International Journal of Pure and Applied Mathematics
Volume 119 No. 10 2108, 485-490
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

KEY FRAME EXTRACTION METHODS

[1] ISRAA HADI ALI, [2]TALIB T. AL - FATLAWI
[1]IT College. Babylon University, Iraq
[2] Ph.D. student - IT College. Babylon University, Iraq
E-mail: [1]israa_hadi1968@yahoo.com, [2] talib.turkey@qu.edu.iq

ABSTRACT

Video structure analysis plays important role in too many fields, such as video summarization, video browsing, compression, analyzing and so on. Existing of tools that capable of analyzing videos' structure is very necessary. Generally, Video is a massive volume object with a high redundancy and insensitive information, it has complex structure consists of scene, shot, frame. One of the fundamental units in the structure analysis is the Key-frame extraction, it provide us with a good video summarization and browsing a large video collections. Key-frame is a frame or set of frames that having a good representation of the entire content of small video clip. It must contain most of the salient features of the represented video clip. In this paper, we introduce the most popular techniques for the key – frames extraction process, their advantages and disadvantages, and further information about key-frame extraction process.

Keywords :- Key-frame (K.F), Shot Boundary Detection(SBD), Content Based Video Indexing and Retrieval (CBVIR).

1. "INTRODUCTION"

With the rapid development of information technologies, the revolution in the digital era, recent advancement in video capture, the development in internet techniques and the appearance of the new application such as video conferencing, video on demand, distance learning, surveillance, forensics, interactive television,…,etc. [1,2,3]. All these factors cause a dramatically increasing in creation and storage of video data and lead to the emergence of a massive data base of video's information. Existing tools that capable of analyzing, indexing, browsing, summarization, management these videos are becoming persistent need.

Video can be defined as a huge volume data object; it contains a high redundancies and intensive information. Video has a complex structure consists of a scene, shot and frames [4], figure 1 shows video's structure. Video structure analysis is important in too many applications. It is the process of splitting the video into its major components, it include scene segmentation, SBD and K.F extraction. It provides users with a comprehensive representation to the videos. K.F extraction is a major topic in the video's structure analysis, it is inspired from the nature of video, where the redundancy is predominant characteristic of a video, the redundant frames can be eliminated to make video more compact and worthy. K.F extraction is the process of extracting frame or set of frames that have a good representation of a shot. It must preserve the salient feature of the shot, while removes most of the repeated frames [3, 5]. The K.Fs considered the major steps in CBVIR, video summarization, since it facilitate the searching of video. If we could to get a compact representation to the video, this facilitate the process of indexing and retrieval that depending on contents.

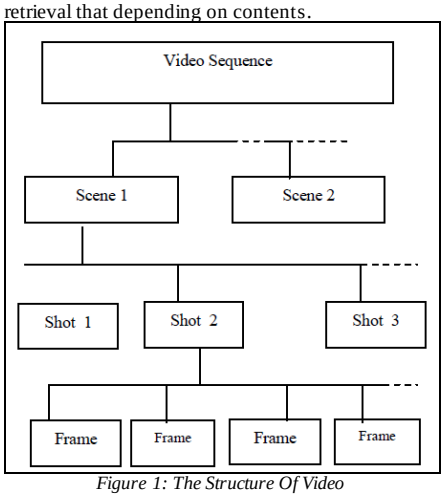

Figure 1: The Structure Of Video

Traditional K.Fs extraction techniques aimed to remove the similar frames while keeping the diversity of the video's visual content [5]. These techniques either consider the video is already segmented into shots, or divided the video frame sequence into set of shot by shot boundary detection. Where the shot can be defined as "a consecutive sequence of frames captured by a camera action that takes place between start and stop operations"[4], where there is a high correlation between frames inside the same shot.

This paper produced a study of most techniques that used for K.F extraction and the major problem that face the researchers in K.F extraction process. This paper structured as follow: - section 2, size of K.Fs set problem of K.F extraction, section 3, the major techniques, and section 4, the conclusion.

.

2. "THE SIZE OF KEY FRAMESS SET"

One of the problems we faced in K.Fs extractions process is determining the size of K.F set. Some methods proposed that for each shot there is only one K.F as in [3], they proposed a K.F is the frame with maximum entropy in each shot. This proposition is not appropriate for big shot, since there are changes in the scene, and some objects may appear while others one may discarded or occluded. Therefore, we must extract a set of K.Fs for each shot to get a good and sufficient representation to the shot's frames. Some others proposed that the K.Fs is the first, middle, or ending frames in the shot, also these methods is simple and low complexity, but the resulted K.Fs may be have low correlation in visual content.

According to [5, 6], there are three ways to determine the size of K.Fs set, these techniques are:

- **"Priori known as a fixed number"**: - In this types, the number of K.Fs is determined a priori as a constant value before the starting the extraction process. It is also known rate constant K.F extraction). If we assume *k* is the number of key frames (the size of K.F set), then the problem of K.F extraction can be rewritten as an optimization problem of finding the key frame set R, where

  $$R=\{F_{i1},F_{i2},\ldots,F_{iK}\}$$

  that is differs from the sequence of the video according to a certain summarization perspective:

  $$\{r_1,r_2,\ldots,r_k\}=\arg\min_{r_i}\{\mathcal{D}(\mathcal{R},\mathrm{V},\rho)\mid 1\le r_i\le n\} \quad (1)$$

  Where n is the video's frames number, ρ represents the summarization perspective, and D represents the distance (dissimilarity) measure. Most of the recently methods of the K.F extraction used ρ as "visual coverage", where it aims to comprise as much as possible of the visual content with a few numbers of key frames. ρ may be represent the number of objects or faces, and so forth.
- **"A posteriori (left unknown)"**:- In this type, the K.Fs' number remains unknown until the process of extraction finishes. Level of visual change determines the size of K.Fs set. If the shot contains a lot of actions and movements, then the required K.Fs to represent that shot is more than static one. Therefore, for the scene with dynamic contents, large number of K.Fs is produced, and this caused inconvenience during interactive tasks. the formulation of this type of K.F extraction problem can be represented as follows:-

  $$\{r_1,r_2,\ldots,r_k\}=\arg\min_{r_i}\{k\mid\mathcal{D}(\mathcal{R},\mathrm{V},\rho)<\varepsilon,1\le r_i\le n\} \quad (2)$$

  Where ε is the dissimilarity tolerance (fidelity level), n, ρ and d have the same meaning as the first type.
- **"Determined"**:- This is in essence of a posteriori, in this category of K.F extraction methods, an appropriate size of K.Fs is determine before the whole extraction process is executed. For example, the methods that depending on clustering techniques.

Also, there exist methods that employed two types of techniques for determining K.F set's size [7, 8]. For example, the procedure will stop when a particular condition is met or K.Fs' number reaches a prior value.

3. "K.F EXTRACTION METHODS"

Several methods and several features are employed to extract the K.Fs from a video sequences. According to [4,5] the methods can be divided in several categories, in this section we explain the categories that used to create or extract K.Fs:-

- **"Sequential Comparison Between Frames":** - In these methods, each frames are compared to the K.F that has been extracted previously, when the differences between a frame in the video sequence and the extracted K.F is high, then this frame is indicated as a new K.F. In [9] a new key frame is extracted by computing the differences between the color histogram of the current frame and the previous key frame. The advantages of these methods are: - these methods are simple in comparison with other

methods, low computational complexity is another merit, and shot's size and shot's content diversity determine the size of K.Fs set. The drawbacks of these methods are the K.Fs set may contain a redundancy, this can be occur when the contents appearing repeatedly in the same shot, also the local properties of the shot presented in the K.F rather than the global properties.

- **"Global Comparison Between Frames"**: - These methods based on minimizing a predefined objective function by using the global difference between frames in a shot. The objective function is selected based on the application. According to [6] the objective function can be one of the following:-

  1) "Minimum correlation": - These methods extract the K.Fs to minimize the sum of correlations between K.Fs; it tries to make the K.Fs uncorrelated with each other as much as possible. In [10] utilized a directed weighted graph to represent the shot's frames and their correlations, they employed A* algorithm to find the shortest path in the graph, then the vertices in that path which represent minimum correlation between frames are selected as K.Fs.
  2) "Minimum reconstruction error" :- in this category, extraction process is done by minimizing the sum of differences between the predicated frame and each frame in the shot, the predicted frame is created by interpolation process of a set of K.Fs. In [11] used an iterative procedure to select a predetermined number of K.Fs, and minimize the short reconstruction error. Also Liu et al. [8] suggest a K.F selection algorithm based on the extent to which K.Fs record the motion in the shot, they also used an interpolation algorithm to create interpolate frames.
  3) "Even temporal variance": In this type of methods, in each shot the frame with equal temporal variance is selected as a K.F for that shot. The objective function can be selected to be the sum of differences between temporal variance of all segments. The temporal variance in a segment can be computed by the cumulative change of the contents over the segments.
  4) "Maximum coverage":- In this category of methods try to maximize the representation coverage of K.Fs. The representation coverage can be defined as the number of video's frame that a specific key frame can represent. These algorithm either used a fidelity criterion if the size of key frame set is not fixed or maximize the key frame's representation if size of key frame set is fixed.

  The advantages of the techniques that based on global comparison are: - the global characteristics are reflected into the extracted K.F, number of the extracted K.Fs is controllable and, in comparison with sequential methods the K.Fs set consider more compact. The drawback of this type is more computational complexity in comparison with sequential methods.

- **"Reference Frame":** - In these methods a reference frame is generated, for K.F extraction process each frame in the shot is compared to that reference frame. The advantage of these methods is that they are easy to implementation and grasp. Their drawbacks are some salient contents and features in the shot may be missed when the reference frame does not represent the shot adequately.

- **"Curve Simplification"**: - In this category of methods, the shot's frames are represented as a points in the feature space. Then the points is linked in a sequential manner to formulate a trajectory curve and try to find the points that give a best representation to the shape of the curve. Calic and Izquierdo [12] Present a real – time method for detection the change that occurs in the scene and extract the K.F, the macro-block features' statistics are analyzed, they dealing with compressed video stream (especially MPEG) to create frame difference metrics. Then a novel discrete contour evolution techniques employed for curve simplification using the frame difference metrics.

- **"Clustering"**:- Methods in this category consider each frame in the video sequences as data points in the feature space, cluster frames, then the frames that have smallest distances (closest) to cluster centers are selected to be K.Fs. Yu et al. [13] present a method for K.F extraction process based on fuzzy K-means. In[14] proposed a method for K.F extraction based on clustering, they first cluster the

motion sequences into two classes based on similarity distances, then used ISODATA algorithm to cluster all frames, and those closest the clusters' centers are selected as K.F. Also Pan et al in [15] present an import shot K.F extraction methods using improved fuzzy C –Means clustering, they employed color feature information, they clustered shot into sub shots, the frame that has a largest entropy is selected as a K.F from each class. In[16] presented a method for extracting K.Fs and isolating foreground, they employed a K-Means algorithm along with mean squared error. The advantages of methods under this category are the K.Fs have the global characteristics of the video, and a generic clustering algorithm can be used. While their drawbacks are they require a high computation cost, the K.Fs set lack to the temporal information of the original video.

- **"Object/Events"**: - Methods that fall into this category consolidate the K.F extraction process with the detection of object/ event to ensure that the information of objects/ events is involved in the extracted K.Fs. The advantage of these methods is that the extracted K.Fs are reflecting the object or the patterns of object motion. Their disadvantage, the heuristic rules that appoint according to the application that detect object/event plays an important role. Therefore, the efficiency of these algorithms depending on experimental settings, where these settings must be chosen carefully to get a good results. .

- **Panoramic Frame**:-In order to get a K.F with a good representative for the features and the content of the shot, and avoid the noise in redundant frame, panoramic frame is the best choice, since all frames or specific frames in the shot are selected to create a new K.F that has a full view of the shot. In [17] adopted a method for video registration based on the computation of homography matrix between frames. The advantage of these methods is that they provide a K.F that has a wide presentation to the shot. The disadvantage is that high computational complexity.

4. CONCLUSION

K.Fs extraction process considers a basic unit in the video's structural analysis; it is remove most of the redundant frames in the video. It has a great significant in too many fields as in video summarization; content based video indexing and retrieval, video searching, video compression and so on. It provides a user with a good representation to the entire shots. This paper presents an extensive survey of the techniques that used to find the K.F set, their advantages and disadvantages, the problem that a user face when trying to extract the K.F. Although there is no uniform evaluation metrics to evaluated the methods of K.Fs extraction, these methods should be high efficiency, robustness, low computational complexity and the extracted K.Fs must be representative to the entire video's sequence frame and must be minimal as much as possible.

REFERENCES:

[1] Y. N. Li, Z. M. Lu, and X. M. Niu, "Fast video shot boundary detection framework employing pre-processing techniques," IET Image Process., Vol. 3, No. 3, pp. 121–134, 2009.

[2] Z. Lu and Y. Shi, "Fast video shot boundary detection based on SVD and pattern matching," IEEE Trans. Image Processing, Vol. 22, No. 12, 2013, pp. 5136-5145.

[3] T.E. Bavisha, 2Ms.M.MadlinAsha, "A Keyword Based User Privacy-Preservation And Copy-Deterrence Scheme For Image Retrieval In Cloud", International Journal of Innovations in Scientific and Engineering Research (IJISER), Vol.4, No.1, pp.30-35, 2017.

[4] Rachida Hannane, Abdessamad Elboushaki, Karim Afdel1, P. Naghabhushan and Mohammed Javed, "An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram", International Journal of Multimedia Information Retrieval, Vol. 5, No. 2, 2016,pp. 89-104.

[5] W. Hu, N. Xie, L. Li, X. Zeng, and S. Maybank, "A survey on visual content-based video indexing and retrieval", IEEE Transactions on Systems, Man, and Cybernetics—part c: Applications and Reviews, Vol. 41, No. 6, 2011, pp. 797 819.

[6] Guangyu Gao, Chi Harold Liu. "Video Cataloguing Structure Parsing and Content Extraction". Taylor & Francis Group, LLC, 2016.

[7] Ba Tu Truong and Svetha Venkatesh, "Video abstraction: A systematic review and classification", ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 3, No. 1, Art. 3, pp. 1–37, 2007.

[8] A. Divakaran, R. Radhakrishnan, and K. A. Peker. "Motion activity-based extraction of key-frames from video shots", In Proceedings of the IEEE International Conference on Image Processing, pp. 932–935, 2002.

[9] T. Liu, X. Zhang, J. Feng, and K. Lo. "Shot reconstruction degree: A novel criterion for keyframe selection", Pattern Recognition Letter, 25(12):1451–1457, 2004.

[10] H. J. Zhang, J. Wu, D. Zhong, and S. W. Smoliar, "An integrated system for content-based video retrieval and browsing," Pattern Recognit., Vol. 30, No. 4, 1997, pp. 643–658.

[11] S. V. Porter, M. Mirmehdi, and B. T. Thomas, "A shortest path representation for video summarization", 12th International Conference on Image Analysis and Processing, 2003, pp. 460–465, Italy.

[12] H.-C. Lee and S.-D. Kim, "Iterative key frame selection in the rate-constraint environment," Signal Processing: Image Communication, Vol. 18, No. 1, 2003, pp. 1–15.

[13] J. Calic and E. Izquierdo, "Efficient key-frame extraction and video analysis" in *Proc. Int. Conf. Inf. Technol.: Coding Comput.*, Apr. 2002, pp. 28–33.

[14] X. D. Yu, L. Wang, Q. Tian, and P. Xue, "Multilevel video representation with application to keyframe extraction," in Proc. Int. Multimedia Modelling Conf., 2004, pp. 117–123. Australia.

[15] Qiang Zhang, Shao-Pei Yu, Dong-Sheng Zhou, and Xiao-Peng Wei" An Efficient Method of Key-Frame Extraction Based on a Cluster Algorithm", Journal of Human Kinetics Vol. 39, No. 1, 2013, pp. 5-13.

[16] Rong Pan, Yumin Tian, Zhong Wang, "Key-frame Extraction Based on Clustering", IEEE Progress in Informatics and Computing, 2010, pp. 867-871, China.

[17] Azra Nasreen, Kaushik Roy, Kunal Roy and G. Shobha "Key Frame Extraction and Foreground Modelling Using K-Means Clustering", 7th International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN), 2015, pp. 141- 145, Latvia.

[18] B. Ghanem, T. Z. Zhang, and N. Ahuja, "Robust video registration applied to field-sports video analysis" In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1–4, 2012.