

A Content Based Video Retrieval Analysis System with Extensive Features by Using Kullback-Leibler

Priya R.

*Research Scholar, Department of Mathematics
Anna University, Chennai, Tamil Nadu, India.
priyarajendranphd@gmail.com, priyacrs@yahoo.com*

Dr. Shanmugam T.N.

*Professor, Department of Mathematics
Anna University, Chennai, Tamil Nadu, India.*

Dr. Baskaran R.

*Associate Professor, Department of Computer Science and Engineering
Anna University, Chennai, Tamil Nadu, India.*

Received 24 November 2011

Accepted 27 July 2012

Abstract

Content-based video retrieval systems have shown great potential in supporting decision making in clinical activities, teaching, and biological research. In content-based video retrieval, feature combination plays a key role. As a result content-based retrieval of all different type video data turns out to be a challenging and vigorous problem. This paper presents an effective content based video retrieval system, which recognizes and retrieves videos with three different types of visual effects. The raw video information is divided into shots and also the object feature, movement feature and also the occlusion options are extracted from these shots and also the feature library is used for the storage method of those options. Advanced on, the Kullback-Leibler distance is computed among the options of the feature library and also the options of the question clip that's extracted within the similar manner. The results show that it is possible to improve a system for content-based video retrieval by using Kullback-Leibler distance model, which takes careful consideration of the structure and distribution of visual features. Hence the final results with the aid of the Kullback-Leibler distance the similar videos are extracted from the collection of videos based on the given query video clip in an effective manner.

Keywords: Video Retrieval, Content-based Video Retrieval (CBVR), Video sequence, shot segmentation, object feature, movement feature, Feature Extraction, Video Indexing, Video Retrieval, kullback-Leibler distance.

1. Introduction

Content Based Image Retrieval (CBIR) is one of the technology that, principle helps to organize digital image collections by their visual content. By means of this explanation, everything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR. An image search based on visual is the most common form of CBIR [1]. A phase of visual information retrieval that has been mostly mistreated is the thought of changing clarifications of visual content by users [3]. As video extends into both the spatial and the temporal domain,

we need techniques for the sequential decomposition of footage so that exact content can be read [4] [2]. Visual information is generally perceived and construed differently by individual users because pictorial demonstration of information is frequently less specific than its textual representation [7] [8]. The growing amount of digitally produced images wants new methods to records and right to use this data. Conventional databases sanction for textual searches on Meta data only.

Video segmentation is first step to the content based video search directing to segment moving objects in video sequences. On the idea of the video contents

through user interactions, content-based video retrieval will with competence aid users to retrieve most well-liked video segments from a large video database [6]. To one side from the usual features like color and texture, a different feature extraction algorithm called edge histogram is introduced. Edges carry essential information to a picture and therefore can be applied to image retrieval. The edge histogram descriptor internments the spatial distribution of edges. A color space is defined as a model for instead of color in terms of intensity values. Normally, a color space defines a one- to four- dimensional space [13]. Though, video will simply become an operative part of everyday computing environments when we can use it with the same ability that we currently use text [10]. Thus, CBVR is important information is automatically removed by using signal processing and pattern recognition techniques for audio and video signals, have been recently widely studied. A color element, or a color channel, is one of the dimensions. Color spaces are connected to each other by mathematical formulas. The two three-dimensional color spaces, RGB and HSV, are inspected [9]. CBIR contains the following four parts in system realization: data collection, build up feature database, search in the data.

Video segmentation splits a video file into shots that is illustrated as Associate in treatment conterminous sequence of video frames that square measure recorded from one camera operation [11]. Edge data primarily based video segmentation, image segmentation video segmentation and alter detection based video segmentation is that the three sorts of Video segmentation [14]. The Feature Extractor may be a period system, is utilized to preprocess all the videos that square measure keep within the info and conjointly store their distinctive options for faster retrieval [9]. For locating contents like a question video stream from video info feature matching is utilized [15]. To represent applicable objects that seem in those shots, the shot options square measure sorted into clusters within the grouping stage [5].

As numbers of pixel values, histogram captures useful characteristics of images but disregards the comparative spatial positions in the images. In other words, histogram does not internment spatial patterns that are imperative to describe image content [17]. To

overcome this restriction, we use histograms of numerous spectral components of an image as the signature, which are used in texture analysis and synthesis [19]. The figures of spectral components hold an excessive quantity of data of images 'marginal distributions and texture patterns, hence providing great discriminating power. Color feature is the most intuitive and obvious feature of the image, and generally adopt histograms to describe it [18]. Color histograms method has the advantages of speediness, low demand of memory space and not sensitive with the images' fluctuations of the size and rotation, it wins wide consideration accordingly. A frame which will characterize the necessary content of a video shot is understood as a video frame. The frames square measure directly extracted from the video sequence; thus the additional machine overhead is not necessary compared with shot-based frame extraction techniques. Video extraction permits quick video water sport and it conjointly offers a strong tool for video content branding and visualization.

Normally, movement feature and object feature plays a categorically significant role in retrieving the similar video clips. However, to enhance the performance of the retrieval, the options have to be compelled to be increased. The prevailing techniques extract the options by completely different means that. During this paper, we have a tendency to enhance the (i) object feature by extracting the special data of the detected object and (ii) movement feature by recognizing the direction of movement of the feature of the subjected video clip. In addition, we have a tendency to extract the occlusion feature for any improvement within the performance of the system.

As a pre-processing step, the raw video material is segmented into frames. Then the frames area unit sorted to their corresponding shots for the consecutive processes. After, for the retrieval of the video, supported the given question video, the options area unit removed supported the article, movement and also the obstruction. So as to get the amount of objects bestowed within the shots, the Fuzzy K means that agglomeration is employed and to spot the placement of the objects the special features additionally combined with it so the movement feature is extracted from the video.

As updated previous paper, the movement feature is principally attentive on the posture and movement of the humans bestowed within the video, the direction and distance is known. Subsequently, the constriction feature is known and so of these options of a video square measure combined as a feature set within the feature library. Inside the planned system, the videos square measure repossessed supported the question clip. For the question video clip the said options object, movement and also the occlusion options square measure extracted and compared with the feature within the feature library. The comparison is achieved via the Kullback-Leibler distance similarity live. Then the similar videos square measure retrieved from the gathering of videos.

The rest of the section of the paper is organized as follows. A related researches review of a few of the existing works in content based video retrieval is presented in Section 2. The proposed effective content-based video retrieval system is detailed in Section 3. The results and discussion are described in Section 4. The conclusions are summed up in Section 5.

2. Review on Related Researches

A motion-based video retrieval technique was proposed by R. Kanagavalli *et al.* [21] have proposed a method in video based content; it was gowned in volume and there was different techniques available to capture, compress, display, store and transmit video while editing and manipulating video based on their content was stilled a non-trivial activity. The proposed surveys various current trends and methods used in video retrieval. The different areas covered in this study are shot segmentation, current approaches used in video retrieval, use of clustering techniques in video retrieval, the video retrieval system, measures used for video retrieval.

Shradha Gupta *et al.* [20] have proposed a method for Video Retrieval system; there video had become an important element of multimedia computing and communication environments, with applications as varied as broadcasting, education, publishing and military intelligence. In Video Retrieval system, each video that is stored in the database has its features extracted and compared to the features of the query image. Proposed worked was to retrieve video from the database by giving query as an object. Video is firstly converted into frames, these frames are then segmented

and an object is separated from the image. Their features was extracted from object image by using SIFT features. Features of the video database obtained by the segmentation and feature extraction using SIFT feature are matched by Mahalanobis Distance. In this paper experiments was done on different samples of video.

Weiming Hu *et al.* [19] have proposed a method for video indexing and retrieval have a wide spectrum of promising applications, motivating the interest of researchers worldwide. In this paper had proposed an overview of the landscape of general strategies in visual content-based video indexing and retrieval, focusing on methods for video structure analysis, including shot boundary detection, key frame extraction and scene segmentation, extraction of features including static key frame features, object features and motion features, video data mining, video annotation, video retrieval including query interfaces, similarity measure and relevance feedback, and video browsing.

B V Patel *et al.* [16] have proposed a method for development of multimedia data types and available bandwidth there was huge demand of video retrieval systems, as users shift from text based retrieval systems to content based retrieval systems. 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. The proposed method was to survey reviews the interesting features that could be extracted from video data for indexing and retrieval along with similarity measurement methods. Hence, identify presented research issued in area of content based video retrieval systems.

Xiao-Ming Chen *et al.* (Chen, Lu and Zhen, 2008) have proposed a Vector Quantisation (VQ)-based video retrieval algorithm by means of global motion features. Their contribution has comprised of two points: first, they designed a VQ-based algorithm for getting rid of the singular points in the motion vector space; second, they have made use of the global motion vector index histogram of all frames in the query video clip to match those video clips in the database. The experimental results have illustrated that their algorithm could efficiently extract the statistical characteristics of the global motion features.

Dyana A. *et al.* (Dyana, Subramanian and Das, 2009) have proposed a system for CBVR based on

shape and motion features of the video object. They have utilized Curvature scale space for shape representation and Polynomial curve fitting for trajectory representation and retrieval. The shape representation has been invariant to translation, rotation and scaling and robust to noise. Trajectory matching has incorporated visual distance, velocity dissimilarity and size dissimilarity for retrieval. The cost of matching two video objects has been on the basis of shape and motion features, in order to retrieve alike video shots. They have tested their system on standard synthetic databases. They have also tested their system on real world databases. Experimental results have illustrated good performance.

Ionuț Mironica et al [18] have proposed a technique for Content-based retrieval in video databases had become an important task with the availability of large quantities of data in both public and proprietary archives. Most of video systems are based on feature classification, but problems appear because of “semantic gap” between high-level human concepts and the machine-readable low-level visual features. In this paper had proposed a technique in relevance feedback approach (RF) to bridge the semantic gap by progressively collecting feedback from the user, which was allowed the machine to discover the semantic meanings of objects or events compared to classic classification-based retrieval.

In Priya Rajendran and Shanmugam [30] have to proposed retrieval system for improve the efficiency an adaptive system with broad features. In this proposed system segmented a video into shots, and then a few characteristic frames were generated from each of the shot on the basis of different features such as color, contour, and texture and motion these frame descriptors were calculated for these shots and were stored in a feature library. When a query for the clip has been given, the features mentioned above were extracted for the query clip and then were compared with the features that are already stored in the feature library. A technique called Latent semantic indexing (LSI) on the basis of similarity measure was used to perform the comparison. Finally, videos similar to the query are obtained from the massive collection of videos. In this system has been estimated with precision-recall and F score technique and we have compared with existing retrieval systems.

3. Proposed Visual Content Based Video Retrieval system

In this section, we offer an efficient CBVR system which retrieves with some different type of videos based on an extensive feature set. The video database, which is a collection of cartoon video, 3D-animated video data. Our proposed model is represented in this section, which illustrates the fundamental operations performed in the system for the retrieval of video clips based on the data.

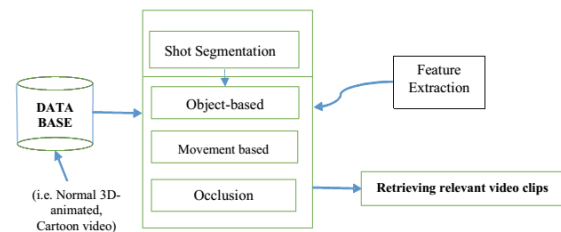


Fig. 1. Proposed method of CBVR system

In previous work we were used in individual video for segmentation and tracking of the objects with carried out throughout the frames for every shot. But, now we are using a three different type of high resolution video for segmentation and features tracking with different object. The primary step in our system is to partition a shot segmentation of cartoon video, 3D-animated video and Normal video respectively. We are structured according to a descending hierarchy of video clips, scenes, shots, and frames. Video structure analysis aims at segmenting a video into a number of structural elements that have semantic contents, including shot boundary detection, key frame extraction, and scene segmentation. As a next step, we are extracting four different types of features, namely, motion, texture, color and contour for all the video frame. After that the extracted features are stored in the feature library. Then the same features (above mentioned) are extracted for a query clip (single clip) and are compared with the features in the feature library. Assistance of LSI similarity measure is used to carry out the comparison. As a final point based on the LSI, the videos are retrieved from the videos collection.

3.1 Shot Segmentation

Dividing the whole video into a number of temporal segments is called shots. A shot may be defined as an incessant sequence of frames generated by a single nonstop click operation. However, from the semantic point of view, its lowest level is a frame followed by shot followed by shot for the whole video. To this method any video, or to extract the options from the

video, it's necessary to section the video by suggests that of various shots. In our segmentation approach, we have divided each frame into several numbers of blocks and DCT will be applied to every block of the frame. The DCT calculation is performed as given in equation (1).

After this calculation, we have DCT values for all the blocks. Hence the vector is formed as $B_{(u,v)_i}^{(j)}$ which is constituted by the DCT values of $m_b \times m_f$ blocks where $i = 1, 2, 3, \dots, m_b$ and $j = 1, 2, 3, \dots, m_f$.

$$S_{(u,v)_i}^{(j)} = \left(\frac{2}{a}\right)^{1/2} \left(\frac{2}{b}\right)^{1/2} \beta_u \beta_v \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} e_{(x,y)_i}^{(j)} \cos \left[\frac{\pi u}{2a} (2x+1) \right] \cos \left[\frac{\pi v}{2b} (2y+1) \right] \quad (1)$$

$$\text{where, } \beta_u = \begin{cases} \left(\frac{1}{2}\right)^{1/2} & \text{if } u = 0 \\ 1 & \text{if } 1 \leq u \leq m-1 \end{cases} \quad (2)$$

$$\beta_v = \begin{cases} \left(\frac{1}{2}\right)^{1/2} & \text{if } v = 0 \\ 1 & \text{if } 1 \leq v \leq n-1 \end{cases} \quad (3)$$

Then as the next step of shot segmentation, Euclidean distance is calculated between two blocks of consecutive frames. This can be performed as follows

$$[F_{ed}^{(j)}(i)] << \sqrt{\sum_{u=0}^{M-1} \sum_{v=0}^{N-1} (S_i^{(j)}(u,v) - S_{i+1}^{(j)}(u,v))^2} \quad (4)$$

By equation (2), a Euclidean distance for each block will be calculated. Then for a single frame, average of Euclidean distance of m_b blocks will be calculated as

$$[M(F_{ed}^{(j)})] = \frac{1}{m_b} \sum_{i=1}^{m_b} F_{ed}(i) \quad (5)$$

Based on the criterion described in the equations (4) and (5), the segments are obtained on the basis of shot change. Each frame is tested with the above mentioned criteria sequentially and then on the basis of the criteria results, segmentation of the video clips based on the shots change is obtained. Now, we have the n_s number of segmented video clips based on the DCT combined with Euclidean distance. These segmented frames will be subjected to the next process of Motion object detection, a process of extracting the moving object from a frame comprised of motion and still object. Therefore, different shots are mined from the subjected video and so $N_s^{(i)}$ shots are attained from every i^{th}

video clip. The attained shots are also an assortment of frames and on the basis of shot, the further process of feature extraction is performed.

3.2 Object based Feature extraction

The work includes extraction of object feature from the video sequence is proposed by [5], an object based video retrieval system. Though, it shows inaccuracy because it does not consider the spatial feature i.e. the location of the object is not taken into account. In this proposed system, we considers the spatial feature and overcomes the aforesaid practical issues. In order to extract the object feature, the frames are identified and assembled to their corresponding shots $s_a, a = 0, 1, 2, \dots, N_s^{(i)}$ and the number of frames in a shot

S_a is determined as $|S_a|$. The shot s_a contains $|S_a|$ frames; the initial frame $f_0^{(i)}$ is assumed as the key frame $f_{key}^{(i)}$ and each shot S_a has its own key frame.

Initially, the frames $f_j^{(i)}$ of shot S_a are in the RGB color is converted to the gray scale component

$$f_{j_{gv}}^{(i)} = 0.2989 * f_{j_r}^{(i)} + 0.5870 * f_{j_g}^{(i)} + 0.1140 * f_{j_b}^{(i)} \quad (6)$$

The developed equation is the Craig's formula for changing RGB color video to grey scale video. Then the cluster method is applied to the born-again grey scale key frame pictures for object identification. The above equation is the Craig's formula for converting RGB color video to gray scale video. Then the clustering process is applied to the converted gray scale key frame images for object identification. The video is dynamic so a static K value cannot be obtained. Hence, the number of objects for the clustering process is determined using the 3D color histogram. After converting the RGB color space $f_{key}^{(i)}$ to the LAB color space, the 3D color histogram is determined for the key frame $f_{key}^{(i)}$. From the obtained

3D color histogram of the selected key frame $f_{key}^{(i)}$, the number of peaks K is also the number of objects is identified. Thus, K clusters are obtained by the clustering process. Fuzzy k means clustering that group facilitates the identification of overlapping teams of objects as a result of it permits the objects to possess membership in additional than one cluster. The pseudo code for the Fuzzy k means clustering that bunch is shown below.

Input: $f_{j_{gv}}^{(i)}$ in a the shot s_a converted into gray scale

K- Number of clusters

Output: set of K clusters

Step 1: For each $f_{j_{gv}}^{(i)} \rightarrow f_{j_{gv|s_a}}^{(i)}$

Select K (pixels P_1, P_2, \dots, P_K from $f_{j_{gv}}^{(i)}$)

Upto no changes in K

Step 2: Use the estimated means to find the degree of membership

$\mu(y, x)$ of $f_{j_{gv}}^{(i)}$ in Cluster η

Step 3: For η from 1 to K

Replace P_η with the fuzzy mean.

For Cluster η

$$P_\eta = \frac{\sum_y \mu(y, x)^2 \cdot f_{j_{gv}}^{(i)}}{\sum_y \mu(y, x)^2}$$

End for

End until

Yu76 End for

Hence K clusters are identified from the gray scale converted frames of the shot s_a .

Track Frame Selection

The extracted clusters of the key frame $f_{key}^{(i)}$ are the objects and they are compared with the other frames of the shot for their presence in the frame. If the object is presented in the remaining frames in the shot then the

corresponding frame index is stored in the $TF_{id}^{(i)}$. This vector contains the index of the track frame index that is having at least a single object. This track frame selection is utilized to reduce the computational time because a frame is said to be a track frame at least a single object in presented in that particular frame. Hence in order to extract the object based features, rather than analyzing all the frames these track frames are analyzed. The following pseudo code details the selection of track frames.

Track Frame Selection

Step 1: For each $f_{j_{gv}}^{(i)}$ to $f_{j_{gv|s_a}}^{(i)}$

Step 2: For $\forall P_\eta, \eta = 1, 2, \dots, K$

Step 3: For $\forall P_\eta, \eta = 1, 2, \dots, K$

Step4:

$$L2_{pq} = \sqrt{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left(P_\eta'(x, y) - P_\eta(x, y) \right)^2}$$

Step 5: If $L2_{pq} \leq 0.6$ then

$$TF_{id}^{(i)} = j$$

End if

End for

End for

End for

Cluster Grouping

After the selection of track frames, the clusters are grouped by analyzing the track frames and then these clusters presented in the remaining frames append to their corresponding clusters. Hence, the K clusters are grouped and the cluster group is used for the further process. The pseudo code shows below details the process.

Cluster Grouping

Step 1: For each $f_{TF_{id}^{(i)}}^{(i)}, id = 1, 2, \dots, N_{TF_{id}^{(i)}} d$

Step 2: For $\forall P_\eta, \eta = 1, 2, \dots, K$

Step 3: For $\forall P_\eta, \eta = 1, 2, \dots, K$

$$L2_{pq} = \sqrt{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left(P_\eta'(x, y) - P_\eta(x, y) \right)^2}$$

$$G_\eta^{(i)} = P_\eta'(x, y)$$

Step 4: If $L2_{pq} \leq 0.6$ then

$$G_\eta^{(i)} \cup P_\eta(x, y)$$

End if

End for

End for

End for

Then the cluster group $G_{\eta}^{(i)}(x, y)$ for the shot S_a and the covariance for this cluster group are computed.

$$C_{s_a}^{(i)}(x, y) = \sum [(x - E(x))(y - E(y))] / (n - 1) \quad (7)$$

where, $E(x)$ is the mean of X and $E(y)$ is the mean of the y value.

Let $Ob_{s_a}^{(i)}(w)$ Hence the object feature is extracted and also the problem during this feature is that it cannot predict the placement of the thing where the thing is known it marks as a feature. Thence the abstraction feature is additionally extracted. The object feature and also the abstraction feature area unit combined and used because the feature set for the any method.

Spatial Feature Extraction

The spatial feature is take out to identify the location of the object. The following pseudo code details the process of obtaining the appropriate objects to be recognized the spatial information.

Spatial Feature Extraction

Step 1: For each $TF_{id}^{(i)}, id = 1, 2, \dots, N_{TF_{id}^{(i)}}$

Step 2: For each
 $Ob_{TF_{id}^{(i)}}^{(i)}(w), w = 1, 2, \dots, N_w$

Step 3: If
 $size(Ob_{key}^{(i)}(w)) > size(Ob_{TF_{id}^{(i)}}^{(i)}(w))$
Resamplesize of $Ob_{key}^{(i)}(w)$

Else
Resamplesize of $Ob_{TF_{id}^{(i)}}^{(i)}(w)$

End if

Step 4: $D_{(w)}^{(i)} = Ob_{key}^{(i)}(w) - Ob_{TF_{id}^{(i)}}^{(i)}(w)$

Normalize $D_{(w)}^{(i)}$ and select minimum $D_{(w)}^{(i)}$

End for

End for

In spatial feature extraction order to complete the spatial information, the clustered key frame $f_{key}^{(i)}(x, y)$ is re-sampled and then $M \times N$ frame is created with each pixel of 0 values ($M = N = 128$). After that the frame is alienated to blocks of $M/4 \times N/4$. Every objective $D_{(w)}^{(i)}$ is placed on the empty white pixel

frame. The quantity of 1's in each block is analyzed and then the index of the block is of maximum number of 1's, is stored in the vector $sp_{s_a}^{(i)}(K)$. Hence the object feature is also combined with the spatial feature and utilized for the subsequent processes. $C_{s_a}^{(i)}(x, y)$ And $sp_{s_a}^{(i)}(K)$ are the extracted object features.

3.3 Movement-based feature extraction

An improved object's movement organization is planned to get rid of the object's movement-based feature from the video sequence. In [28], a posture organization has been planned for constant. The trouble reflects the posture and movement of solely human's gift within the video sequence sometimes. On the opposite hand, it shows impreciseness as a result of it doesn't suppose through the placement of movement i.e. whether or not the human is acquiring front of the screen or removed from the screen and approach of movement i.e. whether or not the movement is from left to right or from right to left further as whether or not the movement is towards the screen or removed from the screen. The above-named sensible problems area unit thought-about and that they area unit awestruck within the planned system. The planned system removes the options like skeleton, center of mass context, object orientation and direction of movement from a middle object. Earlier to extract the options, the video shots area unit subjected to foreground segmentation victimization minimum graph cut algorithmic rule [29]. The foreground divided frames area unit subjected to get rid of the subsequent options.

All the obtained objects are subjected to triangulation so that the objects are filled up with networked triangles. For each triangle, a centroid is determined. The triangulated objects with centroids are given in Fig. 2.

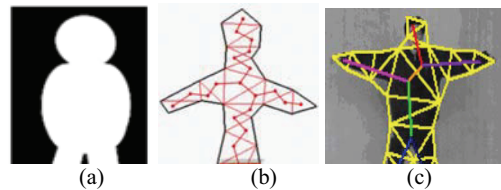


Fig. 2. Extraction of centroid feature: (a) foreground segmented frame, (b) triangulated and centroid marked frame and (c) skeletonized frame.

Hence for each frame, the foreground objects are strong-minded and the centroids are determined. The coordinates of the obtained centroids are stored as centroid features for the particular video clip.

Extraction of orientation and distance of object movement feature:

The distance of object movement is outlined here that at that distance from the screen, the thing exhibits movement within the video whereas orientation is outlined here that the kind of movement of the thing (very specifically humans). So as to accomplish this, the skeleton of the thing is obtained from the Fig. 1. With the reference of the skeleton, the axis of the thing is decided. The obtained skeleton and also the axis formation square measure portrayed in fig. 3.

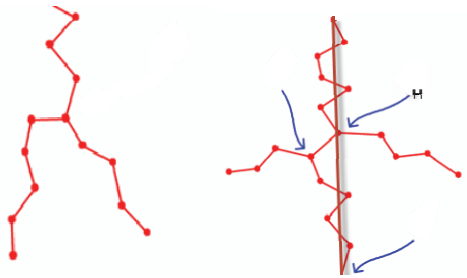


Fig. 3. Extraction of distance of movement: (a) skeleton of the object and (b) axis formation.

From the obtained axis, i.e. altitude of the object, the decision is made. When the axis satisfies the condition, $H \geq H_{TH}$; H is the axis, then it is decided that the object is moving in front of the screen. Otherwise, the object is moving away from the screen at a greater distance. The location is determined by calculating the angle between the axis of the object and ground plane.

Extraction of Direction of movement: The foreground segmented frames are square measure thought-about to extract the direction of movement of the frames. The segmented frames square measure subjected to derive the direction of movement. The direction of movement is obtained by analyzing all the frames gift during a shot in four eventualities. Prior to execute the scenarios, row and column origins D_{xy}^R and D_{xy}^C , respectively are determined as follows

$$D_{xy}^R; \text{ if } F_{xy}(j, i) = 1 \quad (8)$$

$$D_{xy}^C; \text{ if } F_{xy}(i, j) = 1 \quad (9)$$

where, $x, b = 0, 1, 2, \dots, N_f^{(a)} - 1$ and $F_{xy}(i, j)$ is the foreground segmented frame of a shot.

State 1: To determine left to right movement

In this State, only D_{xy}^C is used. If the D_{xy}^C during the course of a shot is in ascending, then it is declared that the movement is from left to right. Once the left to right movement is determined, a direction feature vector is loaded with a binary value 1000 i.e. $D^{feat} = \{1000\}$.

State 2: To determine right to left movement

Similar to State 1, only D_{xy}^C is used. If the D_{xy}^C all through a shot is in descending, then it is declared that the movement is from right to left. Once movement is determined, the direction feature vector is loaded with 0100 i.e. $D^{feat} = \{0100\}$.

State 3: To determine towards the front

In this State, D_{xy}^R is used. If the D_{xy}^R all through a shot is in ascending, then it is declared that the movement is towards front of the screen. Once movement is determined, the direction feature vector is loaded with 0010 i.e. $D^{feat} = \{0010\}$

State 4: To determine away from front

In this State, D_{xy}^R is used. If the D_{xy}^R all through a shot is in descending, then it is declared that the movement is away from front of the screen. Once movement is determined, the direction feature vector is loaded with 0001 i.e. $D^{feat} = \{0001\}$

Hence, established on the object movement, features such as centroid feature, distance of movement feature, orientation feature and direction of movement are extracted and stored in the feature library.

3.4 Occlusion Feature Extraction

In [32], the occluded object's boundaries are detected within the gray-scaled video frames. Changing the video frames from color to grey scale could be a time intense method. Thus the aforementioned downside is overcome in our projected system. Occlusion that happens between two and a lot of objects

provides lean info regarding object structure and form i.e. if the article is retrieved from the occluded objects, then the form details are going to be inconsistent.

The occlusion feature is detected and recognized by processing the consecutive frames of every shot. The proposed occlusion detection is comprised of three major steps. They are as follows: i) Trigon formation, ii) Determining spatial and temporal derivatives, iii) Scum computation

The major steps for detecting the occlusion feature for prior to perform, all the frames are converted from RGB color space to gray scale frame. The gray scale frames are processed further to control the occlusion affected area.

3.4.1. Trigon formation

As primary step, a trigon is formed by resizing the frames into different sizes and processed as per the size of the frame. The frame of size $M \times N$ is initially resizing into a frame of size $\frac{M}{2} \times \frac{N}{2}$. The similar process is repeated by resizing the frame of size $\frac{M}{2} \times \frac{N}{2}$ into $\frac{M}{4} \times \frac{N}{4}$. Repeating the process produces the frames with different sizes. When it is arranged in ascending order based on size, it forms a trigon as follows

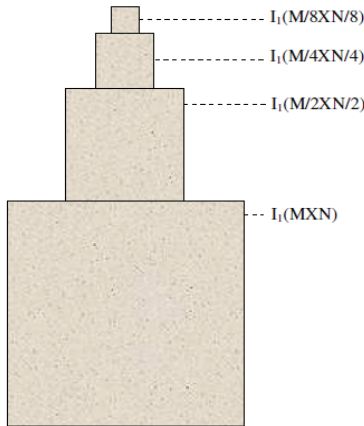


Fig. 4. Trigon formation

To achieve this resizing, a structure-texture decomposition model which has been proposed in [29]. As per the model, the resizing is performed and finally the trigon of frames is formed.

3.4.2. Determining Spatial and temporal derivatives

Spatial by-product is defines the modification of image intensity values as per the modification exhibits in image position. In different words, particularly in videos, the special by-product is capable in decisive the modification of intensity values once an object exhibits modification in position in two frames. The special by-product may be expressed as

$$I'(t, x) = \frac{\partial}{\partial x} I(t, x) \quad (10)$$

The special by-product is obtained solely within the filtered version of the frame, wherever the filtering is performed mistreatment replicate filtering technique.

The temporal by-product refers the modification of intensity values of a picture with regard to a sampling instant of your time. It are often outlined in alternative words as modification of component intensity values in two totally different frames. Two totally different completely different frames ask the frames that are taken at two different instants of your time. The temporal by-product are often given as

$$I'(t, x) = \frac{\partial}{\partial t} I(t, x) \quad (11)$$

Thus attained spatial and temporal derivatives from two completely different frames square measure used additional in determinative occlusion. The temporal derivate square measure obtained when applying linear interpolation between two frames

3.4.3. Scum Computation

The directories of the special derivatives and also the temporal derivatives square measure thought of in decisive the scum of the frames. The scum, here, we have a tendency to represent it because the element indices exhibit motion is set. The scum is set by finding the element indexes with borderline distinction between the element locations of the special spinoff yet because the temporal derivatives. The scum is nothing the occluded portion of the article that's below motion or occluded by the article below motion. This can be finally marked and also the positions of the occluded pixels and also the locations square measure hold on as feature set of the corresponding frame. The feature set is hold on within the feature library for the additional method of the retrieval stage.

3.5 Retrieval of relevant Video clips

The on top of explained video clips are subjected to the feature extraction system and then the most important feature sets, object-based feature set, movement-based feature sets and occlusion feature sets are extracted. The extracted options are keep within the feature library. Once a question clip is given to the CBVR system, the clip is subjected to the feature extraction method and every one the said feature sets are extracted within the similar fashion. Then, every feature set of the info video is analyzed by mensuration its similarity with the feature set of the info videos. This may be accomplished with the help of Kullback Leibler distance (KLD).

KLD based similarity measure: In [33], the KLL based similarity measure has been utilized to compare the sparse multistate image representations.

The similarity between the query video clip and each database video clip is determined by calculating KLD between the feature sets. The KLD can be determined as

$$KLD(F^{query}, F^{data}) = \sum_{z=0}^{|F^{data}|-1} F_z^{query} \log_2 \left(\frac{F_z^{query}}{F_z^{data}} \right) \quad (12)$$

where, $|F^{data}|$ is the feature set determined for the information video clip. Combining Eq. (11) determines the KLD between the feature set of the actual question clip and therefore the feature set of an information video clip. In similar technique, KLD is decided for all the opposite information video clips. A necessary range of information video clips with minimum KLD, are retrieved because the relevant videos for the given video clip. So with the help of the projected CBVR system, relevant videos are retrieved from the information with success

4. Implementation Results and Discussion

In pervious paper, we are taken only 2D normal video for segmentation and validation. Now we have to take three different type of video for segmentation and validation with better solution. The proposed system was implemented in the MATLAB (version 7.8) and the system was validated with the three different datasets that are available in <http://www.open-video.org/>. From this site, the validation datasets, namely, cartoon, 3D-animated and normal game video datasets were obtained with a scale of 60, 40 and 30, respectively. We

have attached some of the intermediate results obtained at the end of each step, by our system when a clip of cartoon, 3D- animated and normal game dataset was applied.

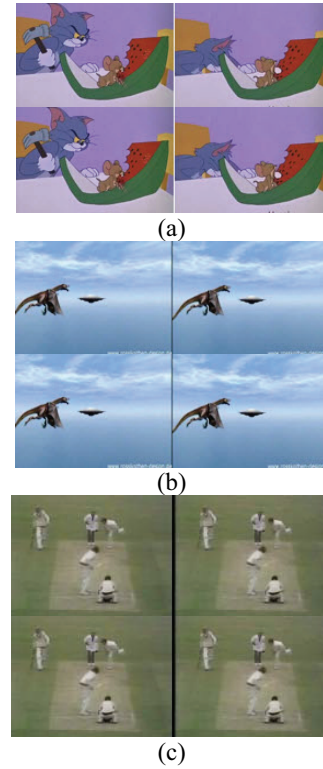
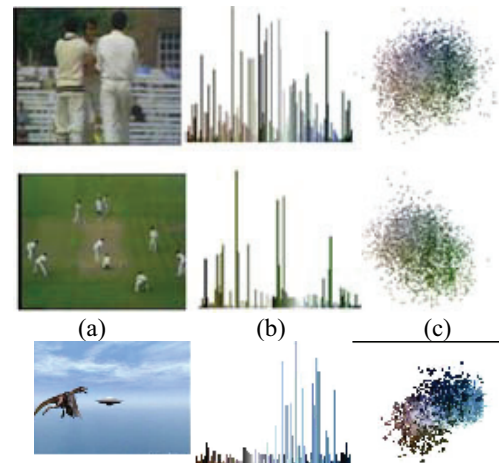


Fig.5.1. (a) (b) & (c) Elementary frames of 3 different shots segmented from 3 different video sequences



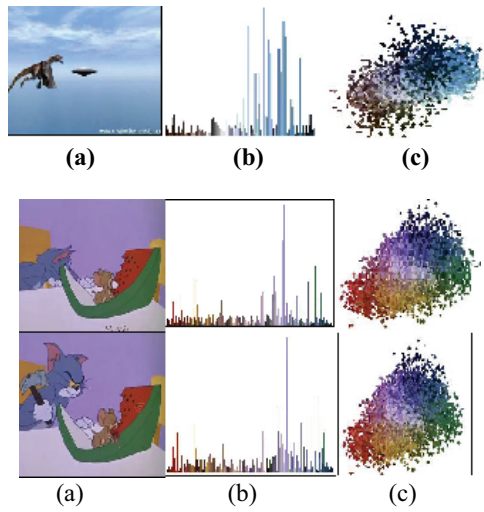


Fig. 5.2. (a) keyframe in RGB colorspace, and their corresponding (b) histogram, (c) 3D color histogram (d) LAB colorspace of keyframe, and their corresponding (e) histogram (f) 3D color histogram

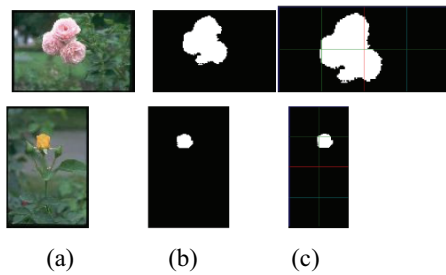


Fig. 5.3 spatial feature extraction (a) keyframe corresponding (b) re-sampled frame (c) re-sampled frame alienated into blocks

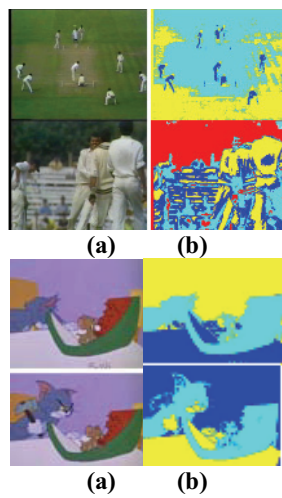


Fig. 5.4. (a) Keyframes and their corresponding (b) clustered frames

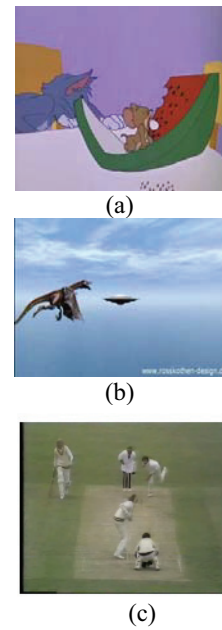
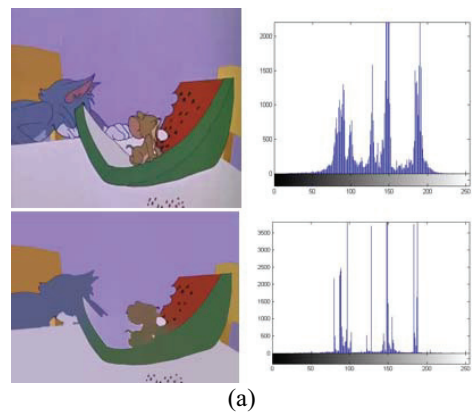


Fig. 5.5. Sample frames and their corresponding color segmented frames of three different video



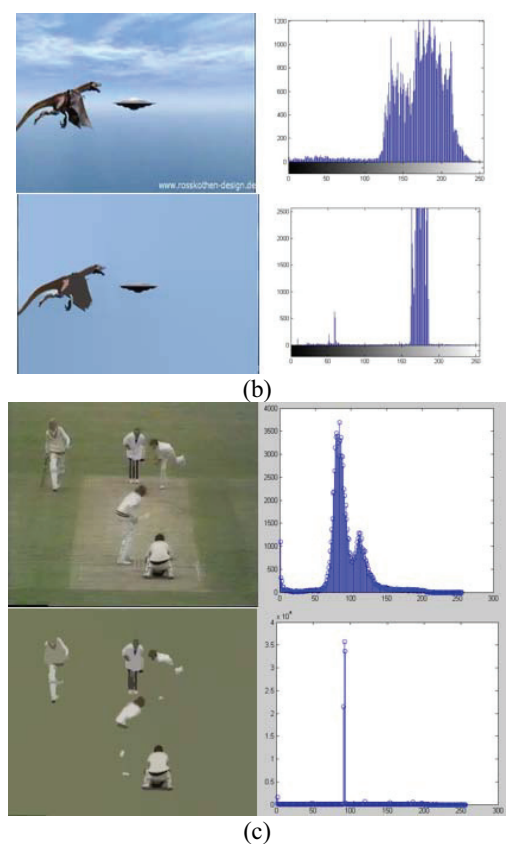


Fig. 5.6. Color histogram for normal frames and color segmented frames of (i.e. Cartoon, 3D-animated and 2D game video)

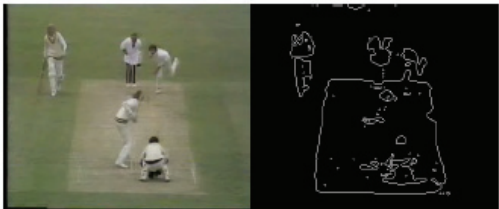
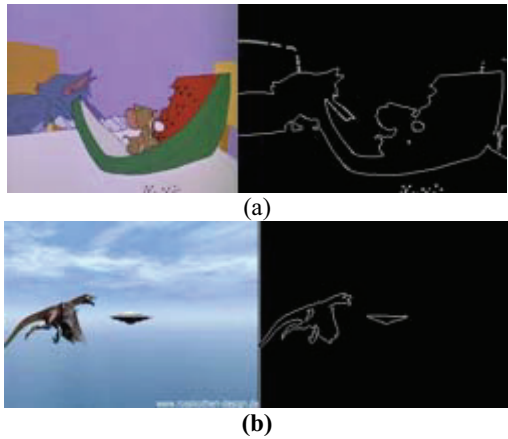


Fig. 6 (a) Results of the proposed CBVR system over games database: (i) input query clip and (ii) some of the retrieved video clips.



Fig. 6.1 (a) Results of the proposed CBVR system over games database: (i) input query clip and (ii) some of the retrieved video clips.



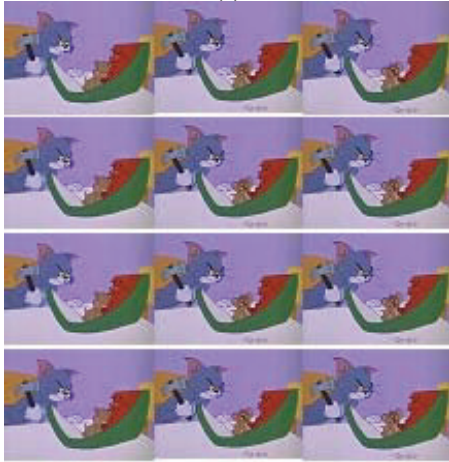


(ii)

Fig. 6.2 (b) Results of the proposed CBVR system over 3D animation database: (i) input query clip and (ii) some of the retrieved video clips



(c)



(ii)

Fig. 6.3 (c) Results of the proposed CBVR system over cartoon database: (i) input query clip and (ii) some of the retrieved video clips

Figure 6 shows some sample frames of the segmented shots which were obtained from a three different video clip. Figure 6.1, 6.2 and 6.3 illustrates extracted objects which exhibits motion in the shots obtained after the application of morphological operator, RGB to YCbCr color space conversion output

for Gaussian mixture model, some sample frames and their corresponding color segmented frames based on Anisotropic diffusion and color histogram for the shots respectively for all three different types of video. Figure 7 depicts the object contour of some sample frames of the shots. Figure 8 details some input query clips and the corresponding video clips retrieved by the proposed content based-video retrieval approach.

$$\text{precision} = \frac{\text{No. of retrieved videos that are relevant to the query clip}}{\text{Total no. of retrieved videos}} \quad (13)$$

$$\text{recall} = \frac{\text{No. of retrieved videos that are relevant to the query clip}}{\text{Total no. of available videos that are relevant to the query clip}} \quad (14)$$

$$F_{\text{score}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (15)$$

The performance of the proposed CBVR system was evaluated by determining the precision-recall values, which was calculated from equation (13) and (14) and F-score by equation (15). The proposed system was compared with CBVR system with motion pattern (Ma and Zhang, 2003), with motion feature and RGB color model (Wen, Chang and Li, 2007), and with histogram features (Lee, Kotani, Chen and Ohmi, 2010). It is compared with normal and 3D- animated and cartoon video. The comparative and mean precision-recall and the F-score values are given in Table 1, 2 & 3 for all three different type of format. In figure. 9, 10 & 11 depicts the performance of the proposed system over the existing systems.

Table. 1. Precision-recall and F-score values for CBVR system (i) with Motion pattern (ii) with motion feature and RGB color model (iii) with histogram features and (iv) with extensive feature set for normal video.

(i)				
Video Data set	Round No.	Precision	Recall	F-score
1	1	0.5	0.08	0.1
	2	0.4	0.133	0.20
	3	0.33	0.21	0.26
	4	0.28	0.28	0.28
2	1	0.2	0.03	0.05
	2	0.3	0.1	0.15
	3	0.35	0.23	0.28
	4	0.45	0.3	0.39
3	1	0.3	0.075	0.12

2	0.25	0.13	0.17
3	0.2	0.15	0.17
4	0.18	0.18	0.18

(ii)

Video Data set	Round No.	Precision	Recall	F-score
1	1	0.6	0.1	0.17
	2	0.55	0.18	0.27
	3	0.5	0.3	0.40
	4	0.42	0.42	0.42
2	1	0.4	0.06	0.10
	2	0.5	0.13	0.21
	3	0.35	0.23	0.28
	4	0.45	0.33	0.38
3	1	0.5	0.125	0.2
	2	0.5	0.25	0.34
	3	0.4	0.3	0.34
	4	0.38	0.38	0.38

(iii)

Video Data set	Round No.	Precision	Recall	F-score
1	1	0.7	0.12	0.20
	2	0.65	0.21	0.32
	3	0.55	0.36	0.44
	4	0.5	0.5	0.5
2	1	0.6	0.1	0.18
	2	0.5	0.16	0.24
	3	0.45	0.3	0.36
	4	0.4	0.4	0.4
3	1	0.6	0.15	0.24
	2	0.55	0.28	0.37
	3	0.47	0.35	0.40
	4	0.45	0.45	0.45

(iv)

Video Data set	Round No.	Precision	Recall	F-score
1	1	0.8	0.133	0.24
	2	0.25	0.25	0.38
	3	0.625	0.45	0.54
	4	0.61	0.61	0.61
2	1	0.8	0.13	0.23
	2	0.7	0.23	0.35
	3	0.55	0.36	0.44
	4	0.69	0.57	0.62
3	1	0.7	0.125	0.28
	2	0.65	0.33	0.433
	3	0.56	0.43	0.48

4	0.53	0.53	0.53
---	------	------	------

Table. 2. Precision-recall and F-score values for CBVR system (i) with Motion pattern (ii) with motion feature and RGB color model (iii) with histogram features and (iv) with extensive feature set for 3D- animated video.

(i)

3D animated video Data set	Round No.	Precision	Recall	F-score
1	1	0.65	0.43	0.51759259
	2	0.58	0.35	0.43655914
	3	0.35	0.21	0.2625
	4	0.26	0.22	0.23833333
2	1	0.42	0.15	0.22105263
	2	0.48	0.37	0.41788235
	3	0.5	0.41	0.45054945
	4	0.7	0.22	0.33478261
3	1	0.55	0.45	0.495
	2	0.33	0.13	0.18652174
	3	0.46	0.25	0.32394366
	4	0.28	0.18	0.21913043

(ii)

3D animated video Data set	Round No.	Precision	Recall	F-score
1	1	0.6	0.32	0.4173913
	2	0.56	0.22	0.31589744
	3	0.7	0.35	0.46666667
	4	0.53	0.22	0.31093333
2	1	0.37	0.28	0.31876923
	2	0.45	0.22	0.29552239
	3	0.25	0.34	0.28813559
	4	0.42	0.36	0.38769231
3	1	0.2	0.15	0.17142857
	2	0.3	0.18	0.225
	3	0.36	0.21	0.26526316
	4	0.48	0.3	0.36923077

(iii)

3D animated video Data set	Round No.	Precision	Recall	F-score
1	1	0.46	0.1	0.16428571
	2	0.6	0.18	0.27692308
	3	0.7	0.28	0.4
	4	0.36	0.14	0.2016
2	1	0.35	0.26	0.29836066
	2	0.45	0.32	0.37402597
	3	0.52	0.38	0.43911111
	4	0.27	0.2	0.22978723
3	1	0.5	0.19	0.27536232
	2	0.5	0.24	0.32432432

3	0.4	0.28	0.32941176
4	0.38	0.15	0.21509434

(iv)

3D animated video Data set	Round No.	Precision	Recall	F-score
1	1	0.33	0.22	0.264
	2	0.26	0.18	0.2127273
	3	0.5	0.34	0.4047619
	4	0.73	0.5	0.5934959
2	1	0.25	0.18	0.2093023
	2	0.2	0.14	0.1647059
	3	0.375	0.27	0.3139535
	4	0.55	0.4	0.4631579
3	1	0.16	0.129	0.1428374
	2	0.13	0.33	0.1865217
	3	0.25	0.19	0.2159091
	4	0.36	0.28	0.315

Table. 3. Precision-recall and F-score values for CBBVR system (i) with Motion pattern (ii) with motion feature and RGB color model (iii) with histogram features and (iv) with extensive feature set for Cartoon video.

(i)

Cartoon video Data set	Round No.	Precision	Recall	F-score
1	1	0.23	0.18	0.2019512
	2	0.33	0.26	0.2908475
	3	0.16	0.13	0.1434483
	4	0.26	0.21	0.2323404
2	1	0.175	0.145	0.1585938
	2	0.25	0.02	0.037037
	3	0.125	0.104	0.1135371
	4	0.2	0.16	0.1777778
3	1	0.11	0.107	0.1084793
	2	0.16	0.14	0.1493333
	3	0.08	0.07	0.0746667
	4	0.13	0.11	0.1191667

(ii)

Cartoon video Data set	Round No.	Precision	Recall	F-score
1	1	0.33	0.28	0.3029508
	2	0.43	0.37	0.39775
	3	0.63	0.542	0.5826962
	4	0.2	0.171	0.1843666
2	1	0.25	0.2	0.2222222

3	2	0.325	0.26	0.2888889
	3	0.475	0.38	0.4222222
	4	0.15	0.12	0.1333333
	1	0.16	0.14	0.1493333
3	2	0.21	0.19	0.1995
	3	0.316	0.279	0.2963496
	4	0.1	0.08	0.0888889

(iii)

Cartoon video Data set	Round No.	Precision	Recall	F-score
1	1	0.23	0.21	0.2195455
	2	0.33	0.3	0.3142857
	3	0.16	0.15	0.1548387
	4	0.26	0.242	0.2506773
2	1	0.175	0.16	0.1671642
	2	0.25	0.23	0.2395833
	3	0.125	0.116	0.120332
	4	0.2	0.186	0.1927461
3	1	0.003	0.11	0.0058407
	2	0.16	0.15	0.1548387
	3	0.08	0.07	0.0746667
	4	0.13	0.12	0.1248

(iv)

Cartoon video Data set	Round No.	Precision	Recall	F-score
1	1	0.43	0.39	0.4090244
	2	0.53	0.48	0.5037624
	3	0.37	0.33	0.3488571
	4	0.2	0.18	0.1894737
2	1	0.32	0.3	0.3096774
	2	0.4	0.36	0.3789474
	3	0.275	0.255	0.2646226
	4	0.15	0.136	0.1426573
3	1	0.21	0.2	0.204878
	2	0.26	0.25	0.254902
	3	0.18	0.17	0.1748571
	4	0.1	0.09	0.0947368

Table. 4. Mean precision, recall and F-score values for different rounds of retrieval obtained by the proposed CBVR system and other existing systems for normal video.

Techniques	Datasets	Round No.	Precision	Recall	F-score
CBVR system with Motion pattern	Normal video dataset	1	0.33	0.06	0.09
		2	0.32	0.12	0.17
		3	0.28	0.20	0.24
		4	0.30	0.25	0.28
CBVR system with motion feature and RGB color model	Normal video dataset	1	0.50	0.10	0.16
		2	0.52	0.19	0.27
		3	0.42	0.28	0.34
		4	0.42	0.38	0.39
CBVR system with histogram features	Normal video dataset	1	0.63	0.12	0.21
		2	0.57	0.22	0.31
		3	0.49	0.34	0.40
		4	0.48	0.45	0.46
CBVR system with extensive feature set	Normal video dataset	1	0.77	0.13	0.25
		2	0.53	0.27	0.39
		3	0.58	0.41	0.49
		4	0.61	0.57	0.59

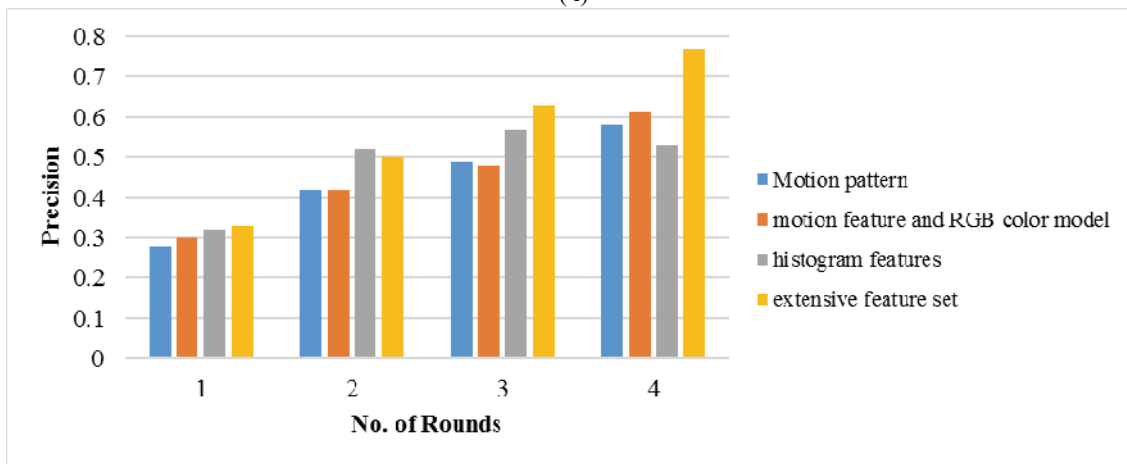
Table. 5. Comparison plot of Mean (i) Precision, (ii) recall and (iii) F-score values for different rounds of retrieval obtained by the proposed CBVR system and other existing systems with 3D- animated video dataset.

Techniques	Datasets	Round No.	Precision	Recall	F-score
CBVR system with Motion pattern	3D-animated video dataset	1	0.23	0.07	0.11
		2	0.282	0.11	0.15
		3	0.228	0.18	0.20
		4	0.310	0.27	0.318
CBVR system with motion feature and RGB color model	3D-animated video dataset	1	0.50	0.10	0.12
		2	0.52	0.16	0.26
		3	0.42	0.28	0.33
		4	0.42	0.37	0.36
caCBVR system with histogram features	3D-animated video dataset	1	0.63	0.11	0.19
		2	0.57	0.22	0.30
		3	0.49	0.34	0.42
		4	0.48	0.44	0.41
CBVR system with extensive feature set	3D-animated video dataset	1	0.77	0.14	0.23
		2	0.53	0.28	0.32
		3	0.58	0.42	0.47
		4	0.61	0.55	0.55

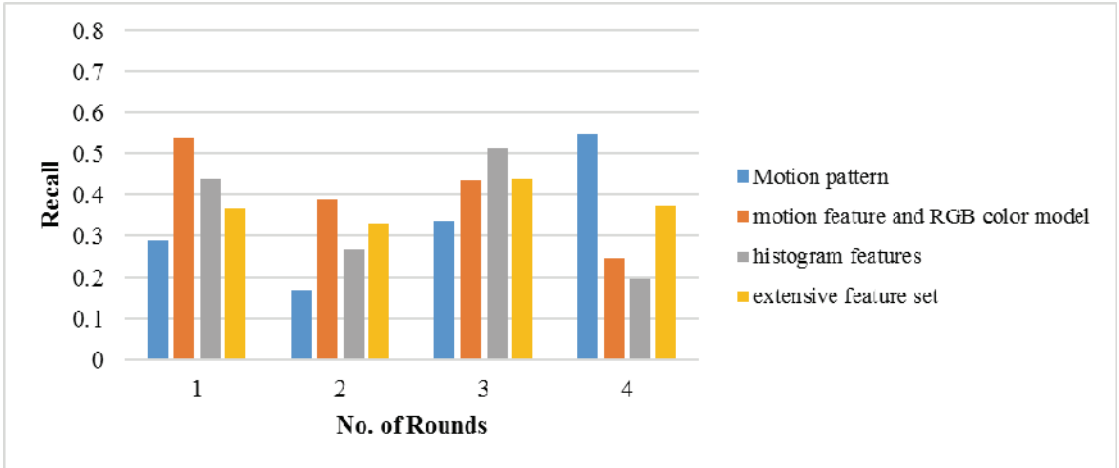
Table. 6. Comparison plot of Mean (i) Precision, (ii) recall and (iii) F-score values for different rounds of retrieval obtained by the proposed CBVR system and other existing systems with cartoon video dataset.

Techniques	Datasets	Round No.	Precision	Recall	F-score
CBVR system with Motion pattern	Cartoon video dataset	1	0.33	0.06	0.09
		2	0.32	0.09	0.15
		3	0.28	0.17	0.21
		4	0.30	0.21	0.26
CBVR system with motion feature and RGB color model	Cartoon video dataset	1	0.50	0.10	0.25
		2	0.52	0.16	0.39
		3	0.42	0.28	0.49
		4	0.42	0.37	0.59
CBVR system with histogram features	Cartoon video dataset	1	0.63	0.12	0.21
		2	0.57	0.22	0.31
		3	0.49	0.34	0.40
		4	0.48	0.45	0.46
CBVR system with extensive feature set	Cartoon video dataset	1	0.77	0.21	0.25
		2	0.52	0.31	0.39
		3	0.57	0.40	0.49
		4	0.60	0.46	0.59

(i)



(ii)



(iii)

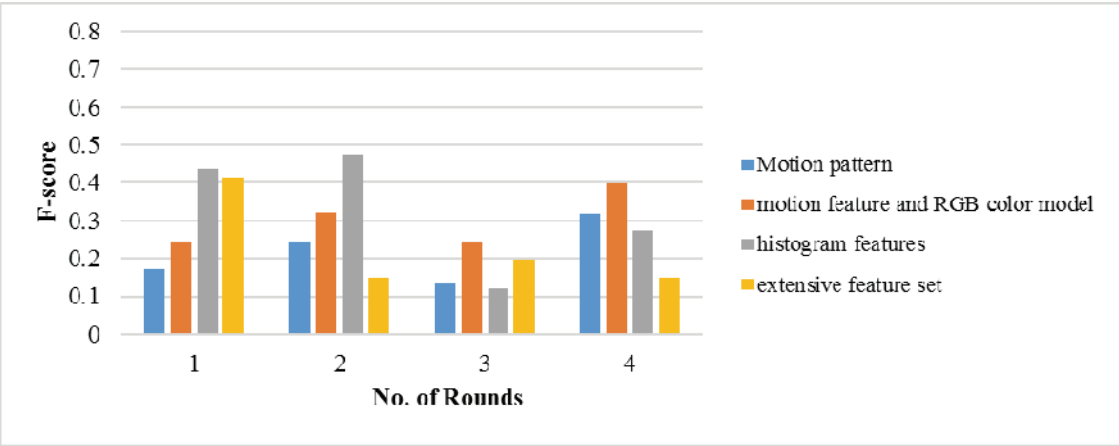
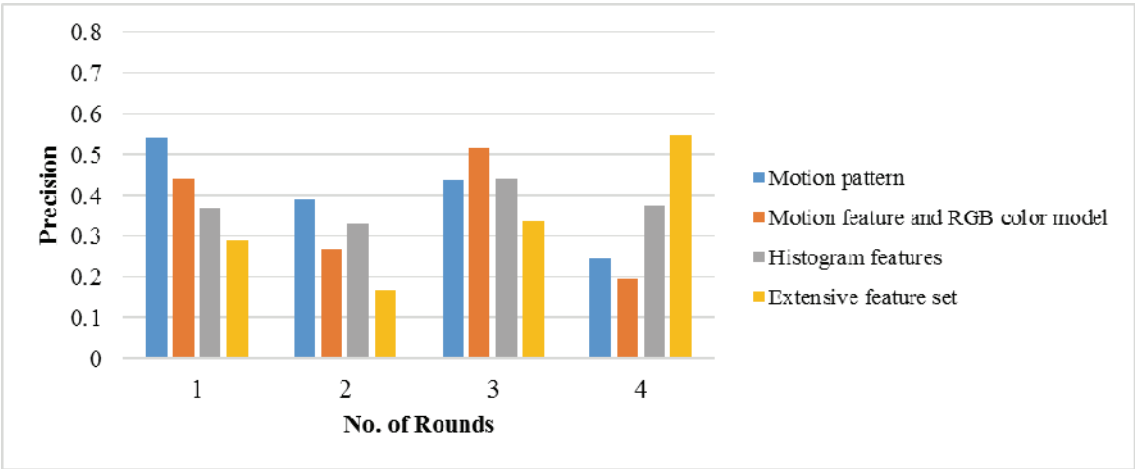


Fig. 7. Comparison plot of Mean (i) Precision, (ii) recall and (iii) F-score values for different rounds of retrieval obtained by the proposed CBVR system and other existing systems



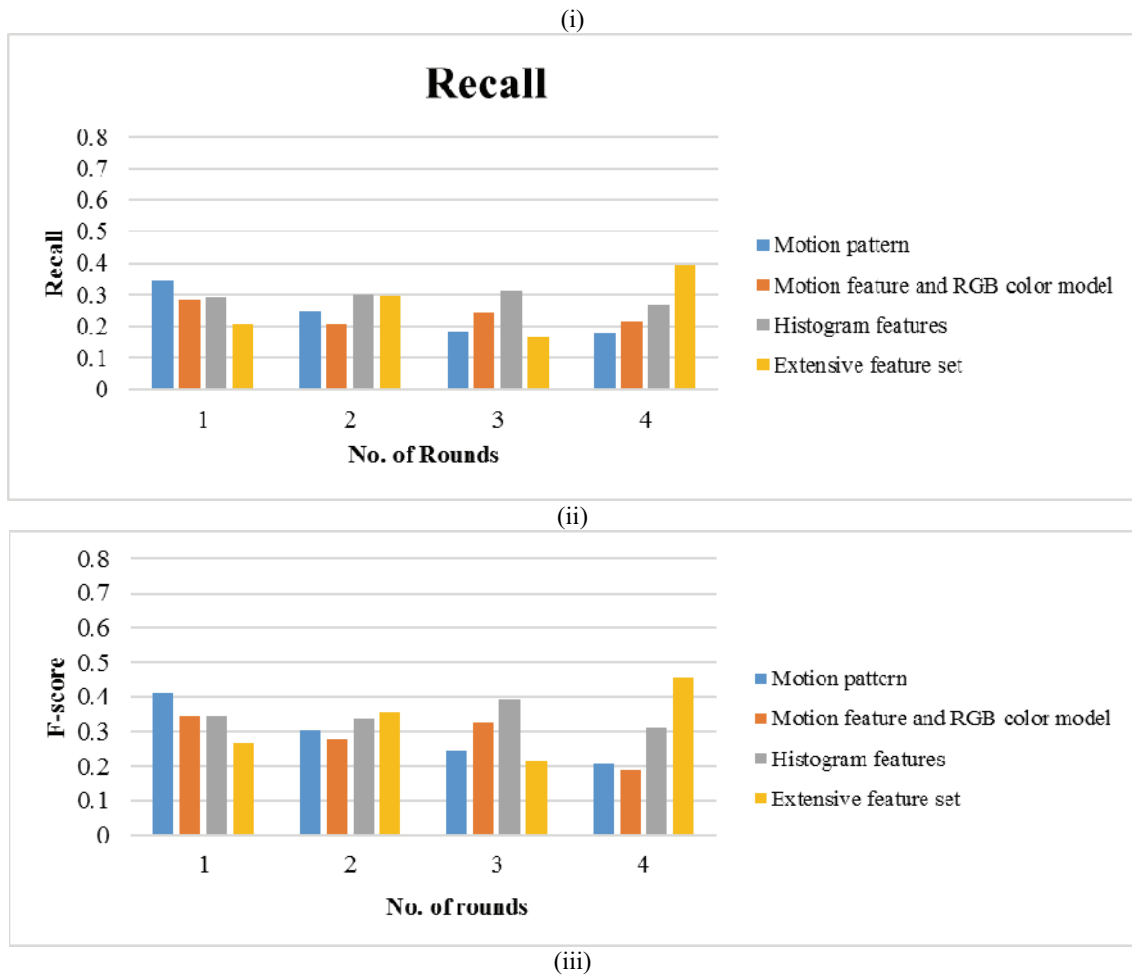
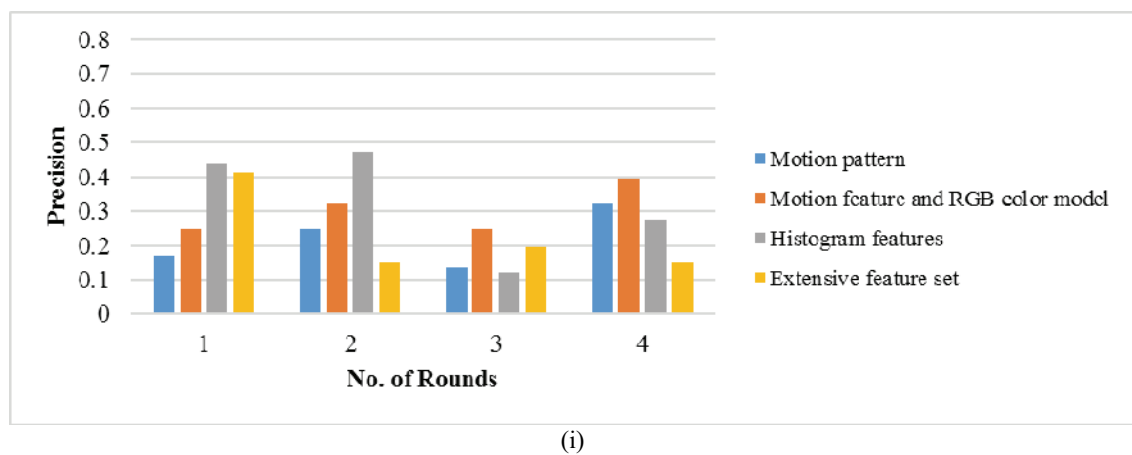
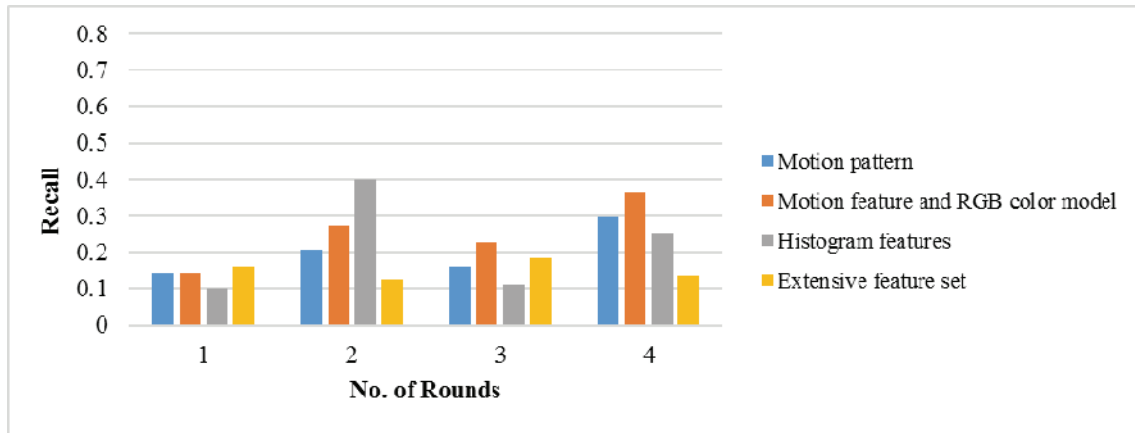
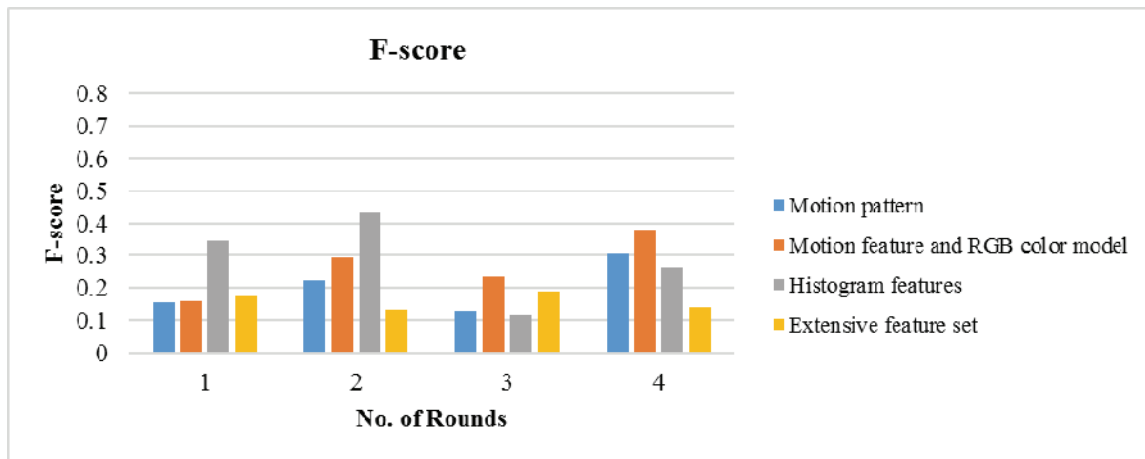


Fig. 8. Comparison plot of Mean of 3D- animated video data set (i) Precision, (ii) recall and (iii) F-score values for different rounds of retrieval obtained by the proposed CBVR system and other existing systems.





(ii)



(iii)

Fig.9. Comparison plot of Mean of Cartoon video data set (i) Precision, (ii) recall and (iii) F-score values for different rounds of retrieval obtained by the proposed CBVR system and other existing systems.

Compared to existing work [31], our proposed work has produced better results. Our proposed work has performed satisfactorily when its movement feature was tested on the gait database. After the inclusion of the movement feature, based on the given query video, the frames of the input videos are retrieved.

5. Conclusion

Visual content based retrieval of video information is a growing research area which has been in attention freshly in the middle of the researchers and experimenters. In this paper, an effective visual content based retrieval of video has presented which implements proficiently. The first and the primary process of the proposed technique is it split the long

video sequence into shots. After that, the object, program and the occlusion features are extracted from the achieved shots and these feature sets are set aside in the feature library for the consequent processes. With the assistance of the Kullback-Leibler distance the correspondence portion is estimated among the features in the feature library and the features of the given query clip is removed in a comparable way. The computed distance selects the retrieval of the similar videos from the collection of videos; hence the effective retrieval of video based on the content has been completed in an actual manner.

References

- [1] Yong Rui and Thomas S. Huang, "Image Retrieval: Current Techniques, Promising Directions, and Open

- Issues", *Journal of Visual Communication and Image Representation*, Vol. 10, pp. 39-62, 1999.
- [2] Elleuch, N. ; Ben Ammar, A. ; Alimi, A.M., "Semi-automatic soft collaborative annotation for semantic video indexing", *EUROCON - International Conference on Computer as a Tool (EUROCON)*, pp. 1- 6, April 2011.
- [3] Jens-Rainer Ohm, F. Bunjamin, W. Liebsch, B. Makai, K. Müller, A. Smolic, D.Zier, "A Set of Visual Feature Descriptors and their Combination in a Low-Level Description Scheme", *Signal Processing: Image Communication*, Vol. 16, No. 1-2, pp. 157-179, September 2000.
- [4] Fong, A.C.M., "Content-based video sequence interpretation", *IEEE Transactions on Consumer Electronics*, Vol. 47, No. 4, pp.873- 879, 2001
- [5] Y. Chung, W. Chin, X. Chen, D. Shi, E. Choi, and F. Chen, "Performance analysis of using wavelet transform in content based video retrieval system", *In Proceedings of the WSEAS International Conference on Computer Engineering and Applications*, pp. 17-19, 2007.
- [6] C. Faloutsos, W. Equitz, M. Flickner, W. Niblack, D.Petkovic, and R. Barber, "Efficient and effective querying by image content", *Journal of Intelligent Information Systems*, Vol. 3, No.3, pp. 231-262, July 1994.
- [7] G. Ahanger and T.D.C. Little, "A Survey of Technologies for Parsing and Indexing Digital Video", *Journal of Visual Communication and Image Representation*, Vol. 7, No. 1, pp. 28-43, 1996.
- [8] Wallapak Tavanapong and Junyu Zhou, "Shot Clustering Techniques for Story Browsing", *IEEE Transactions on Multimedia*, Vol.6, No.4, pp.517-527, Aug 2004
- [9] Nida Aslam, Irfanullah, Kok-Keong Loo and Roohullah, "Limitation and Challenges: Image/Video Search & Retrieval", *International Journal of Digital Content Technology and its Applications*, Vol.3, No.1, pp.98-102, Mar 2009
- [10] Dimitrova, "Multimedia Content Analysis and Indexing for Filtering and Retrieval Applications", *Special Issue on Multimedia Informing Technologies*, Vol.1, No.4, pp.87-100, 1999
- [11] Dagtas, S., Al-Khatib, W. Ghafoor, A. Kashyap, R.L. "Models for motion-based video indexing and retrieval", *IEEE Transactions on Image Processing*, Vol. 9, No. 1, pp. 88- 101, 2000.
- [12] Smoliar, S.W, HongJiang Zhang ; Siew Lian Koh ; Guo Jun Lu, "Interacting with digital video", *IEEE Region 10's Ninth Annual International Conference*, Vol. 2, pp. 852- 856,1994.
- [13] Liang-Hua Chen, Kuo-Hao Chin and Hong-Yuan Liao "An Integrated Approach to Video Retrieval", *In proceedings of ACM International Conference Proceeding Series*, Vol. 313, pp.49-55, 2008
- [14] Xiang Fu and Jie-xian Zeng, "Local Features Based Image Sequence Retrieval", *Journal of Computers*, Vol.5, No.7, P.p:987-994, July 2010
- [15] Tianming Liu, Hong-Jiang Zhang and Feihu Qi, "A Novel Video Key-Frame-Extraction Algorithm Based on Perceived Motion Energy Model", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.13, No.10, pp.1006-1013, Oct 2003
- [16] Alan F. Smeaton, Peter Wilkins, Marcel Worring, Ork de Rooij, Tat-Seng Chua, Huanbo Luan, "Content-based video retrieval: Three example systems from TRECVID", *International Journal of Imaging Systems and Technology*, Vol. 18, No. 2, pp. 195-200, 2008.
- [17] Yannis S. Avrithis, Anastasios D. Doulamis, Nikolaos D. Doulamis, and Stefanos D. Kollias, "A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases", *Computer Vision and Image Understanding*, Vol.75, No.1, pp. 3–24, Aug 1999
- [18] Jinqiao Wang, Hanqing Lu, Lingyu Duan and Jesse S. Jin, "Commercial Video Retrieval with Video-based Bag of Words", *In proceedings of Fifth International Conference on Intelligent Multimedia Computing and Networking*, pp.1-7, July 2007
- [19] Yuxin Peng, Chong-Wah Ngo and Jianguo Xiao, "OM-based video shot retrieval by one-to-one matching", *Multimedia Tools and Applications*, Vol.34, pp.249–266, 2007
- [20] Sang Hyun Kim and Rae-Hong 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*, Vol.12, No.7, pp.592-596, July 2002
- [21] R. Kanagavalli, K. Duraiswamy, "A Study on Techniques used in Digital Video for Shot Segmentation and Content Based Video Retrieval", *European Journal of Scientific Research*, Vol.69, No.3, pp. 370-380, 2012.
- [22] Shradha Gupta, Prof. Neetesh Gupta, Dr. Bhupendra Verma, "Performance of SIFT based Video Retrieval", *International Journal on Computer Science and Engineering (IJCSE)*, Vol. 3 No. 6, pp. 2565- 2571, June 2011
- [23] Weiming Hu, Nianhua Xie, Li Li, Xianglin Zeng, and Stephen Maybank "A Survey on Visual Content-Based Video Indexing and Retrieval", *IEEE Transactions On Systems, Man, And Cybernetics*, Vol. 41, No. 6, 2011.
- [24] Ionuț Mironică, Constantin Vertan, Bogdan Ionescu "A Relevance Feedback Approach to Video Genre Retrieval", *IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, pp. 327- 330, 2011.
- [25] B V Patel and B B Meshram "Content Based Video Retrieval Systems", *International Journal of UbiComp (IJU)*, Vol.3, No.2, pp.13-30, April 2012
- [26] Chen, L.H., Chin, K.H. and Liao, H.Y.M. 'Integration of Color and Motion Features for Video Retrieval,'

- International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 23, No. 2, pp. 313-329, 2009.
- [27] Dyana, A. and Das, S. (2009) 'Trajectory representation using Gabor features for motion-based video retrieval,' *Pattern Recognition Letters*, Vol. 30, No. 10, pp. 877-892.
- [28] Ionuț Mironică, Constantin Vertan, Bogdan Ionescu "A Relevance Feedback Approach to Video Genre Retrieval", *IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, pp. 327- 330, 2011.
- [29] Jean-François Aujol, Guy Gilboa, Tony Chan, Stanley Osher, "Structure-Texture Image Decomposition--Modeling, Algorithms, and Parameter Selection", *International Journal of Computer Vision*, Vol. 67, No. 1, April 2006
- [30] Priya Rajendran and Shanmugam, "A content-based video retrieval system: video retrieval with extensive features", *International Journal of Multimedia Intelligence and Security*, Vol. 2, No.2, pp. 146 - 171, 2011
- [31] Hu and Collomosse, "Motion-sketch based Video Retrieval using a Trellis Levenshtein Distance", *In proceedings of International Conference on Pattern Recognition*, 2010
- [32] Andre N.Stein and Martial Hebert, "Local Detection of Occlusion Boundaries in Video", *in proceedings of BMVC*, pp.407-416, 2006
- [33] Piro, Anthoine, Debreuve and Barlaud, "Image retrieval via Kullback-Leibler divergence of patches of multiscale coefficients in the KNN framework", *In proceedings of International Workshop Content Based Multimedia Indexing*, pp.230-235, June 2008.