Content based Video Retrieval Systems - Methods, Techniques, Trends and Challenges

Aasif Ansari
College of Computers and IT
Taif University
Taif, Saudi Arabia

Muzammil H Mohammed
College of Computers and IT
Taif University
Taif, Saudi Arabia

ABSTRACT
Content Based Video Retrieval (CBVR) has been increasingly used to describe the process of retrieving desired videos from a large collection on the basis of features that are extracted from the videos. The extracted features are used to index, classify and retrieve desired and relevant videos while filtering out undesired ones. Videos can be represented by their audio, texts, faces and objects in their frames. An individual video possesses unique motion features, color histograms, motion histograms, text features, audio features, features extracted from faces and objects existing in its frames. Videos containing useful information and occupying significant space in the databases are under-utilized unless CBVR systems capable of retrieving desired videos by sharply selecting relevant while filtering out undesired videos exist. Results have shown performance improvement (higher precision and recall values) when features suitable to particular types of videos are utilized wisely. Various combinations of these features can also be used to achieve desired performance. In this paper a complex and wide area of CBVR and CBVR systems has been presented in a comprehensive and simple way. Processes at different stages in CBVR systems are described in a systematic way. Types of features, their combinations and their utilization methods, techniques and algorithms are also shown. Various querying methods, some of the features like GLCM, Gabor Magnitude, algorithm to obtain similarity like Kullback-Leibler distance method and Relevance Feedback Method are discussed.

Keyword
CBVR,GLCM, Gabor Magnitude, Kullback-Leibler Distance Method, Relevance Feedback Method.

1. INTRODUCTION
In today’s electronic world huge amount of useful digital information like images, audio and video data apart from textual data exists online and is available to public, government authorities, professionals and researchers very easily and accessible at reasonably cheaper cost due to rapid growth in availability of user friendly and cheaper multimedia acquisition devices at a very large scale like high resolution camera in mobile phones, handy cams and other advanced digital devices, availability of high capacity storage devices like memory cards, hard disks, etc., large scale usage of internet by rapidly growing number of applications used by digital devices to upload huge amount of multimedia information, advanced web technology and internet infrastructure [6], [7]. Video data possesses a lot of information for those using multimedia systems and applications like digital libraries, publications, education, broadcasting and entertainment. Such applications are useful only when video retrieval systems are efficient enough to retrieve videos and other important information from large databases as quick as possible [2]. However, it is extremely challenging for the existing web search engines to search for video over the web so novel methodologies are required that are capable of manipulating the video information according to the content [13]. For multimedia mining, combinations of multimedia data are stored and arranged using techniques like classification and annotation of videos [6], [15], [16]. Most of the web based video retrieval systems work by indexing and searching videos based on texts associated with them but this technique does not perform well because the texts do not contain enough information of the videos [2]. Since video retrieval is not effective using conventional query-by-text retrieval technique, Content Based Video Retrieval (CBVR) is considered as one of the best practical solutions for better retrieval quality [6]. Due to exploitation of rich video content, there is a tremendous scope in area of video retrieval to enhance the performance of conventional search engines [7]. This is leading the area of CBVR into a direction promising to create more effective video search engines in future [12].

In section 2 Processes and components of CBVR systems are elaborated; section 3 shows the methodology to obtain results in CBVR systems. Different types of CBVR systems are given in section 4, problems and challenges posed to information retrieval and CBVR systems are discussed in section 5 and the conclusion is presented in section 6.

2. CONTENT BASED VIDEO RETRIEVAL SYSTEMS PROCESSES AND COMPONENTS

2.1 Formation of a Video
A shot is a set of frames captured by a camera continuously and a clip is the occurrence of such consecutive shots. Consecutive shots showing different students walking in different colleges of a university campus forms a clip of a campus [2].

2.2 Segmentation of Video
The first step in most of existing content based video analysis techniques is to perform segmentation of video into elementary shots. These shots contain a sequence of frames recorded one after another to form a video event or scene continuously varying in time as well as space. These are organized and edited with cut transitions or gradual variation of visual effects forming a video scene or sequence during video sorting [7]. Therefore, process of video segmentation is nothing but converting a video into various smaller video clips representing different scenes where each scene is decomposed again into different shots containing large number of frames in each shot. Features are extracted from these components of video and are then exploited to store, classify, index and retrieve videos from huge databases.

2.3 Classification of Videos

Classification of videos helps to increase efficiency of video retrieval and it is one of the most important tasks [1]. During process of Video classification [24], [25] information is obtained from features extracted out of the video components, videos are then, placed in categories defined earlier. Information including visual and motion features of various components of video like objects, shots and scenes is obtained [1]. Most of the classification techniques are either semantic content classification or non-semantic content classification. The most suitable one is employed as per the type of a video and application and thus, video can be classified to the most suitable and closest among all pre-defined categories. Semantic video classification can be performed at three levels of a video. Video genres, video events and objects in the video [26]. Video genres based classification is to classify videos into one of the pre-defined categories of videos. These categories of videos are kinds of videos commonly exist like videos of sports, news, cartoons, movies, wildlife, documentary movies, etc. Video genres based classification has better and broader detection capability while objects and events have narrow detection range [26]. Event based video classification is based on event detection in a video data and to classify it into one of the pre-defined categories. An event is said to be occurred if it has significant and visible video content. A video can have many events and each event has sub-events. One of the most important steps in content based video classification is to classify events of a video [17]. Shots are most elementary component of a video [7]. Classification of shots determines classification of videos. Shots are classified using features of objects in shots [19]. Different kinds of video features, motion, color, texture and edge for every shot are extracted for video retrieval [7]. Image retrieval methods and techniques can be used for key frame based video retrieval systems [1]. Low level visual features of key-frames are exploited for this purpose [9]. In key-frame based retrieval, as a video is abstracted and represented by features of its key-frames, indexing methods of image database can be applied to shot indexing. Each shot and all its key-frames are linked to each other. For a video retrieval, a shot is searched by identifying its key-frame [3], [4]. Computational cost involved while using all frames of a shot to retrieve a video is much higher than that when only key frames are used to represent a shot. Visual features of these key frames are compared with those of the videos in the database for retrieval [2]. Key-frames are also employed in face [11] and object based video retrieval. A large number of CBVR systems among the existing ones are working with key-frames. Key-frames can deliver a lot of useful information for retrieval purpose and if required, static features of key-frames [20] can also be used to measure video similarity along with motion features [22] and object features [21]. Object based video classification is based on object detection in video data [18]. Faces and texts are also used as a method to classify videos. Four types of TV programs are classified by method proposed by Dimitrova et al. [23]. Faces and Texts are detected and then tracked to each frame of video segment. Frames are labeled for a specific type according to respective faces and texts. An HMM [14] (hidden markov model) is trained to classify each type of frame using their labels. The appearance of textual information while streaming of video frames enables making an automated video retrieval system [10] based on texts appearing in consecutive frames. Video classification using objects such as faces and texts work only in specific environment and this classification for video indexing has the limitation that they are not generic. Object based video classification usually shows poor performance [1].

2.4 Query of a Video

Queries using objects, sketches or example images do not utilize semantic information [1].

Query by Object: The object image is provided. The occurrences of objects in video database are detected and locations of the object determine success of the query [18].

Query by Text: As it is popular for content based image retrieval, example images can be used as query to retrieve relevant videos in a database of videos (query by example) but it has a limitation that motion information of the video being searched is not utilized. It relies only on the appearance information. Also, finding video clip for the interested concept may become too complex using example image. Textual query offers more natural interface and claims to be better approach for querying in video databases [10].

Query by Example: Query by example is better if visual features of the query are used for content based video retrieval [2]. Low level features are obtained from key frames [9] of the query video and then they are compared to separate out the similar videos using their key frames visual features [1].

Query by shot: Some systems utilize the entire video shot as the query instead of key frames [5]. This can be a better option but with a higher computational cost.

Query by clip: A clip can be used for better performance of video retrieval as compared to the technique when a shot is used because a shot do not represents sufficient information about the whole context. All the clips which possess a significant similarity or relevancy with the query clip are retrieved [2].

Query by Faces and Texts: Faces and Texts can also be used as a query to retrieve a video segment containing frames labeled for a specific type according to faces and texts [23]. A suitable algorithm can be used to search the video enquired by the query clip using information obtained from faces and texts in frames of the query clip.

2.5 Features and Features Extraction

For effective video indexing, classification and retrieval visual features embedded in video data is exploited. Three primary features to be extracted are color, texture and motion for effective video indexing. These features are represented by color histogram, Gabor texture features and motion histogram respectively [5]. The most useful information in the videos includes features of the objects, key frames and the motion features [1]. **Key Frame Features:** Key frames in videos contain color, texture and shape based static features. Texture, color and shape are most significant visual properties and are elementary concerns in low level image and computer vision problems. Various color features are color moments, color histograms [75], color correlograms [76] and the color features obtained from some Gaussian models [1]. Different color features are extracted for different types of color spaces such as RGB, HSV, YCbCr and normalized r-g, YUV, and HVC. They play one of the most important roles for video indexing and retrieval. These features are extracted directly from an image or sometimes from sub blocks [77] of the partitioned image. Texture alone is a complicated research problem. It represents an area by roughness, directionality, repeatability and variability features over a certain spatial extent while color is a point property in an image [7]. Texture features are extracted by finding energy distribution in frequency domain by different techniques [39], [40], [41]. Gabor wavelet features are obtained using one such technique to retrieve and classify images and videos [42]. Texture based features are features representing unique occurrence pattern of objects, homogeneity and organization of different objects of various shapes and their own features, independent of intensity and color, with varying background and their correlations with neighboring visual characteristics. Different texture features are orientation features, wavelet transformation based texture features, Tamura features, co-occurrence matrices, simultaneous autoregressive models, etc., [1]. Tamura features are six texture based features corresponding to human visual perception: coarseness, contrast, directionality, line-likeness, regularity, and roughness. The first three features are significant for human perception and are responsible to distinguish different textures [80]. A co-occurrence matrix is a matrix or distribution of co-occurring values for an image [81]. It represents texture in images. The matrix elements are the counts of the number of times a given feature occurs in a particular spatial relation to another given feature [82]. A co-occurrence matrix can use any of the features from the image. GLCM is the co-occurrence matrix when grey level is chosen as a feature. The GLCM is a tabulation of how often different combinations of pixel grey levels occur in an image. An example to find GLCM of a matrix of fig. 1 having grey values 0,1,2,3 are shown here

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Fig 1: Matrix

and its GLCM is shown in fig. 2

Neighbour pixel values / Reference pixel values	0	1	2	3
0	2	2	1	0
1	0	2	0	0
2	0	0	3	1
3	0	0	0	1

Fig 2: GLCM of the matrix of fig. 1

Texture features can be utilized effectively for video retrieval purpose [1]. Hauptmann et al. [38] use Gabor wavelet filters to obtain texture features for video search engine. They design 12 oriented energy filters. A texture feature vector is formed with the mean and variance of the filtered outputs. The image is divided into small blocks and Gabor filter is used to obtain features from these blocks [47]. Hauptmann et al. [46] divide the image into blocks each of size 5 x 5 and compute texture features from each block using Gabor-wavelet filters. Gabor texture features have shown better performance than other texture features [43]. Object shapes and their features are obtained from edges and regional features of various objects using histogram [1]. An Edge Histogram Descriptor (EHD) is designed [78], [79] by dividing an image into 4x4 blocks (16 sub-images). The spatial distribution of edges is obtained and then, categorized into five different orientations of 0, 45, 90, 135 degrees and a 'non-directional' edge in each block. The EHD is the number of pixels forming an edge of a particular category. The output EHD is a 5 bin histogram for each block, getting a total of 80 (5x16) histogram bins. **Motion Features:** The characteristic of dynamic videos that distinguishes them from still images is the motion of objects and motion of background against each other. The foreground motion is caused by moving objects whereas the background motion is caused by camera motion. Visual content with temporal variation is represented by motion features. Tracking of moving object (motion detection) is important in video retrieval systems. It involves separating and finding which pixels belong to moving objects and the pixels belonging to static background over a period of time [83]. The difference between a video and an image is the motion as motion features carry semantic concepts as compared to object and key frame features in an image [1]. Video motion is of two types, background motion and foreground motion caused by camera motion and object's motion respectively. Accordingly, two types of motion features are available. Camera based motion features include features caused by zooming in or out, panning left or right and tilting up or down by camera. Object based motion features are more important as they are able to describe motions of key objects. Motion features are used to classify shots and are employed for shot boundary detection using cuts, gradual and no change frames [84], [85], [86]. Motion features are also employed to obtain key frames by dividing a shot into segments with equal cumulative motion activity using MPEG-7 motion activity descriptor. Key frame is the frame located in the middle of each segment [87]. A triangle model of motion energy for motion patterns in videos was proposed [88] where frames at the turning points of the motion acceleration and motion deceleration are selected as key frames. Motion is the essential visual feature carrying temporal variation of video. The correlation between frame sequences within a video shot is among the motion features. Motion information of a video is obtained by two dimensional motion histogram of the motion vectors and the color histogram [2]. The displacement in horizontal and vertical directions are quantized into 121 bins each (60 bins for positive, 60 for negative and one for zero). Totally, there are 121 x 121 bins for this 2-D motion histogram. Motion vectors are obtained between consecutive frames of MPEG-I video stream. In MPEG video, each frame is partitioned into blocks each of size 16 x 16 pixels called macro blocks (MB). Motion vector is defined as the displacement of the target MB (current frame) from the prediction MB (reference frame). In MPEG format there are I, P and B frames. I frames are not used for motion information. P frames contain forward motion prediction and B frames contain both forward and backward motion prediction. Motion histogram is formed using motion vectors present in P frames. Their average value is obtained for elimination of noise effects by normalizing them using number of frames in a shot [2]. **Object Features:** Objects are represented using features of texture, color and trajectory of the objects [19]. Object features used for object based video retrieval are the color, size, texture features of the regions inside the objects [1]. They can be used to retrieve videos likely to contain similar objects [34]. Faces are also used to retrieve videos as objects in many video retrieval systems. For example, Sivic et al. [35] construct retrieval system of a person that is able to retrieve shots containing that person, given a query face in a shot. Shots are ranked as per the similarity measure. Le et al. [36] propose a method to retrieve faces in broadcast news videos by integrating temporal information into facial intensity information. Texts can also be used as objects and contribute along with faces for video retrieval. Li and

Doermann [37] implement text-based video indexing and retrieval by expanding the semantics of a query and using the Glimpse matching method o perform approximate matching instead of exact matching. Limitation of object based features is that lots of time is consumed for searching and identifying the objects in the videos [1]. Broadly varying types of features are employed by large number of methods to represent [7], classify, enquire and retrieve videos. Among them, most popularly used features [7] are text analysis [30], shape information [28], color histogram [27] and motion activity [29]. A combination of different types of features i.e., object features [21], static features of key frames [32], and motion features [22] can be used to find similar video when demanded by user [1]. Edge histogram and texture features are one of the most reliable data for effective video retrieval application. Textural properties of texts are distinct and distinguish them from its background in the image. This can be exploited by texture based methods to retrieve texts from images. Texture features of the region in an image containing texts can be obtained by techniques using Fourier Transform, spatial variance, Wavelet transform and Gabor filters [10]. **Extraction of Gabor Features:** Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and a specific direction. Expanding a signal using this basis provides a localized frequency description, therefore capturing local features/energy of the signal. Texture features can then be extracted from this group of energy distributions. The scale (frequency) and orientation tunable property of Gabor filter makes it especially useful for texture analysis. The filters of a Gabor filter bank are designed to detect different frequencies and orientations. They can be used to extract features on key points detected by interest operators [72]. From each filtered image, Gabor features can be calculated and used to retrieve images. The algorithm for extracting the Gabor feature vector is shown in fig. 3 and the related equations (1 - 4)are also shown below [73], [89]. For a given image I(x,y), the discrete Gabor wavelet transform is given by a convolution:

$$W_{mn} = \sum_{x_1}\sum_{y_1} I(x_1, y_1) g_{mn} * (x - x_1, y - y_1) \quad - (1)$$

where $*$ indicates complex conjugate and m, n specify the scale and orientations of wavelet respectively. After applying Gabor filters on the image with different orientation at different scale, an array of magnitudes is obtained:

$$E(m,n) = \sum_{x}\sum_{y} |W_{mn}(x,y)| \quad - \quad (2)$$

These magnitudes represent the energy content at different scale and orientation of the image. The main purpose of texture-based retrieval is to find images or regions with similar texture.

The standard deviation σ of the magnitude of the transformed coefficients is:

$$\sigma_{mn} = \sqrt{\frac{\sum_x \sum_y (|W_{mn}(x,y)| - \mu_{mn})^2}{P\ X\ Q}} \quad - \quad (3)$$

Where μ is the mean of magnitude and given as $\mu_{mn} = \frac{E(m,n)}{P\ X\ Q}$

A feature vector f (texture representation) is created using mn as the feature components [74], [68]. M scales and N orientations are used and the feature vector is given in equation (4)

$$f = [\sigma_{00}, \sigma_{01}, \sigma_{02} \ldots \sigma_{(M-1)(N-1)}] \ - (4)$$

$f_{Gabor} = \frac{f - \mu}{\sigma}$ where μ is the mean and σ is the standard deviation of *f*.

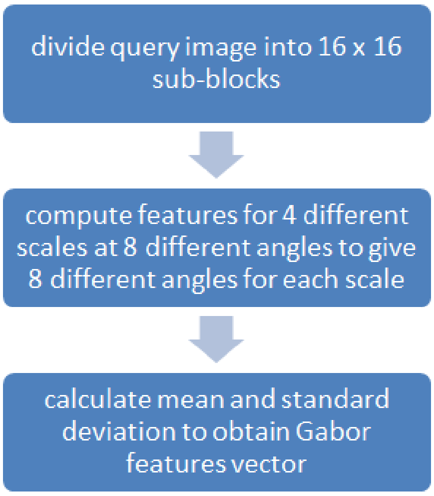

Fig 3: Gabor Filter Algorithm

2.6 Similarity Measure

Queries are classified by categories sorted out according to type of features used or type of example data. The query is found out by calculating similarity between feature vector [44], [45] stored in the database and the query features. The similarity is obtained with the enquired still image, still images from example video clip, objects, texts or a particular face from still images or video clip, motion features from example video [11]. Image similarity matching for example based image retrieval has been studied for many years. The image search engine finds an image from a database with the help of similarity between feature vectors through a distance between them. Usually Euclidean distance is measured to find similarity. Similar images are ranked as per the distance between the query image and images from database. Kullback-Leibler distance method is also employed for the similarity measure between query features and the features from the feature library [7]. Types of features determine the performance of video retrieval system. Once features are generated performance can be enhanced with better results from similarity measure by knowing more accurately about deciding how much close or far is the retrieved result. Euclidean distance and Minkowski type distances are extensively used [7]. Video retrieval result depends greatly on video similarity measures. The videos are retrieved by measuring similarity between the query video and videos from the database. The similarity can be obtained by matching their features, texts, objects, faces, etc. and their combinations. Measuring similarity by matching features is most convenient and direct method [1]. It is measured by the average distance between features of corresponding frames [48]. In query by example similarity measure to find relevant videos usually low level feature matching is used. Video similarity can be measured at different levels of resolution or granularity [49]. A video clip is retrieved by finding key frames occurring sequentially in the video database which are similar to that of the query video [2]. A query frame can also be given to a system to retrieve similar videos from the database. The distance metric is termed as similarity measure whereas in conventional retrieval system, the Euclidean distance between the query and database is calculated to rank the retrieved videos. The video from the database corresponding to the frame similar to the query frame is higher in rank if the Euclidean distance is smaller [4], [10]. The equation for Euclidean distance between the query image Q and an image P is shown in equation(5)

$$\mathrm{ED} = \sum_{i=1}^{n} \sqrt{(V_{pi} - V_{qi}).(V_{pi} - V_{qi})} \quad - \ (5)$$

Where, Vpi and Vqi are the feature vectors of Query image Q and image P respectively of size 'n'. Apart from Euclidean Distance, there are many other methods to measure feature distance between two images like Manhattan distance, the Mahalanobis Distance, Earth Mover's Distance (EMD) and the chord distance [33]. Kullback and Leibler determined similarity measure based on two probability distributions associated with the same experiment [31] i.e., same event space. Kullback-Leibler divergence measure is used to find the difference between two distinct probability distributions [7]. The equation for KL divergence of the probability distributions F, G on a finite set P is given in equation (6).

$$D_{KL}(\mathrm{F}||\mathrm{G}) = \sum_{p \in P} F(p) \log \frac{F(p)}{G(p)} \quad - (6)$$

Below are the steps for Similarity Measure: Let us consider - F as Query clip feature vector, G as Feature library 1st feature vector, i as Element of vector, M as Normalized factor of G

$$V = \frac{F}{Normalization\ (F)} \quad - (7)$$

Then find ((G>0)& (V> 0)) and store that in V_A.

Then similarity measure is carried out using equation (8)

$$D_{KL} = \sum V(\mathrm{V}_A) \log \frac{M * V(\mathrm{V}_A)}{G(\mathrm{V}_A)} \quad - \ (8)$$

Neural Network can also be used to find similar shots. It is used to cluster shots and hence classify videos to the best matching cluster based on features obtained from its shots. The features of color, texture and trajectory of objects in a shot are used to map the shot to the best matching cluster [19] in object-based query. Similarity between the query image I_G and an image I in the video database is obtained by probability of generating the image I given the observation of the query image I_G [1].

3. RESULT EVALUATION

The performance of video retrieval is evaluated with the same parameters as it is evaluated in image retrieval [47]. Recall and precision are the two parameters [2] as given in equations (9) and (10).

$$Recall = \frac{DC}{DB} \quad - (9)$$

$$Precision = \frac{DC}{DT} \quad - (10)$$

$DC = number\ of\ similar\ clips\ detected\ correctly$

$DB = \ number\ of\ similar\ clips\ in\ the\ detabase$

$DT = total\ number\ of\ detected\ clips$

4. CONTENT BASED VIDEO RETRIEVAL SYSTEMS

Content based video retrieval techniques are widely distributed among two types. One of them is comparison of frames and their corresponding features within two clips. A set of frames is obtained which are sequentially matching which helps in the retrieval of videos. This method is simple but the computational cost depends upon the features size and is very high [3], [67], [68], [69], [50], [51], [52], [53], [54], [55], [56]. In addition with that, these techniques have a drawback of synchronization between frames as different clips may have used different rate to encode them. To overcome the drawback of the above techniques a key frame is used to represent an entire shot. Shot matching is done and hence video retrieval is achieved by comparing their features. Drawback of techniques employing key frames matching is that temporal information and the related information between the key frames in a shot is lost. Finding a suitable key frame is difficult to select [57], [58], [59], [60], [61], [62], [63], [64], [65], [66]. To strike a balance between the efficiency and computational cost, more visual features are used from the frames to represent a shot [2]. It is learnt from the evaluation of video information retrieval that good image retrieval leads to good performance of video retrieval system when query is an image or an image from the query video [11]. A large number of approaches have been experimented for indexing, classification and retrieval of videos from huge video databases. The video content is represented by spatial and temporal characteristics of videos. In spatial domain, features are obtained from frames to form feature vectors from different parts of the frames. In temporal domain, video is segmented into its elements like frames, shots, scenes and video clips and features like histograms, moments, textures and motion vectors represent the information content of these video segments [10]. A typical methodology is used in system proposed where a video is retrieved based on a query clip [7]. Here, database is processed offline. They used 2-D correlation coefficient technique along with discrete cosine transform, mean and standard deviation over video sequences for segmentation of videos from database into elementary shots. Each video shot is represented by four types of features. Color, texture, edge and motion feature which is the feature representing temporal information of videos. These features from the query clip are compared with features in the database. Kullback-Leibler method is used to measure similarity. Video sequences are ranked according to the distance measures and similar videos are retrieved. As mentioned above, clip based retrieval yields better results than that when only key frames representing a shot is used. So, it is better to use entire video shot instead of key frames as the query [5]. Broadcast news video database has vast information. The presence of textual captions with audio and video information makes this system an effective textual based automated retrieval system which provides vital information access through retrieving news videos [10]. Face detection is assessed for image and video analysis. It was experimented in a commercial system [70]. It was found that accuracy of face recognition in video collection of the type mentioned in the system [11] was too poor to prove to be useful. Overall a large number of queries do not yield satisfactory results as mentioned [11] about one third of the queries were unanswerable by any of the automatic systems participating in the video retrieval track[71]. No system or method was able to provide relevant results. An integrated video retrieval system is proposed [2] where a video shot is represented not by key frame only but by all frames to extract more visual features of a shot. Color and motion features are integrated to fully exploit the spatio-temporal information contained in a video.

A process flow of a typical CBVR system is shown in fig. 4. A video component i.e., frames, shots or scenes, etc. are extracted from videos and then classified to pre-defined categories. Classification to these categories is done manually. Features are then extracted for each component and stored in features database. Features of the same component from the query video are also extracted and then compared with features stored in the database. The output video is obtained by finding the similarity measure between features of query video and the features stored in the database.

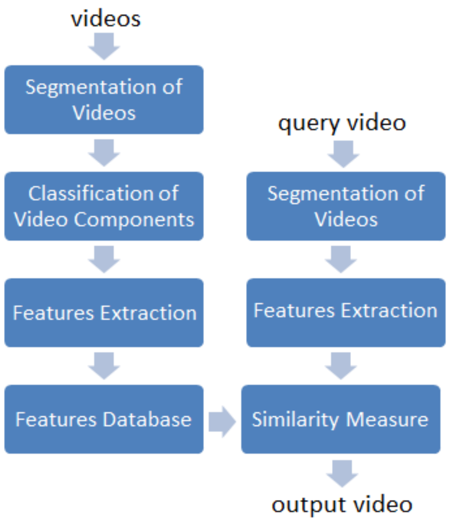

Fig 4: CBVR system

To improve the retrieval performance, relevance feedback technique can be used to resemble human visual judgment and similarity perception up to a certain extent. Systems using relevance feedback are effective in ranking and retrieving similar videos. It removes the difference between low level features and semantic concept of the videos [1]. It relies on feedback obtained by user or can be automatic and accordingly the videos are ranked. The ranking and the feedback is used to improve further searches. A relevance feedback system retrieves initial results by using conventional techniques like query by example image, etc. then, the user will provide feedback to the system regarding relevancy of the retrieved result with the query. The feedback will help to improve the retrieval quality. It is a compromise between a fully automated, unsupervised system and system based on user's feedback because a machine learning algorithm can be used to learn the user's feedback [8]. As it is not easy to fill the gap between low level features and high level concepts for every type of query, video retrieval based on this mapping is difficult. Also, more human involvement yield different results under different circumstances. To tackle these issues a relevance feedback which adjusts its weight according to user's feedback iteratively to fill the gap so that high level concepts can be represented by low level features. Relevance feedback is used in the system [2]. The result is obtained by updating the values of M_u and updating of M_u is done by method shown below.

$$M_u = \begin{cases} M_u + Score_v & \text{if } S_v^u \in S \\ M_u + 0 & \text{otherwise} \end{cases}$$

$$v = 1,2,\dots,L$$

$$u = x,y$$

Weights M_x and M_y are updated using user's feedback. Let S be the set containing the most similar L retrieved video clips, overall similarity value H_v and value of M_x and M_y is 0.5

$$S = [S_1, S_2, \dots\dots, S_L]$$

$$Score = [Score_1, Score_2, \dots\dots, Score_L]$$

be the set containing scores by relevance feedback by the user for each retrieved clips in set S. The scores may have any of the values from -3, -1, 0, +1, +3. Where these values correspond to the feedback as

+3 → highly relevant

+1 → relevant

0 → no opinion

-1 → non-relevant

-3 → highly non-relevant

M_x and M_y are the sets containing the most similar L clips to the query, according to only the color similarity measure and only the motion similarity measure, respectively.

$$S^x = [S_1^x, S_2^x, \dots, S_L^x]$$

$$S^c = [S_1^c, S_2^c, \dots, S_L^c]$$

Weights of M_u are updated using the value of score provided by the user as a feedback. Weights of M_u are more for the more relevant retrieved clips. The weights are then normalized by the total weights to make sum of the normalized weight equal to 1 and if the weight of M_u is < 0, then it is set to 0. The system can be iterated to improve the result for a satisfaction level. As a result, a particular feature representation will represent the semantic concept of the query video.

5. PROBLEMS AND CHALLENGES

With lack of satisfaction from textual based video retrieval, the idea of content based video retrieval has been the attention for researchers since long time. In the beginning of content based video retrieval, they tried to retrieve videos using an image. However, video retrieval using query by image is not successful as it cannot represent a video. A video is a sequence of images and audio. A query video provides rich content information than that provided by a query image. Finding the relevant video by sequentially comparing the low level visual features of key frames of the query video with those of key frames of videos in database provide long pending solution to yield better result[9] of video retrieval. Finding similarity measure requires key frames matching and hence computing key frame features including color histogram, texture and edge features, etc. to calculate distance parameter. These huge computations cause long response time to the users and thus, the problem of high computation cost in computing visual features of videos is persistent. Apart from this, considerations for motion features, temporal, sequence and duration of shots in a video pose a challenge for the research area[6]. The structural and content attributes obtained through content analysis, segmentation, video parsing, abstraction processes and the attributes entered manually are referred to as metadata. Video is indexed on a table using the metadata using clustering process that categorizes video clips or shots. Clustering process categorizes video clips or shots using metadata to form an index table of videos into different visual categories.

Researchers have developed various tools and schemes to index, enquire, browse, search and retrieve videos from large databases but effective and robust tools are still lacking to test with large databases [9]. Due to these limitations [6], [9] a majority of video searches and retrievals still relies on keyword or text attributions.

6. CONCLUSION

It can be concluded from discussions in the previous sections that using a complete video shot yields better result than that using a key frame representing a shot whereas, system using a query clip is superior than that using a single shot instead. Search based on textual information of the video can also be used in CBVR systems. Query by example image is popular for content based image retrieval. Extending this approach for video retrieval has a limitation that motion information of the video is not exploited but only visual information is used. Textual query becomes an option for video retrieval as it provides more natural interface but the result obtained is very poor. An integrated video retrieval system in which video components are represented by more visual features, color and motion features are integrated to fully exploit the spatio-temporal information contained in a video and hence show better results. Automatic retrieval systems should be the focus and it requires more attention from researchers for improved retrieval results. A trend to reduce computational cost is needed to project commercialized systems for video indexing, classification and retrieval to facilitate the availability of low cost, fast and efficient CBVR systems. Capability of these systems can be magnified by reaching huge video databases that exist and are accessible on the web. The accessible databases should empower the users with options to accurately select the desired videos only while filtering out the relevant but undesired as well as irrelevant videos so that valuable, moral, ethical and informative data becomes accessible efficiently, quickly and at low cost.

7. REFERENCES

[1] Weiming Hu, Nianhua Xie, Li Li, Xianglin Zeng, Maybank S., "A Survey on Visual Content-Based Video Indexing and Retrieval", IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41-6,797-819, 11/2011

[2] Liang-Hua Chen, Kuo-Hao Chin, Hong-Yuan Liao, "An integrated approach to video retrieval", Proceedings of the nineteenth conference on Australasian database-Volume 75, 49–55, 2008

[3] Hong Jiang Zhang, Jianhua Wu, Di Zhong, Stephen W. Smoliar, "An integrated system for content-based video retrieval and browsing", Pattern Recognition, Pattern Recognition Society, Published by Elsevier Science Ltd., Vol. 30, No. 4, pp. 643~658, 1997

[4] B V Patel, B B Meshram, "Content Based Video Retrieval Systems", International Journal of UbiComp, vol 3, No. 2, pg 13-30, 2012

[5] Yining Deng, B.S. Manjunath, "Content-based Search of Video Using Color, Texture, and Motion", IEEE, pg 534-537, 1997

[6] Ja-Hwung Su, Yu-Ting Huang, Hsin-Ho Yeh, Vincent S. Tseng , "Expert Systems with Applications", 37, pg 5068-5085, 2010

[7] T.N.Shanmugham, Priya Rajendran, "An Enhanced Content-Based Video Retrieval System Based on Query Clip", International Journal of Research and Reviews in Applied Sciences, Volume 1, Issue 3, 2009

[8] Ritendra Datta, Dhiraj Joshi, Jia Li, James Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age", ACM Computing Surveys, Vol. 40, No. 2, article 5, pg 1-60, 2008

[9] Nicu Sebe, Michael S. Lew, Arnold W.M. Smeulders, "Video retrieval and summarization", Computer Vision and Image Understanding, vol. 92, no. 2-3, pg 141-146, 2003

[10] C. V. Jawahar, Balakrishna Chennupati, Balamanohar Paluri, Nataraj Jammalamadaka, "Video Retrieval Based on Textual Queries", Proceedings of the Thirteenth InternationalConference on Advanced Computing and Communications, Coimbatore, Citeseer, 2005

[11] Alexander G. Hauptmann, Rong Jin, and Tobun D. Ng, "Video Retrieval using Speech and Image Information", Electronic Imaging Conference (EI'03), Storage Retrieval forMultimedia Databases, Santa Clara, CA, January 20-24, 2003.

[12] Steven C.H. Hoi, Michael R. Lyu, "A multimodal and multilevel ranking frame work for content-based video retrieval", ICASSP, 2007

[13] M. Petkovic, W. Jonker, "Content-Based Video Retrieval by Integrating Spatio-Temporal and Stochastic Recognition of Events", Proceedings of IEEE Workshop on Detection and Recognition of Events in Video, pp. 75-82, 2001.

[14] Ara V. Nefian, Monson H. Hayes III, "Hidden Markov Models for Face Recognition", IEEE International Conference on Acoustics, Speech and Signal Processing, 1998

[15] V.S. Tseng, J-H Su, J.-H. Huang, & C-J. Chen,"Integrated mining of visual features, speech features and frequent patterns for semantic video annotation." IEEE Transactions on Multimedia, 10(1), 2008

[16] Virga, P., Duygulu, P., "Systematic evaluation of machine translation methods for image and video annotation", In Proceedings of the fourth international conference on image and video retrieval (pp. 487–496), Singapore, 2005.

[17] G. Lavee, E. Rivlin, and M. Rudzsky, "Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in video," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 39, no. 5, pp. 489–504, Sep. 2009.

[18] J. Sivic and A. Zisserman, "Video Google: Efficient visual search of videos," in Toward Category-Level Object Recognition.. Berlin, Germany: Springer, pp. 127–144, 2006.

[19] G. Y. Hong, B. Fong, and A. Fong, "An intelligent video categorization engine," Kybernetes, vol. 34, no. 6, pp. 784–802, 2005.

[20] Y. Wu, Y. T. Zhuang, and Y. H. Pan, "Content-based video similarity model," in Proc. ACM Int. Conf. Multimedia, pp. 465–467, 2000.

[21] A. Anjulan and N.Canagarajah, "Aunified framework for object retrieval and mining," IEEE Trans. Circuits

Syst. Video Technol., vol. 19, no. 1, pp. 63–76, Jan. 2009.

[22] W. M. Hu, D. Xie, Z. Y. Fu, W. R. Zeng, and S. Maybank, "Semantic based surveillance video retrieval," IEEE Trans. Image Process., vol. 16, no. 4, pp. 1168–1181, Apr. 2007.

[23] N. Dimitrova, L. Agnihotri, and G. Wei, "Video classification based on HMMusing text and faces," in Proc. Eur. Signal Process. Conf., Tampere, Finland,pp. 1373–1376, 2000

[24] M. Roach, J.Mason, L.-Q. Xu, and F. Stentiford, "Recent trends in video analysis: A taxonomy of video classification problems," in Proc. Int. Assoc. Sci. Technol. Develop. Int. Conf. Internet Multimedia Syst. Appl., Honolulu, HI, pp. 348–354, Aug. 2002.

[25] D. Brezeale and D. J. Cook, "Automatic video classification: A survey of the literature," IEEE Trans. Syst., Man, Cybern., C, Appl. Rev., vol. 38, no. 3, pp. 416–430, May 2008.

[26] Y. Yuan, "Research on video classification and retrieval," Ph.D. dissertation, School Electron. Inf. Eng., Xi'an Jiaotong Univ., Xi'an, China, pp. 5–27, 2003.

[27] A. M. Ferman, A. M. Tekalp, and R. Mehrotra, "Robust color histogram descriptors for video segment retrieval and identification," IEEE Transactions on Image Processing, Vol. 11, No. 5, pp 497-508, 2002.

[28] B. Erol, and F. Kossentini, "Shape-based retrieval of video objects," IEEE, Trans. on Multimedia, Vol. 7, No. 1, pp 179-182, 2005.

[29] C.W. Ngo, T.C. Pong, H.J. Zhang, "Motion-based video representation for scene change detection," Int. Journal Computer Vision, pp 127-142, 2002.

[30] L. Chen and T.S. Chua, "A match and tiling approach to content-based video retrieval," Proc. ICME, pp. 301-304, 2001.

[31] J. Meng, Y. Juan, S.F. Chang, "Scene Change Detection in a MPEG Compressed Video Sequence", SPIE Symposium on Electronic Imaging: Science and Technology - Digital Video Compression: Algorithms and Technologies, SPIE Vol. 2419, San Jose, Feb. 1995.

[32] Y. Wu, Y. T. Zhuang, and Y. H. Pan, "Content-based video similarity model," in Proc. ACM Int. Conf. Multimedia, pp. 465–467, 2000.

[33] H.B.Kekre, Sudeep D. Thepade, "Boosting Block Truncation Coding using Kekre's LUV Color Space for Image Retrieval", WASET International Journal of Electrical, Computer and System Engineering (IJECSE), Volume 2, Number 3, pp. 172- 180, Summer 2008.

[34] R. Visser, N. Sebe, and E. M. Bakker, "Object recognition for video retrieval," in Proc. Int. Conf. Image Video Retrieval, London, U.K., pp. 262–270, Jul. 2002.

[35] J. Sivic, M. Everingham, and A. Zisserman, "Person spotting: Video shot retrieval for face sets," in Proc. Int. Conf. Image Video Retrieval, pp. 226–236, Jul. 2005.

[36] D.-D. Le, S. Satoh, and M. E. Houle, "Face retrieval in broadcasting news video by fusing temporal and intensity information," in Proc. Int. Conf. Image Video Retrieval, (Lect. Notes Comput. Sci.), 4071, pp. 391–400, Jul. 2006.

[37] H. P. Li and D. Doermann, "Video indexing and retrieval based on recognized text," in Proc. IEEE Workshop Multimedia Signal Process., pp. 245–248, Dec. 2002.

[38] A. G. Hauptmann, R. Baron, M. Y. Chen, M. Christel, P. Duygulu, C. Huang, R. Jin, W. H. Lin, T. Ng, N. Moraveji, N. Papernick, C. Snoek, G. Tzanetakis, J. Yang, R. Yan, and H. Wactlar, "Informedia at TRECVID 2003: Analyzing and searching broadcast news video," in Proc. TREC Video Retrieval Eval., Gaithersburg, MD, 2003. Available: http://www-nlpir.nist.gov/projects/tvpubs/tvpapers03/ cmu.final.paper.pdf

[39] J. Zhang, M. Marszałek, S. Lazebnik, C. Schmid, "Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study", International Journal of Computer Vision, vol.73 no.2, pp.213-238, June 2007

[40] Engin Avci, "An expert system based on Wavelet Neural Network-Adaptive Norm Entropy for scale invariant texture classification", Expert Systems with Applications: An International Journal, vol.32 no.3, pp.919-926, April, 2007

[41] S. Arivazhagan, L. Ganesan, "Texture segmentation using wavelet transform", Pattern Recognition Letters, vol.24 no.16, pp.3197-3203, December 2003.

[42] Jeff E. Tandianus, Andrias Chandra, Jesse S. Jin, "Video Cataloguing and Browsing", Proceedings of the Pan-Sydney area workshop on Visual information processing, vol. 11, pp. 39 - 45, 2001.

[43] B.S.Manjunath and W.Y.Ma, "Texture features for browsing and retrieval of image data", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 18, no. 8, pp 837-842, Aug. 1996.

[44] Swain M.J. and Ballard, B.H. "Color Indexing," Int'l J. Computer Vision, vol. 7, no. 1, pp. 11-32, 1991.

[45] Hafner, J. Sawhney, H.S. Equitz, W. Flickner, M. and Niblack, W. "Efficient Color Histogram Indexing for Quadratic Form Distance," IEEE Trans. Pattern Analysis and Machine Intelligence, 17(7), pp. 729-736, July, 1995.

[46] A. Hauptmann, M. Y. Chen, M. Christel, C. Huang, W. H. Lin, T. Ng, N. Papernick, A. Velivelli, J. Yang, R. Yan, H. Yang, and H. D. Wactlar, "Confounded expectations: Informedia at TRECVID 2004," in Proc. TREC Video Retrieval Eval., Gaithersburg, MD, 2004. Available: http://www-nlpir.nist.gov/projects/tvpubs/tvpapers04/cmu.pdf

[47] Aljahdali, S.; Ansari, A.; Hundewale, N., "Classification of image database using SVM with Gabor Magnitude", International Conference on Multimedia Computing and Systems (ICMCS), 2012 , vol., no., pp.126,132, 10-12 May 2012

[48] P. Browne and A. F. Smeaton, "Video retrieval using dialogue, keyframe similarity and videoobjects," in Proc. IEEE Int. Conf. Image Process., vol. 3, pp. 1208–1211,Sep. 2005.

[49] R. Lienhart, “A system for effortless content annotation to unfold the semantics in videos,” in Proc. IEEE Workshop Content-Based Access Image Video Libraries, pp. 45–49, Jun. 2000.

[50] R. Mohan, "Video sequence matching", in Proceedings of International Conference on Acoustic, Speech and Signal Processing, pp. 3697–3700, 1998.

[51] Y. Tan, S. Kulkarni, P. Ramadge, "A framework for measuring video similarity and its application to video query by example", in ‘International Conference on Image Processing’, pp. 106–110, 1999.

[52] M. Naphade, M. Yeung, B. Yeo, "A novel scheme for fast and efficient video sequence matching using compact signature", in ‘SPIE Conference on Storage and Retrieval for Media Database’, pp. 564–572, 2000.

[53] T. Hoad, J. Zobel, "Fast video matching with signature alignment", in ‘ACM SIGMM InternationalWorkshop on Multimedia Information Retrieval’, Berkeley, CA, pp. 262–269, 2003.

[54] W. Ren, S. Singh, "Video sequence matching with spatio-temporal constraints", in ‘International Conference on Pattern Recognition’, pp. 834–837, 2004.

[55] C. Kim, B. Vasudev, "Spatiotemporal sequence matching for efficient video copy detection", IEEE Transactions on Circuits and Systems for Video Technology 15(1), 127–132, 2005.

[56] M. Toguro, K. Suzuki, P. Hartono, S. Hashimoto, "Video stream retrieval based on temporal feature of frame difference", in ‘Proceedings of International Conference on Acoustic, Speech and Signal Processing’, Volume 2, pp. 445–448, 2005.

[57] X. Liu, Y. Zhung, Y. Pan, "A new approach to retrieve video by example video clip", in ‘ACM International Conference on Multimedia’, pp. 41– 44, 1999.

[58] A. Jain, A. Vailaya, X. Wei, "Query by video clip", Multimedia Systems 7, 369–384, 1999.

[59] R. Lienhart, W. Effelsberg, R. Jain, "VisualGREP: A systematic method to compare and retrieve video sequences", Multimedia Tools and Applications 10(1), 47–72, 2000.

[60] S. Kim, R. Park, "An efficient algorithm for video sequence matching using the modified Hausdorff distance and the directed divergence", IEEE Transactions on Circuits and Systems for Video Technology 12(7), 592–596, 2002.

[61] N. Diakopouos, S. Volmer, "Temporally tolerant video matching", in ‘ACM SIGIR Workshop on Multimedia Information Retrieval’, Toronto, Canada, 2003.

[62] Y. Peng, C. Ngo, Q. Dong, Z. Guo, J. Xiao, "Video clip retrieval by maximal matching and optimal matching in graph theory", in ‘International Conference on Multimedia and Expo’, pp. 317–320, 2003.

[63] Y. Peng, C. Ngo, "Clip-based similarity measure for hierarchical video retrieval", in ‘ACM SIGMM InternationalWorkshop on Multimedia Information Retrieval’, pp. 53–60, 2004.

[64] K. Sze, K. Lam, G. Qiu, "A new key frame representation for video segment retrieval", IEEE Transactions on Circuits and Systems for Video Technology 15(9), 1148–1155, 2005.

[65] Y. Ho, C. Lin, J. Chen, H. Liao, "Fast coarse-to-fine video retrieval using shotlevel spatial-temporal statistics", IEEE Transactions on Circuits and Systems for Video Technology 16(5), 642–648, 2006.

[66] H. Luo, J. Fan, S. Satoh, W. Ribarsky, "Large scale news video database browsing and retrieval via information visualization", in ‘ACM symposium on applied computing’, Seoul, Korea, pp. 1086–1087, 2007.

[67] Y. Gong, H. J. Zhong, H. C. Chuan and M. Sakauchi, "An image database system with content capturing and fast image indexing abilities", Proc. Int. Conf. Multimedia Computing and Systems, Boston, Massachusetts, 121-130, 1994.

[68] C. Faloutsos, R. Barber, M. Flicker, J. Hafner, W. Niblack, D. Petkovic and W. Equitz, "Efficient and effective querying by image content", J. lntelL Inf. Systems 3, 231-262, 1994.

[69] A. Pentland, R. W. Picard and S. Scarloff, "Photobook: Tools for content-based manipulation of image databases", Proc. IS and T/SPIE. Conf Storage and Retrieval for Image and Video Databases 11, San Jose, California, 34-47, 1994.

[70] Visionics Corporate Web Site, FaceIt Developer Kit Software, http://www.visionics.com, 2002.

[71] The TREC Video Retrieval Track Home Page, http://www-nlpir.nist.gov/projects/trecvid/

[72] Arti Khaparde, B. L. Deekshatulu, M. Madhavilath, Zakira Farheen, Sandhya Kumari V, "Content Based Image Retrieval Using Independent Component Analysis”, IJCSNS International Journal ofComputer Science and Network Security, VOL.8 No.4, April 2000.

[73] H.B. Kekre, V.A. Bharadi, S.D. Thepade, B.K. Mishra, S.E. Ghosalkar, S.M. Sawant, "Content Based Image Retreival Using Fusion of Gabor Magnitude and Modified Block Truncation Coding," icetet, pp.140-145, 2010 3rd International Conference on Emerging Trends in Engineering and Technology, 2010.

[74] Flickner M. et al, “Query by image and video content: the QBIC system”, IEEE Computer 1995, Volume 28, Number 9, pp 23-32, 1995.

[75] A. Amir, W. Hsu, G. Iyengar, C. Y. Lin, M. Naphade, A. Natsev, C. Neti, H. J. Nock, J. R. Smith, B. L. Tseng, Y. Wu, and D. Zhang, “IBM research TRECVID-2003 video retrieval system,” in Proc. TREC Video Retrieval Eval., Gaithersburg, MD, 2003. Available: http://wwwnlpir.nist.gov/projects/tvpubs/tvpapers03/ibm.smith.paper.final2.pdf

[76] J. Adcock, A. Girgensohn, M. Cooper, T. Liu, L. Wilcox, and E. Rieffel, “FXPAL experiments for TRECVID 2004,” in Proc. TREC Video Retrieval Eval., Gaithersburg, MD, 2004. Available: http://wwwnlpir.nist.gov/projects/tvpubs/tvpapers04/fxpal.pdf

[77] R. Yan and A. G. Hauptmann, “A review of text and image retrieval approaches for broadcast news video,” Inform. Retrieval, vol. 10, pp. 445–484, 2007.

[78] C. Foley, C. Gurrin, G. Jones, H. Lee, S. McGivney, N. E. O’Connor, S. Sav, A. F. Smeaton, and P. Wilkins, “TRECVID 2005 experiments at Dublin city university,” in Proc. TREC Video Retrieval Eval., Gaithersburg, MD, 2005. Available: http://wwwnlpir.nist.gov/projects/tvpubs/tv5.papers/dcu.pdf

[79] E. Cooke, P. Ferguson, G. Gaughan, C. Gurrin, G. Jones, H. L. Borgue, H. Lee, S. Marlow,K.McDonald,M. McHugh, N. Murphy,N.O’Connor, N. O’Hare, S. Rothwell, A. Smeaton, and P. Wilkins, “TRECVID 2004 experiments in Dublin city university,” in Proc. TREC Video Retrieval Eval., Gaithersburg, MD, 2004. Available: http://wwwnlpir.nist.gov/projects/tvpubs/tvpapers04/dcu.pdf

[80] H. Tamura, S. Mori, T. Yamawaki, "Textural Features Corresponding to Visual Perception", IEEE Trans. on Systems, Man and Cyber., vol. 8, no. 6, p. 460–473. 2, 4, , June 1978.

[81] www.wikipedia.org

[82] Steven W. Zucker, Demetri Terzopoulos. "Finding Structure in Co-Occurrence Matrices for Texture Analysis", computer graphics and image processing 12, 286 - 308, 1980.

[83] Che-Yen Wen, Liang-Fan Chang, Hung-Hsin Li,"Content based video retrieval with motion vectors and the RGB color model", Forensic Science Journal, volume 6, issue 2, pages 1-36, 2007.

[84] S. Bruyne, D. Deursen, J. Cock, W. Neve, P. Lambert, and R. Walle, “A compressed-domain approach for shot boundary detection on H.264/AVC bit streams,” J. Signal Process.: Image Commun., vol. 23, no. 7, pp. 473–489, 2008.

[85] H. Koumaras, G. Gardikis, G. Xilouris, E. Pallis, and A. Kourtis, “Shot boundary detection without threshold parameters,” J. Electron. Imag., vol. 15, no. 2, pp. 020503-1–020503-3, May 2006.

[86] Z.-C. Zhao and A.-N. Cai, “Shot boundary detection algorithm in compressed domain based on adaboost and fuzzy theory,” in Proc. Int. Conf. Nat. Comput., pp. 617–626, 2006.

[87] A. Divakaran, R. Radhakrishnan, and K. A Peker, “Motion activitybased extraction of key-frames from video shots,” in Proc. IEEE Int. Conf. Image Process., vol. 1, Rochester, NY, pp. 932–935, 2002.

[88] T.MLiu, H.-J. Zhang, and F. H. Qi, “A novel video key-frame-extraction algorithm based on perceived motion energy model,” IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 10, pp. 1006–1013, Oct. 2003.

[89] Sinora Banker Ghosalkar, Vinayak A.Bharadi, Sanjay Sharma, Asif Ansari, "Feature Extraction using Overlap Blocks for Content based Image Retreival" International Journal of Computer Applications (0975-8887), Volume 28-No.7, August 2011.