www.ijaceeonline.com ISSN: 2456 - 3935

**International Journal of Advances in Computer and Electronics Engineering**
Volume 2, Issue 6, June 2017, pp. 26–34

# Comprehensive Survey on Content Based Video Retrieval

Nalini S. Jagtap
Assistant Professor, Computer Engineering, Dr. D. Y. Patil Institute of Engineering, Management and Research, Savitribai Phule Pune University, India
Email: yadav.nalini@gmail.com

**Abstract:** *Digital and networking technologies induced easy generation and storage of multimedia data. Tremendous data getting generated and challenge lays in retrieval this multimedia data efficiently and effectively. Effective choice of extracted features has a major role in content based video retrieval. For a requirement of given application these features are used for selecting, indexing and ranking. Exceptional feature selection also allows the cutback in time and storage costs of the retrieval process. This survey paper reports the interesting features that can be withdrawn from video data for indexing and retrieval along with similarity measurement methods.*

**Keyword:** *Content Based Video Retrieval (CBVR); Feature Extraction; Video Indexing; Video Retrieval*

## 1. INTRODUCTION

Archives are the big asset. History witnessed the importance of archives for knowing the past and motivating the future. Archives are of any form textual or multimedia. In the current digital era, multimedia is heart of elevation in digital era. Current scenario of digital era has enormous multimedia collection. Improvement in digitalized and networking technologies forced in easy generation and storage of multimedia data. Our generation has biggest challenge of efficient access of this enormous multimedia. With all types of multimedia data, a video data is becoming a big anxiety due to its rapid generation and costless availability in market.

Today technological advancement made availability of more and more number of multimedia data like graphics, video, image and other forms of media. Even current generation is using the digital data for the communication, expression of feelings/ emotions, sharing knowledge. These multimedia data is generated from diverse sources like webcam, digital camera, internet, mobiles, graphics tools etc., which has motivated an unexpected technological challenge for technological brains.

Availability of information is not of any use unless and until it can be store/access efficiently. This does not mean only the efficient storage and access of multimedia data but also avoiding the manual process of searching and retrieving the data and making the complete process automatic, powerful. Although the process of storing and transmitting of videos is solved with cheaper storage and high capability networks but the efficient indexing, sorting, filtering, summarizing and retrieval of videos is still under progress.

For retrieval applications meticulous organization of video data is necessary. The skilful storage, indexing and retrieval of video data decide the appropriateness and adoption of search engine from user's perspective. In previous decade video retrieval was managed with help of textual features adhered with video. But the achievement of text based video retrieval system build upon text attached with video which are biased and intuitive with time consuming.

Challenges of traditional techniques of indexing and retrieval of video have guided to elevation of next generation retrieval system based on fundamentally derived features of video like Motion, Color and Speech. Content or features of the video portrays the video better which is not possible with the annotations attached with it. Retrieval systems which works upon the content rather than the metadata of the video is called as Content based retrieval system. Content based video retrieval system works on query video generates the relevant retrieved videos from the corpus.

The ambition of CBVR system is to index the video dataset in more skilled manner for storage aspect and retrieve the same in more accurate and rapid way. Competent retrieval system contributes to access, update and retrieve videos in malleable and less complex ways. Moving sequence of image frames is referred as video. Frame is base element of video and sequence of those images is more important. Videos are summarized in form of key frames. Images differentiating than subsequent image frames are

**Cite this paper:**

Nalini S. Jagtap, "Comprehensive Survey on Content Based Video Retrieval", International Journal of Advances in Computer and Electronics Engineering, Vol. 2, No. 6, pp. 26-34, June 2017.

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 2, Issue 6, June 2017, pp. 26–34

called as Key frames.

### 1.1 Need of Content Based Video Retrieval System

Human perception of the video and thus the image is difficult to automate. Still there were many efforts to do so. One of the efforts for automating a human perception is 'Metadata based image or video retrieval'. In metadata/text based videos/images, the database is represented by adding text describing the content of an image [1, 2, 3, 4, 5]. Based on the resemblance of these text annotations of images, the relevant image from the corpus to that of the query image is found [6, 7, 8].

This annotation based video retrieval has many drawbacks. First drawback is it's a time consuming process. It takes a long to manually analyze the video and then assign a textual metadata with it.

Second drawback of the 'Text based retrieval system' is that it's a subjective process. Every person will have different perceptive about a video/image, thus there can be conflicts about annotating the video by two different persons.

Third major drawback of the system is user querying for the video has to use similar keywords that have been used by annotators. Hence it poses a completion on user to know about annotators thought and replicate the same for query text.

Due to these shortcomings of the 'Text based retrieval system' there is need of Content Based Retrieval System.

## 2. CONTENT BASED VIDEO RETRIEVAL

Video are very rich in the visual content (such as motion, colour, texture and shape), which can be used to exonerate the drawbacks of text based video retrieval system, if can be represented in form of mathematical vectors (feature vectors) accessible by computers. CBVR is middleware between human brains and a computer. The human brain is able to perform complex visual perception, but with limitation of speed in contrary, a computer is capable of restricted visual capabilities at much higher speeds.

In a CBVR, visual video content is represented in form of video features, which are extracted naturally and there is no manual interference, thus omitting the dependency on manual efforts in the feature extraction stage. Computerized feature extrication approaches are computationally extravagant, crucial and domain specific. Thus there is huge scope for curtailing computational complexity making it simplified and generic which gives existence for research in CBVR. Image processing algorithms do help in interpretation and manipulation of visual content of video to generate feature vector out of it. Further image processing domain is researched for making the conversion of visual content of videos in to respective feature vectors in faster, robust, better and more efficient manner with feature vectors being as small as possible.

Block diagram for simplified retrieval procedure is as given in Figure 1.

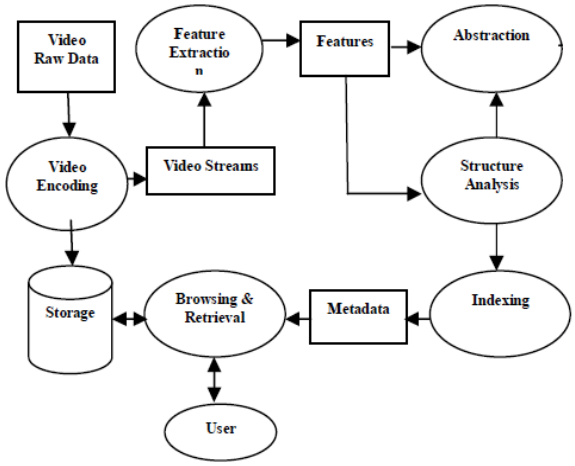


*Figure 1 General Video Retrieval Process [9]*

Typical CBVR systems can organize and retrieve video from video databases, naturally by extracting few features such as texture, color, shape from video and reading for similar videos which have similar feature vectors. Content based video retrieval systems operate in two stages. Feature extraction (FE) is a first stage, a set of features, called feature vector, is generated to meticulously represent the visual content of each video in the database. A feature vector is much miniature in size. In the second stage, similarity measures (SM) searches distance between the query video and each video in the database using their features. Distance computed so that the most relevant similar videos can be retrieved. One problem of this approach is reliance on visual similarity to judge semantic similarity, which creates problems due to semantic gap between low-level features and high level concepts. Even with the subsistence of this problem, if aggressive attempts are made CBVR can be used for real life applications. For example despite of the open problems like robust text understanding, Google and Yahoo have become most popular for searching.

### 2.1 Applications of CBVR

Although some of the intrinsic and core problems in the field are not answered so far, even then CBVR holds enough commitment and maturity to be useful for real world applications. YouTube, what's app has brought the new rebellion in multimedia by means of the video-sharing and distribution.

Now a days, video and image processing techniques are used in multiple applications, such as medicine (to measure and understand some parame-

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 2, Issue 6, June 2017, pp. 26–34

ters of the human body), astronomy (to enhance the quality of images from satellite), compression (to reduce the storage requirement for image database), sports (to seal the motion of an player/sport in order to interpret and improve the performance), rehabilitation (to assess the mobile abilities), motion pictures (to produce special effects based on graphics to action of actors), surveillance (monitor, detect and track persons and vehicles), production industries (to measure the products quality), robotic control (to detect entity and their position so a robot detect them), Televisions (mixing graphics with streaming of live video, e.g., weather forecast), biometrics (to understand some differentiating parameters of a people), photography (improving or adding effects to clicks), etc. Most of these applications make use of the same video and image processing tools and techniques [10].

Novel attempts are made to detect abnormal videos from large web video meta-object database. The Inter Quartaile Range (IQR) outlier detection technique was employed with large scale data, so that abnormal web videos based on their meta-objects are found effectively [11]. Video based traffic surveillance and monitoring systems also requires the efficient analysis of videos. Vehicle detection was divided into two main categories based on vehicle representation, namely, techniques based on motion cues and techniques that employ appearance features [12].

A broad range of applications for CBVR has been described. Some of these are listed below:

#### 2.1.1 Video Search Engines

Search engines have biggest challenge as video search. The task for video search engine is to search the desired video as per requirement of user. The term "search engine" is often used extensively to describe crawler-based search engines, but it's always not same. Each type of "search engine" gathers and ranks listings in different ways. The problem with video search engines is enormous size of videos resulting in higher space consumption and increase in time. Here efficient CBVR technique may help with searching.

#### 2.1.2 Medical Diagnosis

Medical applications have exclusive features (interpreting patient's data, experimenting medical inventions and dealing with complex complicated data) that require the development of particular appliance like for getting expected video clips with special requirement.

The increasing dependence of modernized medicine on diagnostic techniques such as robotic surgical system, histopathology, Magnetic Resonance Imaging, Positron Emission Tomography and computerized tomography has resulted in burst in the number. There is increasing interest for CBVR is resulting into huge scope of research. Current trends of research are of storing and retrieving videos of patient cases for past case history and searching for the specific purpose.

#### 2.2.3 Military

Due to privacy protection military applications of imaging technology are not popular but probably the best-developed. 3d-Radar for applications such as detection of landmines, UXO (UneXploded Ordnance) and IEDs (Improvised Explosive Devices), identification of enemy aircraft from radar screens, detection of targets from satellite photographs, and guidance systems for cruise missiles are examples. Video clip surveillance techniques of crime prevention is also relevant example to the military field which is more focusing on the exact data matching with surveillance query.

## 3. LITERATURE SURVEY

Various sources are getting rise to abandon amount of videos and they are getting stored in the database. Significance gives increased in the endless work of searching a required relevant video. Thereupon it has enlarged a complication of searching process in huge video database. The Content Based Video Retrieval is inspired by searching growing video dataset with potential, effective, speed and more meticulous search method.

Video is a combination of audio and visual data. Visual data of Videos is progression of frames (images) having its own identity of color, shapes, edges and transformed contents. These optical contents are mostly used as features and these features are used in relevant retrieval [13]. Visual contents are usually represented using video segment representation (VSR) and Optimal key frame representation (OFR) [14]. Features are extracted using VSR with help of texture analysis and color moments. These are continued using OFR with occurrence of the pixel intensity values with respect to the pixel location in every image frame of video. These generated features are used in video registration and query execution phase [14] Hidden Markov Model (HMM) is applied for classification which in turn take advantage of speed of color change for each video frame [15] as feature in video retrieval.

### 3.1 Architecture of Video Retrieval System

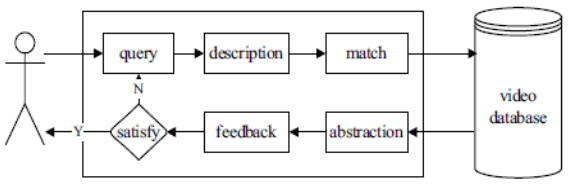


*Figure 2 Architectural framework of retrieval system [16]*

A general video retrieval system usually has modules as: query, description, match, extraction, authen-

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 2, Issue 6, June 2017, pp. 26–34

tication module, as well as video database for storing and collecting video database, video features and indexing information [16]. Figure 1 is shows the architectural framework of a general video retrieval system [17]. Their functions are:

1) *Query module:* It provides the platform for search based on user's different requirements.

2) *Description module:* It converts the user level query into understandable expression and video description.

3) *Match module:* This module matches the input query with database on the basis of features.

4) *Extraction module:* It finds the desired video from the matching videos, and retrieves all the related video automatically for the users.

5) *Authentication module:* This part checks the authenticity of retrieved video with user's requirement. If users are not satisfied, users can update the query condition to re-query. This part can gets users' feedback. It allows the system interact with users and improves retrieval system.

## 3.2 Techniques for Video Retrieval

Content based Video retrieval consists of video content analysis, feature extraction, content designing, indexing and querying [18]. Video instinctively consists of a ranking of units with individual frames at the lower level and at higher level it has segments such as shots, scenes, and episodes. An essential function in understanding video content is to detect segment perimeter.

### 3.2.1 Video Segmentation

A shot is defined a deliberately recorded set of images presenting a continued action in time and space axes by a single camera. Scene is defined as sequence of shots focusing on the same point of interest. Multiple related scenes form an episode [19]. Many camera activities such as tilting, zooming and panning make difficulty in detecting shot changes. Procedures for shot change identification include the following:

- Direct Pixel or Histogram Comparison: In this method pixels of consequent frames are compared. If a multiple pixels differ then a shot change is detected. This is very expensive activity and is prone to minor camera operations like zooming, tilting, etc. A more robust method is histogram comparison. When histograms of two consequent frames differ significantly a shot change is detected [20]. But this method can't handle creeping changes.
- Compressed Domain Features: Compressed video gives added clues such as DCT transform coefficients and motion vectors which is used for cut detection [21]. In MPEG, video compression standard [22] the image is compressed in units of 16 x 16 pixel macroblocks.
- Text Recognition and Closed Captions for Video Indexing: For video indexing newly rising field is using textual information whenever seen. Text available on scenes or images are recognized using Optical character recognition (OCR) [23], [24] is used for segmentation and retrieval.

Using priori knowledge about the scenes, model-based segmentation [25] models of various scenes are built. Using above techniques, firstly the video is divided into different shot. Then the shots are split based on the models

### 3.2.2 Object Detection and Tracking

For video, two aspects are used to retrieve information for detection and tracking of objects: Visual features (such as color and texture) and motion information. A peculiar flow is to segment regions on basis of color and texture information. Then regions with similar motion features can be merged with certain constraints such as adjacency [26], [27].

For distinguishing human faces in compressed video Human skin color and DCT transform coefficients of MPEG, is used [28].

Rules for distinguishing any movements such as walking, playing a scene and attaching/detaching objects using motion vectors are being developed [29]. It is possible to perceive certain facial expressions and body language using models of face or hand movements.

### 3.2.3 Content Modeling, Indexing, and Retrieval

The sensual nature and approximately bigger size of video data requires special browsing and querying functions. For a quick browsing a common approach is followed where main task is to detect shot changes and assign an icon of a key frame for each differentiating shot [30]. Retrieval can also be done on basis of icons, color, text, and image (frame) features.

Ranking of segments such as shots, scenes, and episodes is integral part of hierarchical and compositional model [31]. Very rich set of temporal and spatial operations is a feature of this model promotes querying and composition at different levels. Hierarchical model makes use of Hierarchical Temporal Language (HTL) [32], [33] and uses a video consisting of units such as frames, shots, and subplots.

Another model, an Object-based querying detects and tracks the moving objects and queries provided/selected by the user [34].

The quality of service is an important criterion for

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 2, Issue 6, June 2017, pp. 26–34

the achievement of a video retrieval system, or a Multimedia on Demand (MOD) system. Example of quality of service parameters are delay jitter and the skew between the monomedia streams. Buffering and disk scheduling techniques been proposed and implemented for promising quality of service in such systems various have [35], [36].

### 3.2.4 Color Contents based feature extraction using Block Truncation Coding(BTC) and Thepade's Sorted Ternary Block Truncation Coding(TSTBTC)

Distribution of colors in an image represents Color histogram of image. Color histogram represented as number of components belonging to particular color shade [37]. It outputs the image in frequency domain of shades. For a given image a dimensional layout is called as Color layout. Color layout with color histogram describes the image. Reference image color traits are transformed to destination image using color mapping. For Color mapping color pallets of reference image are matched to destination image and appropriately it will be colored. Thus color contents are vital, dominant and critical part of image and thus the video.

#### *3.2.4.1 Block Truncation Coding (BTC)*

Elevation in digitalization made the professionals and novice users access the digital video data easily. Abundant number of videos are accessible from variety of sources (digital cameras, digital videos, internet, video applications on mobile etc.) has imposed multiple challenges in storage and retrieval of video data.

For color content feature extraction Block Truncation Coding (BTC) is simple and efficient available technique. It's a well proven method for feature extraction in color domain [38]. BTC has played as vital role in antiquity of digital image coding as many leading coding techniques are developed on basis of BTC and they found to be successful [39]. Originally Block Truncation Coding is developed for grayscale images later on it is altered for color images which improves the efficiency of retrieval process as feature vector is boosted [40]. Block truncation process mainly base on dividing the image into multiple blocks. Depending upon count of blocks, BTC is classified into binary and ternary BTC.

Two non-overlapping regions of video frames based on threshold value are for binary BTC. Two values in binary BTC means one for the pixels greater than or equal to the threshold and other for the pixels smaller than the threshold [39, 41]. Two means are calculated for bit maps which gives the feature vector for the image [40].

Three non overlapping regions of video frames rise to ternary BTC. Three differentiating areas of pixels are assembles with help of multimodal characteristic of pixel intensity values [42]. Color BTC makes use of individual color planes. Features are composed of individual Red, Green and Blue planes of an image. Two types of ternary BTC-static or dynamic based on the level decided at runtime. Many compositions for BTC are developed depending on levels and color spaces. Thepade's Sorted Ternary BTC level 2 proves best compared to other variations of BTC [42, 43].

### 3.2.5 Transformed Visual Contents using Orthogonal Transforms

The goal of CBVR system is to miniature the footprint size required for the video and indexing. To curtail the space needed to store the video, it is mandatory to consider only comprehending features of the video. Orthogonal transforms are largely used to excerpt the fastidious features of the image [44]. Many orthogonal transforms have been developed out of them Cosine, Haar and Walsh with partial coefficients has better performance for Content Based Image Retrieval [45].

Assorted orthogonal transforms are accessible in literature like Sine, Haar, Cosine, Walsh, Kekare, Slant and Fourier [46]. Few transforms like Sine, Cosine, Walsh, Haar, Slant and Fourier are amplified in following sections.

#### *3.2.5.1 Discrete Cosine Transform*

Digital signal processing applications has used Discrete Cosine transform (DCT) for pattern recognition, information hiding and content based image retrieval [47]. The DCT can be given 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).

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

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

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

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

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

#### *3.2.5.2 Discrete Sine Transform*

The two dimensional sine transform is given by an Equation (2). Sign transform is widely used in signal and image [48].

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 2, Issue 6, June 2017, pp. 26–34

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

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

*3.2.5.3 Walsh Transform*

Walsh functions are developed as set of normalized orthogonal functions, parallel to sine and cosine functions, but having uniform values ± 1 throughout their segments [49].Walsh matrix, Wj has following properties.

- Wj takes the values +1 and -1.
- Wj[0] set to 1 for all j.
- Wj x WKT =0, for j not equal k and Wj x WKT equal 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.

*3.2.5.4 Haar Transform*

The family of N Haar [50] functions hk(t), (k = 0, 1,2,3,...,N-1) are defined. Interval for Haar Function is $0 \leq t \leq 1$. The value of the specific function $hk$ (t), of a 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 individually determined so that $2^q$ is the highest power of 2 contained in ($2^p < k$) and q-1 is the remainder q-1 = k – $2^p$

When k = 0, the Haar function is gives 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 is seen that p states the amplitude and width of the non-zero part of the function and q decides the position of the non-zero part of the function

.

*3.2.5.5 Fourier Transform*

Expression of a function in terms of the sum of its projections onto a set of basic functions is a Fourier transform. The FT is used in Fourier analysis [51, 52, 53] and it's a specific kind of discrete transform. It transforms into the frequency domain from the original function (which is in the time domain). The discrete Fourier transform F of a two dimensional image is computed 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 $f$ ($m$, $n$) : Original image $F$($k$,$l$) : Transformed image.

*3.2.5.6 Kekre Transform*

Advantage of Kekre Transform is matrix is need not have to be in powers of 2, it can be of any size N x N [54]. 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.

### 3.2.6 Energy Compaction with transformed Visual Content

The important property of Orthogonal Transforms is that when transforms are imposed on image, it distinguishes the high energy and low energy areas from each other. High energy contents of the image have the maximum differentiating features of the image and low energy contents are the non discriminating features of the image. Thus this differentiating high energy contents can only be advised as features representing the image which can separate one image from another image. When all the pixels of the transformed image are used then the feature's size becomes huge, but when few of the high energy coefficients are taken feature vector size reduces excessively and thus lesser number of calculations for comparison of features in retrieval process.

There are two methods for energy compaction viz. fractional energy and partial energy.

*3.2.6.1 Fractional Energy*

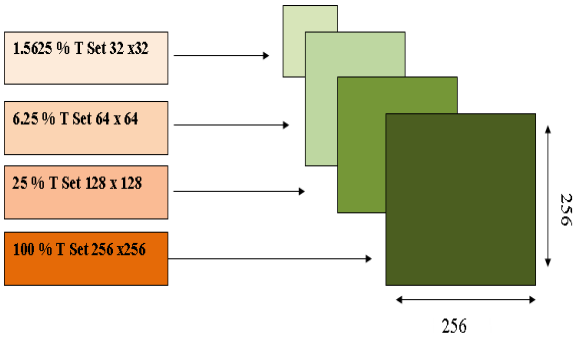


*Figure 3 Feature vector extraction with Fractional energy [55]*

When importance is given to space and speed related issues the fractional energies have shown better

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 2, Issue 6, June 2017, pp. 26–34

performances in transform domain [55]. To get fractional energies video frames are resized into 128 x128 as 25 %, 64 x 64 as 6.25 % and 32 x 32 as 1.5625 % as varying block sizes for energy compaction.

Using these fractional coefficients feature vector database of respective orthogonal transforms is generated. The size of feature vector is N x N for every transform. It can be viewed as shown in Figure 3.

In CBVR varying block sizes of fractional energy coefficients are used with each 'T' transform and the results are compared.

3.2.6.2 *Partial Energy*

For applications of image retrieval partial energy components for feature vector has proved effective and efficient [56]. 3 step process of extraction of partial energies from the all energy coefficients [57] is given as.

a. Create average energy matrix

b. Generate summed energy matrix

c. Indentify Partial energy coefficient table

Steps are summarized in a Figure 4.

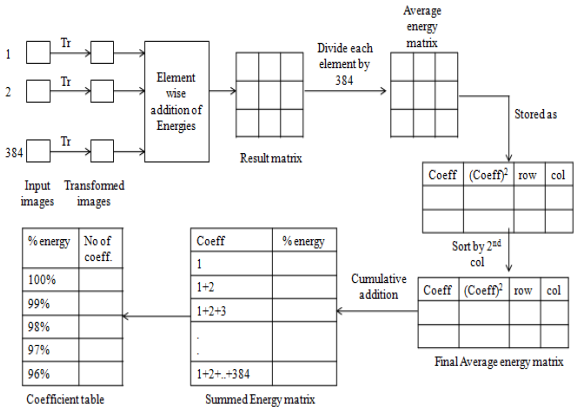


*Figure 4 Energy Compaction using partial energy [57]*

*3.2.6.3 Hybrid Wavelet Transform*

Hybrid wavelet transform is build from the two orthogonal transforms [58]. The hybrid wavelet transform $W_{abxab}$ is created as follows: Consider two matrices I of size axa & J of size bxb.

$$\begin{bmatrix} I_{11} & I_{12} & \dots & I_{1a} \\ I_{21} & I_{22} & \dots & I_{2a} \\ \dots & \dots & \dots & \dots \\ I_{a1} & I_{a2} & \dots & I_{aa} \end{bmatrix} \quad \begin{bmatrix} J_{11} & J_{12} & \dots & J_{1b} \\ J_{21} & J_{22} & \dots & J_{2b} \\ \dots & \dots & \dots & \dots \\ J_{b1} & J_{b2} & \dots & J_{bb} \end{bmatrix}$$

*Figure 5 Two Orthogonal Transforms Iaxa and Jbxb*

A combination of orthogonal matrix I & matrix J generates Hybrid Wavelet Transform ($W_{abxab}$) as 1st a number of rows generated from product of each element from 1st row of matrix I with every column of matrix J. Secondly remaining rows will be filled in circular way with every row of matrix I [59].

$$I_{11} = \begin{pmatrix} J_{11} \\ J_{21} \\ . \\ . \\ J_{b1} \end{pmatrix} \quad I_{12} = \begin{pmatrix} J_{11} \\ J_{21} \\ . \\ . \\ J_{b1} \end{pmatrix} \quad .. \quad I_{1a} = \begin{pmatrix} J_{11} \\ J_{21} \\ . \\ . \\ J_{b1} \end{pmatrix} \quad I_{21} = \begin{pmatrix} J_{12} \\ J_{22} \\ . \\ . \\ J_{b2} \end{pmatrix} \quad I_{22} = \begin{pmatrix} J_{12} \\ J_{22} \\ . \\ . \\ J_{b2} \end{pmatrix} \quad .. \quad I_{3a} = \begin{pmatrix} J_{1b} \\ J_{2b} \\ . \\ . \\ J_{bb} \end{pmatrix}$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $I_{21}$ | $I_{22}$ | .. | $I_{2a}$ | 0 | 0 | .. | 0 |
| 0 | 0 | .. | 0 | $I_{21}$ | $I_{22}$ | 0 | 0 |
| : | : | : | : | : | : | : | : |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | $I_{2a}$ |
| $I_{31}$ | $I_{32}$ | .. | $I_{3a}$ | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | $I_{31}$ | $I_{32}$ | .. | 0 |
| : | : | : | : | : | : | : | : |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | $I_{3a}$ |
| : | : | : | : | : | : | : | : |
| $I_{a1}$ | $I_{a2}$ | .. | $I_{aa}$ | 0 | 0 | .. | 0 |
| 0 | 0 | .. | 0 | $I_{a1}$ | $I_{a2}$ | .. | 0 |
| : | : | : | : | : | : | : | : |
| 0 | 0 | .. | 0 | 0 | 0 | 0 | $I_{aa}$ |

*Figure 6 Creation of Hybrid Wavelet Transform from two orthogonal transforms Iaxa and Jbxb [56]*

## 4. CONCLUSION

Intention of Content Based Video Retrieval system is to increase the attainment of retrieval system with reduction in features, improved in retrieval speed and elevating accuracy.

There is huge contribution by multiple researches with enormous efforts on video retrieval in multimedia databases. Each contribution has its own way of implementation, technology and constraints. With current decade of multimedia data and advancement of technology, videos play crucial role in information transition. Thus, with the provision for storing and processing multimedia databases, relevant and efficient retrieval of those stored videos is always a challenge of the era.

As a survey paper, a corpus of techniques of CBVR been made to deal with a complete and minute review of the most general traditional and modern video retrieval systems from early annotation based manual systems to content based retrieval. A detailed review is attempted on previous research works mainly focusing on the methods/techniques/approaches used to come up to an effective retrieval system with the constraints/challenges and the appraisal mechanisms used. It has been noted that, many singular visual feature base CBVR systems use color. In contrast the hybrid system uses a merging of almost all visual feature types. Thus, currently focus is on video re-

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 2, Issue 6, June 2017, pp. 26–34

trieval for online system, and hence future is awaiting to accommodate those techniques in enterprise levels. On the other side, many methods didn't use the feed-back and self learning scheme. Generally user's requirement in query form may not give desired result due to new guideline used which is out of scope of the knowledge base collections; in such cases it is advised that the system stores the new knowledge/feature/information from the user and later will be used for other consequent queries.

## REFERENCES

[1] David M. Blei and Michael I. Jordan, "Modeling Annotated Data", In Proceedings of The Annual International ACM SIGIR Conference on Research and Development In Information Retrieval, 2003 Issue 1, Vol. 3, pp.127-134.

[2] Chang, Goh, Sychay and Wu, "CBSA: Content-Based Soft Annotation For Multimodal Image Retrieval Using Bayes Point Machines", IEEE Transactions Circuit Systems Video Technology, Volume 13, Number 1, pp. 26–38, 2003.

[3] Datta, Joshi , Li and Wang , "Tagging Over Time:Real-World Image Annotation By Lightweight Meta-Learning", In Proceedings of The ACM International Conference on Multimedia, 2007. R-6

[4] Jeon, Lavrenko, and Manmatha, "Automatic Image Annotation and Retrieval Using Crossmedia Relevance Models", In Proceedings of The Annual International ACM SIGIR Conference on Research and Development In Information Retrieval, 2003, Pages 119-126.

[5] Jin Y., Khan L., Wang L. and Awad M., "Image Annotations By Combining Multiple Evidence and Wordnet", In Proceedings of The ACM International Conference on Multimedia, 2005, Pages 37-47.

[6] Jing, Li, Zhang and Zhang, "A Unified Framework For Image Retrieval Using Keyword and Visual Features", IEEE Transaction Image Processing, Volume 14, Number 7, pp.978–989, 2005.

[7] Li and Wang, "Real-Time Computer ized Annotation of Pictures", IEEE Transactio on Pattern Analysis and Machine Intelligence, Volume 30, Number 6, pp. 985–1002, 2008.

[8] Abdolraheem Khader Alhassan, Ali Ahmed Alfaki. "Color and texture fusion-based method for content-based Image Retrieval", 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), 2017, 10.1109/ICCCCEE.2017.7867649.

[9] Dr. Sudeep Thepade , Ajay Narvekar , Ameya Nawale, "Color Content Based Video Retrieval Using Discrete Cosine Transform Applied On Rows and Columns of Video Frames with RGB Color Space", International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 11, May 2013.

[10] Dr. Kekre, Dr. Sudeep Thepade, Akshay Maloo, "Comprehensive Performance Comparison of Cosine, Walsh, Haar, Kekre, Sine, Slant and Hartley Transforms for CBIR with Fractional Coefficients of Transformed Image", International Journal of Image Processing (IJIP), Volume (5) : Issue (3) : 2011 336-351.

[11] Siddu P. Algur, Prashant Bhat, (2016) "Abnormal Web Video Detection Using IQR Method", International Journal of Advances in Computer and Electronics Engineering, Volume l, Issue 2, June 2016, pp. 31-36.

[12] Ranjeethkumar Chandran, Naveen Raman, "A Review on Video-Based Techniques for Vehicle Detection, Tracking and Behavior Understanding", International Journal of Advances in Computer and Electronics Engineering, Volume 2, Issue 5, pp. 07 – 13, May 2017.

[13] Patel and Meshram, "Content Based Video Retrieval Systems", International Journal of UbiComp (IJU), vol. 3, no. 2, pp. 13-30, April 2012

[14] Padmakala, "An effective content based video retrieval utilizing texture, color and optimal key frame features", Image Information Processing (ICIIP), 2011 International Conference, pp.1-6, Date 3-5 November 2011.

[15] Hanna, "HMM based classification of sports videos using color feature", Intelligent Systems (IS), 2012 6th IEEE International Conference, pp. 388-390, Date 6-8 September 2012

[16] Fang Haiguang, Wen Zhenyu, Liu Jing, Jiao Baocong, "Research on the retrieval process of teaching video resources based on content", Proceedings of IC-BNMT2009

[17] Luo Guobing. Research on the video retrieval algorithm based on content [D]. Wuhan:Huazhong University of Science and Technology, 2006.

[18] Alp Aslandogan and Clement Yu, Senior Member, IEEE, "Techniques and Systems for Image and Video Retrieval", IEEE Transactions on Knowledge and Data Engineering, Vol. 11, No. 1, January/February 1999

[19] Weiss, Duda, and Gifford, "Composition and Search with a Video Algebra," IEEE Multimedia, pp. 12-25, 1995.

[20] Wactlar, Kanade, Smith, and Stevens, "Intelligent Access to Digital Video: Informedia Project," Computer, pp. 46- 52, May 1996.

[21] Kobla, Doermann, and K.-I. Lin, "Archiving, Indexing, and Retrieval of Video in the Compressed Domain," Multimedia Storage and Archival Systems, pp. 78-89, SPIE, Nov. 1996.

[22] D. Le Gall, "MPEG: A Video Compression Standard for Multimedia Applications," Comm. ACM, pp. 46-58, no. 34, 1991.

[23] Lienhart, "Automatic Text Recognition for Video Indexing," Proc. ACM Multimedia Conf., pp. 11-20, Boston, 1996.

[24] Mohan, "Text-Based Search of TV News Stories," Multimedia Storage and Archival Systems, pp. 2-13, SPIE, Nov. 1996.

[25] Zhang, Tan, Smoliar, and Gong, "Automatic Parsing and Indexing of News Video," Multimedia Systems, vol. 2, pp. 256-266, 1995.

[26] Zhang, Kankanhalli, and Smoliar, "Automatic Partitioning of Full Motion Video," Multimedia Systems, vol. 1, no. 1, pp. 10-28, 1993.

[27] Jain, Martin, and Aggarwal, "Segmentation Through the Detection of Changes Due to Motion," Computer Graphics and Image Processing, vol. 11, pp. 13-34, 1979.

[28] Chang, Smith, and Wang, "Automatic Feature Extraction and Indexing for Content-Based Visual Query," Columbia Univ.,Technical Report No. CU/CTR 414-95-20, Jan. 1995.

[29] Courtney, "Automatic, Object-Based Indexing for Assisted Analysis of Video Data," Proc. ACM Multimedia Conf., Boston, pp. 423-424, 1996.

[30] Nagasaka and Tanaka, "Automatic Video Indexing and Full Video Search for Object Appearances," Proc. Conf. Visual Database Systems, pp. 119-133, 1991.

[31] Weiss, Duda, and Gifford, "Composition and Search with a Video Algebra," IEEE Multimedia, pp. 12-25, 1995.

[32] Sistla, Venkatasubrahmanian, "Similarity Based Retrieval of Videos," Proc. IEEE ICDE, Birmingham, United Kingdom, pp. 181-190, 1997.

[33] Yu and Meng, "Principles of Database Query Processing for Advanced Applications, Data Management Systems, Morgan Kaufmann, 1998.

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 2, Issue 6, June 2017, pp. 26–34

[34] Cascia and Ardizzone, "JACOB: Just A Content-Based Query System for Video Databases," Proc. ICASSP, Atlanta, 1996.

[35] Rangan and Vin, "Efficient Storage Techniques for Digital Continuous Media," IEEE Trans. Knowledge and Data Eng., vol. 5, no. 4, pp. 564-573, 1993.

[36] Krunz and Tripathi, "Impact of Video Scheduling on Bandwidth Allocation for Multiplexed MPEG Streams," Multimedia Systems, vol. 5, pp. 347-357, ACM/Springer, 1997.

[37] Thepade Sudeep and Nalini Yadav, "Novel Efficient Content Based Video Retrieval Method Using Cosine-Haar Hybrid Wavelet Transform with Energy Compaction", 2015 International Conference on Computing Communication Control and Automation, 2015.

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

[39] Guoping Qiu, "Color Image Indexing Using BTC", IEEE Transactions on Image Processing, Volume 12, No. 1, pp.93-101, January 2003.

[40] Dr. Kekre, Sudeep Thepade, "Boosting Block Truncation Coding with Kekre's LUV Color Space for Image Retrieval". International Journal of Electrical, Computer, and Systems Engineering, Volume 2, No. 3, pp. 172-180, 2008

[41] Dr. Kekre, Ms. Swapna Borde, "Content Based Image Retrieval," National Conference on Applications of Digital Signal Processing, January 19-20, 2007.

[42] Sudeep Thepade, Ankur Mali, Krishnasagar Subhedarpage, "Content Based Video Retrieval using Thepade's Ternary Block Truncation Coding and Thepade's Sorted Ternary Block Truncation Coding with various Color Spaces", International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), Volume 8, No. 6, pp. 462-466, March-May, 2014.

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

[44] Dr. Kekre, Dr. Tanuja Sarode, Dr. Sudeep 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

[45] Dr. Kekre, Dr. Sudeep 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.

[46] Thepade Sudeep, and 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), 2015, 10.1109/ICCICT.2015.7045728.

[47] Dr. Kekre, Dr. Tanuja, "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, Pages: 1 - 6, DOI: 10.1109/ICCICT.2012.6398176.

[48] Martucci, "Symmetric convolution and the discrete sine and cosine transforms," IEEE Transactions on Signal Processing, Vol. 42,Issue 5, pp. 1038-1051, 1994.

[49] Walsh, "A Closed Set of Orthogonal Functions," American Journal of Mathematics, vol. 45, pp. 5-24, 1923

[50] Anil Jain, "Fundamentals of Digital Processing", 5th Edition, Page 159.

[51] Gonzalez, Woods, "Digital Image Processing", 1st Edn. Pearson Education.

[52] Jayaraman, Esakkirajan, T Veerakumar, "Digital Image Processing", by Mc Graw Hill, 2008.

[53] Anil Jain , " Fundamentals of Digital Image Processing" By Prentice hall ,2006

[54] Dr. Kekre, Sudeep Thepade, Athawale, Shah, Verlekar and Shirke, "Kekre Transform over Row Mean, Column Mean and Both Using Image Tiling for Image Retrieval" International Journal of Computer and Electrical Engineering, vol. 2(6),pp. 964-971, 2010.

[55] Tonge, Ashvini, and Sudeep Thepade. "Key frame extraction for video content summarization using orthogonal transforms and fractional energy coefficients", 2015 International Conference on Information Processing (ICIP), 2015.

[56] Dr. Kekre, Sudeep 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).

[57] Dr. Sudeep Thepade, Pushpa 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.

[58] Nalini Yadav, Sudeep Thepade. "Content Based Video Retrieval using Self Mutated Hybrid Wavelet Transforms with assorted Color Spaces", International Conference & Workshop on Electronics & Telecommunication Engineering (ICWET 2016), 10.1049/cp.2016.1150

[59] Dr. Kekre, Dr. Tanuja Sarode, Dr. Sudeep 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.

## Authors Biography

**Mrs. Nalini S. Jagtap,** is a Assistant Professor of Department of CSE in Dr. D.Y. Patil Institute of Engineering, Management and Research, SPPU University. She completed her M.E. in department at SPPU University. Her research interests are Image processing and Data Mining.

**Cite this paper:**

Nalini S. Jagtap, "Comprehensive Survey on Content Based Video Retrieval", International Journal of Advances in Computer and Electronics Engineering, Vol. 2, No. 6, pp. 26-34, June 2017