# Analysis of color histogram based similarity search and retrieval strategy of videos in Video on Demand systems

Thesis submitted in partial fulfillment for

the degree of

**Bachelor of Technology**
**in**
**Computer Science and Engineering**

By

**MANABESH MANDAL**

**ROLL NO: 108CS014**

and

**KEDAR SANKAR BEHERA**

**ROLL NO: 108CS016**

Under the Guidance
of
**PROF. BIBHUDATTA SAHOO**


Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela-769008, Odisha, India


National Institute of Technology
Rourkela

# CERTIFICATE

This is to certify that the thesis titled "**Analysis of color histogram based similarity search and retrieval strategy of videos in Video on Demand systems** "submitted by Manabesh Mandal :108CS014 and Kedar Sankar Behera: 108CS016 and in the partial fulfillment of the requirement for the degree of Bachelor of Technology in Computer Science Engineering, National Institute of Technology, Rourkela , has been carried out under my supervision. To the best of my knowledge the matter embodied in the thesis has not been submitted to any other university/institute for the award of any degree or diploma.

Date: 14/05/2012

Place: NIT Rourkela

Prof. Bibhudatta Sahoo
Department of Computer Science and Engineering
National Institute of Technology, Rourkela

# ACKNOWLEDGEMENT

We wish to express our sincere and heartfelt gratitude towards our guide Prof. Bibhudatta Sahoo, Computer Science Engineering Department, for his guidance, inspiration and above all help in all regards during the duration of our project. We would also like to thank all the professors of the department of Computer Science and Engineering, National Institute of Technology, Rourkela, for their constant motivation and guidance and last but not the least, a sincere thanks to our batch mates for their help and cooperation.

Manabesh Mandal
Roll No: 108CS014

Kedar Sankar Behera
Roll No: 108CS016

# TABLE OF CONTENTS

**LIST OF FIGURES**

## LIST OF TABLES

## *Abstract*

*The advent of the internet and smart hand held devices have driven the explosion of multimedia data especially video data. It has become difficult for the end user to get his desired content in a stipulated time as services like video-on-demand systems and video share Web, the major contributors of video data, has led to the ever growing quantity of video databases. This has led to extensive research in the field of video similarity search for content-based video retrieval. Traditional methods of content based retrieval strategies are computationally expensive and do not consider the temporal features of a video. Hence a fast content based scalable similarity search strategy has been an active area of research. There are two primary challenges regarding visual similarity search problem: video similarity measure and fast search method in large database. A compact signature of video is computed according to image histogram by extracting frames of a video. The video similarity is measured by the computation of the distance of signature of video. A search method based on clustering index table by index clustering and scoring using different parameters was analyzed.*

# CHAPTER 1

# Introduction

## 1. Introduction

We now have relatively easy access to large pool of multimedia data, especially digital video which has one of the steepest growth curves. Digital TV and set-top boxes, personal video recorders, DVDs and more recently Internet video such as through YouTube and Video On Demand services have all contributed in placing enormous video archives at our disposal, but the prerogative is that we could navigate them effectively. Of all the technical issues related with the video lifecycle like video capture, storage, compression, transmission and processing now solved for practical deployment, video navigation and retrieval remains a challenging task at hand. The ability to navigate, browse and search it in order to locate videos of interest is the matter of concern.

The primary approach of navigating digital video in large-scale practical applications has been to use video metadata, either automatically determined or manually assigned. While content description from user annotation offers useful navigation possibilities, it is still one step removed from being able to search actual video content directly. Here, we concentrate on the application of direct access to video where user queries are matched directly against video content. Various techniques of visual similarity based matching have been studied and we have implemented a suitable algorithm for similarity search and retrieval of video sequences in a video database.

The video similarity in this context means visual similarity. Similarity measurement and efficient search techniques are two key points in similarity search. An efficient visual similarity search approach is implemented and a new scoring mechanism is used to study the corresponding change in recall and precision. The feature extraction of image and video was

achieved by the statistics based on spatial-temporal distribution. The high-dimensional feature was transformed into compact signature of video and the video similarity was measured by the computation of the distance of signature of video. A scalable search method was also proposed based on Clustering Index Table (CIT). By virtue of index clustering, the similarity search can be carried out very efficiently with satisfying recall and precision rate. The experimental results indicate this approach performs well in large database for video similarity search and it can be employed for similar video retrieval in VOD systems.

### 1.1 Literature review

The feature representation is the basis of similarity measurement and a lot of research efforts have been performed to find effective feature representation in image and video [1-7]. Hoi et al. [6] employed coarse-to-fine video retrieval scheme based on shot-level spatial-temporal statistics. The coarse search calculates the relation between two consecutive shots and fine search step refines the search result by using the local color features extracted from the key frames of the query shots. Shen et al. [7] proposed bounded component system to summarize the distribution of all frame points in feature vector space into a single representation. Besides, several other feature extraction algorithms such as motion texture [2] were also proposed. Although these techniques are valuable for the further studies on image and video feature extraction, they are usually very complicated and are not efficient enough for the feature extraction of large video database, especially when the volume of database now is growing very fast.

Efficient indexing and searching in high dimensional metric space is another key point for similarity search. The indexing techniques based on locality sensitive hashing (LSH) [4] have been widely used for approximate nearest neighbor search problem in high-dimensional spaces. The existing LSH techniques, usually containing hundreds of hash tables, are inefficient for scalable fast search in large database.

## 1.2 Motivation

The research in visual information indexing and retrieval has been one of the most popular research directions in the area of Information Technologies. The reasons for that is the technological maturity of capture, storage and network infrastructure that allows common daily life capturing of devices and recording of video with professional equipment or personal mobile devices. Areas of societal activity like video surveillance, digitization and storage of old feature films, broadcasts, documentaries lead to the production of a humongous amount of video data which has to be accessed and searched. [8]

However, owing to the size of large scale video databases and complexity of visual interpretation tasks and the semantic gap between the concept the user is looking for and the digital representation of the visual content research directions are widely open. The existing approaches been mentioned in the previous section consume a lot of computational cost because of the higher dimensional feature vector, hence we explore a simpler yet effective method [1] of color histogram based similar video search and retrieval.

## 1.3 Problem Statement

Video similarity search for content-based video retrieval is important in web service and research field. There is a good scope for research in the search method for scalable fast similarity search in large video database like in the case of VOD. Similarity measurement and efficient search techniques are two key points in Video Similarity Search. The feature representation is the basis of similarity measurement and efficient indexing and searching is another key point for similarity search. The feature extraction of image and video can be achieved by the statistics based on spatial-temporal distribution, which can be transformed into a compact video signature and the video similarity can be measured by the computation of the distance of video signature. Scalable search methods based on the video signature as index can be used..

We have explored the possibility of using a simpler method based on statistical parameters of mean and standard deviation of normalized color histogram of a video sequence to decrease the irrelevant results basing on the meta data . We have tried to decrease the spatial cost of the feature representation from multi-dimensional feature vectors to compact signature decreasing the cost.

## 1.4 Organization of the thesis

The thesis is organized as follows. Chapter 2 describes the theory behind video signature generation and indexing .A basic framework is proposed and the following method is discussed. Section 2.3 suggests a method for video signature generation and Section 2.4 mentions the indexing and retrieval strategies. Section 2.4 states the two performance metrics used for evaluation of the search results. Chapter 3 gives the implementation details of the algorithm and section 3.2 gives the results and recall vs precision graphs. A example query is given in section 3.4. Finally, Chapter 4 concludes the result obtained in chapter 3 and future work.

# CHAPTER 2

# Video Signature Generation and Indexing

## 2.1 Introduction

A visual similarity search method[1] is implemented and a new scoring mechanism is used to study the corresponding change in recall and precision. The feature extraction of image and video was achieved by the statistics based on spatial-temporal distribution. The high-dimensional feature was transformed into compact signature of video and the video similarity was measured by the computation of the distance of signature of video. A scalable search method was also proposed based on Clustering Index Table (CIT)[1]. By virtue of index clustering, the similarity search can be carried out very efficiently with satisfying recall and precision rate. The experimental results indicate this approach performs well in large database for video similarity search and it can be employed for similar video retrieval in Video on Demand systems.

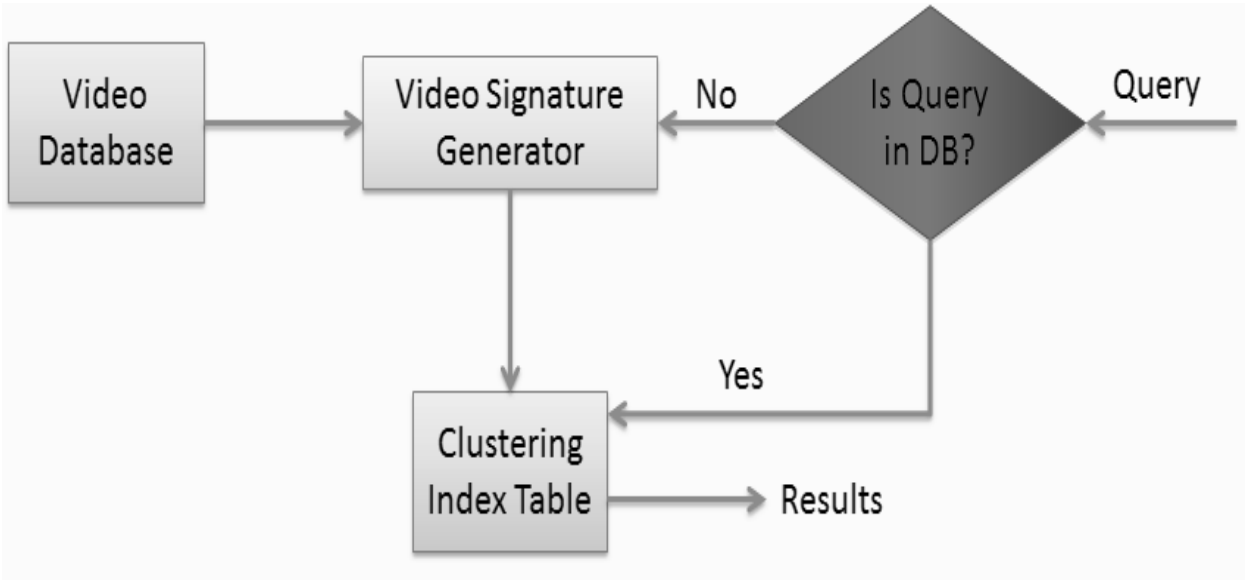


**Figure 2.1: Framework of the visual similarity search engine**

## 2.2 Compact video representation

A framework based on signature-based index structures featuring perceptual visual attributes and a matching and scoring framework is considered. The low-level visual features (color, texture, etc.) in high-dimensional spaces traditionally results in curse of dimensionality.

Hence a compact signature of video is considered by using mean and standard deviation of the feature vector thus reducing its dimension.

### 2.2.1 Signature of video

A framework based on signature-based index structures featuring perceptual visual attributes and a matching and scoring framework is considered [1]. The low-level visual features (color, texture, etc.) in high-dimensional spaces traditionally results in curse of dimensionality.

Hence a compact signature of video is considered by using mean and standard deviation of the feature vector thus reducing its dimension.

$$m_1 = \sum_{kz=1}^{N} \mathrm{i}\,\mathrm{Y_i} \tag{1}$$

$$m_2 = \sum_{i=1}^{N} \mathrm{i}\,\mathrm{Cb_i} \tag{2}$$

$$m_3 = \sum_{i=1}^{N} \mathrm{i}\,\mathrm{Cr_i} \tag{3}$$

$$S = 2m_1 + m_2 + 3m_3 \tag{4}$$

where,

i: no of bins in each image.

$Y_i$: normalized frequency of the luminance.

$Cb_i$: normalized frequency of the blue chrominance.

$Cr_i$: normalized frequency of the red chrominance.

In order to obtain basic information of Spatial-Temporal Distribution of video frame sequences, two statistics are employed to generate compact signature of video: mean ($v_m$) and standard deviation ($v_d$) of image signature of video frame sequences where s is image signature and L is the number of frames in the video.

$$v_m = \sum_{i=1}^{L} S_i / L \tag{5}$$

$$v_d = \sum_{i=1}^{L} [(S_i - v_m)^2 / L]^{\frac{1}{2}} \tag{6}$$

where, L : no of frames.

The Algorithm for generation of signature of video is given below:

1. Read video

2. while (!end_of_video)

    2.1 Capture frame

    2.2 Compute normalized histogram using N bins for each channel.

    2.3Compute $m_1$, $m_2$, $m_3$ using formulae given above.

3. Compute $s=2m_1+m_2+3m_3$

4. Compute $v_m$ and $v_d$ using formulae given above.

$[v_m, v_d]$ represents the compact signature of video which is then stored in a database and later it is used to compute index and ranking of search results.

## 2.3 Indexing and Retrieval

### 2.3.1 Clustering based on compact signature of videos

Given two video clips $v_1$ and $v_2$, their signature of videos are $(v_{m1}, v_{d1})$ and $(v_{m2}, v_{d2})$, respectively, $v_m$ can be used to generate indices and cluster the videos based on these indices and $v_d$ can be used to rank the retrieved results. The indexes of tables are created by the integer quotient of signature of video divided by 10 and the video clips with similar signature are arranged into the same clustering index table [1].

Giving a query video clip, the search is directed to the corresponding index tables. For the completeness of similarity search, the indexes of search tables are decided by the following rules: If the remainder is smaller than 5 and q-1 or q+1 exists, the indexes of tables are q-1 and q. Otherwise, the indexes are q and q+1 [1].

$$q = v_m/10 \tag{7}$$

$$id = \begin{cases} (q-1, q), & r < 5 \\ (q, q+1), & r \geq 5 \end{cases} \tag{8}$$

The similarity of v1and v2 can be defined as following:

$$D = |v_{m1} - v_{m2}| + |v_{d1} - v_{d2}| \quad (9)$$

### 2.3.2 Scoring based on compact signature of videos

When the difference between the symbols exceeds a given value, a negative score is awarded, and when the difference is less than this value, a positive score is given. We refer to this point as the indecision point. In order to manipulate the indecision point, the scoring functions include a scaling factor, k, which operates on the difference [2].

We refer to this scaled symbol difference as δ:

$$\delta = k.|s_q - s_d| \quad (10)$$

The scaling factor, k, allows us to modify the indecision point without affecting the magnitude of the scores awarded by the scoring function.

The linear scoring system uses a simple linear function to award a score based on the difference of the two symbols being compared:

$$s = 20 - \left(\frac{3}{2}\right).\delta \quad (11)$$

From the maximum score of 20, we subtract the scaled symbol difference, δ. The scaled symbol difference is multiplied by a factor of (3/2), which alters the gradient of the function in order to set the indecision point at 12. The choice of 12 as the indecision point was somewhat arbitrary.

## 2.4 Performance metrics

The two performance metrics primarily used in Information retrieval strategies are *precision* and *recall* [2].

Precision (P) is the fraction of retrieved documents that are relevant to the search. Recall(R) in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved. We use both of these parameters to quantify and analyze the results.

$$P = \frac{\text{no of correct matches}}{\text{no of return results}} \tag{12}$$

$$R = \frac{\text{no of correct matches}}{\text{no of data set elements}} \tag{13}$$

## 2.5 Conclusion

The video signature generation method produces a very compact representation of the video i.e. a vector [ $v_m$, $v_d$] which enables it occupy less storage space and fast computation for indexing and retrieval. The indexing used is based on a simple has function and clustering techniques outperforms other complex data structures. Scalability is also achieved as insertion, deletion and modification of the hash table does not consume any appreciable overhead cost.

# CHAPTER 3

# Analysis of compact video signature based method for retrieval

## 3.1 Introduction

The simulation is carried out for a test database, created using MS Access 2007and by a simulator designed in MATLAB R2011. The simulated results are obtained and expressed in terms of performance metrics i.e. Recall(R) and Precision (P) for analyzing the accuracy of the search and for different values of threshold parameter (D) using the clustering and scoring method for determining video similarity.

## 3.2 Simulation Setup

The database in the experiments is made up of 40 video clips downloaded from internet. Each was converted to (.avi) format for uniformity purposes.is data set is also our truth set used to determine our performance metric. The data set was divided into 4 categories .i.e. cricket, formula one, underwater and universe .These categories were chosen randomly and arbitrarily. The experiments were performed on Intel(R) core i5 2.66GHz processor, 4GB memory.

There were 20 queries in tests, computing the average recall rate (RR), precision rate (PR) and run time, respectively. The measurement of recall rate and precision rate is the same as the definition in (12), (13).

| ID | $v_d$ | $v_m$ | index |
|---|---|---|---|
| 1 | 3 | 162 | 16 |
| 2 | 5 | 192 | 19 |
| 3 | 6 | 161 | 16 |
| 4 | 15 | 173 | 17 |

| | | | |
|---|---|---|---|
| 5 | 4 | 174 | 17 |
| 6 | 3 | 183 | 18 |
| 7 | 3 | 184 | 18 |
| 8 | 8 | 191 | 19 |
| 9 | 7 | 174 | 17 |
| 10 | 6 | 174 | 17 |
| 11 | 0 | 175 | 27 |
| 12 | 4 | 192 | 19 |
| 13 | 1 | 182 | 18 |
| 14 | 5 | 162 | 16 |
| 15 | 18 | 168 | 16 |
| 16 | 1 | 150 | 15 |
| 17 | 12 | 185 | 18 |
| 18 | 0 | 161 | 16 |

| | | | |
|---|---|---|---|
| 19 | 6 | 159 | 15 |

**Table 3.1: The table for the Video Signature database**

### 3.2.1 Simulation

The database was created which contains two tables, one having the video id and path; and other having the id ($v_m$, $v_d$) and the video index. We have chosen N no. of videos from the database randomly as a query video their search space was calculated according to eqn (8).

In the clustering based method, D as in the eqn(9). has been calculated. A threshold matrix has been created for D as [10, 15, 20]. The similar videos that satisfy the selected threshold number were calculated and the precision vs recall graph has been created for the random 20 query videos.

In scoring based method, we take the value of k as 1in eqn(10). and D of eqn(9). is used as the difference of signature ( $\delta$) value as in eqn(11). and the indecision value has been taken 12 arbitrarily. So for the videos where the value is greater than 20 are discarded .Now the threshold value of D has been used and the values between the threshold value and 20 is considered to be a similar video.

The recall(R) is calculated by the evaluation of index of the query video and comparing it with the respective truth set i.e. no. of videos returned that belong in that truth set divided by the total no. of videos in that truth set. Precision (P) is calculated by total no. of videos returned that belong to the same truth set of the query divided by the total no. of videos returned.

### 3.3 Result

The precision vs recall graph for different threshold D values for the 2 process are given as follows:

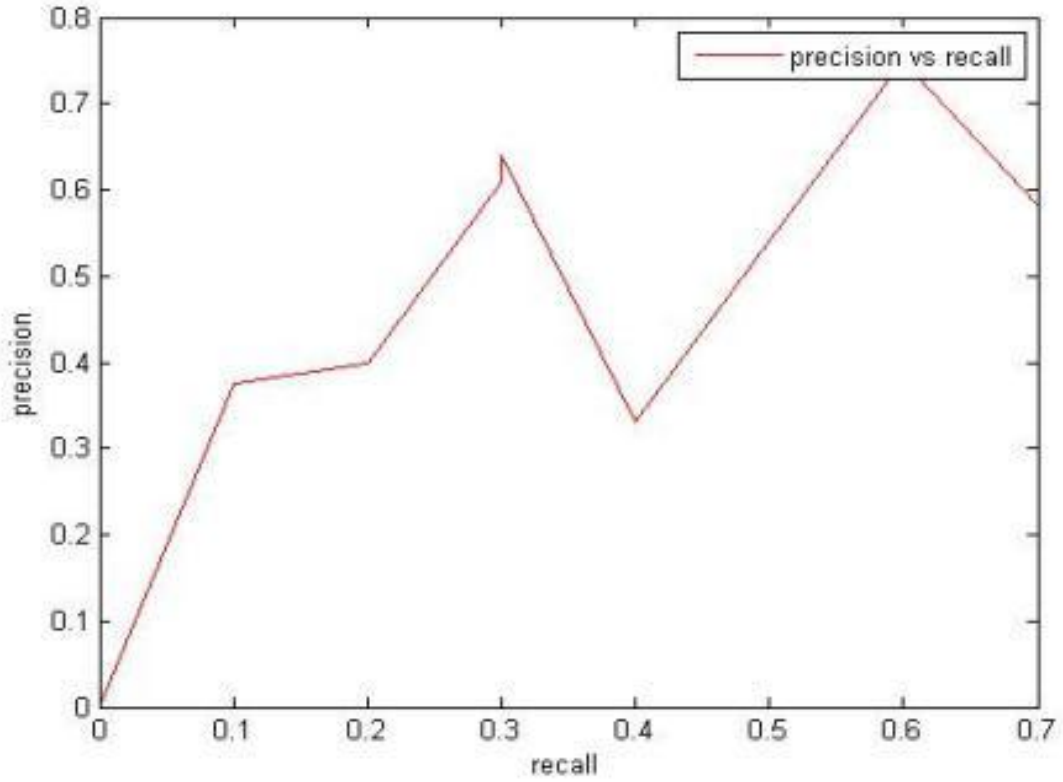


**Figure 3.1: Recall vs Precision graph for D=10 using clustering based indexing**

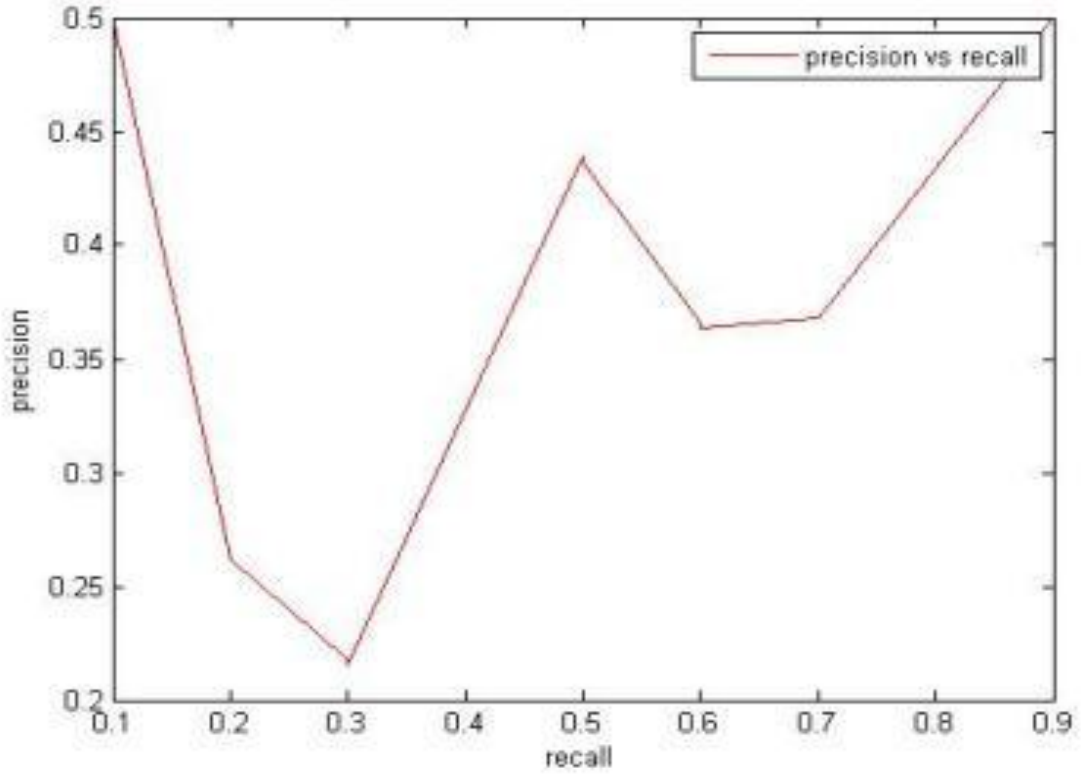


**Figure 2.2: Recall vs Precision graph for D=15 using clustering based indexing**

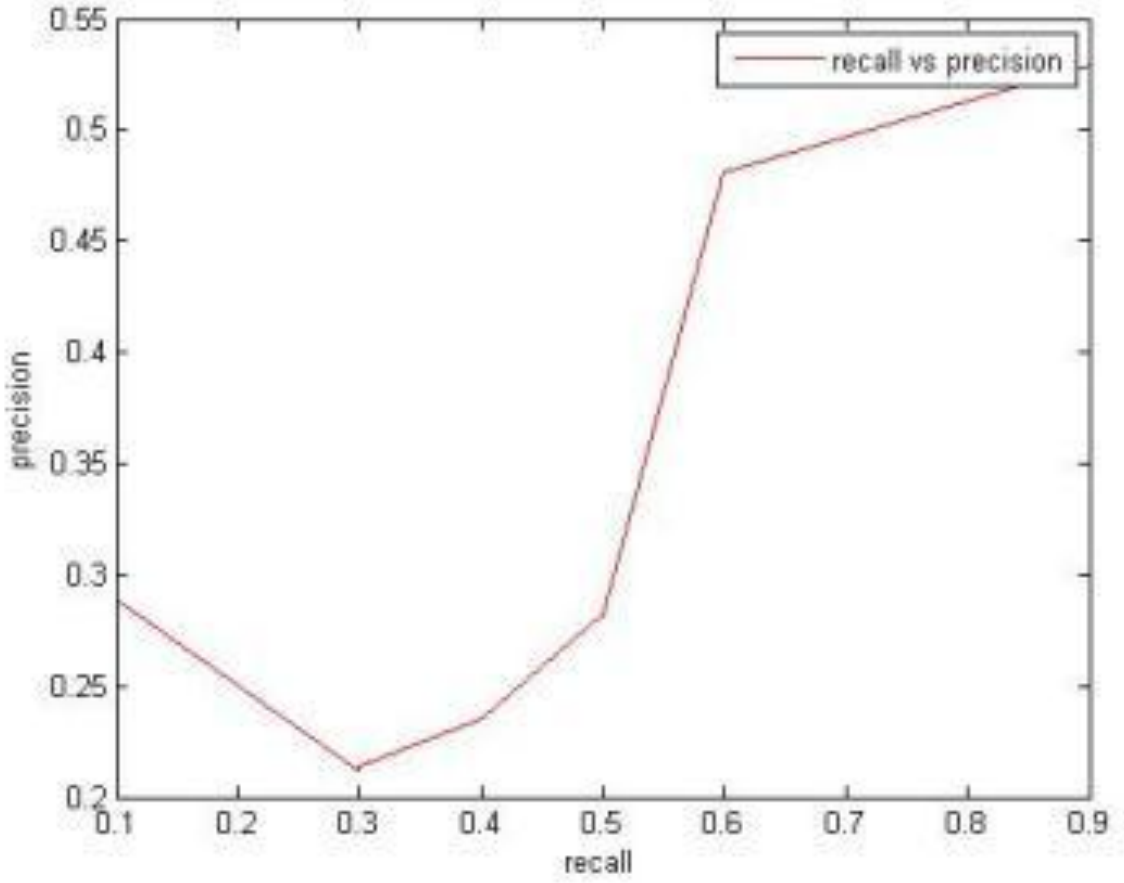


**Figure 3.3: Recall vs Precision graph for D=20 using clustering based indexing**

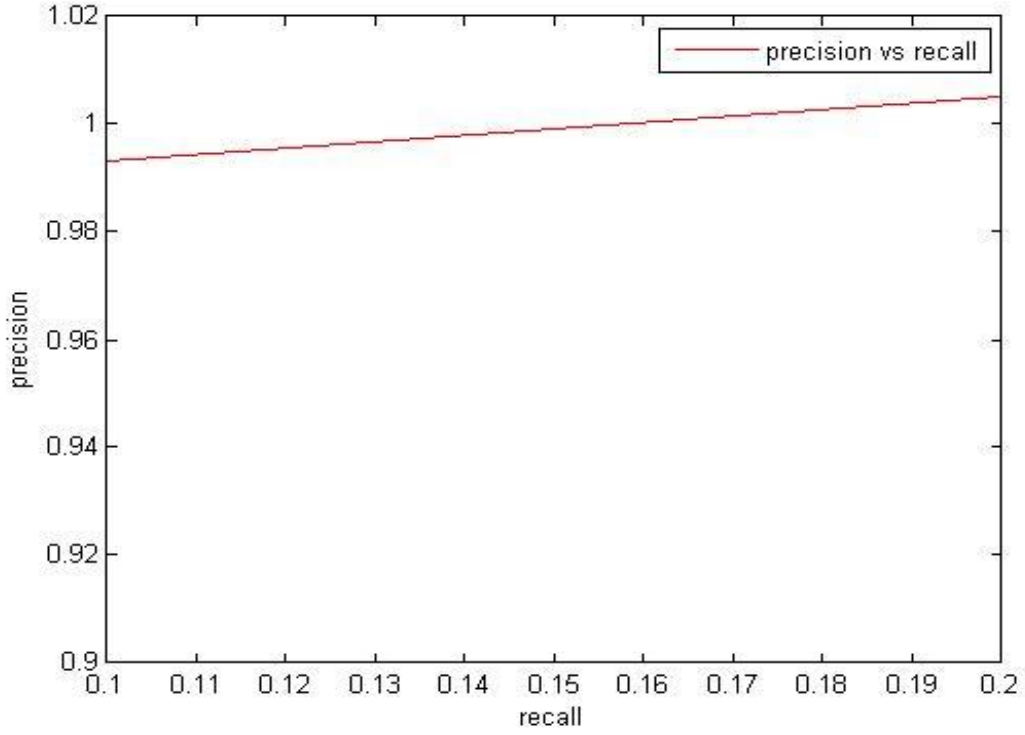


**Figure 3.4: Recall vs Precision graph for D=10 using linear scoring based indexing**

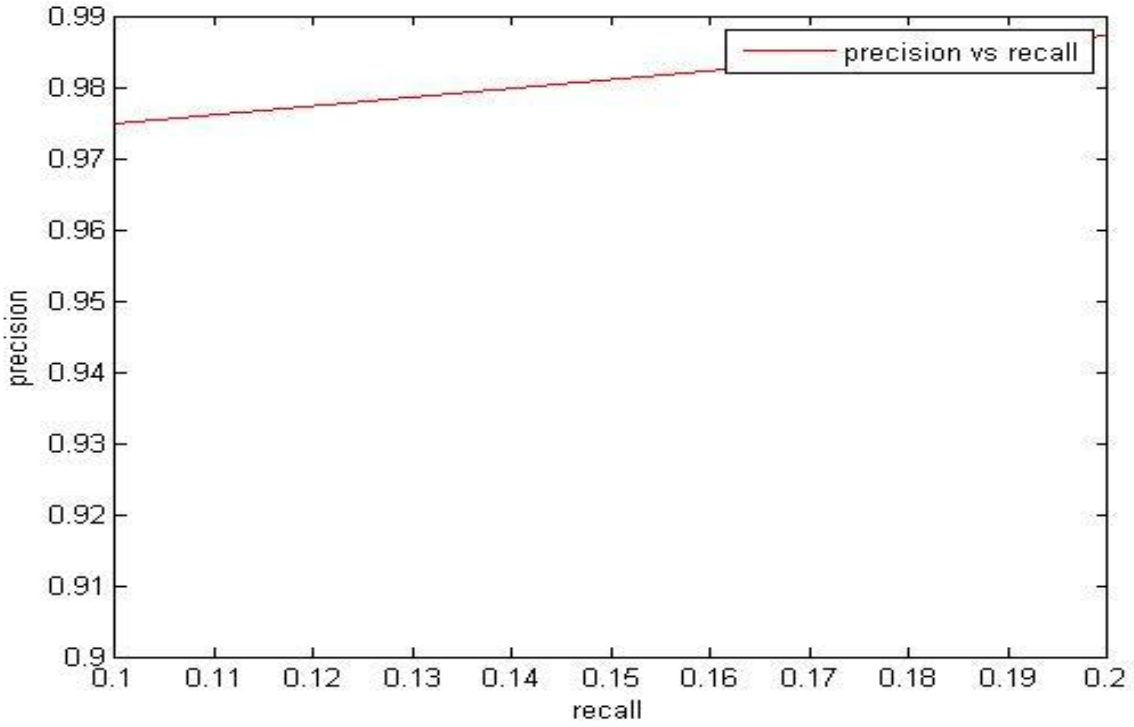


**Figure 3.5: Recall vs Precision graph for D=15 using linear scoring based indexing**

### 3.4 Inference

It has been observed that for D=10, in clustering method the slope rises up at r=0.3 and then falls but for D =15 and 20 in the same method the slope falls at r=0.3 and then rises. With a difference between D=15 and D=20 is that at D=15 the graph is jiggered in comparison to that at D=20.

And the precision steeply rises for the Linear scoring method in comparison with the previous method.

### 3.5 An Example Query

Here is an example query of the selected videos of a formula one event.


**Figure 3.6: A snapshot of a query video**

And the results retrieved are:

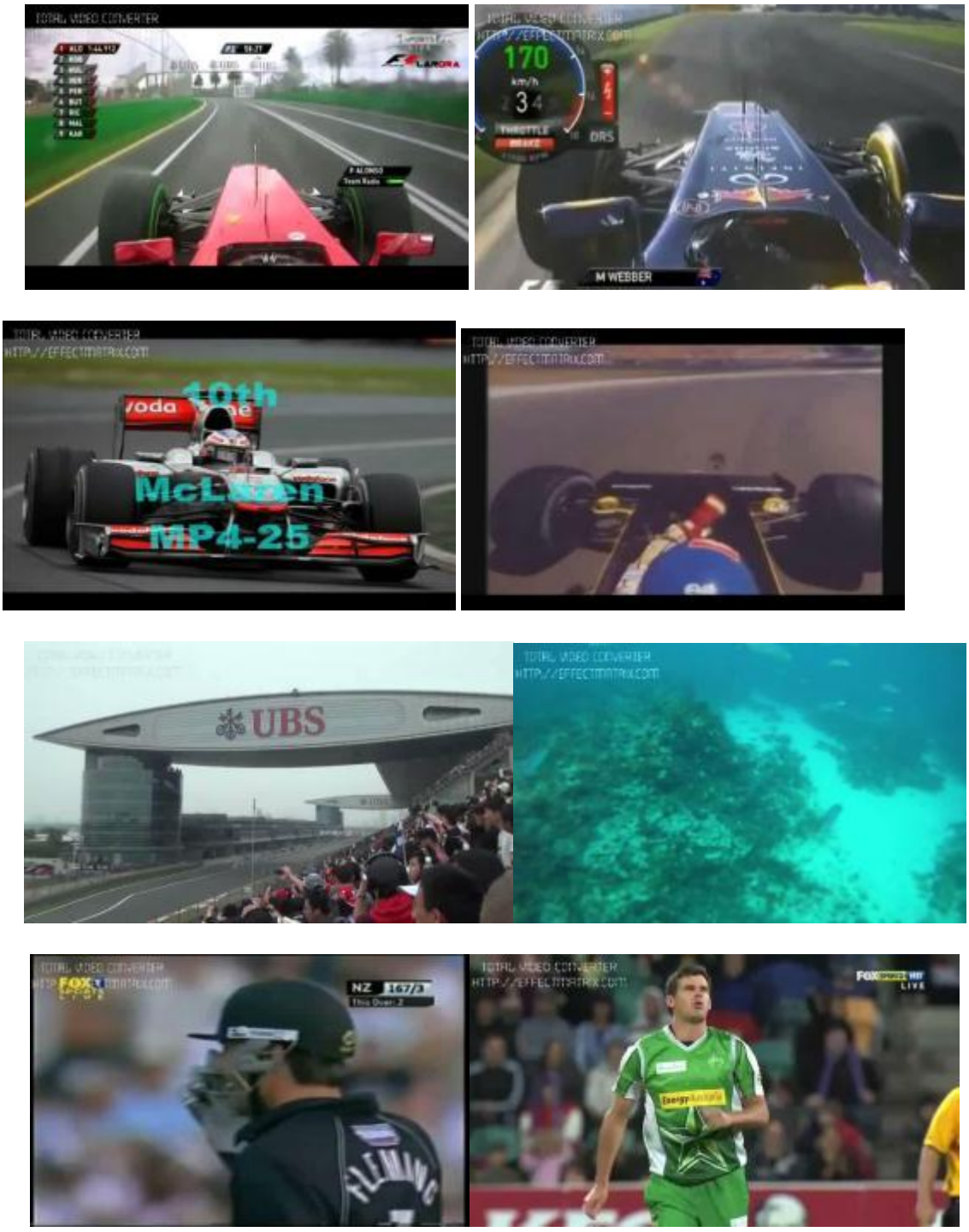


**Figure 3.7: Snapshots of the search results returned.**

# CHAPTER 4

# CONCLUSION AND FUTURE WORK

## 4.1 Conclusion

A video search engine that uses compact signature of videos to represent video clips from a large database has been described. The two major algorithmic designs in this search engine are:-

(i) A feature extraction mapping for fast similarity search.

(ii) A clustering algorithm for grouping signatures into similar clusters.

As only color histogram was used to compute the low level feature vector, other parameters such as texture, motion, luminance centroid etc. can be used to determine low level features. These parameters do not give any indication about the temporal relationships between different video sequences; other methods could be studied and integrated with the existing parameters.

## 4.2 Future work

However, the result can still be improved since the color histogram based scheme is fragile for color distortion problem. To decrease the computation cost the no of frames processed can be reduced by selecting key frames by using a selection strategy. In the future work, other features like color corellogram and texture identification, luminance centroid etc. can be used to tune the video retrieval performance. Also the algorithm can be tested on a very large video database and for more versatile data in the future.

User feedback can be incorporated into the scoring mechanism to refine the search results. The advanced machine learning techniques can be studied and adopted to maximise the effectiveness and efficiency of content based search of videos even further.

**REFERENCES**


[1] Zheng C., Ming Z., "An efficient video similarity search strategy for video-on-demand systems", *Broadband Network & Multimedia Technology, 2009. IC-BNMT '09. 2nd IEEE International Conference on Digital Object Identifier*, pp.174-178.

[2] Hoad TC, Zobel J, "Detection of video sequences using compact signatures", *ACM Trans. on Information Systems, 2006* ,Vol. 24, pp.1-50.

[3] Cheung S, Zakhor A, " Fast similarity search and clustering of video sequences on the worldwide-web", *IEEE Trans. on Multimedia, 2005* ,Vol.7,pp.524-537.

[4] Gionis A., Indyk P., and Motwani R., "Similarity Search in High Dimensions via Hashing", *In Proceedings of the 25th International Conference on Very Large Data Bases, 1999,* pp 12-70 .

[5] Naphade M. R., Yeung M.M., and Yeo,B.L., " A novel scheme for fast and efficient video sequence matching using compact signatures", *In Proceedings of SPIE, Storage and Retrieval for Media Databases, 1999,* Vol. 3972.2000, pp. 564–572.

[6] HOI C.H, "Similarity measurement and detection of video sequences", *Tech. rep. Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong*, 2003.

[7] . Shen H. T, Zhou X., Huang Z., and Shao J., "Statistical summarization of content features for fast near-duplicate video detection.", *Proc of ACM int'l conf. on Multimedia*, 2007 pp.164–165.

[8] Benois-Pineau J., Precioso F., Cord M., Visual Indexing and Retrieval, *Springer Briefs in Computer Science*, 2012.