Comprehensive Performance Comparison of Fourier, Walsh, Haar, Sine and Cosine Transforms for Video Retrieval with Partial Coefficients of Transformed Video

Nalini Yadav
Department of Computer Engineering
Pimpri Chinchwad College of Engineering
Savitribai Phule Pune University
Pune, India

Sudeep D. Thepade
Ph.D., Professor Department of Computer Engineering
Pimpri Chinchwad College of Engineering
Savitribai Phule Pune University
Pune, India

ABSTRACT
The desire of better and faster retrieval techniques has always fuelled to the research in content based video retrieval (CBVR). The extended comparison of innovative content based video retrieval (CBVR) techniques based on feature vectors as partial coefficients of transformed video frames using various orthogonal transforms is presented in the paper. Here the popular transforms are considered like Cosine, Walsh, Haar, Sine, and Fourier transforms. The advantage of energy compaction of transforms in higher energy coefficients is taken to reduce the feature vector size per video by taking partial coefficients of transformed video frames. Reduced feature vector size results in less time for comparison of feature vectors resulting in faster retrieval of videos. The features are extracted in eight different ways from the transformed image. First all the coefficients of transformed image considered as 100% energy and then seven reduced coefficients sets are considered as feature vectors (as 99%, 98%, 97%, 96%, 95%, 90% and 85% energy of complete transformed video coefficients). To extract Gray feature sets the five video transforms are applied on gray image equivalents and the color components of videos. Then these seven reduced coefficients sets are used instead of using all coefficients of transformed videos as feature vector for video retrieval, resulting into better performance and lower computations. The video database of 500 video spread across 10 categories is used to test the performance of proposed CBVR techniques. 500 queries are fired on the database to find average accuracy values for all feature sets per transform for each proposed CBVR technique. The results have shown performance improvement (higher accuracy values) with partial coefficients compared to complete 100% energy of transformed of video frames at reduced computations resulting in faster retrieval. Haar transform surpasses all other considered transforms in performance with highest accuracy values with 90% of partial energy coefficients and size is lowered by 99.93% as compared to other transforms.

General Terms
Content based Video Retrieval.

Keywords
Cosine Transform, Haar Transform, Walsh Transform, Fast Fourier Transform, Sine Transform.

1. INTRODUCTION
The computer systems have been posed with large number of challenges to store/transmit and index/manage large numbers of video effectively, which are being generated from many of the sources. Efficient Storage and transmission is taken care by image compression with significant advancements been made [1, 2, 3]. One of the promising research area for researchers from a wide range of disciplines like computer vision, image processing areas is video indexing and retrieval. The desire of better and faster video retrieval techniques is till enticing to the researchers working in some of important applications for CBVR technology.

The goal of every Content based video retrieval system is to index the videos in dataset with efficiency in both aspects of storage, accuracy and retrieval in faster manner. Efficient video retrieval system provides the way to access, update and retrieve the videos in flexible and less complicated manner. The content of the video is matter of interest here in this paper. Video comprises of frames in predefined sequence. Thus at the bottom line frames of the video are under consideration and contents of the same can be extracted for the research purpose. This research aims at efficient and accurate video retrieval using the contents of video. Main focus is on the energy contents of the video. These energy contents are extracted using various transforms.

2. Literature Survey
Video consist of many frames. Video frame is described with contents like color, edge, shape and intensity. Video is rich in contents thus high storage space required to store video. The purpose of Content Based Video Retrieval system is to minimize the signature size required for the video and index it with efficiency. To minimize the space required to store the video, it is necessary to consider only differentiating features of the video. Orthogonal transforms are extensively used to extract the discriminating features of the image [4]. Multiple orthogonal transforms are available out of them Cosine, Haar and Walsh with partial coefficients have shown better performance for Content Based Video Retrieval [5, 6].

2.1 Orthogonal Transforms Used
Various orthogonal transforms are available in literature which includes Sine, Cosine, Walsh, Haar, Kekare, Slant and Fourier. Transforms used in designed content based video retrieval techniques are Sine, Cosine, Walsh, Haar and Fourier, each of this is elaborated in following sections.

2.1.1 Discrete Cosine Transform

Discrete Cosine transform (DCT) is used in many applications of digital signal processing for pattern recognition, information hiding, and content based image retrieval [7]. The two dimensional DCT can be written in terms of pixel values f(x, y) for x, y= 0, 1, … , N-1 and the frequency-domain transform coefficients F(u, v) as shown in equation(1).

$$F(u,v) = \propto (u) \propto (v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos\left[\frac{(2x+1)\mu\pi}{2n}\right] \cos\left[\frac{(2y+1)v\pi}{2n}\right] \quad (1)$$

$$for\ 0 < u,v < n-1$$

$$Where, \propto (u)|\ \propto (v) = \frac{1}{\sqrt{N}}\ for\ u = 0$$

$$\propto (u)|\ \propto (v) = \frac{\sqrt{2}}{N}\ \ for\ 1 < u < N-1$$

$$for\ 1 < v < N-1$$

2.1.2 Discrete Sine Transform

The two dimensional sine transform is defined by an equation (2).In signal and image processing sine transform is widely used [8].

2.1.3 Walsh Transform

Walsh functions were established as a set of normalized orthogonal functions, resemblance to sine and cosine functions, but having uniform values ± 1 throughout their segments [9].Walsh matrix, Wj has following properties.

- Wj takes on the values +1 and -1.
- Wj[0] = 1 for all j.
- Wj x WKT =0, for j ≠ k and Wj x WKT =N, for
- j=k.
- Wj has exactly j zero crossings, for j = 0, 1... N-1.
- Each row Wj is even or odd with respect to its midpoint.

2.1.4 Haar Transform

The family of N Haar[10] functions hk(t), (k = 0, 1,2,3,...,N-1) are defined on the interval 0 ≤ t ≤1. The shape of the specific function *hk* (t), of a given index k depends on two parameters p and q:

$$k = 2p + q - 1 \quad (3)$$

For any value of $k \geq 0$, p and q are uniquely determined so that 2^q is the largest power of 2 contained in $k(2^p < k)$ and q-1 is the remainder q-1 = k – 2^p

- When k = 0, the Haar function is defined as a constant

$$h0(t) = 1/N \quad (4)$$

- When k > 0, the Haar function is defined as

$$h_k(t) = \frac{1}{\sqrt{N}} \begin{cases} 2^{p/2} & when\ \frac{q-1}{2^p} \leq t < (q-0.5)/2^p \\ -2^{p/2} & when\ (q-0.5)/2^p \leq t < q/2^p \\ 0 & \end{cases} \quad (5)$$

From the definition, it can be seen that p states the amplitude and width of the non-zero part of the function, while q determines the position of the non-zero part of the function.

2.1.5 Fourier Transform

The Fourier transform is simply a method of expressing a function in terms of the sum of its projections onto a set of basis functions. The FT is a specific kind of discrete transform, used in Fourier analysis [11, 12, 13]. It transforms one representation into another, which is called the frequency domain representation, or simply the FT, of the original function (which is often a function in the time domain). Discrete input function is required by the Fourier Transform. The discrete transform F of a two dimensional image is calculated using equation 6.

$$F(k,1) = \frac{1}{\sqrt{MN}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f(m,n) e^{-j2\pi\left(\frac{mk}{M}+\frac{nl}{N}\right)} \quad (6)$$

Where *f* (*m*, *n*) is the original image and *F*(*k*,*l*) is the transformed image.

2.2 Partial Energy

Orthogonal Transforms have the property that when they are applied on image then they separate the high energy and low energy regions from each other. High energy contents of the transformed image have the most distinguishing features of the image and low energy contents are the non distinguishing features of the image. Thus this discriminating high energy contents can only be considered as features representing the image which can discriminate one image from another image. When all the coefficients of the transformed image are considered then the feature vector size becomes very huge, but when few of the high energy coefficients are considered to generate the feature vector then the size of feature vector reduces drastically and this leads to lesser number of computations for comparison of feature vectors in retrieval process. Extracting partial energy components for feature vector has proved efficient for image retrieval [14].

Extraction of partial energies from the all energy coefficients is a three step process [15].

a. Generation of average energy matrix
b. Building a summed energy matrix
c. Extracting Partial energy coefficient table

Steps are summarized in a Figure 1.

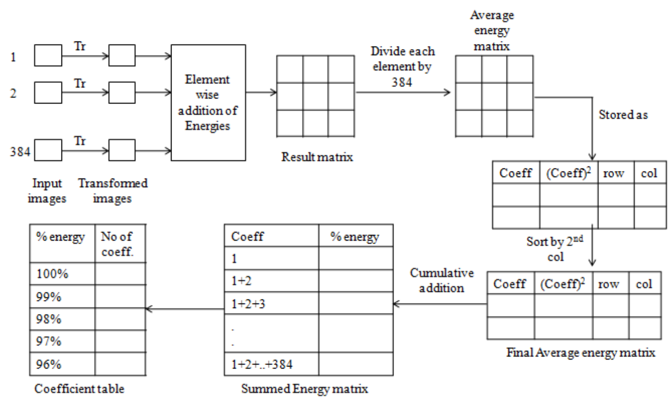

Fig 1: Energy Compaction using partial energy

3. PROPOSED CONTENT BASED VIDEO RETRIEVAL TECHNIQUE

In proposed Content Based Retrieval system videos is represented with partial energy coefficients of transformed video frames.

Content Based Video Retrieval consists of two phases as registration phase and query execution phase.

3.1 Registration Phase

In this phase a feature vector table is built up from transformed visual contents of each video which is to be registered and stored in video database.

Algorithm for registration phase is as below

1. Select a video to be stored.
2. Extract key frames of the respected video. In proposed research every 20^{th} frame is taken as key frame [16, 17].
3. Extract Red, Green and Blue components of each key frame.
4. Apply Transform on individual plane of each key frame.
5. Prepare the feature vector of partial energy coefficients.

Repeat the step 1 to 5 for all videos in database to get the feature vector table

3.2 Query Execution Phase

Algorithm for query execution phase for a given query video is stated as below.

1. Extract key frames of query video. In proposed research every 20^{th} frame is taken as key frame [16, 17].
2. Extract Red, Green and Blue components of each key frame.
3. Apply Transform on individual plane of each key frame.
4. Prepare the feature vector of partial energy coefficients.

Compare the query video feature with feature database using Similarity measure, to get the set of relevant matches from database.

Figure 2 elaborates details about the steps for getting partial energy for a given video from the coefficient table (Figure 1).

1. Decide the percentage of partial energy for feature vector of video.

2. From the coefficient table calculate the number of coefficients required for the given percentage of energy.

3. Extract the corresponding pixel.

4. Prepare the feature vector with the energies from the respective pixel.

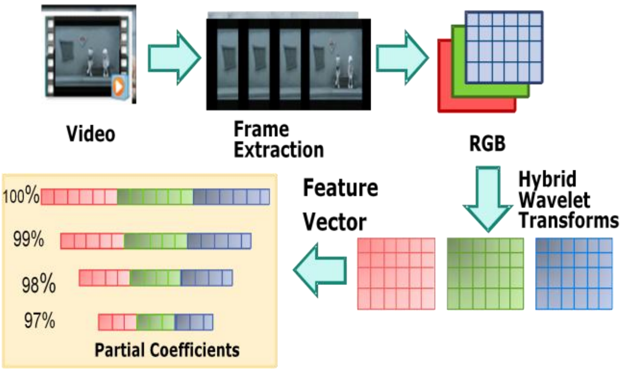

Fig 2: Steps to generate Partial Energy

The proposed technique also explores the different similarity measures for getting higher accuracy of retrieval. In proposed technique query feature vector is compared with feature vectors of database videos with four different similarity measures alias Euclidean Distance, City Block Metric, Sorensen Distance and Kulczynski Distance to evaluate suitability of similarity measure for proposed Content Based Video Retrieval [18].

The proposed research paper aims at increasing the performance of content based retrieval system by reducing the feature vector size with partial energy compaction and increase in accuracy by using efficient similarity measure for matching the query feature vector with database of feature vectors.

To assess the effectiveness of CBVR technique, proposed system has used the average accuracy as statistical comparison parameter. Higher accuracy indicates more accurate method for feature extraction. Accuracy is fraction of retrieved videos that are relevant as given in equation 7.

$$Accuracy = \frac{\{Number\ of\ relevant\ videos\ \ retrived\}}{\{Total\ number\ of\ videos\ retrieved} \quad (7)$$

4. EXPERIMENTATION ENVIRONMENT

In the proposed research, the platform used for experimentation is MATLAB with processor CORE i3.

The experimentation test bed has 500 Videos across 10 categories, each category has 50 videos. Fig. 3 shows the sample from collection of videos considered in data set.

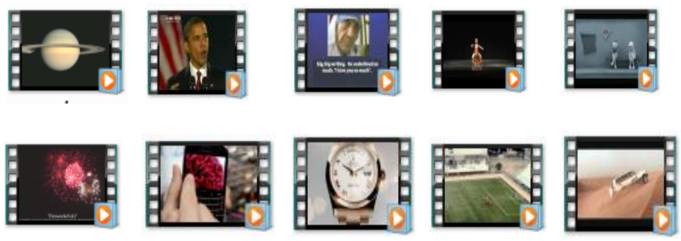

Fig 3: Samples from the Categories of Video Dataset

5. RESULTS AND DISCUSSION

The goal of experimentation of the proposed method is to find out the impact of different orthogonal transforms on Content based Video Retrieval using energy compaction.

The purpose of research work is to reduce the feature vector size by using energy compaction improve the performance of retrieval system with efficient orthogonal transform and similarity measure. The experimentation is carried out with five different orthogonal transform viz. Cosine, Haar, Walsh, Sine and Fourier transform. Energy compaction of transformed visual contents is accomplished with partial energy coefficients viz. 99%, 9%, 97%, 96%, 95%, 90% and 85% of complete energy content. The similarity measures used for comparison of feature vector are Euclidean Distance, City Block Metric, Sorensen Distance and Kulczynski Distance [18].

The experimentation is conducted using video dataset of 500 videos spread across 10 video categories. On the considered database total 500 queries are executed. The average accuracy for four similarity measures - Euclidean Distance, City Block Metric, Sorensen Distance and Kulczynski Distance for five orthogonal transforms – Cosine, Haar, Walsh, Sine and Fourier with 99%, 98%, 97%, 96%, 95%, 90% and 85% partial energy coefficients is considered.

The figure 4 details about the retrieval accuracy with features extracted with Fourier transform. All four similarity measures are used for feature comparison in query execution phase. The figure shows that city block metric outperforms all other three similarity measures.

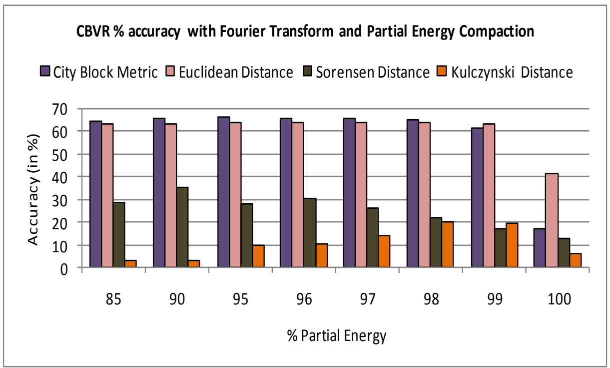

Fig.4. Average accuracy for Fourier Transform for four similarity measures and 99%, 98%, 97%, 96%, 95%, 90% and 85% partial energy coefficients

It is concluded from the figure 4 that city block metric is performing best. Best accuracy for the retrieval purpose is achieved with City Block Metric as a similarity measure for feature comparison. The experimentation with city block metric and Fourier transform with all partial energies is shown in figure 5.

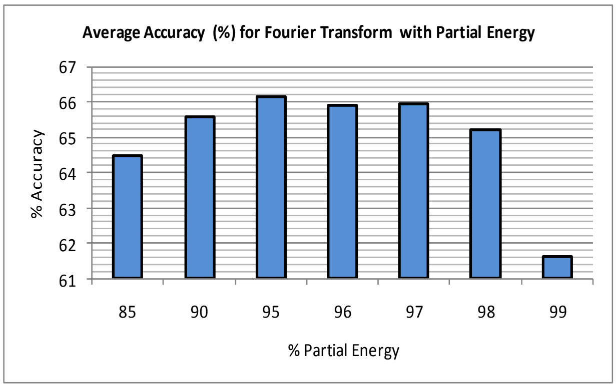

Fig.5. Average accuracy for Fourier Transform with City Block Metric and 99%, 98%, 97%, 96%, 95%, 90% and 85% partial energy coefficients

Figure 5 explains the performance of Fourier Transform with different partial energy for video retrieval. With 95% partial energy coefficients the retrieval accuracy is highest as 66.15%. Thus next experimentation for remaining four transforms is done with city block metric as given in the figure 6 and table I with all partial energies.

Table 1. Average Accuracy Using Five Orthogonal Transforms With Partial Energy

% Partial Energy	Cosine	Haar	Walsh	Fourier	Sine
99	54.38	58.21	50.49	61.62	64.75
98	61.51	58.90	59.64	65.20	65.44
97	63.86	60.63	62.52	65.93	65.68
96	64.60	62.79	64.31	65.91	66.17
95	65.28	64.16	64.57	66.14	66.41
90	65.95	67.36	67.26	65.56	66.32
85	63.53	64.27	64.26	64.48	66.14

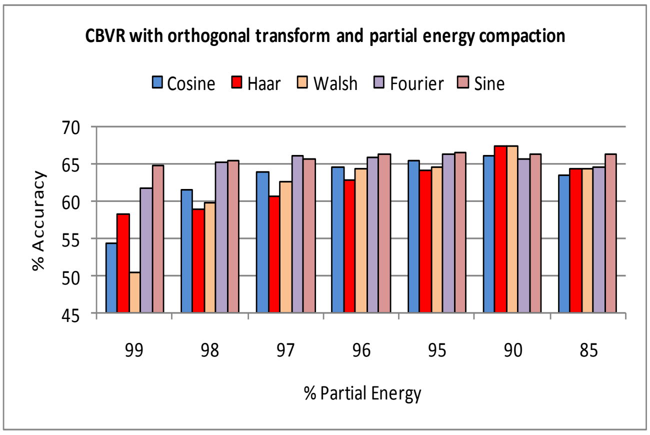

Fig.6. Average accuracy for five orthogonal Transform with City Block Metric and 99%, 98%, 97%, 96%, 95%, 90% and 85% partial energy coefficients

It is observed from the figure 6 and table I, that all transforms performs better at 90% energy coefficients. Comparison of four transforms for 90% energy coefficients is given in figure 7.

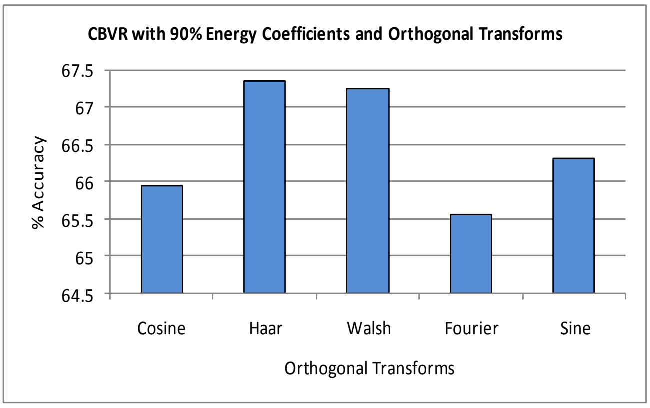

Fig.7. Average accuracy for five orthogonal Transform with 90% partial energy coefficients

Figure 7 concludes that the best performance is given by Haar Transform at 90% energy coefficient with 67.36% of retrieval accuracy. Thus the goal of finding the better retrieval accuracy with less feature vector size the achieved with CBVR with Haar Transform at 90% energy coefficient.

Table II summarizes the performance given by five orthogonal transforms for different energy percentages considered in proposed transformed content based video retrieval.

Table II Summary Of Average Accuracy And Reduction In Feature Vector Size For Different Percentage Of Partial Energy Coefficients

Orthogonal Transform	Energy considered to form feature vector	Number of coefficients considered to form Feature vector	Percentage Reduction in size of Feature vector	Accuracy (In %)
Cosine	90	32	99.51	65.95
Haar	90	43	99.93	67.36
Walsh	90	37	99.94	67.26
Fourier	90	25	99.96	65.56
Sine	90	36	99.95	66.32

From table II, it is can be observed that for the best performing Haar Transform at 90% of partial energy, only 43 coefficients are required compared to 65536 coefficients of 100% energy. Thus there is huge reduction of feature vector size. Hence the goal of reduction in size complexity is achieved through reduction in feature vector size with 90% of partial energy coefficients and highest accuracy of retrieval. When there is reduction in feature vector size then there is less number of comparisons required for matching the query video with database videos which indirectly improves the speed of retrieval. Hence one more aim of CBVR of improving the speed of video retrieval is also achieved with partial energy coefficients.

6. CONCLUSION

Goals of Content Based Video Retrieval system are to increase the efficiency of retrieval system with reduction in feature vector size, increase in retrieval speed and improvising accuracy. The proposed research is focused around these goals. Proposed research evaluated impact of energy compaction with 99%, 98%, 97%, 96%, 95%, 90% and 85% partial energy coefficients on retrieval accuracy and reduction in feature vector size. This paper aims at evaluating the efficiency of the five orthogonal transforms viz. Cosine, Haar, Walsh, Fourier and Sine transforms with Content Based Video Retrieval

Orthogonal transforms are used for Content Based Video Retrieval with partial energy coefficients as 99%, 98%, 97%, 96%, 95%, 90% and 85% and four similarity measures viz. Euclidean Distance, City Block Metric, Sorensen Distance and Kulczynski Distance. After experimentation with dataset of 500 videos, it is observed that Haar Transform is performing best with 90% partial energy coefficients i.e. 99.93% in reduction in feature vector size and 67.36% of accuracy. At 90% of partial energy coefficients Walsh gives second best performance with 67.26% average accuracy with city block metric similarity measure. Thus out of five orthogonal transform Haar and Walsh performs better. With energy compaction, reduction in feature vector size improvises speed thus results into efficient retrieval system..

7. REFERENCES

[1] 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, Volume 2, Issue 1, 2009, pp. 72-79(ISSN: 0974-6285)

[2] Hirata K. and Kato T. “Query by visual example – content-based image retrieval”, In Proc. of Third International Conference on Extending Database Technology, EDBT’92, 1992, pp 56-71

[3] H.B.Kekre, Sudeep D. Thepade, “Rendering Futuristic Image Retrieval System”, National Conference on Enhancements in Computer, Comm. and Information Technology, EC2IT-2009, 20-21 Mar 2009, K.J.Somaiya COE, Vidyavihar, Mumbai-77.

[4] Dr. H. B. Kekre, Dr. Tanuja K. Sarode, Dr. SudeepD.Thepade, Ms.Sonal Shroff, “Instigation of OrthogonalWavelet Transforms using Walsh, Cosine, Hartley, KekreTransforms and their use in ImageCompression”, International Journal of Computer Science and Information Security (IJCSIS), Vol 9, No 6, pp.125-133, June2011

[5] Dr. H.B. Kekre, Dr. Sudeep D. Thepade & Saurabh Gupta, “Content Based Video Retrieval in Transformed Domain using Fractional Coefficients”, International Journal of Image Processing (IJIP), Vol 7, Issue 3, pp 238,274, 2013.

[6] Dr. H. B. Kekre, Dr. Tanuja K., Pratik Bhatia, Sandhya N., “Iris Recognition using Partial Coefficients by applying Discrete CosineTransform, Haar Wavelet and DCT Wavelet Transform”, International Journal of Computer Applications (0975-8887) Volume 32-No.6, October 2011.

[7] Dr. H.B. Kekre, Dr. Tanuja K, ,” Performance Evaluation of DCT, Walsh, Haar and Hartley Transforms on Whole Images and Partial Coefficients in Image Classification”, 2012 International Conference on Communication, Information & Computing Technology (ICCICT), Oct. 19-20, Mumbai, India.

[8] S. A. Martucci, “Symmetric convolution and the discrete sine and cosine transforms,” IEEE Transactions on Signal Processing, Vol. 42,Issue 5, pp. 1038-1051, 1994.

[9] J. L. Walsh, “A Closed Set of Orthogonal Functions,” American Journal of Mathematics, vol. 45, pp. 5-24, 1923 .

[10] Anil K. Jain, “Fundamentals of Digital Processing”, 5th Edition, Page 159.

[11] R. Gonzalez, R. Woods, “Digital Image Processing”, 1st edn.Pearson Education.

[12] S. Jayaraman, S Esakkirajan, T Veerakumar, “Digital Image Processing”, by Mc Graw Hill, 2008.

[13] Anil K.Jain , “ Fundamentals of Digital Image Processing” By Prentice hall ,2006

[14] 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, Volume 2, Issue 1, 2009, pp. 72-79(ISSN: 0974-6285).

[15] Dr. Sudeep Thepade, Pushpa R. Mandal, "Energy compaction based novel Iris recognition techniques using partial energies of transformed iris images with Cosine, Walsh, Haar, Kekre, Hartley Transforms and their Wavelet Transforms.", India Conference (INDICON), 2014 Annual IEEE Conference, Pune, India, pp 1-6, 11-13 Dec. 2014.

[16] Sudeep.D.Thepade, Krishnasagar Subhedarpage, Ankur.A.Mali, Tushar.S.Vaidya, "Performance Augmentation of Video Retrieval using Even-Odd Videos with Multilevel Block Truncation Coding", International Journal of Computer Applications, USA (0975 – 8887) Vol 64, No.9, February 2013.

[17] Sudeep.D.Thepade, Krishnasagar Subhedarpage, Ankur.A.Mali, Tushar.S.Vaidya, "Color Content based Video Retrieval using Block Truncation Coding with Different Color Spaces", International Journal of Computer Applications, USA (0975 – 8887),Vol 64, No.3, February 2013

[18] Sudeep D. Thepade, Nalini Yadav, "Assessment of Similarity Measurement Criteria in Thepade's Sorted Ternary Block Truncation Coding (TSTBTC) for Content Based Video Retrieval", 2015 International Conference on Communication, Information & Computing Technology (ICCICT), Jan. 16-17, Mumbai, India