

Feature Based Video Extraction Using Tree Pruning

R. Manjunath, B. Sreepathi

*Dept. of Computer Science and Engineering ,
Rao Y.Mahabaleswarappa Engineering College
Bellary, Karnataka, India*

Abstract— Not long ago, advances in technology have made enormous amount of multimedia content available. The amount of video content is expanding, due to which the systems that better access the video is needed. Effectiveness of the retrieval system depends on the search method used within the system. The use of unsuitable search method may reduce the performance of the retrieval system. During last few years, multimedia storage is increased and price of storing digital data is also inexpensive. So there is huge volume of videos accessible in the multimedia database. It is very tough task to retrieve the relevant videos from the available huge volume of database, if user needs to access them. Hence a powerful video retrieval system is needed to retrieve the related videos from the available large database. To implement a powerful video retrieval system, visual perception should be taken into consideration. In this paper, we proposed an algorithm for video retrieval based on grouping of the video. This algorithm extracts or removes the key frames based on motion detection, Area of interest with the objects are perceived using bounding box method and are commented over the video and compared with the alike objects from the knowledge base constructed for various classes of videos to recognize the objects. This method also adopts tree based classifier to find out the class of the query video and to retrieve the same group videos from the large video database.

Index Terms—Retrieval, multimedia database, bounding box method. (Key words)

I. INTRODUCTION

Over the past few years the volume of digital content produced as video has grown massively. Main elements that drive for this growth are: rapid increase of affordable digital cameras, admiration of web based video streaming, hand held devices and acquire of digital video by broadcast industry for their distribution service to the public. Day to day a number of video clips are created and added into digital library which leads to increase in the number of video clips in this real world, so there is a need to manage this large amount of content through efficient storage, indexing so that retrieval is not critical these days.

During last few years, methods created for the retrieval of the videos are completely built using their visual features. Such as, texture, motion, shape and spatial-temporal composition are the most commonly used visual features for visual similarity match. In spite of the sustained works in the last few years, the apex provocation remains bridging the semantic gap. It means features at low level are comfortably measured and calculated, but the starting point of the retrieval process is typically the high level query from a human which is relatively complex compared to low level features. Compiling or converting the complex query

by a human to the low level features seen by the computer illustrates the problem in bridging the semantic gap between low level and high level. So the semantic gap is not just converting high level characteristics to low level characteristics. The semantic query is required to understand the meaning for which query is used. This involves in understanding the intelligence and emotion part of the man, the logical portion of the query but the personal importance and emotive tones of the query and the important form of the results. Video retrieval technology has top priority in creating the blueprint of video search engines and retrieval of a primary set of equivalent or related videos from the huge database.

The application of computer vision to the video retrieval problem is Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR). This is used for the searching of video in huge databases. “Content” or query in these circumstances refers to s that is textures, areas, shapes, textures, or other related information that might be obtained from the image. It will be very costly or waste of time without the capacity to test video content, searches should trust on metadata like heading or keywords that is hard and expensive to create. The content-based video retrieval will aid users of the retrieval system to extract desired video from a heavy video database effectively based on the video contents through user interactions. The video retrieval or extraction system is classified as two main parts: a module to extract representative characteristics from video frames and the second tells a suitable similarity model to search alike video frames from huge database. There are many methods used different types of characteristics to represent a video frame, like shape information, histogram, text analysis, texture. Some methods integrate the characteristics to help retrieval process.

II. LITERATURE SURVEY

The literature survey gives a summary of the progress which has taken the theory and practice of video retrieval into the digital age. Research works in the past in content-based retrieval, leads the path towards the automation of these functions in digital image scenarios, and gives the main press of the paper.

Content Based Video Retrieval by Genre Recognition Using Tree Pruning Technology is represented in [1]. It describes a bounding box method and tree pruning technique for key frame extraction and videos are matched based on motion detection. Area of interest with the object is perceived using

bounding box method and a note is written over the video and compared with alike objects from the video database created for different classes of videos.

Video Scene Retrieval Based on Local Region Features is described in [2] contains a method for content extraction and scene retrieval for video sequences based on local region descriptors. The local constant characteristics are acquired for whole frames in a sequence and tracked throughout the shot to extract stable characteristics. The frames in a shot are represented by these stable features rather than features from one or more key frames.

Content Based Video Retrieval is introduced in [3] provides an approach for facilitating the searching and browsing of large image collections over World Wide Web. Here, analysis of video is done on low level visual properties retrieved from video frame. If an effective video retrieval system is to be created, visual perception should play an important role. Authors used a technique which employs many characteristics for indexing and retrieval would be more effective for search tasks of videos.

Content based Video Retrieval using Latent Semantic Indexing and, Motion and Edge Features is represented in [4]. Here authors introduce a video retrieval system based upon the entire different visual hints. The method analyzes each and every frame within the shot to create a compact representation of video shot. In feature extraction step system extract quantized, motion and edge density features. A similarity measure is defined using LSI (Latent semantic indexing) to locate the occurrence of similar video clips in the database.

A Comprehensive Content based Video Retrieval System is introduced in [5]. Here the authors used frills-free method of video retrieval based on sample video input. Characteristics like shape and texture are considered for retrieval. In the new approach the frames are selected as multiples of a number and then the characteristics extraction takes place. Many characteristics should be added to make video retrieval process as precise and accurate as possible.

An Evolving Approach on Video Frame Retrieval Based on, Shape and Region is given in [6]. This system provides a new technology for matching of objects in video based upon the, shape and region. The video shot is segmented and indexed based upon the similarity between the frames. The similarity feature such as, shape and region are measured for the objects between two videos are matched and they are displayed. The similarities between two frames are resulted from three major features such as, shape and region in order to solve the problem of objects retrieval in video.

An Automated Content Based Video Search System Using Visual Cues is introduced in [7]. Here the authors introduce a real-time, interactive system on the Web, based upon the visual model with spatial temporal attributes playing a key role in video retrieval. The system utilizes already developed algorithms for automated video object

segmentation and tracking and use real-time video editing techniques while responding to user queries.

A Technique to Content-Based Video Retrieval Utilizing Diverse Set of Features is introduced in [8]. It presents an effective content based video retrieval system by extracting the characteristic set after converting the raw video into four representation schemes. OAR, OFR, HMSB operator and color are the four important representation schemes used in the proposed system to extract the significant features presented in the raw video. These four different representation schemes can be able to provide the object based feature as well as temporal based features.

Multi feature content based video retrieval using high level semantic concept is presented in [9]. It proposes a system using adaptive threshold for video segmentation and key frame selection as well as using both low level features together with high level semantic object annotation for video representation.

Shot Detection Using Genetic Edge Histogram and Object Based Video Retrieval Using Multiple Features is introduced in [10]. Here authors use an algorithm to detect shot using Genetic Edge Histogram and 2-D discrete cosine transform as a feature and multiple features like, motion, shape and SIFT are used to retrieve the similar shots.

III. PROBLEM DEFINITION

Earlier works in video processing have produced global frame based characteristic extraction of either preassigned number of frames or key frames. A Key frame can be searched as one where one frame differs from its previous frame significantly. Motion detection is used purely to find the key frames.

But the problem is not all video frames is globally similar and may exhibit alike objects. Thus retrieving objects from video and noting those objects in the video is a major step in detecting alike videos.

Segmentation discovers the motion in video as straight frame difference between frames or change in the position of characteristics from one frame to another frame. Secondly techniques desire characteristics like DCT points SIFT features etc and find them in the next frame using block matching or any other searching technique. Every video frame is an area of concern so that there is an independent object from the frames can be identified, thus important objects can be segmented from every frame in contrast with the past where many frames are compared to find a particular object and to extract it.

Further, a classifier which is used in classic digital image processing to separate objects of different size or shape does not suit for video as video frames may be compared with similar corresponding characteristics, frames may not have equal number of objects, and some frames may have common objects and so on. Thus here comes the role of the tree based classifier used to correlate frames through object properties and ultimately video through frames.

Hence problem statement of the project can be defined as "to Segment regions of interest from video frames, extract features and classify them and based on occurrence of region of interests in a video, classify a video into categories to extract matching videos of the category using tree pruning."

IV. CONTENT BASED SEARCH

The proposed system is totally new for video mining and similar video extraction. The proposed system extracts regions of interest from every frame based on motion detection. It then follows it by extracting features from each of these regions of interest. Once features are extracted from regions of interest (ROIs), it indexes them and the features. It also indexes the frame with the ROIs and ultimately video with frames containing ROIs. So in sports video, the video is selected which has most frames with ground object, ball object and advertisement objects. The next example, video is extracted from news video that has anchor objects, text objects and studio objects in most of its frames. The video is given as an input so that frames are extracted and regions of interest are segmented out. Euclidean distance based matching is adopted to match the ROIs with those stored in database and identified ROI names are noted down over the ROIs in the video itself. Finally system runs a tree pruning based technique to find out which objects have appeared in whole video with what probability.

The proposed algorithm is explained in the following steps:

- Step1: The video library contains many types of videos.
- Step2: One video is read at a time and frames are extracted. Each frame is compared with the previous one and motion is detected. Frames with significant motion differences are separated as key frame.
- Step3: Regions containing objects are detected using bounding box method and they are classified.
- Step4: Mean and Standard deviation of RGB and HSV channels are extracted from the regions detected using bounding box.
- Step5: Video features are combined to get a feature vector representing a frame.
- Step6: Query video clip is read, frames are extracted and motion detection is run. Key frames, objects are extracted along with the features.
- Step7: Euclidean distance based matching is adopted to match the objects with those stored in database and identified object names are annotated over the objects in the video itself.
- Step8: Then system runs a tree pruning based technique to identify which objects have appeared in whole video with what probability.

The flowchart is shown in Figure.1

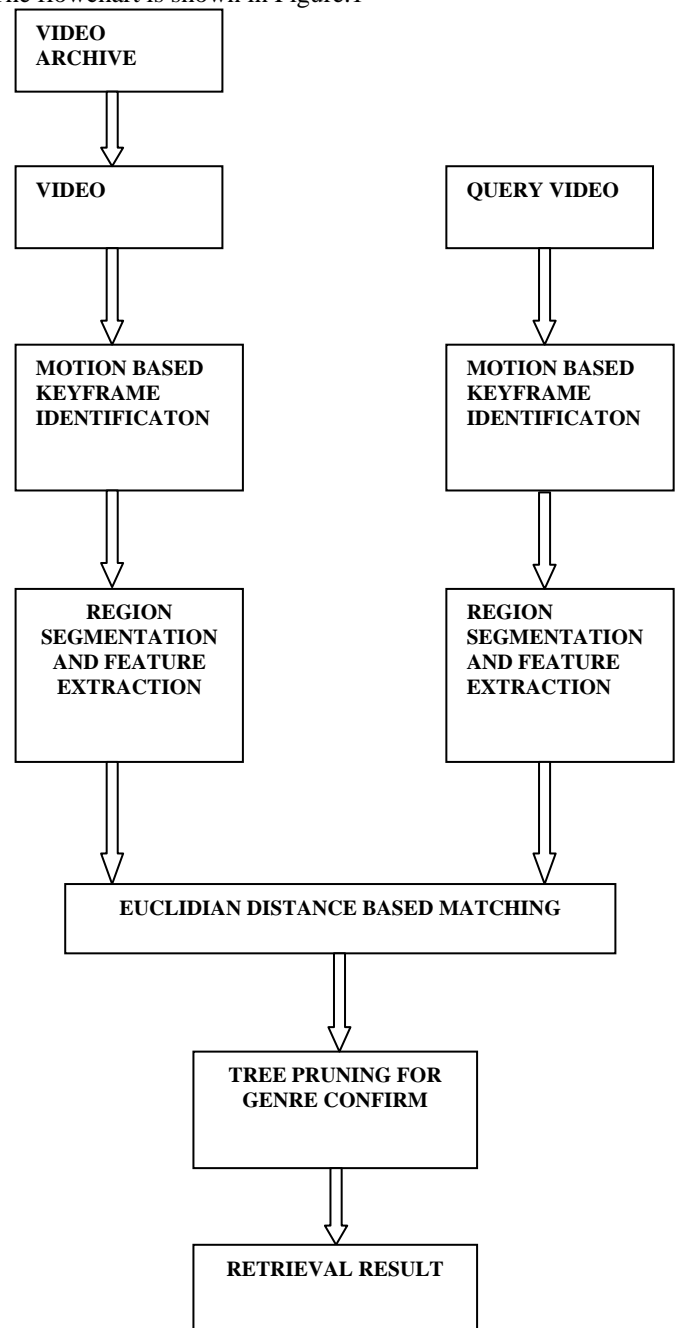


Figure 1 Block diagram of the proposed model

A. Key frame identification

Content-based video extraction and then video analysis are based on two concepts video segmentation and key frame extraction. But key frame extraction is at most important part in video analysis and management, to give related video recap for video indexing, browsing and extraction. The use of key frames reduces the amount of data required in video indexing and provides the framework for dealing with the video content.

Key frame is a selected frame which represents the salient features and information of the shot or series of frames. The key frames extracted from shot should possess the features of the video, and the image features of a video can be

followed by all the key frames in time sequence. Thus the features of the video are identified. But care should be taken to reduce the redundant information while retrieval by rejecting the redundant frames.

B. Identification of region of interest

Human eyes have a perception of a vision due to attention. For example, when we have a perception of an image, our eye movements consists of a series of fixations (repetitive positioning of eyes to parts of the image) and saccades (rapid eye jump). The parts of the image that result in eye fixations and capture primary attention are called regions of interest (ROIs). The Results of Studies in visual attention and eye movement have shown that the human beings generally can concentrate only few ROIs. Recognizing these regions of interest in images is carried out by using bounding box technique.

The 'Bounding Box' method uses a function that draws a rectangle around region of interest. The rectangle containing the region, a 1-by-Q *2 vector, where Q is the number of image dimensions: ndims(L), ndims(BW), or numel(CC.ImageSize), width of the bounding box, upper-left corner.

C. Feature extraction

After the regions of interest with in the video shot is segmented and traced, computing the characteristics of the region of interest is carried out and those are kept within the feature library. For each region of interest Mean and Standard deviation of RGB and HSV channels are retrieved.

D. Matching and extraction

In the extraction process, video clips from the huge video database which are similar to the query clip are extracted by matching the resemblance between the query clip and the database video clips. First the query clip is fed to the proposed extraction system; all the above stated characteristics are extracted as performed for the database video clips. After that, using Euclidean distance mechanism, similarity is measured between every database video clip and the query clip, after that, using this result; related videos are extracted with the help of tree pruning technique.

For measuring the similarity between the query and video under test, distance metric is used, which is the key-component in Content Based Image Retrieval. Here the Euclidean distance between ROIs of the videos in the database and the ROI of query video are computed and used for ranking. The query video's ROI is more similar to the database video's ROI if the distance is smaller. If x and y are feature vectors of database ROIs and query ROI respectively. Then the distance metrics are defined as follows:

The Euclidean distance measure can be defined as

$$d_E(x, y) = \sqrt{\sum_{i=1}^d (x_i - y_i)^2}$$

At last existing video search algorithm utilize a tree-structured hierarchy and subtree(Fig 2(a)) pruning to reduce the search space while traversing the tree from root to leaf nodes for a given query video.

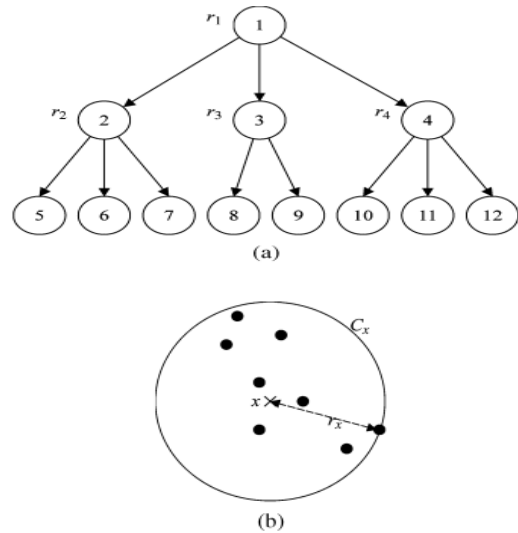


Figure 2. Illustration of the notion of NRC. (a) Tree-structured hierarchy with the NRC r_k associated with a node k , $k = 1, 2, 3, 4$. (b) The NRC r_x of a node x for the cluster C_x

Details are as illustrated in Fig. 2(b), each node x has a feature vector which represents its whole cluster, i.e., subtree C_x within the bound of what we call node radius for cluster (NRC). The NRC, denoted by r_x , is defined as the maximum distance between the node x and the subordinate leaf nodes, i.e., belonging to C_x , and is computed as follows:

$$r_x = \max_{y \in C_x} d(x, y) \tag{1}$$

Where $d(\cdot)$ denotes a distance metric between two feature vectors. To retrieve all similar videos whose distance from the query q is within a threshold value of δ , every node in the tree hierarchy needs to be visited, but some irrelevant clusters can be pruned without degrading the recall rate of retrieval unless the following triangle inequality holds:

$$d(q, x) \leq r_x + \delta. \tag{2}$$

The evaluation of $d(x, y)$ in eq.(2) involves distance computation between two high-dimensional feature vectors.

E. Results and discussion

The proposed content-based video retrieval system is implemented using MATLAB (Matlab2010b) and the performance of the proposed system is analyzed using the evaluation metrics including precision, recall measures. The experiments are conducted in windows XP based system with 3GB RAM on a data set containing 200 videos. 1) **Datasets** : We have performed experiments on a dataset of 200 videos obtained from YouTube website (www.youtube.com). The collected video contains the following categories of objects presented in these videos such as sports, news and movies. The sample snapshot for the input videos of the proposed system is given in figure 3.



Figure 3: A sample snapshot for the input database

2) *Identification of region of interest:* This representation contains the important visual contents in a series of frames by processing every frame presented in it. The IOROI for the SPORTS video is constructed as shown in figure 4.



Fig 4 a) Snapshot of query video

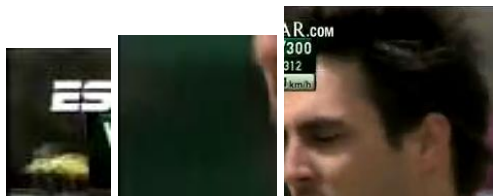


Figure b).A sample snapshot of region of interests

3) *Feature extraction:* For each region of interest, Mean and Standard deviation of RGB and HSV channels are extracted.

4) *Video retrieval using tree pruning:* The characteristics Mean and Standard deviation of RGB and HSV channels are extracted from the ROI of query video and then they are matched with the characteristics of ROIs stored in video library using the Euclidian distance mechanism. Finally for extraction, the proposed system uses tree pruning technique, where a tree-structured hierarchy is used, in which each node is associated with a ROI image or a feature vector which represents all of the images belonging to its subtree. Like that, each child represents a disjoint subset of the images and thus partitions the subtree rooted at its parent node into smaller units. Each leaf node corresponds to a single video in the database. With the node radius for cluster (NRC), which is defined as the maximum distance

between the node and its descendants or cluster, stored at each intermediate node, the triangle inequality is applied to reduce the search space by pruning irrelevant clusters. The matching score compute is used to retrieve the videos from the dataset and the retrieved video for the corresponding input videos is given in figure 5.



Figure5. A sample snapshot of retrieved videos.

5) *Quantitative analysis:* The performance of the proposed approach system is evaluated on the input dataset using the precision and recall measure. Graph-1 shows the Precision-Recall plot: For quantitative analysis, videos from each category are given to the proposed system and results are evaluated as follows.

$$\text{Precision} = \frac{|{\text{relevant videos}} - {\text{retrieved videos}}|}{|{\text{retrieved videos}}|} \quad (3)$$

$$\text{Recall} = \frac{|{\text{relevant videos}} - {\text{retrieved videos}}|}{|{\text{relevant videos}}|} \quad (4)$$

TABLE 1 EXPERIMENTAL RESULTS

VIDEO GENRE	RECALL	PRECESSION
SPORTS	.625	1.00
MOVIES	.538	.92
NEWS	.625	1.00

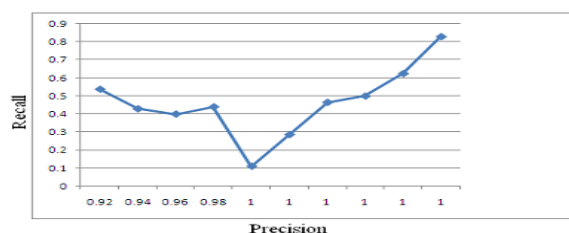


Figure 6.Graph between precision and recall

V. CONCLUSION

In this proposed system, the algorithm for content based video retrieval is designed and experimented on sufficient number of different genres of videos .The algorithm is implemented in Matlab 2010b and executed on Intel core2duo, 2.66 GHz processor with 3 GB of RAM. The algorithm initially extracts Regions of interest through motion estimation. Features of these ROIs are extracted and

are matched with the ROIs of query video. Finally retrieval of videos is carried out using tree pruning. The proposed method has been experimented on different genres of videos like sports, movies and news clips. The performance of the video summarization algorithm is evaluated by the precision and recall measures. The experimental results on standard video datasets reveal that the proposed model is robust and extracts videos from variant genres efficiently.

The system can be further updated with more complicated features that include both shape and texture descriptors like wavelet moments.

REFERENCES

- [1] Amit Fegade and Vipul Dalal ,2014 "Content Based Video Retrieval by Genre Recognition using Tree Pruning Technique" international journal of Engineering and Computer Science ISSN: 2319-7242 volume 3 issue 7.
- [2] Arasanathan Anjulan And Nishan Canagarajah, 2006 "Video Scene Retrieval Based On Local Region Features", Proceedings of the IEEE International Conference on Image Processing, Atlanta, GA., DOI: 10.1109/ICIP.2006.313044
- [3] B. V. Patel and B. B. Meshram,2012 "Content Based Video Retrieval", The International Journal of Multimedia & Its Applications (IJMA) Vol.4, No.5, October 2012
- [4]. Kalpana S Thakare, Archana M Rajurkar, R R Manthalkar, 2012 "Content based Video Retrieval using Latent Semantic Indexing and , Motion and Edge Features", International Journal of Computer Applications (0975- 8887) Volume 54- No.12, 2012
- [5] Shimna Balakrishnan, Kalpana S. Thakre , 2010 "Video Match Analysis: A Comprehensive Content based Video Retrieval System", International Journal of Computer Science and Application Issue ISSN 0974-0767
- [6] D.Shanmuga Priyaa, T.Nachimuthu, Dr.S.Karthikeyan, 2007 "An Evolving Approach on Video Frame Retrieval Based on , Shape and Region", International Journal of Computer Science & Engineering Technology (IJCSET) ISSN : 2229-3345 Vol. 2 No. 4 :
- [7] Shih-Fu Chang William Chen Horace J. Meng Hari Sundaram Di Zhong, 2007"VideoQ: An Automated Content Based Video Search System Using Visual Cues",
- [8] S. Padmakala, G. S. AnandhaMala, 2012 "A Technique to Content-Based Video Retrieval Utilizing Diverse Set of Features", European Journal of Scientific Research ISSN 1450-216X Vol.83 No.4 , pp.558 – 575
- [9] Hamdy K. Elminir, Mohamed Abu ElSoud, Sahar F. Sabbeh, Aya Gamal, 2012 "Multi feature content based video retrieval using high level semantic concept", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 2, ISSN (Online): 1694-0814
- [10] R. Kanagavalli and K. Duraiswamy, 2012 "Shot Detection Using Genetic Edge Histogram and Object Based Video Retrieval Using Multiple Features", Journal of Computer Science 8 (8): 1364-1371, ISSN 1549-3636
- [11] Arasanathan Anjulan., Nishan Canagarajah, 2007"Object based video retrieval with local region tracking", Image Commun., 22: 607-621. DOI: 10.1016/j.image.
- [12] R. Kanagavalli, K. Duraiswamy, 2012 "An Enhanced Algorithm for Shot Change Detection and Object based Video Retrieval using Multiple Features", European Journal of Scientific Research ISSN 1450-216X Vol.79 No.2 , pp.225-243
- [13] Kalpana Thakre, Archana Rajurkar and Ramchandra Manthalkar, 2011 "An Effective Cbvr System Based On Motion, Quantized And Edge Density Features", International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 2
- [14]. Yuxin Peng1,2 and Chong-Wah Ngo2, 2005"EMD-Based Video Clip Retrieval by Many- to-Many Matching", W.-K. Leow et al. (Eds.): CIVR , LNCS 3568, pp. 71–81, Springer-Verlag Berlin Heidelberg 2005
- [15]. Shradha Gupta, Neetesh Gupta, Shiv Kumar, 2011 "Evaluation of Object Based Video Retrieval Using SIFT", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-2
- [16] Shang-Li Yu, JunWei Hsieh, and Yung-Sheng Chen, 2003 "Motion-based Video Retrieval by Trajectory Matching", 16th IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP 2003) , Kinmen, ROC
- [17] JeongKyu Lee,JungHwan Oh,Sae Hwang, 2005 "STRGIndex: SpatioTemporal Region Graph Indexing for Large Video Databases", SIGMOD , Baltimore, Maryland, USA.
- [18] R. Kanagavalli, Dr. K. Duraiswamy, 2012 " OBJECT BASED VIDEO RETRIEVAL", International Journal of Communications and Engineering Volume 06- No.6, Issue: 01
- [19] F. Chevalier *, J.-P. Domenger, J. Benois-Pineau, M. Delest, 2007"Retrieval Of Objects In Video By Similarity Based On Graph Matching", 0167-8655/\$-Elsevier doi:10.1016/j.patrec
- [20]. Arslan Basharat, Yun Zhai, Mubarak Shah, 2007 "Content based video matching using spatiotemporal volumes", 1077-3142/\$ - Elsevier Inc. doi:10.1016/j.cviu.