International Journal of Electrical and Electronics Research ISSN 2348-6988 (online)
Vol. 4, Issue 2, pp: (35-40), Month: April - June 2016, Available at: www.researchpublish.com

Analysis of Various Keyframe Extraction Methods

Priyanka U. Patil[1], Dr. Krishna K. Warhade[2]

[1]M.E. student at Department of Electronics & Telecommunication, MITCOE, Pune, India
[2]Professor at Department of Electronics & Telecommunication, MITCOE, Pune, India

Abstract: **Automatic video summarization is absolutely necessary for fast browsing speed and effective management of large video libraries. Recent advancement in the technology have seen tremendous increase in the usage of multimedia information. Among all these multimedia information Video is most challenging one as it combines all other forms of multimedia information. As we know there is exponential increase of web videos and it's become time-consuming to get a query result from the huge data. So it requires efficient and effective video management system. If we search for particular video in any search engine for ex. Google server it will give lakhs of results within few seconds which are may not be of our interest. Extracting a small number of keyframes that can compact the whole content of video is very important for efficient browsing and retrieval system. The number of key frames required to abstract a shot will depend on visual content complexity within the shot.**

Keywords**: Keyframe extraction, Shot, Scene, Entropy, Color histogram, augmented 3-D keyframe extraction, Triangle model of perceived energy, RPCA-KFE.**

I. INTRODUCTION

There is massive growth in the digital video content which demands effective techniques for fast browsing and efficient management of data. Video summarization technique provides tools for selecting the most informative sequences of still or moving pictures that will help users to take quick look through the whole video clip in a constrained amount of time. Automatic summarization and indexing techniques of video will give users an opportunity to browse and select multimedia document of their choice. On the web, there are an increasing number of websites allowing users to broadcast videos themselves. As an example, YouTube is a video sharing website where users can upload, browse, and share video clips which may be of different format and different lengths. We know video structurally composed of many scenes and each scene composed of many shots as shown in Fig. 1. A scene can be defined as a subdivision of a film in which it presents continuous action in one place, or the setting is fixed.

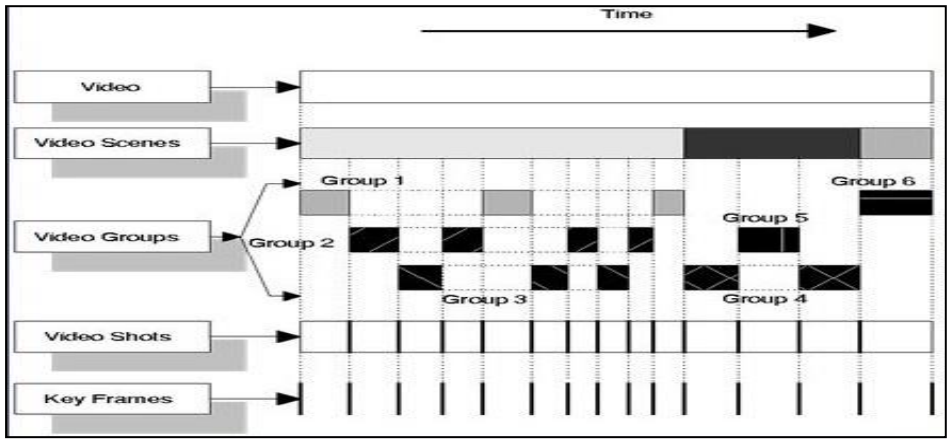

Fig 1: Structural decomposition of a Video [13].

Whereas shot is the Sequence of frames captured by single continuous operation of camera. Several shots are combined to make a scene. Keyframe is one type of video abstraction technique for visual indexing. It is defined as a subset of a video sequence that can represent the video visual content as closely as possible. The keyframe can be a good visual index because its organization allows for easy browsing and navigating.

If we search for particular video in any search engine for ex. Google server it will give lakhs of results within few seconds which may not be of our interest so this problem can be solved by using key frame extraction. Hence Key frame extraction is one of the important research issues in the field of video information retrieval.

II. RELATED WORK

Keyframes are the representative frames which are used to provide a suitable abstraction and framework that will help for indexing, browsing and retrieval of video [1]. With the help of keyframe user will be able to quickly browse over the video with the brief view of few high-lighted frames. Key-frames will reduce the amount of data required for video indexing and gives an organizational framework which deals with video content [2]. Because of importance of keyframe extraction in video retrieval many researchers are working in this area. There are many approaches which are used for keyframe extraction .One of them is shot based structuring of video, in which first shot is detected then each shot is represented by a fixed or variable number of frames.

Zhuang *et al.* proposed a clustering based approach based on an image histogram which has been widely used in the research community for keyframe extraction [3]. It was introduced to determine the keyframes within the shot boundary. In this a video segmented into shot by using a shot boundary detection algorithm. The resemblance of two frames is defined by using their visual content, where the visual content are measured in the terms of color, texture, shape of the salient object of the frame, or the combination all. Here author used color histogram of a frame as visual content, although other visual contents are readily integratable into the algorithm.

Fig 2 shows set of keyframes extracted by using the Histogram-based method which is the most commonly used method to calculate frame difference. Because color histograms do not relate pixels of a given color with spatial information it only records the amount of color information, images with similar color histograms can have different appearances.

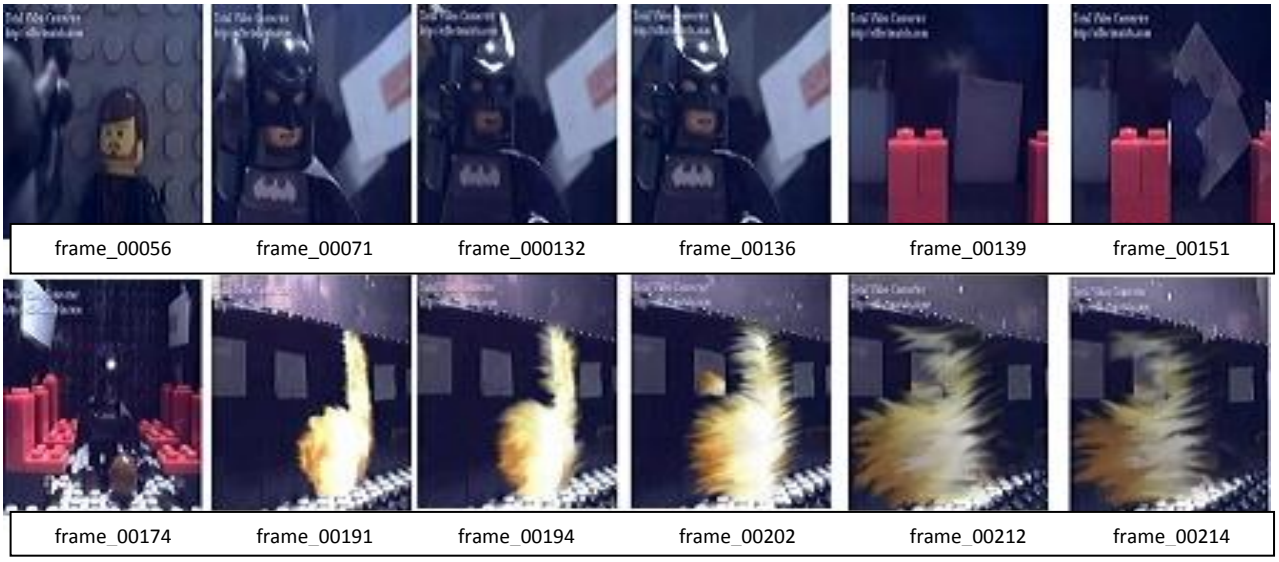

Fig 2: Set of Keyframes Extracted from a video by using image histogram method

Keyframe extraction algorithm using Entropy Difference has been proposed by Markos Mentzelopoulos et al. [4] Entropy gives the information that can be coded for by a compression algorithm. In this algorithm author used the entropy as local operator instead of global feature for the total image. Entropy is one of the good way of representing the impurity or unpredictability of a set of data because it is dependent on the context in which the measurement is taken. Here author said that first distribute the entropy among the image, then the higher distribution of entropy will give us the regions containing the salient objects of a video sequence. Therefore any change in the object appearance of salient regions will affect its relevant semantic information in the entire story sequence.

Zeeshan Rasheed and Mubarak Shah [5] have proposed the color histogram based method of UCF. This algorithm uses the color histogram for measuring the intersection similarity to extract key frames. This algorithm proposes a method to select the first key frame and then based on first frame and next frame difference, keyframe will be decided. There will be comparison with the next frames and new frame will be selected based on difference between frames if there is significant difference, the current frame will be the key frames.

TABLE I: Review comparison of different keyframe extraction methods

ALOGORITHMS	FEATURES	MERITS	DEMERITS
Clustering based approach. [3]	Determine the keyframes within the shot boundary.	Efficient, Effective, On-line processing.	For low-activity shots, it extract less key frames or one single key frame while for high-activity shots, it extract multiple key frames depends on the visual complexity of the shot.
Keyframe Extraction Using Entropy difference. [4]	Entropy gives the impurity or unpredictability of a set of data.	Entropy used as local operator instead of global feature for the total image.	If there is transient change such as flashes the performance Drops to the average algorithms performances.
Color histogram based method of UCF. [5]	Uses the color histogram intersection similarity measure to extract key frames.	Perform a high-level Segmentation of videos into scenes.	The given method is superior to graph-based approaches where all shots are considered to captures the global similarities.
Motion analysis based approach. [6]	Optical flows of frames are calculated.	The key frames are identified by stillness.	Minimum two frames are selected.
Triangle model of perceived motion energy (PME). [7]	It models motion patterns in video.	Combines motion-based temporal segmentation and color-based shot detection. The turning point of motion acceleration and deceleration of each motion pattern is selected as a key frame.	This algorithm doesn't include the integration of color-change analysis and audio cues.
Augmented 3-D keyframe. [8]	The cameras are fixed. This method processes simply a video clip.	This method is used to convert surveillance video clip to a meaningful and compact keyframe.	It is computationally expensive method.
Optimal key frame representation scheme. [9]	Construct key frame which is temporally maximum occurrence frame.	Each pixel is constructed by considering the probability of occurrence of those pixels at the corresponding pixel position among the frames in a video shot.	Computationally complex.
Efficient Video Indexing Scheme for Content-Based Retrieval. [10]	Used to find a compact set of key frames that can represent a video segment for a given degree of accuracy.	Gives multilevel abstract of the video.	Due to de correlated frame samples, they lose the continuous motion information of the Original video.
Motion based key frame extraction. [11]	This method dedicated to summarizing consumer video clips acquired from digital cameras.	Based on analysis of spatio-temporal changes in the scene.	Requires improvement in the accuracy by incorporating high-valued, computable semantic information and image quality metrics.
Robust Principal Component analysis Key Frame Extraction Method. [12]	Input data is decomposed into low rank component and sparse component.	Applicable to the consumer videos which are unstructured.	This algorithm works on the assumption that low rank component contains systematic information and sparse component contains the distinct information.

Wolf has proposed a motion based approach to key frame extraction [6]. He first computes the optical flow for each frame [4], and then simple motion metric is calculated based on the optical flow. This method selects the keyframe at the local minima of motion as a function of time. The justification of this approach shows that in many shots, the key frames are recognized by stillness – means either when the camera stops on a new position or the hold gestures to emphasize their importance [6].

" A Triangle Model of Perceived Motion Energy" (PME) method for extracting keyframes has been proposed by Tianming Liu et al. [7] and is used to model motion patterns in video to extract key frames. The frames are selected as key frames based on the motion acceleration and motion deceleration turning points. This algorithm gives key-frame selection process which is threshold free and fast. In this paper author used motion patterns for measuring the visual content complexity of a scene. Motion pattern is usually consists of a motion acceleration process, followed by deceleration process. It shows an action in events. For extracting key frames based on motion patterns, it is required to build a motion model which reflects the motion activities in video shots will provide the guidance for the selection of key frames. This is objective of triangle model of perceived motion energy (PME).

"Augmented 3-D keyframe" algorithm has been proposed by Gwo-Cheng Chao et al. [8] It gives more meaningful and compact keyframes with representative objects, important contents, motion information, and some marks of moving objects captured by a static camera extracted from a surveillance video clip. This method is useful as it also be integrated with a user interface for an interactive video retrieval and browsing system. There are two important factors of this method .Here the cameras are mainly fixed in the surveillance systems. The second factor is that this method processes simply a video clip. There are many ways for the video segmentation, such as shot boundary detection and motion detection. Any of video segmentation method can be used according to need of the application.

Kin-Wai Sze et al. proposed optimal key frame representation scheme based on global statistics for video shot retrieval [9]. Each pixel is constructed by considering the probability of occurrence of those pixels at the corresponding pixel position among the frames in a video shot. Therefore, this constructed key frame is temporally maximum occurrence frame, which is an optimal representation of all the frames in a video shot.

Hyun Sung Chang et al considered key frame extraction problem is from a set-theoretic point of view, and gives systematic algorithms which are used to find a compact set of key frames that can represent a video segment for a given degree of accuracy [10]. The proposed extraction algorithms can be hierarchically applied to obtain a tree-structured key frame hierarchy which is a multilevel abstract of the video. The key frame hierarchy enables an efficient content-based retrieval.

Motion based key frame extraction (MKFE) method has been proposed by Jiebo Luo [11]. This method was based on analysis of spatio-temporal changes that will occur over time while capturing meaningful information from the scene and camera motions. In particular, a video clip is first segmented into comparable parts based on camera motion types, e.g. pan, zoom, pause, and steady. A set of applicant key frames has been extracted from the each segment based on the motion. Finally, from the initial candidate key frames final keyframes are extracted based on conviction measure of each frame.

A new approach for the keyframe extraction has been proposed by the Chinh Dang and Hayder Radha [12] which is based on the Robust Principal Component Analysis (i.e. RPCA-KFE).This method gives good result even for the consumer videos which are unstructured. This algorithm works on the Robust Principal Component Analysis in which input data is decomposed into the low rank components which give the systematic information of the input and a set of sparse components which provides distinct information of input data. This algorithm is efficient as compared to previous methods as stated by the author. The table 1 provided below gives brief review of different keyframe extraction methods and it's comparison with respect to features, merits and demerits are described.

III. EVALUTION CRITERIA FOR THE DIFFERENT KEY FRAME EXTRACTION TECHNIQUES

3.1 Shot boundary based approach:

In the shot boundary, method first video is segmented into shots. First frame of each shot is selected as a keyframe for that particular shot. Regardless of the visual complexity of the shot of keyframes for each shot is limited to one.

3.2 Visual content based approach:

Shot based criteria:

The first frame of the shot will always be selected as the first key frame; more than one keyframes can be selected depends on other criteria

Color feature based criteria:

In this present frame of the shot will be compared with the last keyframe. If significant changes occur then that frame will be the new keyframe of the shot.

3.3 Motion based criteria:

When there is a zooming-like shot, then minimum two frames will be selected: the first and last frame, among one will represent a global, while the other will represent a more focused view.

IV. APPLICATIONS

Distance learning:

The goal of the distance learning is to provide quality of learning that is effective and comparable with the traditional classroom environment. With the help of keyframe extraction it is possible for the student to querying the topics of interest directly rather than watching whole video.

Telemedicine:

Telemedicine is used to provide the good health care services where direct services are difficult to provide due to some geographical environment. It is combination of audio, video, electronic information which provides diagnosis, consultations, and procedures to the patients at remote site here also keyframe will provides a major help.

Interactive televisions:

Keyframe extraction is useful for enhancing the functionality of the interactive televisions. It can be done by using shot boundary detection in which video is broken up into shots. By comparing with adjacent frames shot boundary changes can be identified

Multimedia news:

As we know there is very much impact of the multimedia news on the social life. So it is important to extract and analyze the important information required for user. There are other applications of the keyframes such as digital libraries, video retrieval etc.

V. CONCLUSION

In the world of multimedia information video is most important one as it combines other multimedia information such as audio etc within it so it requires efficient video management technique. The increased demand of multimedia services requires the development of techniques to store, navigate and retrieve visual data. The key frame is a simple but effective form of summarizing a long video sequence. Keyframe will provide more compact and meaningful video summary. So with the help of keyframes it will become easy to browse video and it will give the appropriate search result. As we know there is an exponential increase of web videos and it's become time-consuming to get a query result from the huge data. Hence there is requirement of the efficient keyframe extraction methods for effective video summarization. This paper gives the brief review on the different keyframe extraction methods their features, merits and demerits. In future we wanted to develop robust algorithm for key frame extraction.

REFERENCES

[1] H.J.Zhang, J. Y. A. Wang, and Y. Altunbasak, "Content-Based Video Retrieval and Compression:A Unified Solution," in Proc. IEEE Int. Conf. on Image Proc., 1997.

[2] Tianming Liu, Hong-Jiang Zhang, and Feihu Qi "A Novel Video Key-Frame-Extraction Algorithm Based on Perceived Motion Energy Model" IEEE Transactions on Circuits and Systems for Video Technology, vol.13, no.10, October 2003.

[3] Yueting Zhuangt, Yong Rui, Thomas S. Huang and Sharad Mehrotra "Adaptive Key Frame Extraction Using Unsupervised Clustering", IEEE Int. conf. on Image Proc., 1998.

[4] Markos Mentzelopoulos and Alexandra Psarrou "KeyFrame Extraction Algorithm using Entropy Difference," Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, MIR 2004, October 15-16, 2004, New York, NY, USA

[5] Zeeshan Rasheed and Mubarak Shah, "Detection and Representation of Scenes Videos" IEEE Transactions on Multimedia, vol.7, no.6, December 2005.

[6] W. Wolf, "Key Frame Selection by Motion Analysis," in Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc., 1996.

[7] Tianming Liu, Hong-Jiang Zhang, and Feihu Qi "A Novel Video Key-Frame-Extraction Algorithm Based on Perceived Motion Energy Model" IEEE Transactions on Circuits and Systems for Video Technology, vol.13, no.10, October 2003.

[8] Gwo-Cheng Chao, Yu-Pao Tsai, and Shyh-Kang Jeng "Augmented 3-D Keyframe Extraction for Surveillance Videos", IEEE Transactions on Circuits and Systems for Video Technology, vo l. 20, no .11, November 2010.

[9] Kin-Wai Sze, Kin-Man Lam, and Guoping Qiu "A New Key Frame Representation for Video Segment Retrieval," IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 9, September 2005

[10] Hyun Sung Chang, Member, Sanghoon Sull, and Sang Uk Lee "Efficient Video Indexing Scheme for Content-Based Retrieval," IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no.8, December 1999.

[11] Jiebo Luo, Christophe Papin, and Kathleen Costello "Towards Extracting Semantically Meaningful Key Frames From Personal Video Clips: From Humans to Computers" IEEE Transactions on Circuits and Systems for Video Technology vol. 19, no. 2, February 2009.

[12] Chinh Dang and Hayder Radha "RPCA-KFE: Key Frame Extraction for Video Using Robust Principal Component Analysis," IEEE Transactions on Image Processing, vol. 24, no. 11, November 2015.

[13] http://image.slidesharecdn.com/vorlesungmedienwissenschaften-unipotsdam-20130121-130121043740-phpapp01/95/from-structural-media-analysis-to-exploratory-search-21-638.jpg?cb=1358743491