

CONTENT BASED VIDEO RETRIEVAL SYSTEM

Madhav Gitte¹, Harshal Bawaskar², Sourabh Sethi³, Ajinkya Shinde⁴

¹B.E. Scholar, Department of Information Technology, Sinhgad College of Engineering Pune-41, University of Pune, Maharashtra, India

²B.E. Scholar, Department of Information Technology, Sinhgad College of Engineering Pune-41, University of Pune, Maharashtra, India

³B.E. Scholar, Department of Information Technology, Sinhgad College of Engineering Pune-41, University of Pune, Maharashtra, India

⁴B.E. Scholar, Department of Information Technology, Sinhgad College of Engineering Pune-41, University of Pune, Maharashtra, India

⁵B.E. Scholar, Department of Information Technology, Sinhgad College of Engineering Pune-41, University of Pune, Maharashtra, India

Abstract

Video retrieval is a young field which has its genealogy rooted in artificial intelligence, digital signal processing, statistics, natural language understanding, databases, psychology, computer vision, and pattern recognition. However, none of these parental fields alone has been able to directly solve the retrieval problem. In this paper we present a System that supports video mining from multimedia warehouse using multimodal feature has two stages, the first one is building the multimedia warehouse and the second one is retrieving the video from that multimedia warehouse. The Video Retrieval system includes various steps: Video Segmentation, Key frames Selection, Feature Extraction. For retrieving the video from warehouse, the retrieval subsystem processes the presented query, performs similarity matching operations and this can be done using Euclidian Distance Algorithm, and finally displays the result to end user.

Keywords: Video Segmentation, Key-frame Selection, Feature Extraction, Similarity Matching etc.

1. INTRODUCTION

There is amazing growth in the amount of digital video data in recent years, Lack of tool to classify and retrieve the video. Duplicate contents in video frustrate the user. Universally accepted video retrieval and indexing technique are not well defined or available. Most of the multimedia search systems rely on available metadata or contextual information in text form. These all challenges motivate us to present video mining from multimedia warehouse using multimodal features. A common first step for most content-based video analysis techniques available is to segment a video into elementary shots, each comprising a continuous in time and space. These elementary shots are composed to form a video sequence during video sorting or editing with either cut transitions or gradual transitions of visual effects such as fades, dissolves and wipes. The distance between adjacent frames can be based on statistical properties of pixels, compression algorithms, or edge differences. The most widely used method is based on histogram differences.

In this paper We are presenting Content Based Video Retrieval (CBVR) System it includes various steps: Video Segmentation: Segments the video into shots, Key frame Selection: Selects the key frame to represent the shot using Euclidian Distance Algorithm, Feature Extraction: Features are extracted for the key frame and stored into feature vector. Features are of two types that are spatial and temporal. Spatial features are further classified as color,

shape and edge; similarly temporal features are also further classified as motion and audio. Indexing: Hierarchical Clustering Tree Algorithm is used to index the key frames. For retrieving the video from warehouse, the retrieval subsystem processes the presented query, performs similarity matching operations and this can be done using Euclidian Distance Algorithm, and finally displays the result to end user.

2. RELATED WORK DONE

The need for content-based access to image and video information from media archives has captured the attention of almost all researchers in recent years. Research efforts have led to the development of methods that provide access to image and video data. These methods have their roots in Computer Vision and Pattern Recognition. The methods are used to determine the similarity in the visual information content extracted from low level features. These features are then clustered for generation of database indices. This section discussed a comprehensive literature survey on the use of these pattern recognition methods which enable image and video retrieval by content.

- Oscar D. Robles et. al. are propose the two new primitives for representing the content of a video in order to be used in a Content-Based Video Retrieval System. The techniques presented in the paper titled "Towards A Content-Based Video Retrieval System Using Wavelet-Based Signature" compute first a

multi-resolution representation using the Haar Transform. Two types of signatures are extracted afterwards, one based on multi-resolution global color histograms and the other one based on multi-resolution local color histograms. The tests performed in the experiments include the recall measure achieved with the proposed primitives[14].

- A system for “Recognizing Objects in Video Sequences” is presented by Visser et al. They use the Kalman filter to obtain segmented blobs from the video, classify the blobs using the probability ration test, and apply several different temporal methods, which results in sequential classification methods over the video sequence containing the blob[12].
- “A Semantic Video Retrieval Approach using Audio Analysis” is presented by Bakker and Lew in which the audio can be automatically categorized into semantic categories such as explosions, music, speech, etc. In their research literature, significant attention has been given to the visual aspect of video, however, relatively little work directly uses audio content for video retrieval. The paper gives an overview of our current research directions in semantic video retrieval using audio content. In this paper discuss the effectiveness of classifying audio into semantic categories by combining both global and local audio features based in the frequency spectrum. Furthermore, introduce two novel features called Frequency Spectrum Differentials (FSD), and Differential Swap Rate (DSR)[13].

3. ARCHITECTURAL BLOCK DIAGRAM

The following figure shows the architectural block diagram; there are two blocks shown, first block indicates the **off-line processing** and second block indicates **on-line processing**.

3.1. Off Line Processing:

In off-line processing administrator uploads the various video clips/data and gives it to the media descriptors the media descriptor performs the feature extraction of that video and then the key frame is chosen from the available frame and indexing is done on that key frame and these indexes are stored on data warehouse along with indexes and various other features also gets stored on data warehouse.

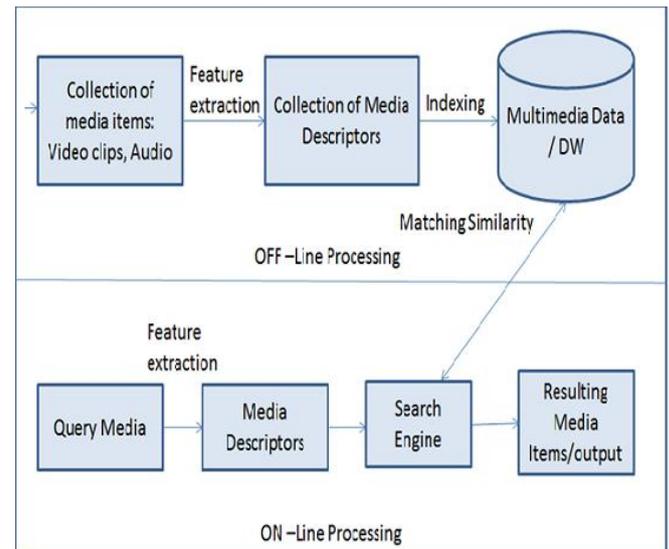


Fig -1: Architectural block diagram

3.2. On-Line Processing

In on-line processing block user takes a one short video clip and it is given to the query media then query media give it to the media descriptor, media descriptor performs feature extraction based on the given video then these features are given to the search engine, then search engine request to the data warehouse taking that various features and data warehouse match features of requested video with stored video on data warehouse and final matched result is given to the user.

4. FLOW OF SYSTEM

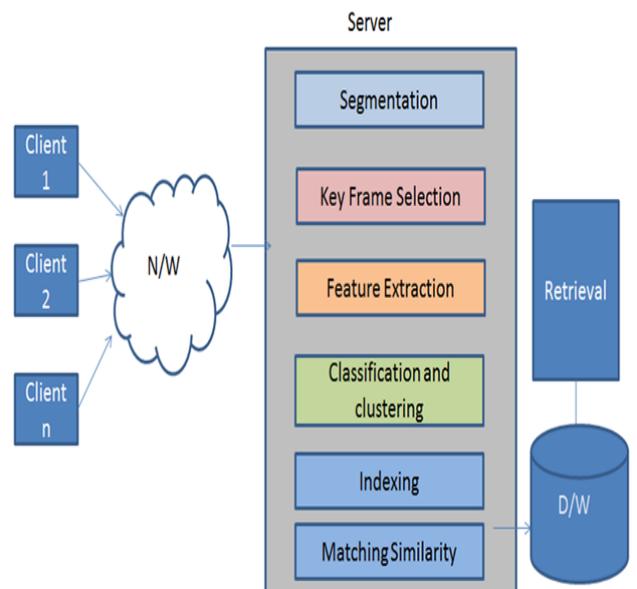


Fig -2: Execution flow of system

The above figure shows the flow of system; in this system different clients are interacting with the server through the network. The server system also interacted with data warehouse, the video data stored in warehouse. The server

system performs various operations on video clips these are segmentation, key frame extraction, feature extraction, classification and clustering, indexing and matching similarity by using various algorithms. These features are matched with features stored in data warehouse and user gets the final retrieved result from data warehouse.

4.1 Video Segmentation

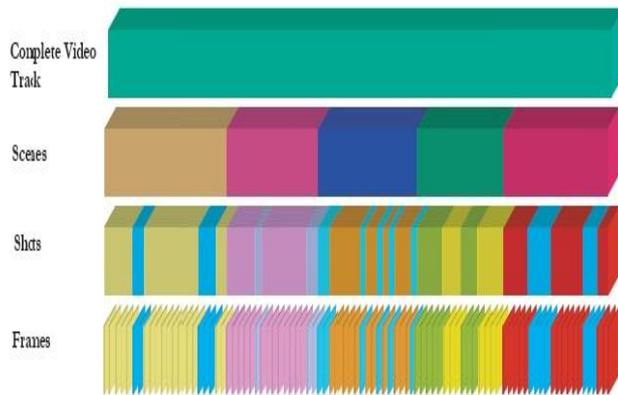


Fig -3: Video Segmentation

Video segmentation is first step towards the content based video search aiming to segment moving objects in video sequences. Segmentation of Video done with the help of step by step process of video segmentation, The complete video is first converted into scenes, then scenes are converted into shots and finally shots are converted into various frames[10].

4.2 Key-Frame Selection

Selects the key frame among the extracted frames of the video, to represent the shot using Euclidian Distance Algorithm.

4.2.1 Euclidian Distance

Euclidean distance is used as a similarity measure between two feature vectors and minimum Euclidean distance yields the best similarity.

4.3 Feature Extraction

Features are extracted for the key frame and stored into feature vector. Features are of two types that are **spatial** and **temporal**.

4.3.1 Spatial

Spatial features are further classified as *color*, *shape* and *edge*;

- For **Color** feature we are using Local Color Histogram (LCH) and Global Color Histogram (GCH), Average RGB Algorithms.
- For **Shape and Edge** we are using Eccentricity Algorithm.

4.3.2 Temporal

Temporal features are also further classified as *motion* and *audio*.

4.4 Classification and Clustering

4.4.1 Classification

Classification of video contents is done based on Support Vector Machines (SVM). Automatic Content Based Retrieval and Semantic Classification of the Video Contents [15] presented a learning framework where construction of a high-level video index is visualized through the synthesis of its set of elemental features. This is done through the medium of **Support Vector Machines** (SVM). The support vector machines associate each set of data points in the multidimensional feature space to one of the classes during training.

4.4.2 Clustering

For Clustering we are using **K- Mean's clustering algorithm**.

This is a widely used clustering algorithm. It assumes that we know the number of clusters k . This is an iterative algorithm which keeps track of the cluster centers (means). The centers are in the same feature space as x .

1. Randomly choose k centers μ_1, \dots, μ_k .
2. Repeat.
3. Assign $x_1 \dots x_n$ to their nearest centers, respectively.
4. Update μ_i to the mean of the items assigned to it.
5. Until the clusters no longer change.

Step 3 is equivalent to creating a Voronoi diagram under the current centers. k -means clustering is sensitive to the initial cluster centers. It is in fact an optimization problem with a lot of local optimal. It is of course sensitive to k too. Both should be chosen with care.

4.5 Indexing

For Indexing purpose we are using B+ Tree to store index of the specified frames. Indexing is done with the help of Hierarchical Clustering Tree (HCT) algorithm.

4.6 Matching Similarity

In retrieval stage of a Video search system, features of the given query video is also extracted. After that the similarity between the features of the query video and the stored feature vectors is determined. That means that computing the similarity between two videos can be transformed into the problem of computing the similarity between two feature vectors [11]. This similarity measure is used to give a distance between the query video and a candidate match from the feature data database as shown in Figure. 6.

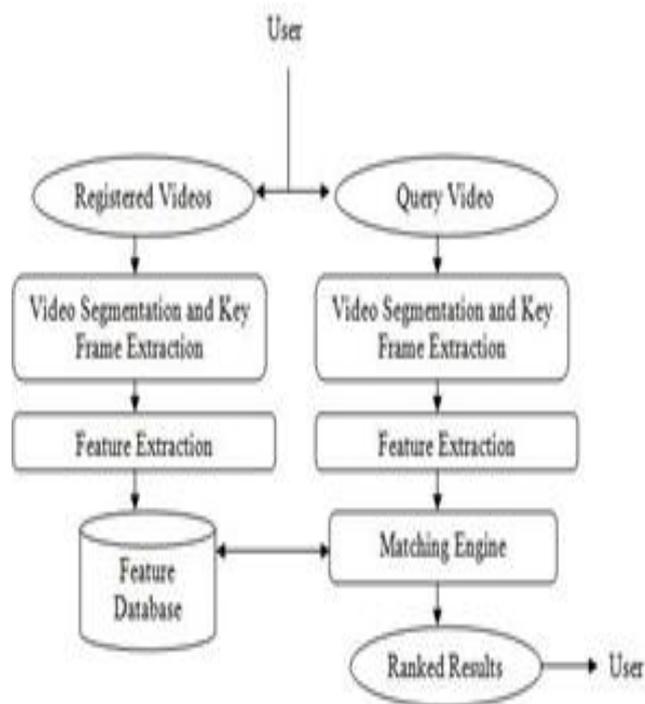


Fig-6: similarity Matching and Retrieval of Result

5. PERFORMANCE EVALUATION

Performance of the system is evaluated based on the Precision and Recall values. Table-1 shows the how much Precision and Recall is calculated for a given query video.

Precision =

$$\frac{\text{No. of retrieved videos that are relevant to the query clip}}{\text{Total no. of retrieved videos}}$$

Recall =

$$\frac{\text{No. of retrieved videos that are relevant to the query clip}}{\text{Total no. of relevant videos available in database}}$$

Table-1: Precision and Recall

QUERY VIDEO	Precision	Recall
1.mp4	0.75	0.4
2.mp4	0.714	0.71
3.mp4	0.625	0.44
4.mp4	1	0.6
5.mp4	0.428	0.25
6.mp4	0.6	0.33
7.mp4	0.4	0.75
8.mp4	0.44	0.44
9.mp4	0.285	0.12
10.mp4	0.272	0.07

Following graph shows the performance of the system according to the values obtained for Precision and Recall.

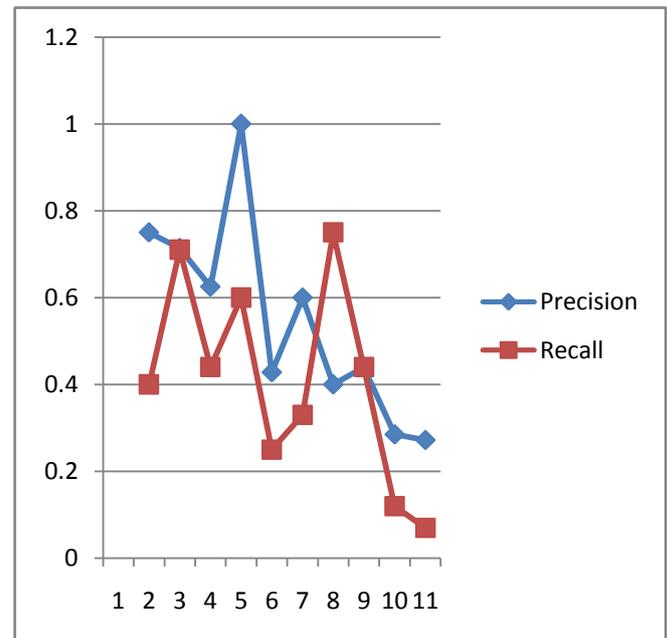


Fig-7: Graph of Precision Vs Recall

6. RESULT ANALYSIS

6.1. TIME ANALYSIS (Similarity Matching)

Time Analysis for similarity matching is done based on the amount of time required for retrieving the video from the database and the percentage match of the query video with video files stored in database.

Following table shows the video length and retrieved time:

Table-2: Video length and retrieved time

QUERY VIDEO	TIME IN SEC	PERCENT MATCHING	RETRIEVE TIME
1.mp4	10sec	30	76sec
2.mp4	4sec	50	70sec
3.mp4	116sec	50	81sec
4.mp4	2sec	40	100sec
5.mp4	10sec	30	69sec
6.mp4	2sec	30	78sec
7.mp4	10sec	20	83sec
8.mp4	10sec	30	82sec
9.mp4	10sec	20	73sec
10.mp4	15sec	30	72sec

6.2. Time Analysis (Feature Extraction)

Time Analysis for feature extraction is done based on various features like Color, Shape, Edge. In feature extraction the required time for different features is based on their related algorithms, For Color feature Average RGB, Local Color Histogram (LCH), For Shape Eccentricity and For Edge Edge Frequency algorithm is used.

Table-3: Time analysis

QUERY VIDEO	Avg. RGB	LCH	ECCENTRICITY	EDGE FREQUENCY
1.mp4	0.11sec	0.11sec	0.811sec	0.562sec
2.mp4	0.047sec	0.047sec	0.281sec	0.187sec
3.mp4	0.146sec	0.156sec	1.809sec	0.843sec
4.mp4	0.062sec	0.047sec	7.426sec	0.312sec
5.mp4	0.093sec	0.094sec	0.936sec	0.406sec
6.mp4	0.093sec	0.125sec	0.952sec	0.53sec
7.mp4	0.109sec	0.109sec	0.905sec	0.546sec
8.mp4	0.062sec	0.062sec	0.406sec	0.312sec
9.mp4	0.109sec	0.125sec	0.358sec	0.624sec
10.mp4	0.125sec	0.124sec	0.359sec	0.608sec

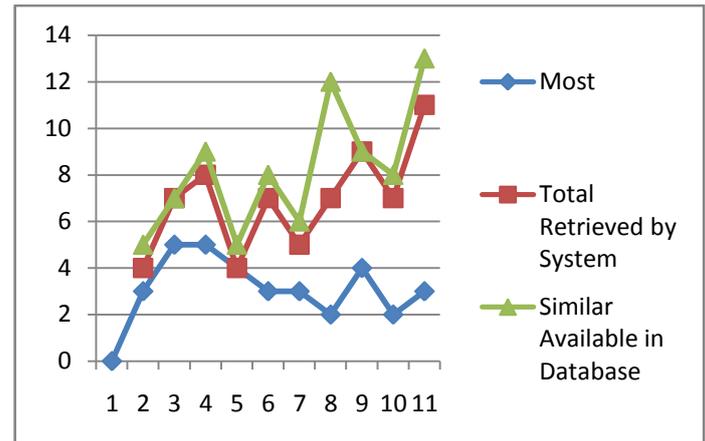
6.3. Final Retrieval

After final retrieval of the video from the database, the various results are shown when user queries the video to the database. The total no of videos retrieved by the system, No of similar videos are available in the database and most matched videos from database with the query video.

Table-4: Final Retrieval

QUERY VIDEO	Most Matched	Total Retrieved by System	Similar Available in Database
1.mp4	3	4	5
2.mp4	5	7	7
3.mp4	5	8	9
4.mp4	4	4	5
5.mp4	3	7	8
6.mp4	3	5	6
7.mp4	2	7	12
8.mp4	4	9	9
9.mp4	2	7	8
10.mp4	3	11	13

Following Graph shows the No. of videos most matched with query video, total videos retrieved by the system and similar videos available in the database.

**Fig-8.**Graph of most matched, retrieved and available video

7. CONCLUSIONS

This Paper has been envisioned for the purpose of retrieving the video from the Multimedia Database by using efficient algorithms to increase the performance of the system which is difficult in traditional video retrieving system. We are implementing Content Based Video Retrieval System.

REFERENCES

- [1] Avinash N. Bhute, B.B. Meshram "Automated Multimedia Information Retrieval using Color and Texture Feature Technique" IJECCE Volume 3, Issue 5, ISSN (Online): 2249-071X, ISSN (Print): 2278-4209
- [2] Ashok Ghatol "Implementation of Parallel Image Processing Using NVIDIA GPU framework." Advances in Computing Communication and Control. Springer Berlin Heidelberg, 2011. 457-464.
- [3] Journal article – NianhuaXie, Li Li, XianglinZeng, and Stephen Maybank "A Survey on Visual Content-Based Video Indexing and Retrieval" IEEE Transactions On Systems, Man, And Cybernetics—Part C:Applications And Reviews, Vol. 41, No. 6, November 2011.
- [4] Journal article – Hyun Sung Chang, SanghoonSull, Sang Uk Lee "Efficient Video Indexing Scheme for Content-Based Retrieval" IEEE Transactions On Circuits And Systems For Video Technology, Vol. 9, No. 8, December 1999.
- [5] Lijie Liu, GuoliangFan, "Combined Key Frame Extraction and Object Based Video Segmentation" IEEE Transactions On Circuits And Systems For Video Technology, Vol. 15, No. 7, July 2005.
- [6] Hang-Bong Kang, "Spatio-Temporal Feature Extraction From Compressed Video Data", IEEE Tencon, 1999.
- [7] O. Chapelle, P. Haffner, and V. Vapnik, "Svms for histogram based image classification," IEEE Trans. Neural Netw., vol. 10, pp. 1055-1064, 1999.
- [8] Rachid, Benmokhtar and Benoit Huet, Sid- Ahmed Berrani and Patrick Lechat " Video Shots Key Frames Indexing And Retrieval Through Pattern Analysis and fusion Techniques", CRE 46134752.

- [9] Bosque, José L., Oscar D. Robles, Luis Pastor, and Angel Rodríguez. "Parallel CBIR implementations with load balancing algorithms." *Journal of parallel and distributed computing* 66, no. 8 (2006): 1062-1075.
- [10] Dimitrovski, Ivica, et al. "Video Content-Based Retrieval System." *EUROCON, 2007. The International Conference on &# 34; Computer as a Tool&# 34;*, IEEE, 2007.
- [11] H. Farid and E. P. Simoncelli, Differentiation of discrete multidimensional signals, *IEEE Trans Image Processing*, vol.13(4), pp. 496--508, Apr 2004.
- [12] Visser, R., Sebe, N., Bakker, E.: Object recognition for video retrieval. In: *International Conference on Image and Video Retrieval, Lecture Notes in Computer Science*, vol. 2383, Springer (2002) 250–259.
- [13] Bakker, E., Lew, M.: Semantic video retrieval using audio analysis. In: *International Conference on Image and Video Retrieval, Lecture Notes in Computer Science*, vol. 2383, Springer (2002) 260–267
- [14] Robles, Oscar D., et al. "Towards a content-based video retrieval system using wavelet-based signatures." *7th IASTED International Conference on Computer Graphics and Imaging-CGIM. 2004.*
- [15] Ankush Mittal, Sumit Gupta(2006), "Automatic content-based retrieval and semantic classification of video content", *Int. J. on Digital Libraries* 6(1): pp. 30-38.

BIOGRAPHIES



Madhav V. Gitte, is a Final year Graduate student, Pursuing his bachelor degree in Information Technology at Sinhgad College of Engineering Pune-41, University of Pune, India. His area of interest is in image compression, Database and Algorithms



Harshal P. Bawaskar, is a Final year Graduate student, Pursuing his bachelor degree in Information Technology at Sinhgad College of Engineering Pune -41, University of Pune, India. His area of interest is in Video processing, Data Warehousing and Data Mining.



Sourabh Sethi, is a Final year Graduate student, Pursuing his bachelor degree in Information Technology at Sinhgad College of Engineering Pune-41, University of Pune, India. His area of interest is in Image processing, XML, Web Mining.



Ajinkya V. Shinde, is a Final year Graduate student, Pursuing his bachelor degree in Information Technology at Sinhgad College of Engineering Pune-41, University of Pune, India. His area of interest is in android developing ,c#, java,c, Database and Algorithms