Content Based Video Retrieval in Transformed Domain using Fractional Coefficients

Dr. H. B. Kekre *hbkekre@yahoo.com*
Senior Professor Computer Engineering Department
Mukesh Patel School of Technology, Management & Engineering
NMIMS University
Mumbai, India

Dr. Sudeep D. Thepade *sudeepthepade@gmail.com*
Professor & Dean (R&D),
Pimpri Chinchwad College of Engineering,
University of Pune,
Pune, India

Saurabh Gupta *saurabh.gupta761@gmail.com*
M.Tech (Comp. Engg.) Student
Mukesh Patel School of Technology, Management & Engineering
NMIMS University
Mumbai, India

Abstract

With the development of multimedia and growing database there is huge demand of video retrieval systems. Due to this, there is a shift from text based retrieval systems to content based retrieval systems. Selection of extracted features play an important role in content based video retrieval. Good features selection also allows the time and space costs of the retrieval process to be reduced. Different methods[1,2,3] have been proposed to develop video retrievals systems to achieve better performance in terms of accuracy.

The proposed technique uses transforms to extract the features. The used transforms are Discrete Cosine, Walsh, Haar, Kekre, Discrete Sine, Slant and Discrete Hartley transforms. The benefit of energy compaction of transforms in higher coefficients is taken to reduce the feature vector size by taking fractional coefficients[5] of transformed frames of video. Smaller feature vector size results in less time for comparison of feature vectors resulting in faster retrieval of images. The feature vectors are extracted and coefficients sets are considered as feature vectors (100%, 6.25%, 3.125%, 1.5625%, 0.7813%, 0.39%, 0.195%, 0.097%, 0.048%, 0.024%, 0.012%, 0.006% and 0.003% of complete transformed coefficients). The database consists of 500 videos spread across 10 categories.

Keywords: Key Frame, Feature Extraction, Similarity Measures, Orthogonal Transforms.

1. INTRODUCTION

The amount of video content being uploaded to the internet is increasing day by day. Hence, extracting specific videos from the massive amount of multimedia data has been the focus of attention over the past few years. The traditional text-based search has drawbacks like manual annotations are time consuming and costly to implement [1]. With the increase in the number of media in a database, the complexities in determining the required information also increases.

1.1 Content Based Video Retrieval (CBVR)

Content-Based Video Retrieval System (CBVR) is defined as the search which will retrieve video from database based on contents. Content relates to color, shapes, textures, or any other information that can be obtained from the video directly.

CBVR mainly contains three stages. Firstly key frames are extracted from video. Key-frames are still images extracted from original video data that best represent the content of video [9]. Secondly, feature extraction (FE) is done, where a set of features, called feature vector, is generated to accurately represent the content of each video in the database. After this, similarity measurement is the done where a distance between the query video and each video in the database using their feature vectors is used to retrieve the "closest" videos. Features such as Motion features [2,3], Color features and edge using histograms [2] and DCT transforms[3]. Video Retrieval Based on Textual Queries [10] presented an approach that enables search based on the textual information present in the video. Regions of textual information are indented within the frames of the video.

Here transforms are used to extract features as they give high energy compaction in transformed domain, hence frames from video in transformed domain are used for feature extraction in CBVR. The energy compaction is in few elements, so large number of the coefficients of transformed image can be neglected to reduce the size of feature vector [4].

In this paper, Discrete Cosine, Walsh, Haar, Kekre, Discrete Sine, Slant and Discrete Hartley transforms are used with reduced size feature vector using fractional coefficients of transformed frames. Mean Squared Error(MSE) is used as similarity measure as shown in equation(1.1)

$$\text{MSE}(x, y) = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - y_i\right)^2 \qquad \text{.....(1.1)}$$

2. ORTHOGONAL TRANSFORMS

2.1 DCT(Discrete Cosine Transform) [6]

The NxN cosine transform matrix C={c(k,n)},also called the Discrete Cosine Transform(DCT),is defined as

$$c(k,n)=\begin{cases}\frac{1}{\sqrt{N}} & k=0, 0\le n\le N-1\\ \sqrt{\frac{2}{N}}\cos\frac{\Pi(2n+1)k}{2N} & 1\le k\le N-1, 0\le n\le N-1\end{cases} \qquad \text{.....(2.1.1)}$$

The one-dimensional DCT of a sequence {u(n),0≤n≤N-1} is defined as

$$v(k) = \alpha(k)\sum_{n=0}^{N-1} u(n)\cos\left[\frac{\Pi(2n+1)k}{2N}\right] \quad 0 \le k \le N-1 \qquad \text{.....(2.1.2)}$$

Where,

$$\alpha(0) = \frac{1}{\sqrt{N}}, \alpha(k) = \sqrt{\frac{2}{N}}\ for 1 \le k \le N-1$$

The inverse transformation is given by

$$u(n) = \sum_{k=0}^{N-1}\alpha(k)v(k)\cos\left[\frac{\Pi(2n+1)k}{2N}\right], 0 \le n \le N-1 \qquad \text{.....(2.1.3)}$$

2.2 DST(Discrete Sine Transform)

The NxN sine transform matrix $\psi = \{\Psi(k,n)\}$, also called the Discrete Sine Transform(DST), is defined as

$$\Psi(k,n) = \sqrt{\frac{2}{N+1}} \sin\frac{\Pi(k+1)(n+1)}{N+1} \qquad 0 \leq k, n \leq N-1 \qquad \text{.....(2.2.1)}$$

The sine transform pair of one-dimensional sequences is defined as

$$v(k) = \sqrt{\frac{2}{N+1}} \sum_{n=0}^{N-1} u(n) \sin\frac{\Pi(k+1)(n+1)}{N+1} \qquad 0 \leq k, n \leq N-1 \qquad \text{..... (2.2.2)}$$

The inverse transformation is given by

$$u(n) = \sqrt{\frac{2}{N+1}} \sum_{n=0}^{N-1} v(k) \sin\frac{\Pi(k+1)(n+1)}{N+1} \qquad 0 \leq n \leq N-1 \qquad \text{.....(2.2.3)}$$

2.3 Haar Transform

The Haar wavelet's mother wavelet function φ(t) can be described as:

$$\varphi(t) = \begin{Bmatrix} 1 & ,0 \leq t \leq \frac{1}{2} \\ -1 & ,\frac{1}{2} \leq t \leq 1 \\ 0 & ,Otherwise \end{Bmatrix} \qquad \text{..... (2.3.1)}$$

And its scaling function $\varphi(t)$ can be described as,

$$\varphi(t) = \begin{Bmatrix} 1 & ,0 \leq t \leq 1 \\ 0 & ,Otherwise \end{Bmatrix} \qquad \text{.....(2.3.2)}$$

2.4 Walsh Transform

Walsh transform matrix is defined as a set of N rows, denoted W_j, for j = 0, 1, ..., N - 1, which have the following properties:

- W_j takes on the values +1 and -1.
- $W_j[0] = 1$ for all j.
- $W_j \times W_{KT} = 0$, for $j \neq k$ and $W_j \times W_{KT}$ W_j has exactly j zero crossings, for j = 0,1....N-1.
- Each row W_j is even or odd with respect to its midpoint.

Walsh transform matrix is defined using a Hadamard matrix of order N. The Walsh transform matrix row is the row of the Hadamard matrix specified by the Walsh code index, which must be an integer in the range [0... N-1]. For the Walsh code index equal to an integer j, the respective Hadamard output code has exactly j zero crossings, for j = 0, 1... N - 1.

2.5 Kekre Transform

Kekre Transform matrix can be of any size NxN, which need not have to be in powers of 2. All upper diagonal and diagonal values of Kekre's transform matrix are one, while the lower diagonal part

except the values just below diagonal is zero. Generalized NxN Kekre Transform matrix can be given as

$$K_{NxN} = \begin{bmatrix} 1 & 1 & 1 & .. & 1 & 1 \\ -N+1 & 1 & 1 & .. & 1 & 1 \\ 0 & -N+2 & 1 & .. & : & : \\ : & 0 & .. & .. & 1 & 1 \\ 0 & 0 & 0 & .. & -N+(N-1) & 1 \end{bmatrix} \quad \text{.....(2.5.1)}$$

The formula for generating the term K_{xy} of Kekre Transform matrix is

$$K_{x,y} = \begin{cases} 1 & , x \le y \\ -N+(x+1), & x = y+1 \\ 0 & , x > y+1 \end{cases} \quad \text{.....(2.5.2)}$$

2.6 Hartley Transform [7]

The Discrete Cosine Transform(DCT) utilizes cosine basis functions, while Discrete Sine Transform(DST) uses sine basis function. The Hartley transform utilizes both sine and cosine basis functions. The discrete 2-dimensional Hartley Transform is defined as,

$$F(u,v) = \frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\,Cas\left\{\frac{2\pi}{N}(ux+vy)\right\} \quad \text{.....(2.6.1)}$$

Inverse discrete 2-dimensional Hartley Transform is,

$$f(x,y) = \frac{1}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} F(u,v)\,Cas\left\{\frac{2\pi}{N}(ux+vy)\right\} \quad \text{.....(2.6.2)}$$

where, $Cas\theta = \cos\theta + \sin\theta$

2.7 Slant Transform [8]

The Slant transform is a member of the orthogonal transforms. It has a constant function for the first row, and has a second row which is a linear (slant) function of the column index. The matrices are formed by an iterative construction that exhibits the matrices as products of sparse matrices, which in turn leads to a fast transform algorithm.

The Slant transform matrix of order two is given by

$$S_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad \text{.....(2.7.1)}$$

$$S_N = \frac{1}{2^{1/2}} \begin{bmatrix} 1 & 0 & \vdots & & \vdots & 1 & 0 & \vdots & \\ a_N & b_N & \vdots & 0 & \vdots & -a_N & b_N & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & 0 & \vdots & I_{(n/2)-2} & \vdots & & 0 & \vdots & I_{(n/2)-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & \vdots & & \vdots & 0 & -1 & \vdots & \\ -b_N & a_N & \vdots & 0 & \vdots & b_N & a_N & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & 0 & \vdots & I_{(n/2)-2} & \vdots & & 0 & \vdots & -I_{(n/2)-2} \end{bmatrix} \begin{bmatrix} & \vdots & \\ S_{N/2} & \vdots & 0 \\ & \vdots & \\ & \vdots & \\ \cdots & \vdots & \cdots \\ & \vdots & \\ & \vdots & \\ 0 & \vdots & S_{N/2} \\ & \vdots & \end{bmatrix}$$

The matrix $I_{(n/2)-2}$ is the identity matrix of dimension (N/2)-2. The constants a_N, b_N may be computed by the formula

$$a_{2N} = \left(\frac{3N^2}{4N^2 - 1} \right)^{(1/2)}, \qquad b_{2N} = \left(\frac{N^2 - 1}{4N^2 - 1} \right)^{1/2} \qquad \text{.....(2.7.2)}$$

3. FRACTIONAL ENERGY

3.1 Feature Vector Database 'Fractional T-RGB' [4,5]

Frame of Size (NxN)

'T' transformed frame of Size (NxN)

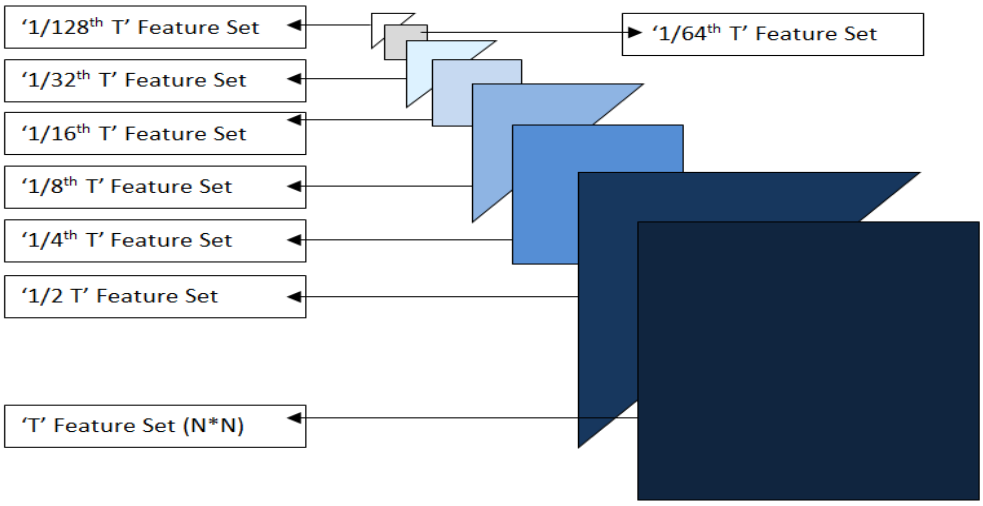

FIGURE 1: Proposed CBVR Techniques using fractional Coefficients of Transformed Frames.

The fractional coefficients of transformed frame as shown in Figure 1, are considered to form 'fractional T' feature vector databases. Here first 50% of coefficients from upper triangular part of feature vector 'T' are considered to prepare the feature vector database '50%' for every frame. Thus DCT, Walsh, Haar, Kekre, DST, Slant, DHT feature databases are used to obtain new feature vector databases as 50%-DCT, 50%-Walsh, 50%-Haar, 50%-Kekre, 50%-DST, 50%-Slant, 50%-DHT respectively. Then for each frame in the database, fractional feature vector set for DCT, Walsh, Haar, Kekre, DST, Slant, DHT using 25%, 12.5%, 6.25%, 3.125%, 1.5625% ,0.7813%, 0.39%, 0.195%, 0.097%, 0.048%, 0.024%, 0.012%, 0.006% and 0.003% of total coefficients are formed.

4. PROPOSED CBVR TECHNIQUES

4.1 Extraction of Key Frames

A Video is read from database. Then, every 10th frame of each video is stored in database as a key frame in RGB color space. Up to 100th frame of every video is stored in the database.

4.2 Feature Extraction for feature vector 'Fractional T-RGB'

Here the feature vector space of the frame of size NxNx3 has NxNx3 number of elements. This is obtained using following steps of T-RGB:

i. Read Videos from Database.
ii. Read every 10th frame of Videos up to 100th frames.
iii. Separate Red, Green and Blue planes of the color frame.
iv. Apply the Transform 'T'[4] on individual color planes of frame.
v. Combine Red, Green and Blue planes of the color frame.
vi. Extract features of color planes of frames.
vii. Concatenate feature of frames of Video.
viii. The result is stored as the complete feature vector 'T-RGB' for the respective video.

4.3 Query Execution

Query Execution is done as follows:

i. Read Query Videos from Database.
ii. Read NxNx3 every 10th frame of Query Videos up to 100th frames.
iii. Separate Red, Green and Blue planes of the color frame.
iv. Apply the Transform 'T'[4] on individual color planes of frame.
v. Combine Red, Green and Blue planes of the color frame.
vi. Extract features of color planes of frames.

vii. Concatenate feature of frames of Query Video.
viii. This feature set is compared with other feature sets in feature database using Mean Squared Distance as similarity measure.

5. IMPLEMENTATION

Database contains 50 videos of each category Mountains, Elephant, Desert, Zoozoo, Ocean, Firework, Dance, Lecture, Obama and Lawn Tennis. Hence, total of 500 videos are stored in the database. Precision and recall are used as statistical comparison parameters[4,5] for the proposed CBVR technique. The standard definitions of these two measures are given by following equations.

$$\Pr ecision = \frac{Number\_of\_relevant\_videos\_retrieved}{Total\_number\_of\_videos\_retrieved}$$

$$\mathrm{Re}\, call = \frac{Number\_of\_relevant\_videos\_retrieved}{Total\_number\_of\_relevant\_videos\_in\_database}$$

6. RESULT AND DISCUSSION

The performance of each proposed CBVR technique is tested by firing 500 queries per technique on the database of 500 videos. Average Mean Square Error distance is used as similarity measure. The average precision and average recall are computed by grouping the number of retrieved videos sorted according to ascending average Mean Square Error distances with the query video. In all transforms, the average precision and average recall values for CBVR using fractional coefficients are higher than CBVR using full set of coefficients.

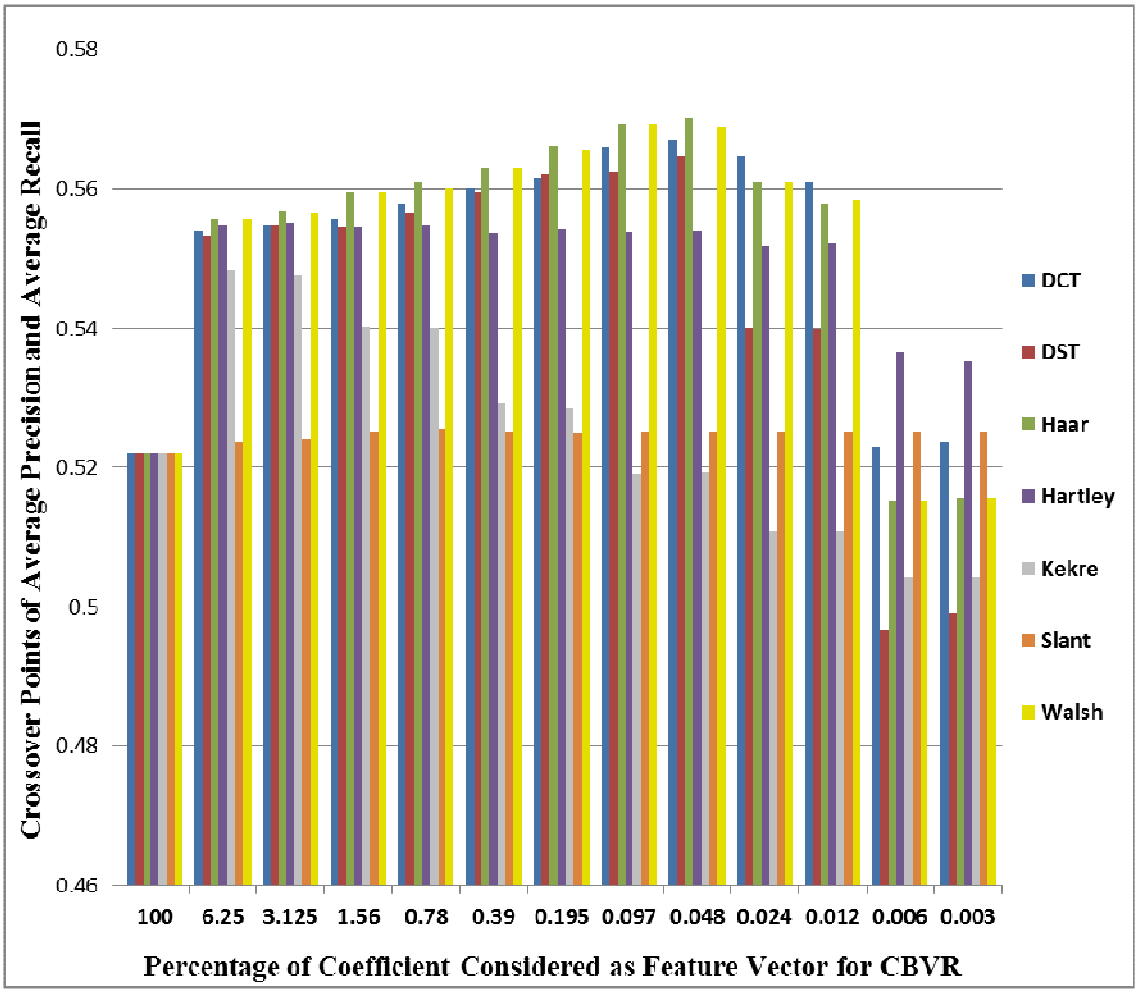

FIGURE 2: Performance Comparison Using Fractional Coefficients across Transforms.

Figure 2 shows performance comparison using Fractional Coefficients across Transforms. As shown in figure 2, Haar transform performs better than all discussed transforms till 0.024% of coefficients. Walsh transform performs second best compared to all discussed transforms till 0.024% of coefficients. DCT Transform performs best result when 0.012% coefficient is considered. But, Hartley Transform performs best result when 0.006% and 0.003% coefficient is considered.

In all, Haar Transform gives highest average precision and average recall showing it outperforms all transforms.

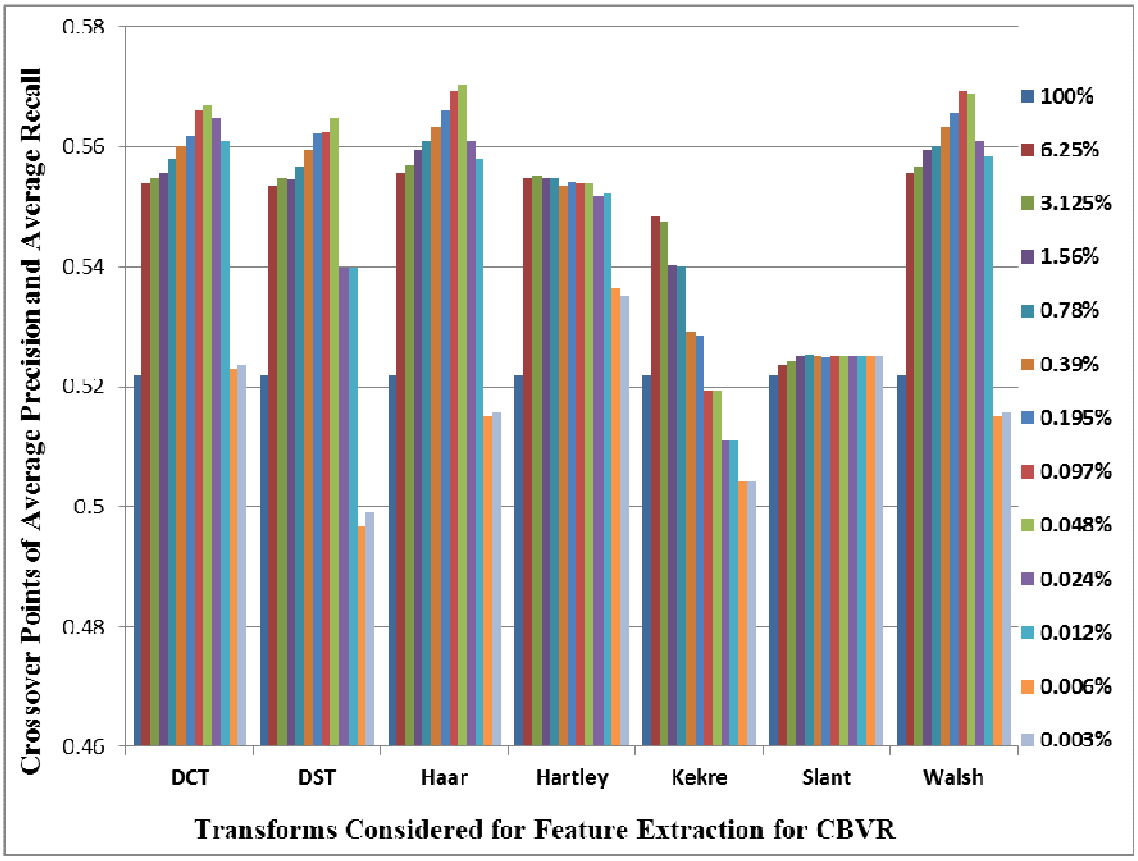

FIGURE 3: Performance Comparison of Transforms w.r.t Percentage of Coefficient Taken into Consideration.

As shown in Figure 3, when 0.048% coefficient is taken in consideration DCT, DST, Haar and Walsh transforms gives best performance as compared to all percentage coefficients taken. DCT, DST, Haar and Walsh transforms shows continuous increase in performance from 100% to 0.048%. Performance of Hartley and Slant does not show much variation with reduction of number of coefficients.

7. COMPARATIVE ANALYSIS

The following are the comparative analysis of the proposed method with other researches:

- In the current trend, every first frame of the shot of a video is taken as a key frame. To compute first frame of every shot of a video, firstly a video is divided into shots. This is done by taking the difference between every two consecutive frames. Whenever the difference is large then, those consecutive frames are kept in separate shot. When a video is divided into shots then, every first frame of the shot is taken as a key frame. This leads to increase in time complexity.

  This is avoided in the proposed method by taking every 10^{th} frame of a video and hence reducing the time complexity.

- In other researches, feature vector is formed by extracting Textures, comparing blocks of frames to get motion feature, color histograms to get color features, which increases computational complexity because size of the feature vector is large.

- In the proposed method, feature vector is formed by taking frames in Transformed domain with variable sizes of coefficients and hence reducing computational and as well time complexity by reducing feature vector size.

- In others researches, only 5 videos are taken into consideration and with only 1 category of video, whereas in the proposed method, results are displayed using 500 videos with 10 different categories.

- In the proposed method, the highest crossover point of average precision and recall is 0.5702 using Haar Transform when 0.048% coefficient is taken into consideration. This means that accuracy is 57.02% with reduction in time complexity by 2048 times compared to full feature set.

8. CONCLUSION

In the information age where the size of video databases is growing exponentially, more precise retrieval techniques are needed, for finding relatively similar videos. Computational complexity and retrieval efficiency are the key objectives in the video retrieval system.

Average Precision and Recall Crossover points is taken as performance index since its values varies between 0 and 1. When precision and recall is 1, then all the videos, similar to the query video are fetched from the database. But, when precision and recall is 0, then none of the videos which are similar to the query video are fetched from the database.

Here, the performance of video retrieval is improved using fractional coefficients of transformed frames of video at reduced computational complexity. In all transforms (DCT, DST, Haar, Hartley, Kekre, Slant and Walsh), the average precision and average recall values for CBVR using fractional coefficients are higher than using full set of coefficients. Hence, the feature vector size for video retrieval could be greatly reduced, which ultimately will result in faster query execution in CBVR with better performance.
Haar Transform performs best compared to all other transforms that is DCT, DST, Haar, Kekre, Slant and Walsh. Crossover points of Precision and Recall of Kekre reduces as size of coefficients reduces. Hartley and Slant does not show any variation as size of coefficients decreases whereas for other transforms increases due to energy compaction. When 0.048% coefficient is taken then DCT, DST, Haar and Walsh transforms give best performance as compared to all percentage coefficients.

In future, wavelets generated from these transforms can be implemented and compare results with the transforms.

9. REFERENCES

[1] B.V.Patel and B.B.meshram, "Content based Video Retrieval Systems", IJU, vol.3, No.2, April 2012.

[2] T.N.Shanmugam and Priya Rajendran, "An Enhanced Content-Based Video Retrieval System Based On Query Clip", International Journal of Research and Reviews in Applied Sciences, ISSN: 2076-734X, EISSN: 2076-7366 ,vol.1, Issue 3(December 2009).

[3] Kalpana Thakre, Archana Rajurkar and Ramchandra Manthalkar, "An effective CBVR system based on motion, quantized color and edge density features", IJCSIT, vol.3, No 2, April 2011.

[4] H.B.Kekre, Sudeep D. Thepade, " Improving the Performance of Image Retrieval using Partial Coefficients of Transformed Image", International Journal of Information Retrieval (IJIR), Serials Publications, vol. 2, Issue 1, 2009, pp. 72-79(ISSN: 0974-6285)

[5] Dr.H.B.Kekre, Dr. Sudeep D. Thepade and Akshay Maloo, "Comprehensive Performance Comparison of Cosine, Walsh, Haar, Kekre, Sine, Slant and Hartley Transforms for CBIR with Fractional Coefficients of Transformed Image", IJIP, vol.5, Issue (3) : 2011

[6] Ahmed, N.; Natarajan, T. ; Rao, K.R. "Discrete Cosine Transform", IEEE TRANSACTIONS ON COMPUTERS, vol.C-23, Issue: 1, pp 90 – 93, Jan. 1974.

[7] R. N. Bracewell, "Discrete Hartley transform," Journal of the Optical Society of America, vol.73, Issue 12, pp 1832-1835, Dec. 1, 1983.

[8] Maurence M. Angush and Ralph R. Martin, ‘‘A Truncation Method for Computing Slant Transforms with Applications to Image Processing’’, IEEE TRANSACTIONS ON COMMUNICATIONS, vol.43, No.6, June 1995

[9] P.Geetha and Vasumathi Narayan, ‘‘A Survey of Content Based Video Retrieval’’, Journal of Computer Science, vol. 4 (6),pp 474-486, 2008

[10] C.V.J Jawahar, Balakrishna Chennupati, Balamanohar Paluri, Nataraj Jammalamadaka, ‘‘Video Retrieval Based on Textual Queries’’, International Conference on Advanced Computing and Communication, 2005