A Simple Review On Content Based Video Images Retrieval

Dr P. N Chatur[1], R. M. Shende[2]
Department of Computer Science and Engineering,
Government College of Engineering,
Amravati (Maharashtra), India.
Head *of Department*[1], M.Tech 2nd year[2]

Abstract

Content-based retrieval allows finding information by searching its content rather than its attributes. The challenge facing content-based video retrieval (CBVR) is to design systems that can accurately and automatically process large amounts of heterogeneous videos. Moreover, content-based video retrieval system requires in its first stage to segment the video stream into separate shots. Afterwards features are extracted for video shots representation. And finally, choose a similarity/distance metric and an algorithm that is efficient enough to retrieve query – related videos results. There are two main issues in this process; the first is how to determine the best way for video segmentation and key frame selection. The second is the features used for video representation. Various features can be extracted for this sake including either low or high level features. This paper proposes a survey for a content based video retrieval systems that tries to address the issues for video segmentation and key frame selection as well as using both low level features for video representation. The major themes covered by the paper include shot segmentation, key frame extraction, feature extraction, clustering, indexing and video retrieval-by similarity. The aim of this paper is to review and analyze the interesting features that can be extracted from video data for indexing and retrieval along with similarity measurement methods.

1. Introduction

A system that supports video content-based indexing and retrieval has, in general, two stages: The first one, the database population stage, performs the following tasks: Video segmentation: Segment the video into constituent shots, Key frame selection: Select one frame frames, Hence shots. The second stage, the retrieval subsystem processes the presented query (usually in form of QBE), performs similarity matching Operations, and finally displays results to the user. [1]. Every year video content is growing in volume and there are different techniques available to capture, compress, displayer more to represent each shot and Feature extraction: Extract low-level and other features from key frames or their interrelationships in order to represent these, store and transmit video while editing and manipulating video based on their content is still a non-trivial activity. Recent advances in multimedia technologies allow the capture and storage of video data with relatively inexpensive computers. However, without appropriate search techniques all these data are hardly usable. Users want to query the content instead of the raw video data. Today research is focused on video retrieval 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. Features of the query object are compared with the shot Features for retrieval. [2]. The Internet forms today's largest source of Information containing a high density of multimedia objects and its content is often semantically related. The identification of relevant media objects in such a vast collection poses a major problem that is studied in the area of multimedia information retrieval. Before the emergence of content-based retrieval, media was annotated with text, allowing the media to be accessed by text-based searching based on the classification of subject or semantics. In typical content-based retrieval systems, the contents of the media in the database are extracted and described by multi-dimensional feature vectors, also called descriptors.[3].A novel video retrieval system using Generalized Eigen value Decomposition The system contains two major subsystems: database creation and database searching. In both subsystems, we propose new methods for shot-feature extraction, feature dimension transformation and feature similarity measuring base on GED.[4]

Large collections of publicly available video data grow day by day, the need to query this data efficiently becomes significant knowledge-based methods focuses on three techniques namely, rules, Hidden Markov Models (HMMs),and Dynamic Bayesian Networks.[5] Video processing always is performed on frames which are basic block of video. Group of frames captured together is called shot. Few minutes shot may contain hundreds of frames, which makes video large in size. Storing and processing theses individual frames are memory and computational expensive.

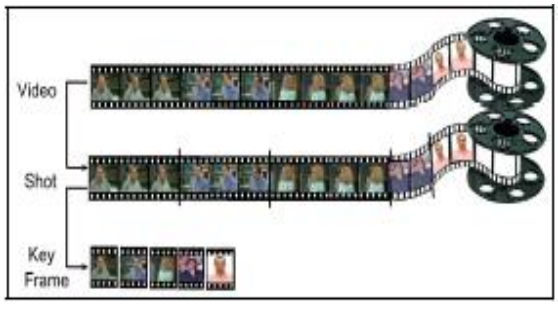

Figure 1: Video Segmentation

2. Related Work

2.1. Video Structure Analysis

The video takes into consideration four different levels which are frame, shot, scene, and story level. In frame level, each frame is treated separately as static image, set of contiguous frames all acquired through a continuous camera recording make shot level, set of contiguous shots common semantic significance make scene level and the complete video object is story level.[9] A typical structure of video is shown in Fig.2

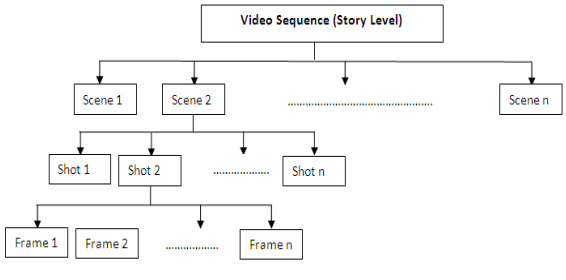

Figure 2: Video Structure

A hierarchical video browser consists of a number of levels, from the video title, to groups of shots, to shots, and to individual frames. Representative frames are displayed at each level. Subsequent levels are displayed when selected. Using the hierarchical video browser, the user can find relevant shots or frames quickly. For example, in Figure 3, the user first selected video 3 based on the title and cover frame and a collection of video shot groups were displayed to the user. Based on the visual information displayed in each video shot group, the user found group 1 interesting and selected it. All the shots in group 1 were displayed with an r frame for each shot. [1]

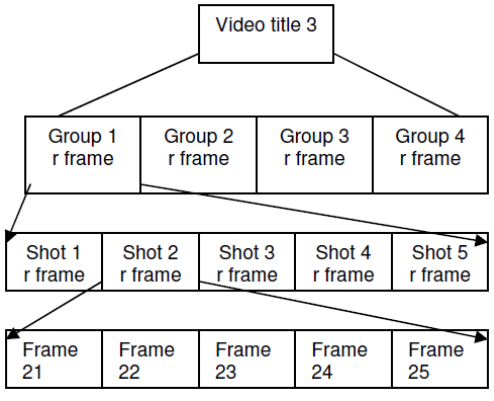

Figure 3: Video browser

2.2. Shot Boundary Detection

2.2.1 Shot boundary detection scheme based on rough fuzzy set. Han *et al*. [1] describe a technique for video shot boundary detection using rough fuzzy set. The selected low-level features are essential to achieve high accuracy for shot boundary detection. But there are too many features available in the frame or video, such as pixel values of different color channels, statistic features, intensity and color histogram etc. To detect the video shot boundaries, 12 candidate features, classified into 5 types, are usually extracted for common use [7]. The first is the RGB space model, the changes of three colors during shot transition can be measured, the 2ndis HSV space model, the component of which can be measured to the changes of hue, saturation and value between adjacent frames. In computation, we compute the mean of every component of each frame in the RGB or HSV model. The histogram features is categorized into two types: gray histogram and color histogram, which are our third and forth types of Features.

2.2.2The Hidden Markov model Technique. Boreczky and Lynn [11], describe a technique for segmenting video using hidden markov model. It uses three types of features for video segmentation-the standard histogram difference, an audio distance measure and an estimate of object motion between two

adjacent frames. The histogram feature measures the distance between adjacent frames based on the distribution of luminance levels. The hidden markov model has the following states cut, fade, dissolve, pan, zoom and shot. Each state of the HMM has an associated probability distribution that models the distribution of image, audio and motion features conditioned on that state. The parameters of the HMM are learned through a training phase. Once the parameters are trained, segmenting the video into its shots, camera motions and transitions is performed using the viterbi algorithm.

2.2.3 Shot change detection based on sliding window method. Li and Lee [8], describe a technique for shot change detection based on sliding window method. The Conventional Sliding Window (CSW) method has long been used in video segmentation for its adaptive thresholding technique. A hard cut is detected based on the ratio between the current feature value and its local neighbourhood in the sliding window. This method uses possibility values produced from different thresholds to measure the degree of possibility of a cuts presence. Detecting a cut based on possibility values, the method is more robust to camera/object motions. Each step produces a likelihood value, which measures the possibility of the presence of a cut. However, one of the purposes is to relax the threshold/parameter selection problem, i.e., to make the intermediate parameters to be valid for a wide range of video programs and to diminish the influence of the final threshold on the overall detection Performance.

2.2.4 Histogram based detection. Colin *et al*. [19], present a detailed evaluation of a histogram-based shot cut detector. The algorithm was specifically applied to large and diverse digital video collection consisting of eight h of TV broadcast video. It was found that the selection of similarity thresholds for determining shot boundaries in such broadcast video was difficult and necessitates the development of adaptive thresholding in order to address the huge variation of characteristics. The histogram creation technique used compared successive frames based upon three 64-bit histograms (one of luminance and two of chrominance). These three histograms were then concatenated to form a single N dimensional vector, where N was the total number of bins in all three histograms. The cosine measure was used for comparing histograms of adjacent frames [19]. A low cosine value indicates similarity. Setting the threshold involves a trade-off between two apparently conflicting point's sufficiently high threshold level to insulate the detector from noise and low enough threshold to make the

2.2.5 Shot Segmentation by graph partitioning. Cernekova, *et al*. [20], on the detection of gradual transitions such as dissolves and wipes, which are the most difficult to be detected. Unlike the abrupt cuts, the gradual transition spreads across a number of frames. In this method an automated shot boundary detection based on the comparison of more than two consecutive frames is used. Within a temporal window we calculate the mutual information for multiple pairs of frames. This way we create a graph for the video sequence where the frames are nodes and the measures of similarity correspond to the weights of the edges. By finding and disconnecting the weak connections between nodes we separate the graph to sub-graphs ideally corresponding to the shots. The major contribution of the algorithm is the utilization of information from multiple frames within a temporal window, which ensures effective detection of gradual transitions in addition to abrupt cut detection [21, 22]. The method relies on evaluating mutual information within a certain temporal frame window. The video frames are represented as nodes in a graph, whose edge weights signify the pair wise similarities of data points. Clustering is realized by partitioning the graph into disjoint sub-graphs. The method is able to detect efficiently abrupt cuts and all types of gradual transitions, such as dissolves, fades and wipes with very high accuracy.

2.3 Key Frame Extraction

Video consists of a number of levels, from the video title, to groups of shots, to shots, and to individual frames. One of the best ways to represent Frames is by its low level features. [1]

2.3.1 Color-Based Features. Color-based features include color histograms, color moments, color correlograms, a mixture of Gaussian models, etc. The exaction of color-based features depends on color spaces such as RGB, HSV. Color features can be extracted from the entire image or from image blocks into which the entire image is partitioned. Color-based features are the most effective image features for video indexing and retrieval. In particular, color histogram and color moments are simple but efficient descriptors [12].

2.3.2 Texture-Based Features. Texture can be defined as the visual patterns that have properties of homogeneity that do not result from the presence of only a single color or intensity. Shape Histogram of Image Texture is also an important visual feature that

refers to innate surface properties of an object and their relationship to the surrounding environment.[14] Many objects in an image can be distinguished solely by their textures without any other information. There is no universal definition of texture. Texture may consist of some basic primitives, and may also describe the structural arrangement of a region and the relationship of the surrounding regions In our approach we have used the texture features using gray-level co-occurrence matrix (GLCM). [12]

2.3.3 Shape Based Features. Shape-based features that describe object shapes in the image can be extracted from object contours or regions. A common approach is to detect edges in images and then describe the distribution of the edges using a histogram. Xiang-Yang Wang, Yong-Jian Yu [15] first divides the image into blocks and then extracts a edge histogram for each block. Shape e-based features are effective for applications in which shape information is salient in videos. However, they are much more difficult to extract than color or texture-based features.

2.4. Query and Video Retrieval

Video contains multiple types of audio and visual information, which are difficult to extract, combine in general video information retrieval.

2.4.1 Similarity Measure. Tan *et al*. [24] employed dynamic programming to align two video sequences of different temporal length. Global matching, on the other hand, measures the similarity between two shots by computing the distance between the two representative features of shots. For retrieving similar videos, besides computing similarity among shots, the temporal order of similar shots between two videos are also taken into account [22]. Ngo and Pong [21], Video retrieval is still at its preliminary state, despite the fact that videos in addition to image information, consist of extra dimensional information. Three problems that have been attempted are retrieve similar videos [22], locate similar video clips in a video [23] retrieve similar shots. In general, similarity measure can be done by matching features either locally or globally. Local matching requires aligning and matching frames (or key frames) across time. The most direct measure of similarity between two videos is the average distance between the features of the corresponding frames [25]. Query by example usually uses low-level feature matching to find relevant videos. However, video similarity can be considered in different levels of resolution or granularity. According to different user' demands, static features of key frames, object features, and motion features [26] all can be used to measure video similarity.

2.4.2 Query Types:

2.4.2.1 Query by Example. This query extracts low-level features from given example videos or images and similar videos are found by measuring feature similarity. The static features of key frames are suitable for query by example, as the key frames extracted from the example videos or exemplar images can be matched with the stored key frames.

2.4.2.2 Query by Sketch. This query allows users to draw sketches to represent the videos they are looking for. Features extracted from the sketches are matched to the features of the stored videos. Hu *et al*. [26] propose a method of query by sketch, where trajectories drawn by users are matched to trajectories extracted from videos.

2.4.2.3 Query by Objects. This query allows users to provide an Image of object. Then, the system finds and returns all occurrences of the object in the video database [24]. In contrast with query by example and query by sketch, the search results of query by objects are the locations of the query object in the videos.

2.4.2.4 Query by Keywords. This query represents the user's query by a set of keywords. It is the simplest and most direct query type, and it captures the semantics of videos to some extent. Keywords can refer to video metadata, visual concepts, transcripts, etc. In this paper, we mainly consider visual concepts.

2.4.2.5 Video Retrieval using visual information

While analyzing the video imagery, we considered the colour similarity of images and the presence of faces and text that was readable on the screen. Using the TREC video Collection and the automatic known-item queries, we compared our probabilistic image retrieval model against two other vector-based image retrieval algorithms, namely the well-known QBIC image search engine and a Munsell-color histogram based image retrieval algorithm. Both of these two algorithms represent an image as a vector of features and compute the similarity between images based on the Euclidean distance between their representation vectors. The main finding from the results on individual features is probabilistic image retrieval provided the best result for any single metadata type. It is not too surprising that the results indicate that image retrieval was the single biggest factor in video retrieval for this evaluation. Good image retrieval was the key to good performance

in this evaluation, which is consistent with the intuition that video retrieval depends on finding good video images when given queries that include images or video. One somewhat surprising finding was that the speech recognition transcripts played a relatively minimal role in video retrieval for the known-item queries in our task. This may be explained by the fact that discussions among the track organizers and participants prior to the evaluation emphasized the importance of a video retrieval task.

2.4.2.6 Textual query for video retrieval. Jawahar and Chennupati [18], we present an approach that enables search based on the textual information present in the video. Regions of textual information are identified within the frames of the video. Video is then annotated with the textual content present in the images. An advanced video retrieval solution could identify the text present in the video, recognize the text and compute the similarity between the query strings and pre-indexed textual information present in the video. However, success of this technique depends on two important aspects:

• Quality of the input video

• Availability of an OCR for robustly recognizing

The text images

2.4.2.7 Refinement and relevance feedback. Several Relevance Feedback (RF) algorithms have been proposed over the last few years. The idea behind most RF-models is that the distance between images labelled as relevant and other similar images in the database should be minimal. The key factor here is that the human visual system does not follow any mathematic metric when looking for similarity in visual content and distances used in image retrieval systems are well-defined metrics in a feature space.

. 3. Proposed Plan

The challenge facing content-based video retrieval is to design systems that can accurately and automatically process large amounts of heterogeneous videos. In our proposed video retrieval system requires in its first stage to segment the video stream into separate shots. Afterwards features are extracted for video shots representation and finally, choose a similarity/distance metric and an algorithm that is efficient enough to retrieve query related videos results. There are two main issues in this process: the first is how to determine the best way for video segmentation and key frame selection. The second is the features used for video representation. Video is divided into images. An image is uniformly divided into 16 coarse partitions as a first step. After the above coarse partition, the centroid of each partition is selected as its dominant color. Texture of an image is obtained by using Gray Level Co-occurrence Matrix .Color and texture, features are normalized. Weighted Euclidean distance of color, texture and Histogram features is used in retrieving the similar video images.

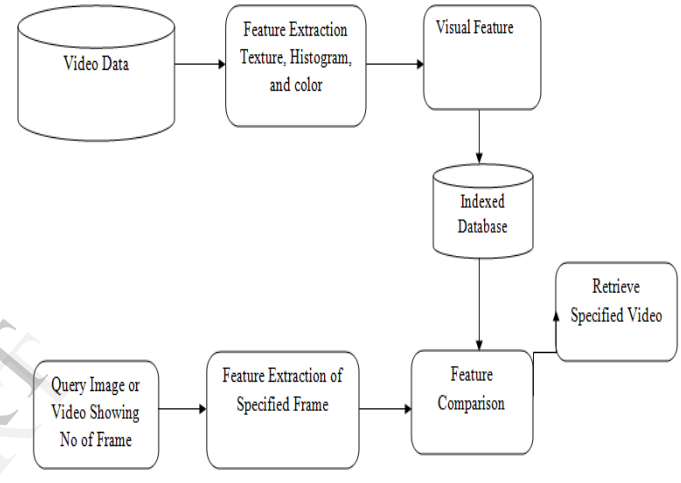

Figure 4: Block Diagram of Proposed Video Retrieval System

4. Conclusions

We probably covered only a small part of all existing video retrieval systems but we can still draw some conclusions from this survey we have presented a review on recent developments in content-based video Retrieval. Different aspect of video retrieval System has been described with the focus on the following tasks: video structure analysis including shot boundary detection, key frame extraction and scene segmentation, extraction of features video search including interface, similarity measure, Video Retrieval using visual information, texture query and At the end of this survey, I have also given Proposed video retrieval system which based on feature extraction from feature like color ,texture and Histogram of an Video Image.

References

[1] Shweta Ghodeswar, B.B.Meshram Technicians "Content Based Video Retrieval"

[2] R. Kanagavalli, Dr. K. Duraiswamy "Object Based Video Retrievals" International Journal of Communications and Engineering Volume 06– No.6, Issue: 01 March2012.

[3] Dr. S. D. Sawarkar , V. S. Kubde "Content Based Video Retrieval using trajectory and Velocity features" International Journal of Electronics and Computer Science Engineering ISSN- 2277-1956

[4] Ali Amiri, Mahmood Fathy, and Atusa Naseri "A Novel Video Retrieval System Using GED-based Similarity Measure" International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 2, No.3, September 2009

[5] S.Thanga Ramya 1 P.Rangarajan "Knowledge Based Methods for Video Data Retrieval" International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 5, Oct 2011

[6] B. V. Patel A. V. Deorankar, B. B. Meshram "Content Based Video Retrieval using Entropy, Edge Detection, Black and White Color Features" Computer Engineering and Technology (ICCET), 2010 2nd International Conference on

[7] Gao, X. and X. Tang, 2002. Unsupervised video shot segmentation and model-free anchorperson detection for news video story parsing. IEEE Trans. Circuits Syst. Video Technol., 12: 765-776.

[8] Shan Li, Moon-Chuen Lee, 2005. An improved sliding window method for shot change detection. Proceeding of the 7th IASTED International Conference Signal and Image Processing, Aug. 15-17, IIonolulu, IIawaii, USA, pp: 464-468.

[9] Hamdy K. Elminir, Mohamed Abu ElSoud "Multi feature content based video retrieval using high level semantic concept" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 2, July 2012

[10] John, S., Boreczky and D. Lynn, 1998. "A hidden Markova model framework for video segmentation using audio an image features". In: Proceedings of IEEE International conference on Acoustics, Speech and Signal Processing, May 12-15, 6:

[11] Weiming Hu, Niangua Xie, Li Li, Xianglin Zeng, and Stephen Mayban" Survey on Visual Content-Based Video Indexing and Retrieval" IEEE Transaction on System, Man, and cybernetics —part C: application and Reviews, Vol. 41, no. 6, November 2011

[12] P. Geetha and Vasumathi Narayanan" A Survey of Content-Based Video Retrieval" Journal of Computer Science 4 (6): 474-486, 2008 ISSN 1549-3636

[13] M.Babu Rao Dr.B.Prabhakara Rao Dr.A.Govardhan "Content Based Image Retrival using Dominant Colour and Texture Features" International Journal of Computer science and information security Vol. 9,No.2 2011

[14] Xiang-Yang Wang , Yong-Jian Yu , Hong-Ying Yang "An effective image retrieval scheme using color, texture and shape features "Computer Standards and Interface Science Direct

[15] T.N.Shanmugam and Priya Rajendran "An Enhance Content based video retrieval system based on query clip" International Journal of Research and Reviews in Applied Sciences ISSN: 2076-734X, EISSN: 2076-7366 Volume 1, Issue 3

[16] Fan Hui-Kong "Image Retrieval using Both Colour and texture features " Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009

[17] M.Babu Rao R.B.Prabhakara Rao Dr.A.Govardhan "Content based image retrieval using dominant colour and texture feature" International Journal of Computer Science and Information Security,Vol. 9, No. 2, February 2011

[18] C V. Jawahar, BalaKrishna Chennupati, Balamanohar Paluri and Nataraj Jammalamadaka, "Video Retrieval Based on Textual Queries", in Proceedings of the Thirteenth International Conference on Advanced Computing and Communications, Coimbatore, December 2005.

[19] O'Toole, C., A. Smeaton, N. Murphy and S. Marlow, 1999. Evaluation of automatic shot boundary detection on a large video suite. In: 2nd U.K. Conference Image Retrieval: The Challenge f Image Retrieval, Feb. 25-26, Newcastle, U.K.

[20] C.-W. Ngo, H.-J. Zhang and T.-C. Pong, "Recent Advances in Content based video analysis," International Journal of Image and Graphics, vol.1, no. 3, pp. 445–468, 2001.

[21] Yi Wu, Y. Zhuang, and Y. Pan: "Content-Based Video Similarity Model," In Proc. of the 8th ACM Int. Multimedia Conf. on Multimedia, USA,

[22] Anil Jain, Aditya Vailaya, Wei Xiong., "Query by video Clip," In Proceedings of Fourteenth International Conference on Pattern Recognition, vol.1. pp: 909-911, 16- 20 Aug 1998.

[23] Jang-Hui Kim, Hye-Youn Lim, and Dae-Seog Kang,"An Implementation of the Video Retrieval System by Video Segmentation".Proceedingsof APCC 2008.