge is proposed for video indexing and retrieval. A key feature of LSI is the ability to establish associations between similar kinds of information, so the probability or chances of producing accurate index is very high. Latent semantic indexing (LSI), it is a technique used for intelligent information retrieval (IR).

In this method, it is stated that LSI is a method that exploits the idea of vector space model and singular value decomposition (SVD). LSI uses SVD to reduce noise and dimensionality in the initial term document representation and to capture latent relationships between the terms and the document. Here the proposed system works by analyzing the key frames in video shots and extracting the different visual features from these frames. Then the feature matrix is formed by combining the different types of features from all shots of a video. Then the latent semantic indexing is performed on the feature matrix.

The proposed method performs well when there is complex background and it becomes more reliable as the scene contains more edges in the background. The proposed method is robust to different in respective feature characteristics like character size, position, contrast and color.

## 3. PROPOSED ALGORITHM

A. The design considerations using visual approach:

- ➢ Video detachment or segmentation which includes shot boundary detection.
- ➢ Feature Extraction includes extracting feature from segmented video clips.
- ➢ Video mining to the output of extracted feature.
- ➢ Video interpretation or annotation to build a semantic index.
- ➢ User query.
- ➢ Retrieval of expected video i.e. the output.

### 3.1 Video Retrieval Using Semantic Approach

In the multimedia community, the high-level context is mostly used for improving concept detection and retrieval accuracy. Traditionally concepts are retrieved using trained concept detectors and then the high-level context is used for refining the results. Semantic gap meaning to the gap between low level and high level features semantic meanings of content of the input. Combining two types of information which are semantically expressed at different levels such as texts and images is an instance of the "semantic gap" problem.

### 3.1.1 Semantic Concepts

Semantic means the studying of the content which focuses on the relation between signifiers or say tokens in the language of artificial intelligence like phrases ,signs ,symbols, words and what they are stand for such as their denotation. The proper definition of a set of atomic semantic-concepts (objects, scenes, and events) which is assumed to be broad enough to cover the semantic query space of interest high level concepts. The linguistic semantic meaning the actual concept which helps to understand human perceptions.

### 3.2 VIDEO RETRIEVAL BASED ON VISUAL APPROACH

As there are different clues are hidden in the video content can be of immense use in indexing. Textual information which is used to present captions associated with the video or appearing on the video or say handwritten or printed text in the scene. If we can identify the textual content, it can be effectively used for characterizing the video clips, as a complementary measure to the visual appearance-based features.

Due to rapid changes in digital technologies, in the recent years many people wish to publish their digital information on the internet community such as text, image, video, audio etc. Hence, it requires effective indexing and searching tools for web. The visual approach includes feature extraction, which includes parameters like color, shape, texture and size.

## 4. PROPOSED METHODOLGY

The aim of the proposed system is to introduce a set of techniques called semantic combination in order to efficiently integrate text and image retrieval systems in the context of multimedia information access.

The proposed system consists of two approaches as mentioned above as semantic and visual approach.

### 4.1 Semantic Based Video Retrieval

The semantic approach is split into four parts:

1) Query by extraction of words and qualified phrases,

2) Calculate the semantic concept relatedness scores or weight factor between extracted query terms and concepts,

3) Build the synonym dictionary and representative words dictionary

4) Retrieval of resultant videos.

### 4.1.1 Query to Concept Mapping

In this stage, each query term's concept weight vector is being computed which is determined from a term's similarity to each concept. From this, the computed vector is fused with shot concept confidence vector in the third stage to produce a shot's relevance to the query. This process works like as the Lesk semantic relatedness score to compute the similarities between the pair of words.

The semantic similarity measures not only consider the Lesk semantic relatedness measure i.e. *is-a* relationships between words i.e. synonyms but also relationships as well(e.g. has-a). This method allows to compute more general semantic relatedness scores between pairs of words or phrases, which are better suited for query expansion purposes e.g., the term US flag is semantically associated with the term American flag through is-a relationship but is also associated to other terms, such as or stars and stripes, through different relationships.

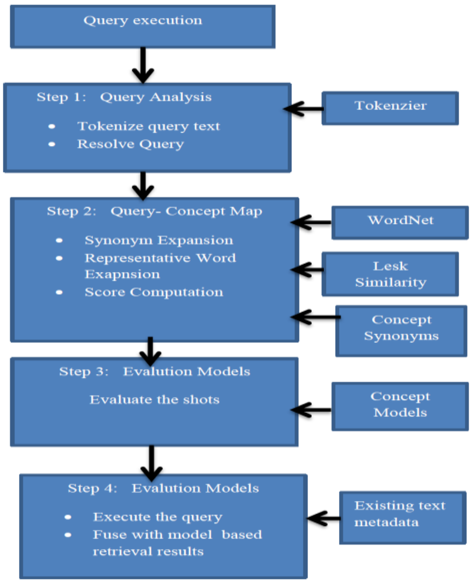

**Fig 1:** Overview of Semantic Based Video Retrieval

### 4.1.2 Retrieval of Video

The extracted query terms are compared to all synonyms of a concept and representative words. For each term, concept combination includes concept plus weight factor, i.e. compute the highest similarity score (that of the best matching synonym) as the concept weight for the given query term, and aggregate these scores over all query terms.

### 4.2 VIDEO RETRIEVAL BASED ON VISUAL APPROACH

### 4.2.1 Video Segmentation

Segmentation of images into homogeneous regions representing sub-regions of the objects depicted .Determining a contiguous set of such regions in order to identify those objects .The Segmentation of video process includes video segmentation i.e. the complete video is first converted into scenes, then scenes are converted into shots and finally shots are converted into various frames.

### 4.2.2 Feature Extraction

Extracting features from the output of video segmentation. Feature extraction is the time consuming task in TBVR. This can be overcome by using the multi core architecture [16]. These mainly include features of key frames, objects, and audio/text features.

### 4.2.3 Key Frame Extraction

There are great redundancies among the frames in the same shot; therefore, certain frames that best reflect the shot contents are selected as key frames [18] to succinctly represent the shot. The features used for key frame extraction include colors (particularly the color histogram), edges, shapes, optical flow.

In this stage, from the input query after segmentation the key frame is being selected among the extracted frames of the video, for the similarity matching using Euclidian Distance Algorithm.

### 4.2.4 Features of Key Frames

### A. Colors

Color spaces shown to be closer to human perception and used widely in RBIR include, RGB, HSV (HSL), YCrCb and the hue-min-max-difference (HMMD).

### B. Shape

For shape we use canny edge detector and sobel algorithm.

### C. Texture

For texture we use LBP [local binary pattern] and COOC algorithm.

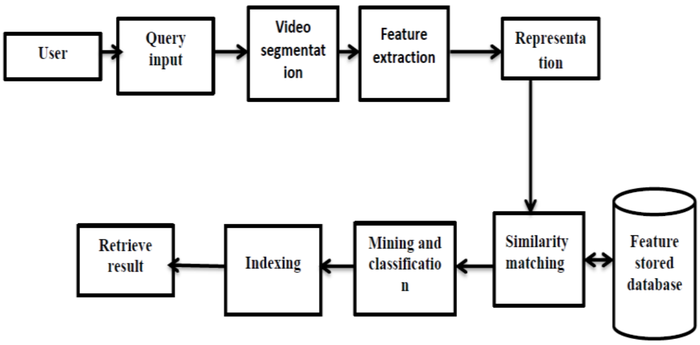

**Fig 2:** Framework of Visual based Video Retrieval

## 5. Simulation Results

The proposed system is implemented in the JAVA platform (Netbeans IDE) and tested using the database video clips of MP4, and AVI format. As there is no standard dataset specified in the literature, I have prepared a dataset which contains collection of 40 videos of varying size.

Success in the search task is measured through precision and recall as the central criteria to evaluate the performance of retrieval algorithms. Precision is the fraction of the retrieved video that is relevant. Recall is the fraction of relevant videos that is retrieved.

The most common evaluation measures used in IR are 'precision' and 'recall'. The same is used to measure the performance of proposed system.

$$\text{Precision}=\frac{\textit{no of relevant videos retrieved}}{\textit{total no of videos retrieved}} \quad \ldots \quad (5.1)$$

$$\text{Recall} = \frac{\textit{no of relevant videos retrieved}}{\textit{total no of relevant videos in the collection}} \quad \ldots \quad (5.2)$$

## 6. PERFORMANCE EVALUATION

The overall system is evaluated using values of precision and recall. Table 1- shows the how much precision and recall is calculated for the particular query.

**Table 1 :** Precision Vs Recall

| Input query | precision | recall |
|---|---|---|
| 1.mp4 | 0.6 | 0.33 |
| 2.mp4 | 0.625 | 0.44 |
| 3.mp4 | 0.714 | 0.71 |
| 4.mp4 | 0.75 | 0.4 |
| 5.mp4 | 1 | 0.6 |
| 6.mp4 | 0.428 | 0.25 |
| 7.mp4 | 0.274 | 0.07 |
| 8.mp4 | 0.285 | 0.12 |
| 9.mp4 | 0.44 | 0.44 |

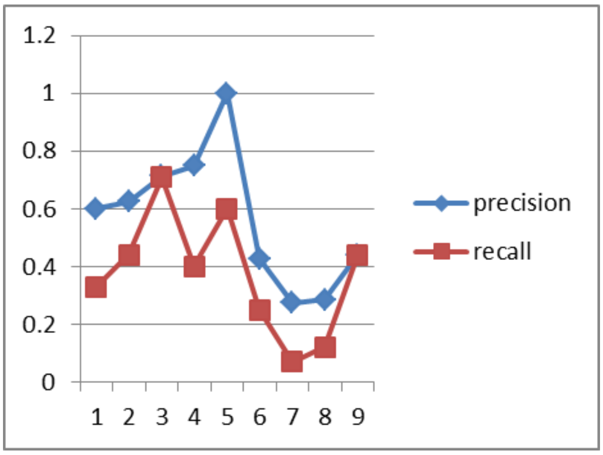

**Fig 3:** Graph for precision and recall

### 6.1 FINAL RETRIEVAL

The final retrieval meaning the overall system performance with the input query which shows the results, when user inputs the query , the final retrieval graph shows no of similar videos are available in the database ,the total no of videos retrieved by the system, and most matched videos from database with the input query .

**Table 2:** Table for most matched, retrieved and available video

| Input Query | Most matched | Total Retrieved by the System | Similar available in the database |
|---|---|---|---|
| 1.mp4 | 3 | 5 | 6 |
| 2.mp4 | 5 | 8 | 9 |
| 3.mp4 | 5 | 7 | 7 |
| 4.mp4 | 3 | 4 | 5 |
| 5.mp4 | 4 | 4 | 5 |
| 6.mp4 | 3 | 7 | 8 |
| 7.mp4 | 3 | 11 | 13 |
| 8.mp4 | 2 | 7 | 8 |
| 9.mp4 | 4 | 9 | 9 |

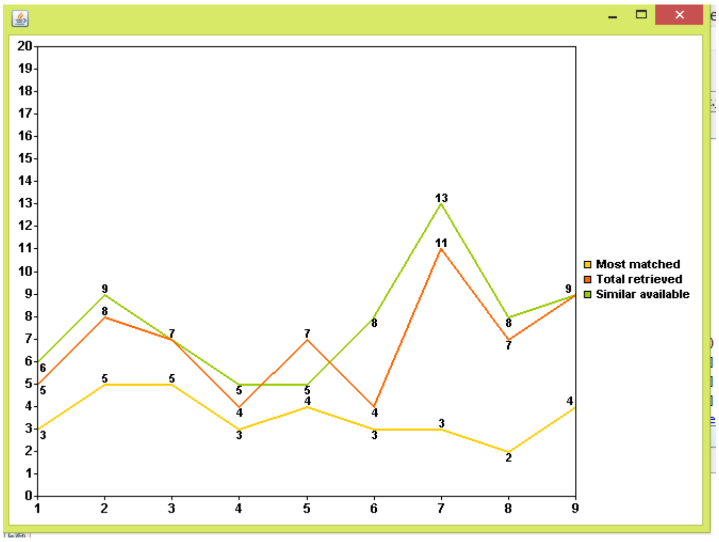

**Fig 4:** Graph of available, most matched and retrieved video

### 7. CONCLUSION AND FUTURE WORK

In this paper, we have presented the two approaches for text based video retrieval namely visual and semantic approach. The experimental results show that the proposed method outperforms the other retrieval methods in terms of average precision and recall. Fast video search for the different hierarchical indices are all interesting research questions. In this proposed system the indexing time for 1000 image database takes 5-6 minutes. In the cloud computing environment, for video indexing and retrieval approach where the individual videos to be searched and the dataset of videos are both changing dynamically, will form a new and flourishing research direction in video retrieval in the very near future. Affective computing describes human psycho-logical feelings such as sadness, anger, romance, pleasure and violence. Hierarchically organizing and visualizing retrieval results are all interesting research issues.

Future work on the extension of the system to audio visual material indexing and temporal and special layout retrieval has started, with additional features extracted from the image sequences based on motion and audio track. The work done in the developed system for CBVR can be used fully in the new version of the system based on the key-frames extracted from the video sequences.

### ACKNOWLEDGEMENT

I am indeed thankful to my guide Prof. Avinash N. Bhute for her able guidance and assistance to complete this paper; otherwise it would not have been accomplished. I extend my special thanks to Head of Department of Information Technology, Dr. Nilesh Uke who extended the preparatory steps of this paper-work. I am also thankful to the head & Principle of STES'S, Sinhgad College of Engineering, and Dr. S. D. Lokhande for his valued support and faith on me.

## REFERENCES

[1]. Di Zhong and Shih-Fu Chang, 1999, "*An Integrated Approach for Content-Based Video Object Segmentation and Retrieval*", IEEE Transactions on Circuits and Systems for Video Technology, Vol.9, No.8, pp.1259-1268.

[2]. Oh J.H., and Bandi B., 2002, "*Multimedia Data Mining Framework for Raw Video Sequences*", in Proceedings ACM International Workshop Multimedia Data Mining, Edmonton, AB, Canada, pp.18–35.

[3]. Alexander Haubold , Apostol (Paul) Natsev, Milind R. Naphade "*Semantic Multimedia Retrieval Using Lexical Query Expansion And Model-Based Reranking*" Department of Computer Science Columbia University, New York, NY 10027, IBM Thomas J. Watson Research Center Hawthorne, NY 10532

[4]. Divakaran A., Radhakrishnan R., and Peker K.A., 2002,"*Motion Activity Based Extraction of Key-Frames from Video Shots*", in Proc. IEEE International Conference of Image Process., Vol.1, Rochester, NY, pp.932–935.

[5]. Hanjalic A., Lagendijk R. L., and Biemond J., 1999, "*Automated Highlevel Movie Segmentation for Advanced Video-Retrieval Systems*", IEEE Transaction Circuits System Video Technology, Vol.9, No.4, pp.580–588.

[6]. Changsheng Xu, Yi-Fan Zhang, Guangyu Zhu, Yong Rui, Hanqing Lu and Qingming Huang, 2008, "*Using Webcast Text for Semantic Event Detection in Broadcast Sports Video*", International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-3, Issue-4, April-2015

[7]. Noboru Babaguchi, Yoshihi koKawai, and Tadahiro Kitahashi, 2002, "*Event Based Indexing of Broadcasted Sports Video by Intermodal Collaboration*", IEEE Transactions on Multimedia, Vol.4, No.1, pp.68-75.

[8]. Rashmi M, Roshan Fernandes "*Video Retrieval Using Fusion of Visual Features and Latent Semantic Indexing*" Dept. of Computer Science and Engineering, Nitte Mahalinga Adyanthaya Memorial Institute of Technology, Nitte, Udupi Dist., Karnataka, India, International Journal of Innovative Research in Computer and Communication Engineering ,Vol.
2, Issue 5, May 2014

[9]. Zhao L., Qi W., Wang Y.J., Yang S.Q., and Zhang H.J., 2001, "*Video Shot Grouping Using Best First Model Merging*", in Proc. Storage Retrieval Media Database, pp.262–269

[10]. Li H.P., and Doermann D., 2002, "*Video Indexing and Retrieval Based on Recognized Text*", in Proceedings IEEE Workshop Multimedia Signal Process, pp.245–248.

[11]. Milind Ramesh Naphade and Thomas S. Huang, 2001, "*A Probabilistic Framework for Semantic Video Indexing, Filtering and Retrieval*", IEEE Transactions on Multimedia, Vol.3, No.1, pp.141-151.

[12]. Yoshitaka A., Hosoda Y., Hirakawa M., and Ichikawa T., 1998, "*Content-Based Retrieval of Video Data Based on Spatio temporal Correlation of Objects*", in Proceedings IEEE Multimedia Computing and Systems, pp.208-213.

[13]. Dr. Sunitha Abburu "*Multi Level Semantic Extraction For Cricket Video By Text Processing*" Professor & Director, Department of Computer Applications Adhiyamaan College of Engineering, Hosur,pin-635109,Tamilnadu, India, International Journal of Engineering Science and Technology Vol. 2(10), 2010, 5377-5384