International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN No.:2349-9745, Date: 2-4 July, 2015

Video Retrieval : A Review

Kainjan Sanghavi[1], Dr. Rajeev Mathur[2], Mahesh Sanghavi[3]
[1]*Computer Engineering, PAHER,Udaipur,kainjan@gmail.com*
[2]*Principal, LMCST,Jodhpur, rajeev.mathur69@gmail.comEmail id*
[3] *Computer Engineering, SNJB's KBJ COE,Chandwad, sanghavi.mahesh@gmail.com*

Abstract— The significance of content based video retrieval has swelled out. Video retrieval aims at fetching videos from a video repository which are pertinent to the query in a flexible manner. Many algorithms and approaches have been designed for solving the problems of video retrieval, classification, and indexing. This paper provides an overview of current research in video retrieval and outline of areas for future research.
Keywords - Content based Video Retrieval, Indexing, Scene change detection, Classification, Video Repository

I. INTRODUCTION

The availability of huge network bandwidth, rapid growth of video technologies and easy access to complex standards has encouraged simultaneous retrieval of videos from large video repositories. Thus, proficient techniques for preprocessing, classification, retrieval and indexing of videos have become more significant. Video is a significant mode of communication and multimedia information. Videos are more affluent than images, and carry bulk of information with very modest structure. These distinctiveness make video retrieval and indexing a challenging area of research. Prior, the video databases were reasonably small, and the retrieval were major based on manual keywords. However, with the advent of the field of big data processing the database have increased in size as well as incorporated advanced technologies for storing and securing data. Thus, content based video indexing and retrieval based from large video repository is further research confronts.
Widespread use of content-based video retrieval and indexing subsist in searching and browsing large video archives, indexing and archiving multimedia presentation(such as news event analysis, TV broadcast summarization), simple and effortless access to educational information and video surveillance, video content filtering [1].
A video is composed of audio, metadata as well as visual information. Video retrieval aims at extracting video using the above mentioned information. Fig.1 shows the general process of CBVR. Content based Video Retrieval systems require video segmentation, feature extraction, querying and retrieval using appropriate indexing. Video segmentation includes finding of the shot boundaries, keyframes and scenes from the video. Secondly, features are extracted from these scenes or shots. These features may be low level or high level features for visual or auditory channels. The query can be in the form of keywords, example, sketch, objects, natural language, emotions or a combination of any of these. Videos retrieved on the basis of comparison with database videos are then indexed according to their ranks.
The significance and recognition of Content Based Video Retrieval has promoted many researches in this arena. This paper reviews about various CBVR components, limitations and virtues of different techniques and the future directions. Section II discusses about various video preprocessing analysis. Section III briefly reviews the work for feature extraction. Section IV describes the different approaches for video query, retrieval and indexing. Finally, Section V concludes the survey.

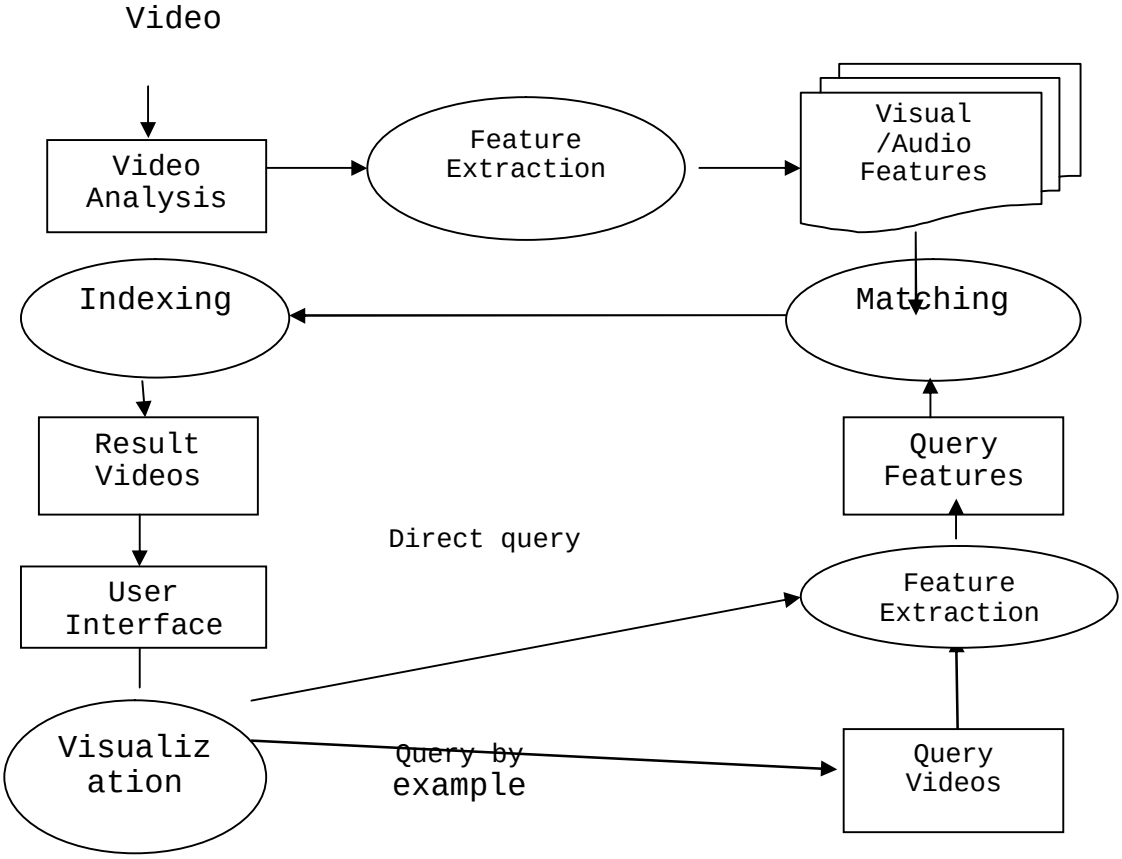

Fig.1 Framework for content based video retrieval

II. VIDEO PARSING

Video parsing is the detection and identification of meaningful segments of video i.e shots, scenes and frames as shown in Fig. 2.

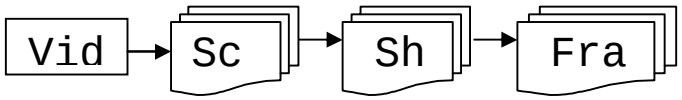

Fig. 2 Video parsing process

• Shot: A sequence taken by a single camera
• Scene: single dramatic event taken by a small number of related cameras.
• Frame: A still image
One of the main research areas in this direction is about the identification of video sequences i.e. finding scenes/shots/keyframes.

2.1 Shot Transition Detection:

Shot transition detection is utilized to separate up a video into basic temporal units referred to as shots; a shot could be a series of reticulated consecutive footage taken contiguously by one camera and representing never-ending action in time and space[2]. It is an elementary step for automatic classification and content-based video retrieval or summarization applications which offer an

proficient access to very large video repositories, e.g. an application might query for a representative image from every scene that make a summary of a movie. In this context the search engine can search things like "Show display all films wherever there is a scene with a mickey mouse in it." These are usually grouped into two types:

2.1.1 Abrupt Transitions: These are unexpected transitions from one shot to another, i. e. one frame belongs to the primary shot, the subsequent frame belongs to the second shot. They are conjointly known as hard cuts or merely cuts.

2.1.2 Gradual Transitions: Here the two shots are joined using chromatic, spatial or spatial-chromatic effects that progressively swap one shot by another. These are soft transitions are classified as wipes, dissolves, fades.

Many papers present the various algorithms to detect shot boundaries [3-6]. Two major steps include in shot boundary detection: i) Extracting features ii) Finding dissimilarities among the adjacent frames.If the dissimilarities are major, a shot is detected.

2.2 Scene Change Detection

A single shot generally results from a solitary continuous camera operation. This partitioning is typically achieved by consecutive measurement of inter-frame variations and learning their variances, e.g., identifying spiky crest. This method is usually known as scene change detection (SCD). Scenes have higher level semantics than shots.
Most of the prevailing effort on SCD relies on full-image video analysis. The variations between the varied SCD methods are the measurement utility used, feature chosen, and therefore the subdivision of the frame images. Several use the intensity feature [7-11] or motion information [12-14] of the video information to calculate the interframe difference sequence. The setback of intensity-based approaches is that they might fail once a peak exists by associate object or camera motion. Motion-based algorithms are also costly in terms of computation, since they perform matching of two frames block-by-block. As soon as the differences are computed, a global threshold is used to determine a scene change. But this is not efficient as this may not essentially imply that there's a scene change reported. In fact, scene changes with globally low peaks represent often cause the failure of the algorithms. Scene changes, either abrupt or gradual, are localized processes, and should be checked consequently.

2.3 Keyframe Analysis

The magnified demand for intelligent process and analysis of multimedia system data has led to the augmentation of various techniques for intelligent video management. Among these approaches, shot transition detection is the primary commencement of content based video analysis and secondly, key frame also is a straightforward technique, however known as an economical type of video abstract.
Key frames can be identified using histogram difference [15] between two successive frames. Janko Calic and Ebroul Izquierdo proposed an algorithm which uses frame difference [16]. Borth et. al[17] provide work for key frames using K-means clustering algorithm. The complexity however is increased using this approach.

III. FEATURE EXTRACTION

Extracting the features means to reduce the quantity of resources needed to elucidate a vast set of information. A Content Based Video Retrieval (CBVR) system achieves its internal representation of content through feature extraction. The CBVR features include low level features as Color , Texture ,Shape and Motion. High level features as spatial relationship, semantic primitives, Text & domain concepts. Some features also include objective and subjective attribute.

The color-based feature extraction depends upon color spaces as RGB,HSV,YCbCr etc. Color histograms are generally used to represent the color features of video frames [18] [19]. Texture is specified in order to retrieve a specific pattern appearing in an image.

The energy distribution within the frequency domain is utilized by several methods for texture retrieval and classification in order to acknowledge the texture. These techniques include Neural Network, Wavelet, Gabor filter[20]. The illustration of the content of the image considers shape of the image as a significant feature. Human eyes percept image by being sensitive to edge features. Object recognition can be effectively performed using edge histograms. The Edge Histogram (EH) uses the Sobel Operator to capture the spatial distribution of edges [21].

Motion estimation techniques[22] figure out the core of video processing applications. Motion estimation extracts motion data from the video sequence. However, the majority researchers apply Block Matching Algorithm to extract motion information.

IV. VIDEO QUERY, INDEXING AND RETRIEVAL

Once the feature vector is formed as a function of time, a video database can also be searched in order to retrieve frames that possess specific properties like gloomy frames, frames with many trees or motion and so on. The feature vector space is idyllic for such comparisons, since it contains all essential frame properties, whereas its dimension is far less than that of the image space. Thus, query types can be different as reviewed here.

Query By Example : It is considered to be a promising approach since it provides a user with an intuitive way of query representation and the form of expressing a query condition close to that of the data to be evaluated

Query By Sketch : Using sketch-based interaction to create new content is an intriguing idea: users could assemble an object from existing parts or even create a complete scene from existing objects, using only rough sketching strokes. Berlin et.al [23] ask participants to create an input query sketch given the name of a category only (e.g. "airplane"), without providing an example rendering.

Query by Objects : Here, the users provide an image of object as a query. The system searches and returns the images those match with that object. As compared to the previous two discussed types the search results of query by objects are the locations of the question object within the videos.

Query by Keywords : This query is represented by a collection of keywords. It is the only and most direct query type. It captures the linguistics of videos to some extent. Keywords may refer to video information, visual ideas, transcripts.

Video indexing is the process before querying. It provides watchers some way to access and navigate contents easily; almost like book indexing. This management can be done by: Metadata-based, Text based Audio – based, Content – based and Integrated [24].

Along with the various features, distance metrics is also one of the important aspect for video similarity matching. M.C.Lee et.al. [25] provide a study of distance measure for Video sequence similarity matching**.** Based on the edit operations, the vstring edit distance,Leveinsthein Distance, Manhattan distance are used as similarity measure for video sequence matching.

CONCLUSION

Review of the previous study reveals that as manual annotations are tedious in text-based retrieval, the content-based retrieval systems are more desirable than the text-based approach. Content-based image retrieval techniques also cannot be applied on real-time videos that consist of the motion information. Thus, Content based video retrieval systems (CBVR) need to be explored. This survey gives the ways in the state of field of CBVR in the areas of feature extraction, representation and users interaction. CBVR uses the visual contents of video in the process of retrieving the videos from the database.

REFERENCES

[1] Dimitrova, Nevenka, et al. "Applications of video-content analysis and retrieval." *IEEE multimedia* 9.3 (2002): 42-55.
[2] R. Hamid, S. Maddi, A. Bobick, and M. Essa, "Structure from statistics Unsupervised activity analysis using suffix trees," in Proc. IEEE Int. Conf Comput. Vis., Oct., 2007, pp. 1–8
[3] Amel, Abdelati Malek, Ben Abdelali Abdessalem, and Mtibaa Abdellatif. "Video shot boundary detection using motion activity descriptor." arXiv preprint arXiv:1004.4605 (2010).
[4] Ganesh.I.Rathod ,Dipali A. Nikam," An Algorithm for Shot Boundary Detection and Key Frame Extraction Using Histogram Difference", International Journal of Emerging Technology and Advanced Engineering ,ISSN 2250 - 2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 8, August 2013
[5] Chen, Shu-Ching, Mei-Ling Shyu, and Chengcui Zhang. "Innovative shot boundary detection for video indexing.", Video data management and information retrieval (2005): 217-236.
[6] Mr. Sandip T. Dhagdi, Dr. P.R. Deshmukh , "Keyframe Based Video Summarization Using Automatic Threshold & Edge Matching Rate" , International Journal of Scientific and Research Publications, Volume 2, Issue 7, July 2012 ,ISSN 2250 3153
[7] Nagasaka A, Tanaka Y (1991) Automatic Video Indexing and Fall video Search for Object Appearances. In: Knuth E, Wegner L (eds) Second Working Conference on Visual Database Systems (Budapest, Hungary), IFIP WG 2.6, North-Holland, New York, NY, USA, pp 119–133
[8] Otsuji K, Tonomura Y (1993) Projection Detecting Filter for Video Cut Detection. In: Rangan P (ed) Proc First ACM International Conference on Multimedia, ACM, New York, NY, USA, pp 251–257
[9] Otsuji K, Tonomura Y, Ohba Y (1991) Video Browsing Using Brightness Data. SPIE 1606 (Visual Communications and Image Processing): 980–989
[10] Zhang HJ, Kankanhalli A, Smoliar SW (1992) Automatic Partition of Animate Video. Tech. Report, Institute of System Science, National University of Singapore, Singapore
[11] Zhang HJ, Kankanhalli A, Somilar SW (1993) Automatic Parsing of Full-Motion Video. Multimedia Syst 1:10–28
[12] Akutsu A, Tonomura Y, Hashimoto H, Ohbak Y (1992) Video Indexing Using Motion Vectors. In: Maragos P (ed) Proc of SPIE: Visual Communication and Image Processing 92. SPIE, Bellingham, WA, USA, pp 1522–1530
[13] Hsu PR, Harashima H (1994) Detecting Scene Changes and Activities in Video Databases. In: ICASSP'94, Vol. 5, IEEE, Piscataway, NJ, USA, pp 33–36
[14] Shahraray B (1995) Scene Change Detection and Content-Based Sampling of Video Sequences. SPIE 2419 (Digital Video Compression: Algorithms and Technologies): 2–13
[15] S. Thakare, "Intelligent Processing and Analysis of Image for shot Boundary Detection", International Journal of Engineering Research and Applications, Vol. 2, Issue 2, Mar-Apr 2012,pp.366-369.
[16] Calic and E. Izquierdo, "Efficient Key-frame Extraction And Video Analysis", International Symposium On Information Technology, April 2002,IEEE.
[17] D. Borth, A. Ulges, C. Schulze, T. M. Breuel, "Key frame Extraction for Video Tagging &Summarization", 2008.
[18] Nievergelt, H. Hinterberger, and K.C.Sevcik."The grid file: An adaptable, symmetric multikey file structure", the ACM Transactions on Database Systems , 9(1):38–71,March 1984
[19] G. Pass, R. Zabih, and J.Miller ," Comparing images using color coherence vectors", In Proc. Of the F
ourth ACM Multimedia Conference , pages 65–74 , New York, NY.USA, November 1996
[20] J. Zhang, M. Marszałek, S. Lazebnik, C. Schmid, "Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study", International Journal of Computer Vision, vol.73 no.2, pp.213-238, June 2007
[21] Patil, Bhagya, Anupama Pattanshetty, and Suvarna Nandyal. "Plant classification using SVM classifier." (2013): 519-523.

[22]Milind Phadtare, "Motion estimation techniques in video processing", Electronics Engineering Times, August 2007
[23] Eitz, Mathias, et al. "Sketch-based shape retrieval." *ACM Trans. Graph.* 31.4 (2012): 31.
[24] Dr. Dimitrios Tzovaras, "Audio-Visual Content search and retrieval in a distributed P2P repository" FP6 Summary of Projects , Information Society. Networked Audio-Visual Systems.
[25]D.A. Adjeroh, M.C.Lee, I.King, "A distance measure for video sequence similarity matching", *Proc. MINAR'98,* Hong Kong, August 1998.

