Interactive Video Retrieval Using Semantic Level Features and Relevant Feedback

Sadagopan Padmakala[1] and Ganapathy AnandhaMala[2]
[1]Department of Computer Science, Anna University, India.
[2]Department of CSE, Easwari Engineering College, India.

Abstract: *Recent years, many literatures presents a lot of work for content-based video retrieval using different set of feature. But, most of the works are concentrated on extracting features low level features. But, the relevant videos can be missed out if the interactive with the users are not considered. Also, the semantic richness is needed further to obtain most relevant videos. In order to handle these challenges, we propose an interactive video retrieval system. The proposed system consists of following steps: 1) Video structure parsing, 2) Video summarization and 3) Video Indexing and Relevance Feedback. At first, input videos are divided into shots using shot detection algorithm. Then, three features such as color, texture and shape are extracted from each frame in video summarization process. Once the video is summarized with the feature set, index table is constructed based on these features to easily match the input query. In matching process, query video is matched with index table using semantic matching distance to obtain relevant video. Finally, in relevance feedback phase, once we obtain relevant video, it is given to identify whether it is relevant for the user. If it is relevant, more videos relevant to that video is given to the user. The evaluation of the proposed system is evaluated in terms of precision, recall and f-measure. Experiments results show that our proposed system is competitive in comparison with standard method published in the literature.*

Keywords: *shot detection, color, shape, texture, video retrieval, relevant feedback.*

Received January 31, 2013; accepted June 17, 2014

1. Introduction

Video is the technology of electronically capturing, recording, processing, storing, transmitting, and reconstructing a sequence of still images representing scenes in motion to be very precise. Due to the technology evolution in multimedia, digital TV and information highways. Today, a giant amount of video data is widely available. Due to the absence of proper search method all these data are almost not usable [21]. An advanced technology for representing, modeling, indexing, and retrieving multimedia data is indispensable in order to increase the customer demand for visual information. Yet, the relational or object oriented data model based conventional database management system is inept in providing adequate facility for managing and retrieving video contents [33]. For the reason that, this conventional system is inadequate due to three major reasons:

1. Absence of facilities for handling spatiotemporal relations.
2. Absence of knowledge-based techniques for defining raw data into semantic contents.
3. Dearth of query representations [20].

The multimedia data, unlike text, necessitates several stages of pre-processing to obtain indices related to inquiry so the searching for a multimedia content is somewhat complex., there is a lack of commonly agreed upon vocabulary so the image or a video sequence can be depicted in several ways [18]. To allow access to extremely giant databases of images and videos by modern scalable browsing algorithms, and to provide semantic visual interfaces a robust Content-Based image and Video Indexing and Retrieval (CBVIR) system is crucial to index/retrieve and compress visual information [16, 25, 27].

Content-Based Video Retrieval (CBVR) [7] systems seem to be an inherent combination of Content-Based Image Retrieval (CBIR) systems [12]. while handling images several aspects that should be addressed when using videos are ignored [1, 8, 17, 32, 35]. The four important processes involved in content-based video indexing and retrieval [2, 6, 13] are:

- Video content analysis: For key information in a video visual content plays an important role. However, other media components such as text and audio also carry useful information in characterizing the video program for both customer and professional applications The integration of these components would be more effectual [11, 24].
- Video structure parsing: In the process of video structure parsing segmenting the video into individual scenes is a vital step. Using appropriate shot boundary detection algorithms the video can be segmented into frames with similar visual contents [9] and also the processing of speech components that occur with them have been proven effective in realizing this objective [13].
- Video summarization: Video summarization refers to the presentation of visual information about the

structure of video, which should be much shorter compared to original video [30].

- Video indexing: The metadata refers to the structural and content attributes located in content analysis, video parsing, and abstraction processes, or the attributes that are manually entered. Video indices and the table of contents can be constructed based on these attributes via, for example, a clustering process that classifies sequences or shots into diverse visual categories or an indexing structure [25].

For efficient retrieval of digital video information the content based video matching is very essential but it is a challenging and difficult task. One major reason for this is the amount of intra-class variation where the same semantic content can take place under diverse illumination, appearance, and scene settings. For determining whether two videos are analogous or not various factors have been considered. These factors include resemblance of the foreground objects, motion of object, background appearance, camera movement, etc., [25].Video holds various types of audio and visual information that are knotty to extract, combine or trade-off using common video information retrieval system [7, 10]. A variety of approaches have been proposed for CBVR [34]. For constructing a hierarchical motion description based motion index tree that serves as a classifier to identify the analogous motions for the sample query A content-based 3D motion retrieval algorithm [14] is developed. Various key frame-based retrieval methods for video have also been proposed. Recently, text retrieval techniques based on vocabulary has been used for object matching in videos [26]. Color segmentation, optical flow computation dependant motion segmentation and object tracking methods are also available [4].

This paper presents an effective content based video retrieval system by extracting the feature set after converting the raw video into three feature extraction schemes such as color, texture and shape. Color, texture and shape are the three important representation schemes used in the proposed system to extract the significant features presented in the raw video. In initial stage the person informs about the required video. The video can be of any form of signal. The clarity depends upon the pixel value. As the customer suggests the suggestion is send to the query section and in the query section the required video is analyzed and then the suggestion is sent to the matching section. In the video matching section the matching is carried on to check whether the required video is present or not. And then the relevant video is sent to the person and if the required video is not obtained then the user feedback is sent to the video matching section and the procedure is repeated until the required video is obtained.

The main contributions of our proposed system are:

1. Under this system, we can improve certain shortages that are presented the existing video retrieval system and obtain an improvement in a precision, recall and f-measure.
2. Video summarization process is done by color, shape and texture features.
3. Relevant Video Clips is retrieved using indexing process and kernel based fuzzy C-means clustering.

The rest of the paper is organized as follows: section 2 reviews the recent research works with respect to content-based video retrieval, section 3 describes the proposed system for content-based video retrieval, section 4 explains result and discussion and section 5 concludes the paper.

2. Related Work

The literature presents numerous algorithms and techniques to retrieve significant videos from the database due to the widespread interest of content-based video retrieval in a large number of applications. Here, we discuss some recent researches related to content-based video retrieval.

Sze *et al.* [29] have developed an optimal key frame representation scheme based on global statistics for video shot retrieval. Each pixel in that optimal key frame was constructed by considering the probability of occurrence of those pixels at the corresponding pixel position among the frames in a video shot. Therefore, that constructed key frame was called Temporally Maximum Occurrence Frame (TMOF), which was an optimal representation of all the frames in a video shot. The retrieval performance of that representation scheme was further improved by considering the pixel values with the largest probabilities of occurrence and the highest peaks of the probability distribution of occurrence at each pixel position for a video shot.

Cotsaces *et al.* [5] have presented a method for performing fast retrieval in video based on the output of face detectors and recognizers. The developed method was both robust because it was based on a convolution-like video content similarity computation and fast because it made extensive use of database indexing. The retrieval performance of their algorithm had been verified by the implementation of a real system that used face detection and recognition to index real videos.

Su *et al.* [28] have presented the use of motion vectors embedded in MPEG bitstreams to generate so-called “motion flows”, which were applied to perform video retrieval. They used the information about motion vectors in an MPEG bitstream directly to generate some trajectory-like motionflows to describe local motion. Since motion vectors were uniformly distributed in each video frame, they processed a case of multiple moving objects in a shot.

Ramya and Rangarajan [22] have addressed the specific aspect of inferring semantics automatically from raw video data using different knowledge-based methods. In particular, this paper focuses on three techniques namely, rules, Hidden Markov Models (HMMs), and Dynamic Bayesian Networks (DBNs).

First, a rule-based approach that supports spatio-temporal formalization of high-level concepts was introduced. Then the focus of this paper was towards stochastic methods and also demonstrates how HMMs and DBNs could be effectively used for content-based video retrieval from multimedia databases. Although this work has not compared the two stochastic approaches between each other, an intuitive conclusion was that the DBN approach has been more suitable for fusing multimodalities in retrieval. Based on this conclusion on the property of DBNs that each feature could influence the decision with a specific probability. In HMM approach the process of quantization, which leads to discrete HMMs, has treated all features equally. However, operations with HMMs were less time-consuming than with DBNs.

Patel *et al*. [19] have proposed that the Selection of extracted features play an important role in content based video retrieval regardless of video attributes being under consideration. These features were intended for selecting, indexing and ranking according to their potential interest to the user. Good features selection also allows the time and space costs of the retrieval process to be reduced. This survey reviews the interesting features that could be extracted from video data for indexing and retrieval along with similarity measurement methods. They have also identified present research issues in area of content based video retrieval systems. Another possibility was to develop an interactive user interface based on visually interpreting the data using a selected measure to assist the selection process. Extensive experiments comparing the results of features with actual human interest could be used as another method of analysis. Since user interactions are indispensable in the determination of features, it was desirable to develop new theories, methods, and tools to facilitate the user's Involvement.

Anh *et al.* [1] have introduced an approach based on Scale-Invariant Feature Transform (SIFT) feature, a new metric and an object retrieval method. Their algorithm was built on CBIR method in which the key frame database includes key frames detected from video database by using their shot detection method. Experiments show that the approach of their algorithm has fairly high accuracy. Because of the ability of the segmentation process to separate main objects from their correlative background with acceptable accuracy and the ability of being invariable under the changing of geometry transforming and rate, the scheme of key frame segmentation, calculating SIFT feature and object retrieving could recognize similar main objects from different shots with good accuracy.

Thornley *et al*. [31] have provided a discussion and analysis of methodological issues encountered during a scholarly impact and bibliometric study within the field of computer science. The purpose of their paper was to provide a reflection and analysis of the methods used to provide useful information and guidance for those who may wish to undertake similar studies, and was of particular relevance for the academic disciplines which have publication and citation norms that may not perform well using traditional tools. The results have indicated that multi-word frequency provides a promising means to track research trends in certain disciplinary areas by chronologically identifying frequency of terms. Comparing bibliographic data fields such as 'Keywords', 'Title' and 'Abstract they could see for instance that 'Concept Detection', ' Video Search' and 'Shot Boundary Detection' holds similar ranking in frequency across fields.

Luan *et al.* [15] have introduced an effective interactive video retrieval system named VisionGo. It jointly explores human and computer to accomplish video retrieval with high effectiveness and efficiency. It assists the interactive video retrieval process in different aspects:

1. It maximizes the interaction efficiency between human and computer by providing a user interface that supports highly effective user annotation and an intuitive visualization of retrieval results.
2. It employs a multiple feedback technique that assists users in choosing proper method to enhance relevance feedback performance.
3. It facilitates users to assess the retrieval results of motion-related queries by using motion-icons instead of static key frames. Experimental results based on over 160 h of news video have shown to demonstrate the effectiveness of the VisionGo system.

Rooij and Worring [23] have implemented a system for active bucket-based video retrieval, evaluate two different learning strategies. Their method showed that it used in video retrieval with an evaluation using three groups of non-expert users. One base line group used only the categorization features of Media Table such as sorting and filtering on concepts and fast grid preview, but no online learning mechanisms. One group used on-demand passive buckets. The last group used fully automatic active buckets which autonomously added content to buckets.

3. Interactive Video Retrieval Using Semantic Level Features and Relevant Feedback System

With the rapid growth of video information, video retrieval becomes very vital and necessary for collecting useful information from video database. The aim of video retrieval system is defined that searching and retrieving videos relevant to a user defined query. It is one of the most popular topics in both real life applications and multimedia research. Thus, there is a huge need for generating a video search system, which can perform search based on the semantic level features and relevant feedback. In this paper, we propose a video retrieval system using combining

color, shape, texture features and relevant feedback system, which is summarized in Figure 1.

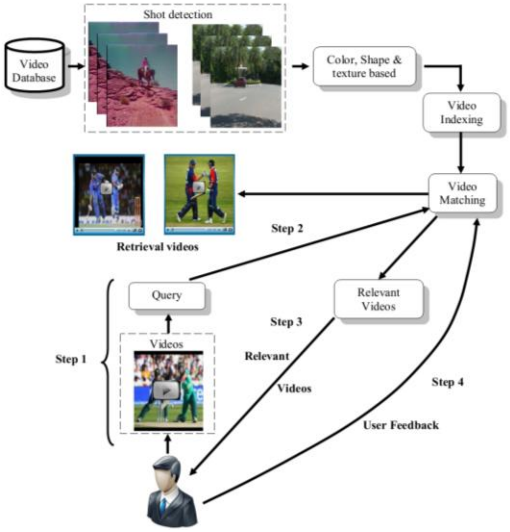

Figure 1. Overall block diagram of proposed system.

3.1. Video Structure Parsing

Video shot detection is an important and challenging process for many video processing applications like video indexing, summarization, video retrieval and more. It is the process of capturing a sequence of frames without any major change occurring among them. Several researches are available in the literature for video shot detection using different techniques. Here, discrete cosine Transform and correlation measure have been applied to find the number of frames present in each shot. Initially, the 1st and 2nd frames are partitioned into a set of blocks and then DCT is applied to each block of the frame. The corresponding output frame Y after applying DCT on the input frame X can be defined as:

$$Y_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X_{mn} \cos\frac{\pi(2m+1)p}{2M} \cos\frac{\pi(2n+1)q}{2N}, 0 \le p \le M-1, 0 \le q \le N-1 \quad (1)$$

Where $\alpha_p = \begin{cases} 1/\sqrt{M}, & p = 0 \\ \sqrt{2/M}, & 1 \le p \le M-1 \end{cases}$; $\alpha_q = \begin{cases} 1/\sqrt{N}, & q = 0 \\ \sqrt{2/N}, & 1 \le q \le N-1 \end{cases}$

Then, the correlation coefficient is computed in between the frame 1 and 2 using the following equation [3].

$$r = \frac{\sum_m \sum_n \left(Y^1_{mn} - \overline{Y^1}\right)\left(Y_{mn}{}^2 - \overline{Y^2}\right)}{\sqrt{\left(\sum_m \sum_n \left(Y_{mn}{}^1 - \overline{Y^1}\right)^2\right)\left(\sum_m \sum_n \left(Y_{mn}{}^2 - \overline{Y^2}\right)^2\right)}} \quad (2)$$

Where $\overline{Y^1} = mean(Y^1),\ and\ \overline{Y^2} = mean(Y^2)$

After finding the correlation for the first and second frame, the same procedure is repeated for the consecutive frames presented in the video. Then, the frames within a shot can be identified by maximizing the cross correlation term which gives a measure of the degree of similarity between two frames of video.

3.2. Video Summarization

After shot segmentation, each frame is described by their color, texture and shape features. The detailed feature extraction process is described in following section:

3.2.1. Semantic Color Feature

Color is one of the main visual cues, and it has been frequently used in image processing, analysis and retrieval. According to the strong relationship between colors and human emotions, an emotional semantic query model based on color semantic description is proposed in this section.

- At first, each frame is converted into HSV color space which is a visual property and usefulness in content based image retrieval applications. Generally, HSV color space represents a visual perception of the variation in Hue, Saturation and Intensity values of an image pixel.
- Subsequently, HSV values are normalized to the range [0, 1]. Hue value is usually quantized into a small set of about 10-20 base color names. We uniformly quantize the Hue value into 10 base colors, red, orange, yellow, green, aqua, aquamarine, blue, violet, purple, and magenta. Saturation and Value are quantized (not uniformly) into 4 bins respectively as adjectives signifying the saturation and luminance of the color.
- We generate features by using the above properties of the HSV color space based on the color naming model given is in Table 1.

Table 1. Color naming model

Value of HSV	Base color names
0 - 0.1	Orange
0.1-0.2	Yellow
0.2-0.3	green
0.3-0.4	aqua
0.4-0.5	aquanmarine
0.5-0.6	blue
0.6-0.7	violet
0.7-0.8	purple
0.8-0.9	magenta
0.9-1.0	red

- Based on the color naming model is given in table 1, the count of the Hue, Saturation and Value is taken as a color feature. The count says that how many times each color presented in the frame.

3.2.2. mBm based Texture Feature Extraction

In this section, we develop an efficient feature extraction algorithm to extract the texture features using multi-fractal Brownian motion (mBm). Here, mBm is used to analyze the irregular texture variations of input image for robust texture based feature extraction. The texture based feature can be generated by following steps as describe below:

- At first, the each frames or images are converted into 8×8 blocks.
- Then, the DWT is applied on each block to extract high and low frequency information, which gives

first level decomposed image of one approximate image (LL) and three detail images (LH,HL,HH). Here, haar wavelet is used to detect the high and low frequency characteristics from each block. Haar wavelet transform is capable of detecting and characterizing specific phenomena in time and frequency planes. Also, it is a smooth and quickly vanishing oscillating function with a good localization in both frequency and time. We calculate the mBm value for four images or sub bands, such as LL, LH, HL and HH. For each sub bands, mBm value calculates as follows:

$$H(t) = lim_{\alpha \to 0} \frac{log\left(\frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{M-1}|W_z(a,t)|^2\right)}{2loga} \quad (3)$$

Where,

$H(t)$ ➔The time-varying scaling (or Holder) exponent.

α ➔Scaling factor, The value for α is Pre-determined between 1 and 2.

W_z➔Wavelets transform coefficients

Finally, mean of the mBm value is calculated from four subbands as texture feature:

$$mBm\ feature = \frac{LL + HH + HL + LH}{4} \quad (4)$$

3.2.3. Level-Set-Based Shape Feature Extraction

In this section, we implement a binary level set method in an iterative procedure, which consists of the following steps:

- Initialize the binary level set function ϕ
- Evolve the level set function ϕ according to Equation 1.
- Smooth the level set function by a Gaussian filter, i, e $\phi = G_{ij} * \phi$.
- Let ϕ take 1 if $\phi' \geq 0$; otherwise, let ϕ take-1. Go to step 2.

In computer vision and computer graphics Level-set-based shape modeling is an important research topic. In this study, we implement a more recent work [36] on binary level-set representation for object shape detection. Consider the basic dentition of level set given as

$$\begin{gathered} \varphi_t + F|\nabla\varphi| = 0,\ given \varphi(x,t=0) \\ \phi_t + F_0|\nabla\phi| + v(x,y,t)\nabla\phi = \in k|\nabla\phi| \\ \frac{\partial\phi}{\partial t} = vg(I)\|\nabla\phi\| \end{gathered} \quad (5)$$

Where, $\frac{1}{1+\|\nabla(G_\sigma * I)\|}$ and G_σ denotes the Gaussian filter with standard deviation σ .

Where $F_o|\nabla\varphi|$ - is the motion of the curve in the direction normal to front, $U(x,y,t)\nabla\varphi$ - is the term that moves the curve across the surface and $\in k|\nabla\varphi|$ - is the speed term dependent upon curvature.

In our study, $U(x,y,t)$ is the gradient of image and $\in k|\nabla\varphi|$ is approximated using a central difference. We first convert the MRI to binary image. The level set is used on these binary images to track the shape at the boundary of images. Note for binary images, only digital derivative approximations exist at the boundary. We initialize the level-set function using the gradient of the image. We propagate this gradient across the surface given as [3].

$$\phi_{ij}^{t+1} = \phi^t{}_{ij} - \Delta t[\max(G_{ij},0)\Delta^+ + \min(G_{ij},0)\Delta^-] \quad (6)$$

Where ϕ_{ij}^t is the value of ϕ at pixel i at time t, Δt is the time step (or scaling factor), G_{ij} is a Gaussian filter to smooth the edges, and Δ^+ and Δ^- describe the normal component and are given as

$$\Delta^+ = [max(D_x^-,0)^2 + min(D_x^+,0)^2 + max(D_y^-,0)^2 + min(D_y^+,0)^2]^{1/2} \quad (7)$$

$$\Delta^- = [max(D_x^+,0)^2 + min(D_x^-,0)^2 + max(D_y^+,0)^2 + min(D_y^-,0)^2]^{1/2} \quad (8)$$

Where D_x^- , D_x^+ , D_y^- and D_y^+ are the forward and backward derivative approximation in x-and y-directions. The Gaussian filter creates a larger attraction range allowing the level sets to be attracted to the boundary. These steps iterate and stop when the boundary is completed upon convergence.

3.3. Video Indexing and Relevance Feedback

3.3.1. Video Indexing

The videos presented in the input training dataset are subjected to the feature extraction system and so the important features, such as texture, color, and shape are extracted. The extracted features from the input videos are stored in the feature library as indexing. The purpose of storing an index is to optimize speed and performance in finding relevant documents for a search query. This process computes the image feature vectors which are then used by distance calculation and helps in video retrieval process. Subsequently, each extracted feature for query video clip (texture, color and shape) is matched with the corresponding feature set presented in the feature library. At first, each feature is converted into a vector form and these vectors are clustered using kernel based Fuzzy C-Means (FCM) clustering. Generally, FCM is an algorithm of clustering which enables a particular piece of data to be a member of multiple clusters. It tends to be considerably restricted to spherical clusters only. To solve this problem, kernel fuzzy c-means algorithm is employed by mapping data with nonlinear relationships to appropriate feature spaces. Here, input training features are clustered and grouped based on the input dataset. Each cluster contains individual centroid.

- KFCM: Any point X_i has a set of coefficients giving the degree of being in the i^{th} cluster o_i. With kernel based fuzzy *c*-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster:

$$o_i = \frac{\sum_{l=1}^{n} u_{il}^{m} k(x_l, o_i) x_l}{\sum_{l=1}^{n} u_{il}^{m} k(x_l, o_i)} \quad (9)$$

Where,

$$u_{ij} = \frac{(1-k(x_j, o_i))^{-1/m-1}}{\sum_{l=1}^{c} (1-k(x_j, o_i))^{-1/m-1}} \quad (10)$$

In above equation, radial basis function kernel can be used:

$$k(x_i, x_j) = \exp\left(- \frac{\left\| x_i - x_j \right\|^2}{r^2} \right) \quad (11)$$

Finally, the required number of video clips relevant to the query video is retrieved from the database effectively using centroid value. The indexing process is detailed in Figure 2, which specifies that each cluster consists of its centroid value with corresponding videos. The training dataset is fully trained and stored as indexing format. Now, this way says that we can easily retrieve accurate videos in testing phase.

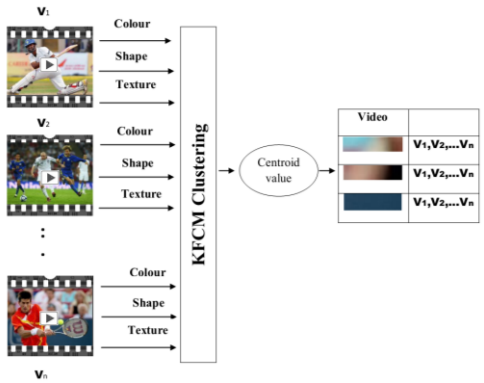

Figure 2. Indexing operation.

3.3.2. Relevance Feedback Process

In this phase, at first, the user suggests the suggestion is send to the query section and the required video is analyzed in the query section. Then, the features are extracted through feature extraction process and then the suggestion is sent to the matching section. Then, using Euclidean distance, each extracted feature for query video clip (texture, color, and shape) is matched with the respective feature set presented in the indexing process. The matching process computes the similarity between a query and video data in the dataset. Here, we delineate the similarity as their distance. Here, the computed distance is utilized to identify the similarity between the query video clip and each database video clip. Thus, the matching process is carried on to check whether the required video is present or not. And then the relevant video is sent to the person and if the required video is not obtained then the user feedback is sent to the video matching section through indexing. If video is presented in first cluster, the video is retrieved for user query and otherwise video matching process is moved to second cluster.

Comparing a query with video data distributes three distances: TD_C, distance of trained color feature, TD_s distance of trained shape feature and TD_T, distance of trained texture feature. Calculate the distance D between the query and trained videos as follows:

$$D = [(Q_C - TD_C) + (Q_S - TD_S) + (Q_T - TD_T)] \quad (12)$$

Where, D-is the distance Q_C-Color feature of the input query Q_S-Shape feature of the input query Q_T- Texture feature of the input query TD_T-Centroid of trained texture feature set TD_s-Centroid of trained shape feature set TD_C-Centroid of trained color feature set

After distance calculated using the Equation 12 for query video, the required number of video clips corresponding to the query video is retrieved from the database successfully based on the threshold (T). The retrieval process is done by following equation:

$$V^{retrived} = D[Q_V, TD] \quad D < T \quad (13)$$

Where,

$V^{retrived}$ → Retrieved videos.

$D[Q_v, TD]$ → Distance between query and trained feature set.

T → Threshold.

The threshold T is fixed based on the variation of distance between query and trained feature set. Applying the user relevance feedback, the system improves the query result is retrieved easily and more accurately.

4. Results and Discussion

The results and discussion of the proposed video retrieval system is given in this section. The proposed video retrieval system has been implemented using MATLAB (Matlab7.10) and the performance of the proposed system is analyzed using the evaluation metrics including precision, recall and F-measure.

4.1. Dataset Description

Experimental dataset of our proposed system obtained from YouTube website (www.youtube.com). The collected database contains 23 videos, which includes that different categories of objects presented in these videos such as, cricket, football, golf, tennis, walking man and the length of these videos usually contains 150 to 250 frames. The sample snapshot of the input videos are presented in Figure 3.

Figure 3. A sample snapshot for the input database.

4.2. Experimental results

In this section, we report experimental results from our video retrieval system. The experimental results of the feature extraction are presented in Figures 4, 5 and 6.

4.2.1. Color based Feature

Figure 4 illustrates the color based features.

a) Frame corresponding to the input walk video.

H	S	V
3287	5634	1059
3335	983	1377
454	1137	140
69	788	162
164	384	181
1359	176	243
171	94	471
54	69	710
468	55	423
141	23	1209

b) Output of color feature.

Figure 4. Color based feature.

4.2.2. Texture based Feature

Figure 5 shows the texture textures.

a) Frame corresponding to the input walk video.

2.789525
2.789525
4.900235
3.497088
-17.5918
11.08182
1.996956
10.63392
12.84817
1.991274

b) Sample output of texture feature.

Figure.5. Texture based feature.

4.2.3. Shape based Feature

Figure 6 shows the shape based features.

a) Frames corresponding to the walk video.

b) Output of shape feature.

Figure 6. Shape based feature.

4.2.4. Video Retrieval results

The features including texture, color, and shape are extracted from the input videos and it is stored in the indexing storage. For retrieving, the video clips are given to the proposed system that extracts the features and the features are matched with the indexing storage using the designed video matching. The matching score computed is used to retrieve the videos from the dataset and the retrieved video for the corresponding input videos is given in the following Figures 7 and 8.

a) Query video.

b) Retrieval for the video clip in figure 7-a.

Figure 7. Retrieved videos.

a) Query video.

b) Retrieval for the video clip in figure 8-a

Figure 8. Retrieved videos.

4.3. Quantitative Analysis

The performance of the proposed approach system is evaluated on the input dataset using the precision, recall and F-measure. For quantitative analysis, videos from each category are given to the proposed system and results are evaluated with the defined as follows:

$$Precision = \frac{\{SimilarVideos\} \cap \{Retrieved\ videos\}}{\{Retrieved\ videos\}} \quad (14)$$

$$Recall = \frac{\{SimilarVideos\} \cap \{Retrieved\ videos\}}{\{Similar videos\}} \quad (15)$$

$$F-measure = \frac{2*PR}{P+R} \quad (16)$$

Table 2 is showed for the query video clips and corresponding precision, recall and F-measure. The results obtained employing these two input parameters are plotted as graph shown in Figures 9, 10 and 11. The graph shows the performance of the proposed system in retrieving the relevant videos and it clearly differentiate the results obtained for different videos. To effectiveness and comparison of our proposed system, we compare our proposed system against one standard video retrieval system [29]. The author [29] presented an optimal key frame representation scheme based on global statistics for video shot retrieval. Their techniques with our database are taken for our comparative analysis. The performance of the existing system and proposed system is shown in Figures 9, 10 and 11 and Table 2. From the table and figures, our proposed system achieves the better performance when compared with existing system [29].

Table 2. Precision, Recall and F-measure for the input video clips.

Query video clip	Retrieve		Manual		Correct		Precision		Recall		F-measure	
	Proposed	Existing [29]	Proposed	Existing [29]	Proposed	Existing [29]	Proposed	Existing [29]	Proposed	Existing [29]	Proposed	Existing [29]
	6	6	3	3	3	2	0.5	0.33	1	0.67	0.67	0.44
	7	7	5	5	4	3	0.57	0.42	0.8	0.60	0.66	0.50
	7	7	5	5	4	2	0.57	0.28	0.8	0.40	0.66	0.33
	7	7	5	5	4	3	0.57	0.42	0.8	0.60	0.66	0.50
	7	7	5	5	4	4	0.57	0.57	0.8	0.80	0.66	0.67
	6	6	5	5	5	5	0.83	0.83	1	1	0.91	0.91
	7	7	10	10	6	5	0.86	0.71	0.6	0.50	0.71	0.59
	7	7	10	10	7	7	1	1	0.7	0.70	0.82	0.82
	8	8	10	10	7	7	0.87	0.87	0.7	0.70	0.77	0.78
	2	2	3	3	2	2	1	1	0.67	0.67	0.80	0.80
	2	2	3	3	2	2	1	1	0.67	0.67	0.80	0.80
	2	2	2	2	2	2	1	1	1	1	1	1
	2	2	2	2	1	1	0.5	0.5	0.5	0.50	0.5	0.5
	10	10	10	10	9	10	0.9	1	0.90	1	0.9	1
	9	9	10	10	9	9	1	1	0.9	0.90	0.95	0.95
	8	8	10	10	8	7	1	0.87	0.80	0.70	0.89	0.77
	7	7	10	10	7	7	1	1	0.70	0.70	0.82	0.82
	10	10	10	10	9	9	0.9	0.9	0.90	0.90	0.9	0.9
	8	8	10	10	8	8	1	1	0.80	0.80	0.89	0.88
	7	7	10	10	6	6	0.86	0.85	0.60	0.60	0.71	0.705
	3	3	10	10	2	1	0.67	0.33	0.20	0.10	0.31	0.15
	6	6	10	10	5	5	0.83	0.83	0.50	0.5	0.62	0.62
	3	3	2	2	2	2	0.67	0.67	1	1	0.80	0.80

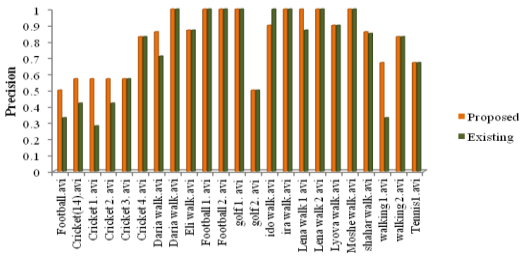

Figure 9. Precision graph plotted for different videos.

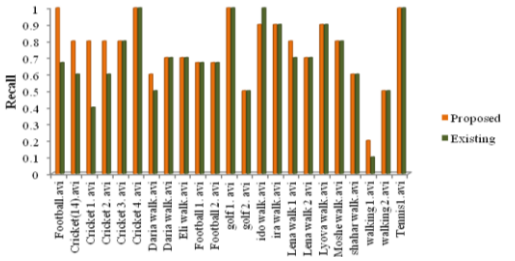

Figure 10. Recall graph plotted for different videos.

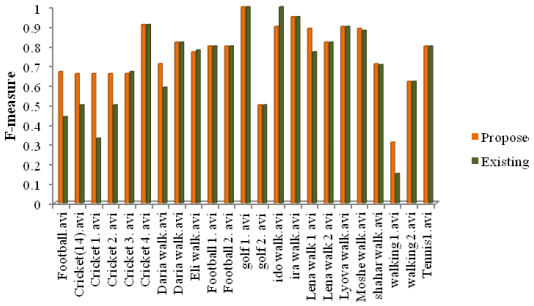

Figure 11. F-measure graph plotted for different videos.

5. Conclusions

We have developed an efficient content based video retrieval system based on the color, texture and shape. At first, texture feature is extracted using multi-fractal Brownian motion (mbm), the color feature is obtained using semantic color model and shape feature is extracted using level set method. Then, those features are stored in the indexing format. Finally, relevance feedback system provides a better and accurate retrieval way in the proposed video retrieval system.

References

[1] Anh T., Bao P., Khanh T., Thao B., Tuan T., and Nhut N., "Video Retrieval using Histogram and Sift Combined with Graph-based Image Segmentation," *Journal of Computer Science*, vol. 8, no. 6, pp. 853-858, 2012.

[2] Bole R., Yeo B., and Yeung M., "Video Query: Research Directions," *IBM Journal of Research and Development*, vol. 42, no. 2, pp. 233-252, 1998.

[3] Bourke P., http://local.wasp.uwa.edu.au/~pbourke/miscellaneous/ correlate/, Last Visited 1996.

[4] Chang S., Chen W., Meng H., and Sundaram H., and Zhong D., "An Automated Content based Video Search System using Visual Cues," *in Proceeding of the 5th ACM*

International Conference on Multimedia, Washington, pp. 313-324, 1997.

[5] Cotsaces C., Nikolaidis N., and Pitas I., "Face-Based Digital Signatures for Video Retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 549-553, 2008.

[6] Dimitrova N., Zhang H., Shahrara B., Sezan I., Huang T., and Zakhor A., "Applications of Video-Content Analysis and Retrieval," *IEEE Multimedia*, vol. 9, no. 3, pp. 42-55, 2002.

[7] Geetha P. and Narayanan V., "A Survey of Content-Based Video Retrieval," *Journal of Computer Science*, vol. 4, no. 6, pp. 474-486, 2008.

[8] Ghodeswar S. and Meshram B., "Content Based Video Retrieval," *in Proceeding of International Symposium on Computer Engineering and technology*, Punjab, 2010.

[9] Hanjalic A., "Shot-Boundary Detection: Unraveled and Resolved," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 2, pp. 90-105, 2002.

[10] Hauptmann A., Ng T., and Jin R., "Video Retrieval Using Speech and Image Information," *in proceeding of Electronic Imaging Conference, Storage and Retrieval for Multimedia Databases*, Santa Clara, 2003.

[11] Hauptman A., Ng T., Baron R., Lin W., Chen M., Derthick M., Christel M., Jin R., and Yan, R., "Video Classification and Retrieval with the Informedia Digital Video Library System," *in proceeding of Text Retrieval Conference*, Gaithersburg, 2002.

[12] Lee J., Oh J., and Hwang S., "Strg-index: spatio-temporal region graph indexing for large video databases," *in Proceeding of the ACM SIGMOD International Conference on Management of Data*, Maryland, pp. 718-729, 2005.

[13] Lew M., Sebe N., and Gardner P., *Video Indexing and Understanding*, Springer, 2001.

[14] Liu F., Zhuang Y., Wu F., and Pan Y., "3D Motion Retrieval with Motion Index Tree," *Computer Vision and Image Understanding*, vol. 92, no. 2, pp. 265-284, 2003.

[15] Luan H., Zheng Y., Wang M., and Chua T., "VisionGo: Towards Video Retrieval with Joint Exploration of Human and Computer," *Journal of Information Science*, vol. 181, no. 19, pp. 4197-4213, 2011.

[16] Malik F. and Baharudin B., "The Statistical Quantized Histogram Texture Features Analysis for Image Retrieval Based on Median and Laplacian Filters in the DCT Domain," *The International Arab Journal of Information Technology*, vol. 10, no. 6, pp. 616-624, 2013.

[17] Misra C. and Sural S., *Computer Vision-ACCV2006*, Springer Link, 2006.

[18] Mittal A., "An Overview of Multimedia Content-Based Retrieval Strategies," *Informatica*, vol. 30, pp. 347-356, 2006.

[19] Patel B., Meshram B., Shah and Kutchhi A., "Content Based Video Retrieval Systems", *International Journal of UbiComp*, vol. 3, no. 2, pp. 13-30, 2012.

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

[21] Petkovic M., "Content-based Video Retrieval," *in proceeding of the PhD Workshop on EDBT*, Konstanz, 2000.

[22] Ramya S. and Rangarajan P., "Knowledge Based methods for Video Data Retrieval," *International Journal of Computer Science and Information Technology*, vol. 3, no. 5, pp. 165-172, 2011.

[23] Rooij O. and Worring M., "Active Bucket Categorization for High Recall Video Retrieval," *IEEE Transactions on Multimedia*, vol. 15, no. 4, pp. 898-907, 2013.

[24] Rooij, O. and Worring, M., "Active Bucket Categorization for High Recall Video Retrieval," *IEEE Transactions on Multimedia*, vol. 15, no. 4, pp. 898-907, 2013.

[25] Sebe N., Lew M., and Smeulders A., "Video Retrieval and Summarization: Editorial Introduction," *Computer Vision and Image Understanding*, vol. 92, no. 2-3., pp. 141-146, 2003.

[26] Sivic J. and Zisserman A., "Text retrieval approach to object matching in videos," *in Proceeding of 9th IEEE International Conference on Computer Vision*, Washington, pp. 1470-1477, 2003.

[27] Subramanian M. and Sathappan S., "An Efficient Content Based Image Retrieval Using Advanced Filter Approaches," *The International Arab Journal of Information Technology*, vol. 12, no. 3, pp.229-236, 2015.

[28] Su C., Liao H., Tyan H., Lin C., Chen D., and Fan K., "Motion Flow-Based Video Retrieval," *IEEE Transactions on Multimedia*, vol. 9, no. 6, pp. 105-112 2007.

[29] Sze K., Lam K., and Qiu G., "A New Key Frame Representation for Video Segment Retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 9, pp. 1148-1155 2005.

[30] "http://www.open-video.org/, Last Visited 2013.

[31] Thornley C., McLoughlin S., Johnson A., and Smeaton A., "A Bibliometric study of Video Retrieval Evaluation Benchmarking (TRECVid): a Methodological Analysis,"

Journal of Information Science, vol. 37, no. 6, pp. 1-19, 2011.

[32] Xu L. and Wang K., "Extracting Text Information for Content-Based Video Retrieval," *in Proceeding of the 14th international Conference on Advances in Multimedia Modeling*, Kyoto, pp. 58-69, 2008.

[33] Yoshitaka A. and Ichikawa T., "A Survey on Content-Based Retrieval for Multimedia Databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, no. 1, pp. 81-93, 1999.

[34] Zhai Y., Liu J., Cao X., Basharat A., Hakeem A., Ali S., and Shah M., *Video understanding and content-based retrieval*, TREC Video Retrieval Evaluation, 2005.

[35] Zhaoming L., Xiangming W., Xinqi L., and Wei Z., "A Video Retrieval Algorithm Based on Affective Features," *in proceeding with the 9th IEEE International Conference on Computer and Information Technology*, Larnaca, pp. 134-138, 2009.

[36] Zhu G., Zhang S., Zeng Q., and Wang C., "Boundary-based Image Segmentation using Binary Level Set Method," *Optical Engineering*, vol. 46, no. 5, pp. 050501, 2007.

Sadagopan Padmakala has received the B.E Degree, from the Department of Computer Science ,Bharath Institute of Technology, University of Chennai, Chennai, India and M.E degree from the Department of Computer Science Anna university, Chennai, in 1997 and 2006 respectively. She is currently pursuing the Ph.D degree in Anna university, Chennai, working closely with Dr.G.S.Anandha Mala. Presently, she is working as Associate Professor, at St.Josephs's Institute of Technology, Chennai, India.

Ganapathy AnandhaMala received B.E degree from Bharathidhasan University in Computer Science & Engineering in 1992, M.E degree in University of Madras in 2001 and Ph.D degree from Anna University in 2007. Currently she is working as Professor in Easwari Engineering College, Chennai, India, and heading the department of Computer Science and Engineering. She has published more than 40 technical papers in various international journal / conferences. She has 20 years of teaching experience on graduate level. Her area of interest includes Image Processing and Grid Computing.