ijecs open access www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 3 Issue 7 July, 2014 Page No. 7088-7092

Content Based Video Retrieval by Genre Recognition Using Tree Pruning Technique

Amit Fegade[1], Vipul Dalal[2]

[1]Alamuri Ratnamala Institute of Engineering and Technology, Shahapur, Mumbai University, India
amit.fegade121@gmail.com

[2]Vidyalankar Institute of Technology, Wadala, Mumbai University, India
vipul.dalal@vit.edu.in

Abstract: ***Recent advances in technology have made tremendous amount of multimedia content available. The amount of video content is increasing, due to which the systems that improve the access to the video is needed. Efficiency of the retrieval system depends upon the search method used in the system. The use of inappropriate search method may degrade the performance of the retrieval system. During past years, multimedia storage grows and cost of storing digital data is also cheaper. So there is huge amount of videos available in the multimedia database. It is very difficult task to retrieve the relevant videos from the large database as per the user needs. Hence an effective video retrieval system is required to retrieve the relevant videos from the large database. In order to create an effective video retrieval system, visual perception must be taken into account. In this paper, we have proposed an algorithm for video retrieval based on genre of the video. This algorithm extracts the key frames based on motion detection, Regions of interest with the objects are detected using bounding box method and are annotated over the video and compared with the similar objects from the knowledge based prepared for various genre videos for recognition of objects. It also uses a tree based classifier to identify the genre of the query video and to retrieve the same genre videos from the video database.***

Keywords: Genre recognition, Key frame extraction, Motion detection, ROIs, Tree pruning, Video retrieval.

1. INTRODUCTION

The world as a living space is shrinking, are we really shrinking or have we found a new horizon to live in. It is true we are expanding leaps and bounds in GBs and terabyte world. Recent advances in technology have made tremendous amounts of multimedia information available to the general population. A video in simplest of words is agglomeration of data. With the ever escalating videos the systems for processing these videos need to be developed. Analyzing these videos as small data packets for the simplicity of human effort is the need of the hour.

During recent years, numbers of techniques have been developed for video retrieval. Videos are retrieved based on similarity between their visual features like shape, texture, motion, spatial-temporal composition. These are the most common visual features used in visual similarity match. Despite the sustained efforts in the last years, the paramount challenge remains bridging the semantic gap. Low level features of the videos are easy to measure and compute, but starting point of the retrieval process is the high level query from the human. Translation of the high level query posed by the human to low level features by the computer clarifies the problem in bridging the semantic gap. However, the semantic gap is not merely translating high level features to low level features. The significant feature of a semantic query is to understand the meaning behind the query. This can involve understanding both the intellectual and emotional sides of the human, personal preferences and emotional sub tones of the query and the preferential form of the results. Thus video retrieval is the important technology used in the design of video search engines and retrieval of the relevant set of videos from the database.

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision to the video retrieval problem, that is, the problem of searching for video in large databases. “Content” in this context can refer to the various information obtained from the image such as shape, texture etc. Without examine the visual content, the search should depend on the metadata such as caption or keywords which are very expensive to produce.

2. TERMINOLOGY

It is worth understanding some of the basic terminologies coming across the research in video analysis, processing, and retrieval.

2.1. Scripted/Unscripted Content:

A video that is carefully produced according to a script or plan that is later edited, compiled and distributed for consumption is referred to as scripted content. News videos, dramas & movies are examples of scripted content. Video content that is not scripted is then referred to as unscripted. In unscripted content, such as surveillance video, the events happen spontaneously. One can think of varying degrees of “scripted-ness" & “unscripted-ness" from movie content to surveillance content.

2.2. Video Shot:

A consecutive sequence of frames recorded from a single camera. It is the building block of video streams.

2.3. Frame:

The pictorial information in video is considered to be the series of images what are called frames. These frames can be extracted.

2.4. Video Scene:

A collection of semantically related and temporally adjacent shots, depicting and conveying a high-level concept or story. While shots are marked by physical boundaries, scenes are marked by semantic boundaries.

2.5. Key Frame:

The frame that represents the salient visual content of a shot. Depending on the complexity of the content of the shot, one or more key frames can be extracted.

The scripted video data can be structured into a hierarchy consisting of five levels: video, shot, scene and key frame which increase in granularity from top to bottom. Similarly, the unscripted video data can be structured into a hierarchy of four levels: play/break, audio-visual markers, highlight candidates, highlight groups, which increase in semantic level from bottom to top.

3. RELATED WORK

The literature presents numerous algorithms and techniques for the retrieval of significant videos from the database due to the widespread interest of content-based video retrieval in a large number of applications. Some recent researches related to content-based video retrieval are discussed in this section. Video Scene Retrieval Based on Local Region Features is represented in [1] describes a novel method for content extraction and scene retrieval for video sequences based on local region descriptors. The local invariant features are obtained for all frames in a sequence and tracked throughout the shot to extract stable features. The scenes 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 [2] proposes an approach for facilitating the searching and browsing of large image collections over World Wide Web. In this approach, video analysis is conducted on low level visual properties extracted from video frame. It is believed that in order to create an effective video retrieval system, visual perception must be taken into account. Authors conjectured a technique which employs multiple features for indexing and retrieval would be more effective in the discrimination and search tasks of videos. Content based Video Retrieval using Latent Semantic Indexing and , Motion and Edge Features is presented in [3]. In this work authors propose a video retrieval system based on the integration of various visual cues. The approach analyses all frames within a shot to construct 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 [4]. In the implemented system, authors used frills-free method of video retrieval based on sample video input. Features like shape and texture are considered for retrieval. In the new approach the frames are selected as multiples of a number and then the feature extraction takes place. More features can be added so that there is precise and accurate retrieval of the videos. An Evolving Approach on Video Frame Retrieval Based on, Shape and Region is introduced in [5] proposes a new methodology for matching of objects in video based on the, shape and region. The objects are segmented and indexed based on 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[6]. Here the authors propose a novel, real-time, interactive system on the Web, based on the visual paradigm, with spatio- temporal attributes playing a key role in video retrieval. A Technique to Content-Based Video Retrieval Utilizing Diverse Set of Features is introduced in[7] presents an effective content based video retrieval system by extracting the feature set after converting the raw video into four representation schemes. OAR, OFR, HMSB operator and colour 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 introduced in[8]. This work 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[9]. In this study a novel algorithm is proposed for shot detection 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. Object based video retrieval with local region tracking is introduced in[10]. This work describes a method for video retrieval system based on local invariant region descriptors. A novel framework is proposed for combined video segmentation, content extraction and retrieval. A similarity measure, previously proposed by the authors based on local region features, is used for video segmentation. The local regions are tracked throughout a shot and stable features are extracted. The conventional key frame method is replaced with these stable local features to characterize different shots. A grouping technique is introduced to combine these stable tracks into meaningful object clusters. Object Based Video Retrieval is introduced in[11]. In this work Edge Detection and DCT based block matching is used for shot segmentation and the region based approach is used for retrieval. In content based Video Retrieval (CBVR) the feature extraction plays the main role. The features are extracted from the regions by using SIFT features. Finally the features of the query object are compared with the shot features for retrieval. Evaluation of Object Based Video Retrieval Using SIFT in introduced in [12]. The local invariant features are obtained for all frames in a sequence and tracked throughout the shot to extract stable features. Proposed work is 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. Then features are 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 Nearest Neighbour Search (NNS). Spatiotemporal Region Graph Indexing for Large Video Databases is introduced in[13]. Here the authors propose a new graph-based data structure and indexing to organize and retrieve video data. Several re-searches have shown that a graph can be a better candidate for modelling semantically rich and complicated multimedia data. Proposed system uses a new graph- based data structure called *Spatio-Temporal Region Graph* (STRG). STRG further provides temporal features, which represent temporal relationships among spatial objects. The STRG is decomposed into its sub graphs in which redundant sub graphs are eliminated to reduce the index size and search time, because the computational complexity of graph matching (sub graph isomorphism) is NP-complete. In addition, a new distance measure, called *Extended Graph Edit Distance* (EGED), is introduced in both non-metric and metric spaces for matching and indexing respectively.

4. PROPOSED SYSTEM

The proposed system is an entirely new for video summarization and similar video extraction. The proposed system extracts regions of interest from every frame based on motion detection and then extracting features from each of these ROIs. Once features are extracted from ROIs, it indexes them and the features. It also indexes the frame with the ROIs and ultimately video with frames containing ROIs. Thus a sports video is one which contains most frames with ground object, ball object and advertisement objects. On the contrary a news video is one which contains text objects, studio objects and anchor objects in most of its frames. User inputs a video from which 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 annotated over the ROIs in the video itself.

The proposed algorithm is described in the following steps,

Step 1: Different types of videos are stored in video database.

Step 2: One video at a time is selected and frames are extracted using motion detection.

Step 3: Regions of interest with objects are identified using bounding box method and are classified.

Step 4: The features i.e. mean and standard deviation of RGB and HSV channels are extracted from ROIs.

Step 5: Video features are combined to get a feature vector representing a frame.

Step 6: Query video clip is given to the proposed system and key frames are extracted using motion detection and objects are extracted along with the features.

Step 7: Objects of the query video are matched with the objects stored in the database using Euclidian distance matching and identified object names are annotated over the objects in the video itself.

Step 8: With the help of tree pruning technique, system identifies which objects have appeared in whole video with what probability.

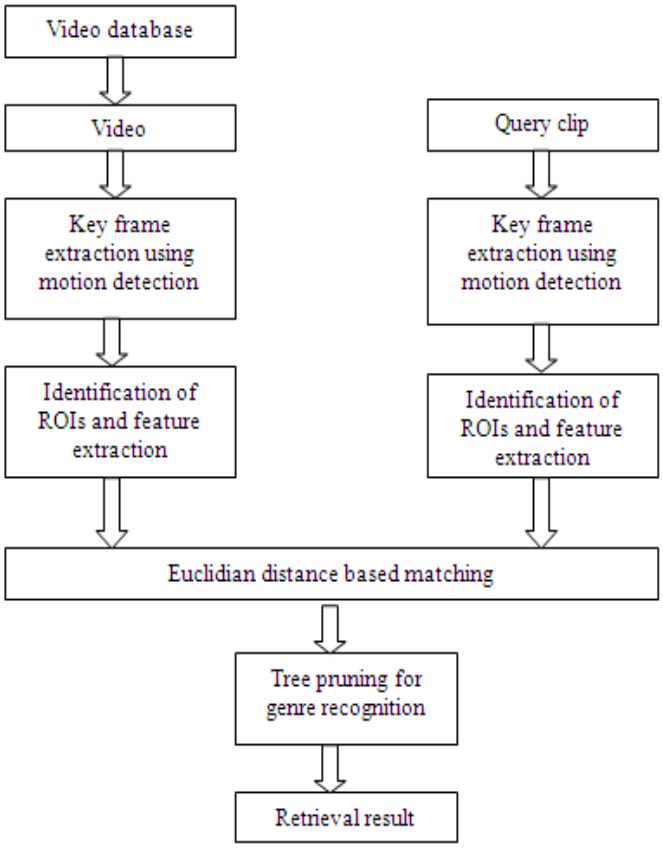


Figure 1: .Block diagram of the proposed model

4.1. Key Frame Identification:

Video segmentation and key frame extraction are the bases of video analysis and content-based video retrieval. Key frames provide a suitable abstract for video indexing, browsing and retrieval. The use of key frames greatly reduces the amount of data required in video indexing & browsing and provides an organization framework for dealing with video content. Key frame is the frame which can represent the salient content and information of the shot. The key frames extracted must summarize the characteristics of the video, and the image characteristics of a video can be tracked by all the key frames in time sequence. Furthermore, the content of the video can be recognized. During key frame extraction, it is necessary to discard the frames with redundant information.

Here the proposed method works on group of frames extracted from a video. Key frames are identified using motion detection. Motion detection is a process of detecting a change in position of an object relative to its surroundings or change in surroundings relative to an object. It takes a list of frames in an order in which they are extracted. It is based on predefined threshold values that specify whether 2 frames are similar. The main function of the proposed system is to choose smaller number of key frames. It starts from first frame from list of files. If consecutive frames are within threshold, then two frames are similar. Repeat the process till frames are similar, delete all the similar frames and take first frame as a key frame. Start with the next frame which is outside of the threshold & repeat the steps for the all video frames.

4.2. Identification of Regions of Interest (ROIs):

Attention plays an important role in human vision. For example, when we look at an image, our eye movements

comprise a succession of fixations (repetitive positioning of eyes to parts of the image) and saccades (rapid eye jump). Those parts of the image that cause eye fixations and capture primary attention are called regions of interest (ROIs). Studies in visual attention and eye movement have shown that humans generally only attend to a few ROIs. Bounding box technique is used to detect these visually attentive regions.

The Bounding Box function 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.

4.3. Extraction of Features:

Once the ROIs are identified and tracked, features of ROIs are extracted and stored into feature library. The features i.e. mean and standard deviation of RGB and HSV channels are extracted for each ROI.

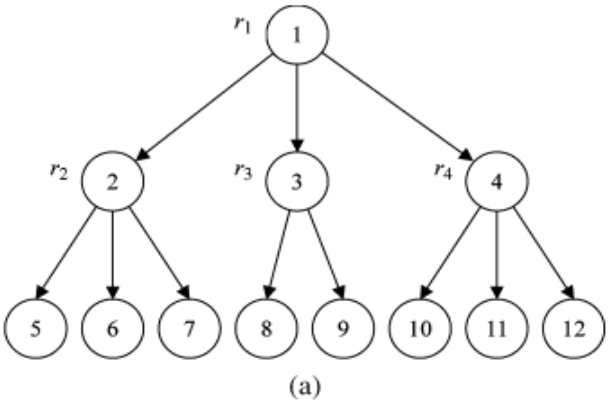
4.4. Retrieval:

In retrieval, videos from the database that are similar to the query clip are retrieved and retrieval is done on the basis of similarities between the videos and query clip. When the query clip is fed to the proposed system, all the features are extracted as performed for the database video clips. With the help of Euclidian distance based matching mechanism, features of query clip are compared with the features of database video clips stored in feature library. Smaller is the distance greater is the similarity. Finally videos are retrieved with the help of tree pruner classifier.

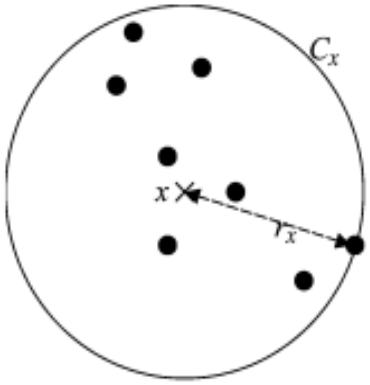
The distance metric is also called as similarity measure. In this, we are calculating the Euclidian distance between the regions of interest of query video clip and regions of interest of videos stored in the database. The regions of interest of query video clip are more similar to the regions of interest of database video clips, if the Euclidian distance is smaller. If x and y are the feature vectors of the database video regions of interest and query clip regions of interest respectively, then Euclidian distance can be calculated as follows:

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

Finally search algorithm makes a use of tree-structured hierarchy and sub tree pruning which reduces search space while traversing from root to leaf node.



(a)



(b)

Figure 2: Illustration of notion of NRC (a) Tree-structured hierarchy with NRC r_1 associated with the node k, k=1,2,3,4. (b) The NRC r_x of a node x for the cluster C_x.

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

$$d(q,x) \leq r_x + \delta \qquad (2)$$

The evaluation of d(q,x) in (2) consist of computing the distance between two high dimensional feature vectors.

CONCLUSION

Most of the previous studies in video processing have proposed global frame based feature extraction of either pre-assigned number of frames or key frames. A Key frame is detected as one where one frame differs from its previous frame significantly. Motion detection is used purely to find the key frames. However every video frame may not be globally homogeneous and may represent different objects. Therefore extracting objects from video, annotating them over video is an important step towards finding similar videos. Segmentation detects motion in video either as direct frame difference between frames or change of position of features from one frame to another frame. Second techniques desire features like DCT points, SIFT features etc and searches them in the next frame using block matching or any other searching technique. Now if we consider that every video frame in itself is a area of concern and can identify independent object from the frames, then important objects can be segmented from every frame unlike the past studies where several frames are needed to track a particular object and to extract it. Further, a classifier used in conventional image processing techniques are not well suited for video as video frames may be correlated, some frames may have common objects, frames may have unequal number of objects and so on. Therefore tree based classifier must be used to correlate frames through object properties and ultimately video through frames.

REFERENCES

[1] Arasanathan Anjulan and Nishan Canagarajah, "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

[2] V.Patel and B.B.Meshram, "Content Based Video Retrieval," The International Journal of Multimedia & Its Applications (IJMA) Vol.4,No.5,October 2012

[3] Kalpana S Thakare, Archana M Rajurkar, R R Manthalkar, "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

[4] Shimna Balakrishnan, Kalpana Thakre, "Video Match Analysis: A Comprehensive Content based Video Retrieval System," International Journal of Computer Science and Application Issue ISSN 0974-0767

[5] D.Shanmuga Priyaa, T.Nachimuthu, Dr.S.Karthikeyan, "An Evolving Approach on Video Frame Retrieval Based on Video Frame Retrieval Based on, Shape and Region," International Journal of Computer Science & Engineering Technology (IJCSET) ISSN 2229-3345 Vol. 2 No. 4

[6] Shih-Fu Chang William Chen Horace J. Meng Hari Sundaram Di Zhong, "VideoQ: An Automated Content Based Video Search System Using Visual Cues,"

[7] S. Padmakala, G. S. AnandhaMala, "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

[8] Hamdy Elminir, Mohamed Abu ElSoud, Sahar Sabbeh, Aya Gamal, "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

[9] R. Kanagavalli and K. Duraiswamy, "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

[10] Arasanathan Anjulan_, Nishan Canagarajah, "Object based video retrieval with local region tracking," Image Commun., 22: 607-621. DOI: 10.1016/j.image.

[11] R. Kanagavalli, Dr. K. Duraiswamy, "OBJECT BASED VIDEO RETRIEVAL," International Journal of Communications and Engineering Volume 06– No.6, Issue: 01

[12] Shradha Gupta, Neetesh Gupta, Shiv Kumar, "Evaluation of Object Based Video Retrieval Using SIFT," International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-2

[13] JeongKyu Lee,JungHwan Oh,Sae Hwang, "STRGIndex: Spatio-Temporal Region Graph Indexing for Large Video Databases," SIGMOD ,Baltimore, Maryland, USA.