A Survey on Scene Detection and Video Mining Techniques

D.Shanmuga Priyaa[1] Dr. S. Karthikeyan[2]

ABSTRACT

Data mining is an emerging field of interest for research in particular video mining has attracted a lot of research minds because of the vast quantity of video data available in recent years. Video data mining plays an imperative role in competent video data management with the ever-increasing amount of video data. All along with the enormous amount of data, a quick method to mining semantic patterns is needed. The foremost problem on video data mining is the appearance of the object in a visual media. The quandary is complicated because the object can come into view in different customs in different parts of the video due to dissimilar imaging conditions, lightening conditions, back grounds and occlusions. It is extremely difficult to extract and combine the multiple types of audio and visual information that are incorporated in a video. The research conducted in the field of visual data mining is not enough. All these variations make visual data mining more challenging compared to other data mining. The study is based on current trends in video retrieval. The survey includes the shot detection, key frame extraction, clustering, indexing, and video object retrieval based on similarity measures for color, shape, region, motion, texture, refinement and relevance feedback. In addition this paper also provides marginal idea for future research in this field.

[1]Asst.Prof at Karpagam University, Coimbatore.

[2]Professor and Director in School of Computer Science and Applications, Karpagam University, Coimbatore.
Email : skaarthi@gmail.com

Keywords : Data Mining, Data Management, Clustering, Video Mining, Scene Detection, Pattern, Segmentation, Video Object Retrieval.

I. INTRODUCTION

Data mining is the process of analyzing data from different perspectives and to find useful patterns or information which can be used in a variety of applications. In other words, Data mining is a technique to discover previously unknown and interesting patterns called 'semantic patterns' from an immeasurable quantity of data. Though the data mining technique is applied in a variety of domains, this survey will give attention on the application video data, called 'video data mining' [1] [2]. Video data mining plays an imperative role in competent video data management with the ever-increasing amount of video data. Along with the enormous amount of data, a quick method to mining semantic patterns is needed.

It is extremely difficult to extract and combine the multiple types of audio and visual information that are incorporated in a video. Accessible work in video data mining can be divided into two categories: mining similar motion patterns and mining similar objects. The first type of systems uses motion information to mine similar event patterns or identify peculiar events. The second category systems aim to group frequently appearing objects in videos. The predicament is difficult because the object can appear in different ways in different parts of the video due to different imaging conditions, lightening conditions, back grounds and occlusions [3].

A common first step for most content-based video analysis techniques available is to segment a video into elementary shots, each comprising a continuous in time and space. These elementary shots are composed to form a video sequence during video sorting or editing with either cut transitions or gradual transitions of visual effects such as fades, dissolves and wipes [4].

The following steps are most significant in case of a content based video retrieval system. They are video segmentation, feature extraction and feature grouping. Generally, a video segmentation algorithm attempts to divide a video sequence into meaningful subgroups termed as 'shots.' In recent times a number of techniques, varying from color histogram to block based approaches with motion compensation have been proposed for video segmentation [5]. The mainstream of the presented feature extraction algorithms decide on one or more key frames as being representative of each shot. Feature extraction techniques such as wavelets or Gabor filters are extensively used to extract the features from these frames. In the final stage, the shot features are grouped into clusters to correspond to relevant objects which appear in those shots. In a query problem, features selected from the query region will be compared with existing cluster features for possible matches [6].

II. Related Work

This section of the paper discusses the earlier works proposed by researchers for effective video data mining. The foremost problem on video data mining is the appearance of the object in a visual media. The object can come into view in different customs in different parts of the video due to dissimilar imaging conditions, lightening conditions, back grounds and occlusions.

Anjulan and N. Canagarajah in [7] proposed an approach for video scene retrieval. Their proposed approach is based on the local region features. The approach is an alternative to the key frame method. They described a novel technique for content extraction and scene retrieval for video sequences based on local region descriptors. Stable features are extracted throughout a shot rather than from a small number of key frames. Local regions are followed throughout a shot with features being extracted from stable tracks. A capable method was proposed for region tracking to steer clear of possible repetition of the features. Their proposed framework is robust to camera and objects motion and can withstand severe illumination changes, spatial editing and noise.

An effective video mining technique was described by Missaoui et al. in [8]. Their paper is dedicated to revisiting image and video mining techniques from the perspective of image modeling approaches, which amount to the theoretical basis for these techniques. The most important areas belonging to image or video mining are: image knowledge extraction, content-based image retrieval, video retrieval, video sequence analysis, change detection, model learning, as well as object recognition. Conventionally, these areas have been developed independently, and hence have not benefited from some common sense approaches which provide potentially optimal and time-efficient solutions. Two different types of input data for knowledge extraction from an image collection or video sequences are considered: original image or symbolic (model) description of the image. Several basic models are described briefly and compared with each other in order to find effective solutions for the image and video mining problems. They include feature-based models and object-related structural models for the representation of spatial and temporal entities (objects, scenes or events).

Compound image is a combination of text, picture and graphs. Noise reduction in compound image is necessary to maintain the quality of images. Noise is added into an image at the time of image acquisition (or) image

capturing. After capturing, image preprocessing is necessarily done to correct and adjust the image for further classification and segmentation. From the literature study different filtering techniques are available to reduce the noise from compound images. Normally the filters are used to improve the image quality, suppress the noise. This paper proposes median filtering technique for removing salt & pepper noise from various types of compound images. Several examples were conducted to evaluate the performance of the median filter on noise. [9]

Chasanis et al. in [10] expressed an approach for video data mining. Video indexing requires the efficient segmentation of the video into scenes. In the method they proposed, the video is first segmented into shots and key-frames are extracted using the global k-means clustering algorithm which represent each shot. They then applied an improved spectral clustering method to cluster the shots into groups based on their visual similarity. Moreover they assigned a label to each shot corresponding to the group to which they belong to. In order to identify the patterns in the sequence of shot labels they compared the label pairs of successive shots. A scene boundary is detected when there occurs' a change in such pattern. Their proposed approach achieved high correct detection rates while preserving a good trade-off between the number of missed scenes and the number of false detected scenes.

The earlier work for color similarity deals with the simplest feature such as mean color vector of a region in the RGB space as the colorimetric feature [11]. They assumed that if two regions strongly differ on one of the features, they are not similar. The property is called as absorbing property [11]. The shape similarity measures benefits from absorbing property. They used the region's oriented bounding box (OBB) properties to characterize a region shape [11]. The approach developed is based on matching of region adjacency graphs (RAG) of pre-segmented objects [11]. The region features (texture, color, shape) are not strongly relevant due to the resolution. The problem of object matching can be expressed in terms of directed acyclic graph (DAG) matching [12]. The mainstream of the presented feature extraction algorithms decide on one or more key frames as being representative of each shot. Fourier descriptors, compactness, and eccentricity features are commonly used for shape retrieval in the content based image retrieval systems [13] [14]. A new method for object based image retrieval, efficient subimage retrieval (ESR) is introduced by [15].

Fang et al. in [16] projected a fuzzy logic approach for detection of video shot boundaries. In their paper, they proposed a fuzzy logic approach to put together hybrid features for detecting shot boundaries inside general videos. These features include color histogram intersection, motion compensation, texture change, and edge variances. The fuzzy logic approach contains two processing modes, where one is dedicated to detection of abrupt shot cuts including those short dissolved shots, and the other for detection of gradual shot cuts. These two modes are unified by a mode-selector to decide which mode the scheme should work on in order to achieve the best possible detection performances. The advantages of their contribution can be highlighted as: (i) a range of features can be integrated by fuzzy logic operation to exploit their individual strength collectively; and (ii) while directly thresholding features remains sensitive to noises, selecting threshold in fuzzy domain provides a buffered operation and thus makes the detection more reliable. Experimental results support that the proposed

algorithm is effective in video segmentation benchmarked by three existing algorithms and measured by precision and recall rates.

An efficient video shots retrieval system based on local feature detection, description and matching was presented by Yuanjia Du et al. in [17]. Initially, they used a face tracker to obtain information on faces in different viewpoints. A visual vocabulary is built off-line using an invariant descriptor computed on tracked character face regions in all shots. The vocabulary is sophisticated in two customs to make the retrieval system more competent. Firstly, the visual vocabulary is minimized by only using facial features selected on face regions which are detected by a precise face detector. Secondly, three criteria, namely Inverted-Occurrence-Frequency Weights, Average Feature Location Distance and Reliable Nearest-Neighbors, are calculated in advance to make the on-line retrieval procedure more resourceful and accurate.

Kimiaki et al. in [18] described as effective approach for video data mining. In their paper, concerning to the temporal localities of the semantic event, they extracted the sequential patterns, each of which is a sequence obtained by connecting temporally close and robustly connected raw level metadata. Subsequently, they proposed a parallel data mining method in order to reduce the expensive computational time. Afterward, they verified whether their parallel algorithm was effective or not against huge amount of data. Finally, extracted patterns are verified in opposition to human interpretation, that is, whether these patterns can be considered as semantic events or not. In other words, they investigated what kind of semantic events the extracted patterns characterize. In order to denote the spatial aspects they extracted a key frame which was represented by the middle video frame in the shot. Moreover, they assigned the class index to each key frame by using some threshold values or by using k-Means algorithm [19] based on a distance metric using histogram intersection. Also in order to represent temporal aspects, they extracted time duration, sound stream and motion vector in a shot.

A new method for multimedia data mining was presented by M. Shyu et al. in [20]. Their paper proposed a subspace-based multimedia data mining framework for video semantic analysis, distinctively video event/ concept detection, by concentrating on two fundamental issues, i.e., semantic gap and rare event/concept detection. Full automation is achieved by the proposed framework by means of multimodal content analysis and intelligent integration of distance-based and rule-based data mining techniques. Additionally, the exclusive domain-free characteristic indicates the immense potential of extending the proposed multimedia data mining framework to an extensive collection of different application domains.

Zhu et al. in [21] projected a video data mining technique using semantic indexing and event detection from the association perspective. The low-level features often have little meaning for naive users, who much prefer to identify content using high-level semantics or concepts. This creates a gap between systems and their users that must be bridged for these systems to be used effectively. Their paper first presents a knowledge-based video indexing and content management framework for domain specific videos (using basketball video as an example). Moreover, the approach makes use of video processing techniques to find visual and audio cues (e.g., court field, camera motion activities, and applause), and introduced a multilevel sequential association mining to investigate

associations among the audio and visual cues, organized the associations by assigning each of them with a class label, and used their appearances in the video to construct video indices.

A method of video mining with frequent itemset configurations was explained by Quack et al. in [22]. The goal of their work is to mine interesting objects and scenes from video data. In other words, to detect frequently occurring objects automatically. Mining such representative objects, actors, and scenes in video data is useful for many applications. Their approach relies on frequent itemset mining algorithms, which have been successfully applied to several other, large-scale data mining problems such as market basket analysis or query log analysis. Also they demonstrated how to integrate spatial arrangement information in transactions and how to select the neighborhood defining the subset of image features included in a transaction. For scenes with significant motion, they defined this neighborhood by the use of motion segmentation. To this end, they also introduced an uncomplicated and very quick technique for motion segmentation on feature codebooks.

Chang et al. in [23] expressed an approach for video shot detection using clustering. In their paper, they presented the development of a new color feature extraction algorithm that addresses this problem, and they also proposed a new clustering strategy using clustering ensembles for video shot detection. Based on Fibonacci lattice-quantization, they developed a novel color global scale-invariant feature transform (CGSIFT) for better description of color contents in video frames for video shot detection. CGSIFT as an initial step quantizes a color image, representing it with a small number of color indices, and then uses SIFT to extract features from the quantized color index image. They also developed a new space description method using small image regions to represent global color features as the second step of CGSIFT. Clustering ensembles focusing on knowledge reuse are then applied to accomplish better clustering results than using single clustering methods for video shot detection. Evaluation of the proposed feature extraction algorithm and the new clustering strategy using clustering ensembles reveals very promising results for video shot detection.

Video mining for creative rendering was described by Chen et al. in [24]. In their paper, a video motion mining framework for creative rendering is presented. The user's capture intent is derived by analyzing video motions, and respective metadata is generated for each capture type. The metadata can be used in a number of applications, such as creating video thumbnail, generating panorama posters, and producing slideshows of video. Their research theme is the outcome of the following observations: people capture home videos for a variety of purposes. One of the main purposes is to capture action and sound, in which case many trophy shots representing a fleeing moment have been captured on home videos. Another one is to capture the environment, such as a panoramic view on top of a mountain. Still another purpose is to capture an object that has caught the attention of the videographer, such as a close-up of a humming bird perching on a tree branch. The goal of their research is to automatically detect these capture types, classify them into the right category, and generate the appropriate metadata that can be used for further rendering.

Pissinou et al. in [25] presented an approach for spatial-temporal composition of video objects. A key attribute of video data is the associated spatial and temporal semantics. It is imperative that a video model models

the characteristics of objects and their relationships in time and space. The intention of their work is to design a model representation for the specification of the spatio-temporal relationships among objects in video sequences. Their model described the spatial relationships among objects for each frame in a given video scene and the temporal relationships (for this frame) of the temporal intervals measuring the duration of these spatial relationships. It also models the temporal composition of an object, which reflects the evolution of object's spatial relationships over the subsequent frames in the video scene and in the entire video sequence. Their model representation also provides an effective and expressive way for the complete and precise specification of distances among objects in digital video. This model is a basis for the annotation of raw video.

Apart from the above mentioned research works a method for video content analysis and retrieval was explained by Dimitrova et al. in [26]. Managing multimedia data requires more than collecting the data into storage archives and delivering it via networks to homes or offices. The paper also explained technologies and applications of video-content analysis and retrieval using specific examples.

III. Future Enhancement

Video data mining is an emerging field of research and the conducted research works to investigate video data management is not sufficient. As the open issue, some symbols in multi-stream are assigned by using some thresholds. This may cause miss-counting although their value are very similar. So the future research necessitates developing a mining algorithm not only using the time constraint but also using the difference of the length of pattern like Dynamic Programming [27]. Future work will consider classifying video objects using the clustered features. In the future, we will further investigate about the potentials of using mining-based scene detection method on other kinds of sports videos. More elaborate audiovisual features would be designed and extracted to enhance the scene detection performance. The future work may focus on medical video mining for efficient database indexing, management and access. The future work may address the challenge of creating descriptors that incorporate color, space, and texture simultaneously, ideally resulting in further increases in performance and more robust operation. Future works include testing on larger datasets, defining more interestingness measures, and stronger customization of itemset mining algorithms to video data. Furthermore, the future research may address the problem of joining constraint information with traditional clustering ensembles. Future work may address the challenge of extracting the low-level features such as color, motion, shape, texture based on similarities using fuzzy.

IV. Conclusion

This paper presents a survey on various data mining techniques adopted earlier in literature for scene detection and video mining. Though the data mining technique can be applied in a variety of domains, this survey primarily gave attention on the application video data, called 'video data mining'. The research conducted in the field of visual data mining is not enough. The foremost problem on video data mining is the appearance of the object in a visual media. The object can come into view in different customs in different parts of the video, which makes it adverse to explore the data mining techniques in case of video. In addition this paper also provides marginal idea for future research in this field. Future works include testing on larger datasets, defining more interestingness measures, and stronger customization of itemset mining

algorithms to video data. Also, the future work may address the challenge of creating descriptors that incorporate color, space, and texture simultaneously, ideally resulting in further increases in performance and more robust operation. The future work may focus on medical video mining for efficient database indexing, management and access.

REFERENCES

[1] K. Shirahama, Y. Matsuo and K. Uehara, *"Video data mining: Extracting cinematic rules from movie,"* In Proceedings of 4th International Workshop on Multimedia Data Mining (MDM/ KDD 2003), pp. 18–27, 2003.

[2] D. Wijesekera and D. Barbara, *"Mining cinematic knowledge: Work in progress,"* In Proceedings of the International Workshop on Multimedia Data Mining (MDM/KDD 2000), pp. 98–103, 2000.

[3] Anjulan and N. Canagarajah, *"A Novel Video Mining System,"* IEEE International Conference on Image Processing, 2007. ICIP 2007, vol. 1, pp. 185-188, 2007.

[4] P. Geetha and Vasumathi Narayanan, *"A Survey of Content-Based Video Retrieval,"* Journal of Computer Science, vol. 4, no. 6, pp. 474-486, 2008.

[5] S. V. Porter, *"Video segmentation and indexing using motion estimation,"* Ph.D. Thesis, University of Bristol, Bristol, 2003.

[6] A. Anjulan, and N. Canagarajah, *"Object based video retrieval with local region tracking,"* Elsevier, Signal Processing: Image Communication, vol. 22, pp. 607-621, 2007.

[7] Anjulan and N. Canagarajah, *"Video scene retrieval based on local region features,"* In Proceedings of ICIP, 2006.

[8] Rokia Missaoui and Roman M. Palenichka, *"Effective image and video mining: an overview of model-based approaches,"* Proceedings of the 6th international workshop on Multimedia data mining: mining integrated media and complex data, pp. 43-52, 2005.

[9] D.Maheswari, Dr.V.Radha, *"Noise Removal In Compound Image Using Median Filter"*, (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 04, 2010.

[10] V. Chasanis, A. Likas, and N. Galatsanos, *"Scene Detection in Videos Using Shot Clustering and Symbolic Sequence Segmentation"*, in IEEE Workshop on Multimedia Signal Processing, 2007.

[11] F. Chevalier, J.P. Domenger, J. Benois-Pineau, and M. Delest, *"Retrieval of objects in video by similarity based on graph matching"*, Elsevier, Pattern Recognition Letters, 2007.

[12] F. Chevalier, J.P. Domenger, and M. Delest, *"A Heuristic for the Retrieval of Objects in Low resolution Video"*, Elsevier, An international Workshop on Content Based Multimedia Indexing, 2007, conducted by IEEE.

[13] Eakins, J.P., Boardman, J.M., and Shields, K., *"Retrieval of trade mark images by shape feature – the ARTISAN project"*, Proceedings of International Conference on Electronic Library and Visual Information Research, pp. 101-109, 1994.

[14] Wang, J., Yang, W., and Acharya, R., *"Efficient access to and retrieval from a shape image database"*, IEEE Workshop on Content- Based Access to Image & Video Libraries, pp. 63-67, 1998.

[15] Christoph H. Lampert, *"Detecting Objects in Large Image Collections and Videos by Efficient Subimage Retrieval"*, International Conference on Computer Vision (ICCV), Kyoto, Japan, 2009.

[16] Hui Fang, Jianmin Jiang and Yue Feng, *"A fuzzy logic approach for detection of video shot boundaries,"* Journal of Pattern Recognition Society, Elsevier, vol. 39, pp. 2092-2100, 2006.

[17] Yuanjia Du, and Ling Shao, *"Video shots retrieval using local invariant features,"* ACM, Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics, pp. 73-78, 2009.

[18] K. Shirahama, Koichi Ideno and K. Uehara, *"Video Data Mining: Mining Semantic Patterns with temporal constraints from Movies,"* In Proceedings of 4th International Workshop on Multimedia Data Mining, 2005.

[19] M. Murty A. Jain and P. Flynn, *"Data clustering: A review,"* ACM Computing Surveys, vol. 31, no. 3, pp. 264–323, 1999.

[20] M. Shyu, Z. Xie, M. Chen, and S. Chen, *"Video semantic event/concept detection using a subspace-based multimedia data mining framework,"* in IEEE Transactions on Multimedia, vol. 10, no. 2, pp. 252-259, 2008.

[21] X. Zhu, X. Wu, A. K. Elmagarmid, Z. Feng, and L. Wu, *"Video Data Mining: Semantic Indexing and Event Detection from the Association Perspertive,"* in IEEE Transaction on Knowledge and Data Engineering, vol. 17, no. 5, pp. 665-677, 2005.

[22] Till Quack, Vittorio Ferrari and Luc Van Gool, *"Video Mining with Frequent Itemset Configurations,"* white paper, 2006.

[23] Yuchou Chang, D. J. Lee, Yi Hong, and James Archibald, *"Unsupervised Video Shot Detection Using Clustering Ensemble with a Color Global Scale-Invariant Feature Transform Descriptor,"* EURASIP Journal on Image and Video Processing, vol. 2008, 2008.

[24] Mei Chen, *"Video Mining for Creative Rendering,"* World Academy of Science, Engineering and Technology, vol. 7, pp. 86-90, 2005.

[25] N. Pissinou, I. Radev, K. Makki, and W. J. Campbell, *"Spatio-Temporal Composition of Video Objects: Representation and Querying in Video Database Systems,"* IEEE Transactions on Knowledge and Data Engineering, vol. 13 no. 6, pp. 1033-1040, November 2001.

[26] N. Dimitrova, H. Zhang, B. Shahraray, I. S. T. Huang, and A. Zakhor, *"Applications of video-content analysis and retrieval,"* in proc. of IEEE Multimedia, vol. 9, no. 3, pp. 42–55, 2002.

[27] G. Navarro, *"A guided tour to approximate string matching,"* ACM Computing Surveys, vol. 33, no. 1, no. 31–88, 2001.

Author's Biography

D.Shanmuga Priyaa was born on 1st January 1974. She received her Bachelor degree in Industrial Electronics from Bharathidasan University in 1995 and Master degree in Computer Applications from University of Madras in 1998. She completed her M.Phil from Bharathidasan University in 2004. She is working as an Asst.Prof at Karpagam University, Coimbatore. Currently she is pursuing Ph.D. Her fields of interest are Data mining, Video Procesing and Image processing.

Karthikeyan S. received the Ph.D. Degree in Computer Science and Engineering from Alagappa University, Karaikudi in 2008. He is working as a Professor and Director in School of Computer Science and Applications, Karpagam University, Coimbatore. At present he is in deputation and working as Assistant Professor in Information Technology, College of Applied Sciences, Sohar, Sulatanate of Oman. He has published more than 14 papers in Natrional/International Journals. His research interests include Cryptography and Network Security