International Journal of Engineering & Technology, 7 (4.19) (2018) 889-892

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Proposed Method for Key Frame Extraction

Israa Hadi Ali[1], Talib T. Al – Fatlawi*[2]

[1,2]IT College. University of Babylon, Iraq
[2]College of Computer Science and Information Technology, University of Al-Qadisiyah, Iraq
**Corresponding Author E-mail: talib.turkey@qu.edu.iq*

Abstract

Video structure analysis can be considered as a major step in too many applications, such as video summarization, video browsing, content-based video indexing,and retrieval and so on. Video structure analysis aims to split the video into its major components(scenes, shots, keyframes). A key frame is one of the fundamental components of video; it can be defined as a frame or set of frames that give a good representation and summarization of whole contents of a shot. It must contain most of the features of the shot that it represented. In this paper, we proposed an easy method for key frame extraction from the video's shot. In the first step of the proposed system, the frames are divided (hashed) into groups (buckets) based on cosine distance, in this step the frame is converted to HSV color space, and angle between frame is computed, the frames that have similar angle are going the same bucket. In the second step, from each groupkeyframe is selected, the results we get can be considered good and reasonable.

Keywords: *Key Frame (KF), Shot Boundary Detection(SBD), Content-Based Video Indexing and Retrieval (CBVIR).*

1. Introduction

As a result of the industrial and digital revolution, which led to the development of means of digital imaging, as well as the advances in internet, the appearance of social media and sites that deal with digital videos, such as YouTube dailymotion and so on, also the appearance of software that have a high capacities to deal with videos. These factors make the video capturing, sharing, creating a very easy process. As a result,massive video databaseshavebeen created. Therefore the presence of tools thatfacilitate the dealing with such massive video's database become urgent need[1,2,3,4].
The video can be defined as "a huge volume data object; it contains high redundancy and intensive information"[4], it has a complex structure that consists of scenes, shots,andframes[5]. Figure 1 shows the structureof the video. Analyzethe structure of videos can play a major role in many fields, as in CBVIR, video compression, video summarization, video management…etc.Itcomprises scene segmentation, SBD and Key Frame extraction. It gives a user a good overview of the videos.

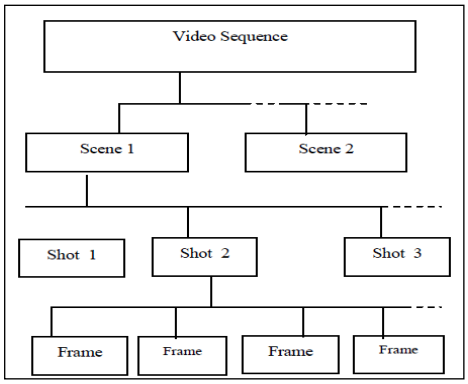

Fig. 1: The Structure of Video

Keyframe extraction is an important step in the video's structure analysis; it inspired by the nature of the video, where the frame's redundancy is the dominant property. Keyframe extraction process aims to extract frame or set of keyframesthat have a good representation of the shot and remove most of the redundant frames to get a more compact representation of a small video clip; it must keep most the feature that provides us with a good representation of the shot[4,6]. Traditional techniques that used for KF extraction trying to remove most of the frame that has similar contents in the shot and preserve those they have diversity contents. Where the shot can be defined as "a consecutive sequence of frames captured by a camera action that takes place between the start and stop operations"[5].
When we try to extract $KF_{(s)}$from the shots, we must determine the size of KF's set that belong to each shot.Some researchers extract one KF for each shot, while the other represents each shot with a set of KF[4]. According to[6,7] the size of KF's set can be one of three choices:-

- **"Priori known as a fixed number"**: - in this category, the number of KFs is determined before the extraction process started.
- **"A posteriori (left unknown)"**:- In this category, KF's number remains unknown until the process of extraction finishes.
- **"Determined"**:- in this category of KF extraction methods, an appropriate size of KFsis determined before the whole extraction process is executed.

In this paper we proposed straightforward method for key frame extraction, our method employed cosine distance to divide (hash) the shot's frames into groups(buckets), later from each group we extract a keyframe, later the extracted frame must besatisfied a threshold to decide if it accepted as a keyframe or discarded. The later of this paper is organizedas follows:- section2 the related work, section3 the proposed system, section4 the results and finally asection5 conclusion.

2. Related work

Several techniques and several features are employed to extract the keyframe from the videos' clip. For example,a sequential comparison between frames is used as in[8]the color histogram of the current frame are compared the histogram of the previously extracted keyframe. In the methods that are similar to this method, the diversity of the shot's content determines the size of keyframe's set. Entropy can also be employed to extract the KF as in [3], the frame that has a high entropy is selected to be the KF of the shot.

Other techniques are based on a global comparison between frames to extract the KF. These methods try to minimize predefined objective function. The objective function may be:- minimum correlation as in[9], the minimum reconstruction error as in[10,11] even temporal variance, or Maximum coverage[4]. These methods suffer from high computational complexity[5].

Some others methods generate a reference frame, then each in the shot is compared to that frame to extract the new KF as in [12], the drawbacks ofthe methods under this category are some salient contents and features in the shot may be missed when the reference frame does not represent the shot adequately.

Other researches proposed each frame in the shot as a point in the feature space; then these points are linked sequentiallyto formulate a trajectory curve and try to find the points that give the best representation to the shape of the curve as in [13].

Clustering techniques also used for the KF extraction purpose. In these methods each frame is considered as data points in the feature space, then these data points are clustered using one of the clustering methods, and the frame that hasa small distance to the cluster center is selected as a keyframe. In[14] presented a new method for KF extraction using 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 KF. Also, Pan et al. in [15] present an import shot KF extraction methods using improved fuzzy C –Means clustering, they employed color feature information, they clustered shot into sub-shots, the frame that has the largest entropy is selected as a KF from each class. In[16] presented a method for extracting KFs and isolating foreground, they employed a K-Means algorithm along with the mean squared error. The advantages of methods under this category are the KFs 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 KFs set lack to the temporal information of the original video.

Panoramic Frame is another choice, it provides a good and broad representation of the shot's frame and avoids noise and contents redundancy.

In[17] present a method for constructing a panoramic key frame using homography matrix between frames. The major disadvantage of these methods is high computation complexity[4].

3. The Proposed Method

As we mentioned previously, our methods extract set of KFsfor each shot based on cosine distance;later from each group, we extract a keyframe that hasgood representation to that group,Figure2 shows the structure of the proposed system. As we know that there is a high correlation between the frames in the same shot, since the adjacent frames usually contain the same foreground and background, thereforethey have a very similar histogram. Also,the changes in contents of frames are smooth;if there arechanges in the scene, this leads to variations in the histogram. We employed this fact in our work.

Themajor steps of the proposed system consist of the following steps:-

1. Resize the frames to 96*96 pixels to speed up the computation. Then convert each frame to HSV color space, and compute the histogram to the value V.after that we normalize the resulted vector by dividing each bin on the total number of pixels in the image.

2. Compute the cosine distance (and angle) between the first frame and all frames in the shot. Then the frames that have the same angle are going to the same group. The cosine distance between two vectors x(first frame) and y (current frame) is computed as following:-

$$\cos(\theta) = \frac{\sum_{i=0}^{d} x_i y_i}{\sqrt{\sum_{i=1}^{d} x_i^2}\sqrt{\sum_{i=1}^{d} y_i^2}} \quad (1)$$

Where d represents the length of vector.

Here we use the angle as a hash function that divides the frames into bucketsaccording to the angle between the first frame and frame under consideration. Each bucket with a small number of frames (usually less than 10) is merged with other buckets since a small number of the video sequence cannot have a great change in details.

3. From each group (bucket) we compute the average histogram (for V value). The resulted vector can be considered as a cluster center.From each group, we find the frame that has a small distance to the average histogram.

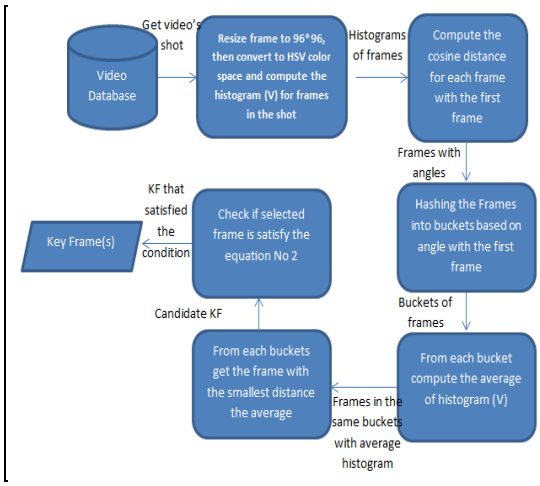

Fig. 2: Block Diagram of the proposed system

4. Each keyframe extracted must satisfy the following condition:-

$$\theta(K_i, K_{i-1}) > 1 \quad (2)$$

Where:-

θ represent the angle between frames.

K_i is the current key frame,

K_{i-1} is the previous key frame.

This means that both keyframes fall in different angles. If this equation is not satisfied then the currentkey frame is discarded.

4. Experimental Results

To evaluate the performance of the proposed system, we have used different video for the testing, the experiment is performed with an MPEGand MP4 format, with frame rate29 and 30 frames per second, all of these videos of different resolution, the performance shows that the results a high performance. The figures below show the extracted keyframe from different videos. In figure(3) we show the video shot that started with a frame that has a number 15242 and ending with 15428, after we apply the proposed system on this shot we get two key frames (15350 and 15411), as we see these key frames provide us with a broad

representation of the entire shot. In figure 4the shot is started with frame0 and ending with frame 70, since there is no significant change in this shot, the system produces a single key frame (frame at position 70). The same thing in figure 5, the systemproducesa single key frame for the entire shot.

5. Conclusion

KFs extraction process considers a basic unit in the video's structural analysis; it provides the user with a good representation of the whole shot, while removes most of the redundant frames. It playsanessential role in too many applications such as video summarization, CBVIR, video compression, video management, and so on. In this paper we adopt a new method for key frame extraction, we employ the cosine distance to divide the shot's frame into buckets. Later, from each bucket we compute the average of the histogram, consider this frame as a center of the bucket.Then from each bucket, we find the frame that has the smallest distance to the center. This method considers an easyand provide excellent results as we show in experimental results. Also the number of key frames is determined automatically withoutthe intervention of users, where the diversity of shot content determines the number of keyframes.

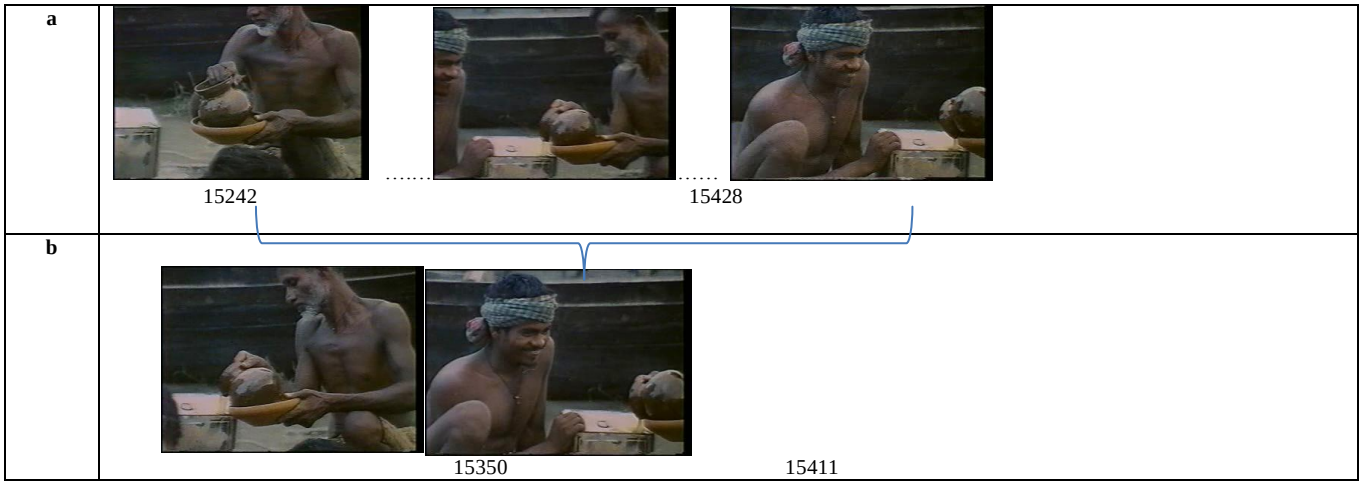

Fig. 3: Result of the proposed system. a) shows the original frames of the shot starting from frame 15242 and ending at 15428. b) shows the resulting keyframes, the keyframes give a good overview of the entire shot

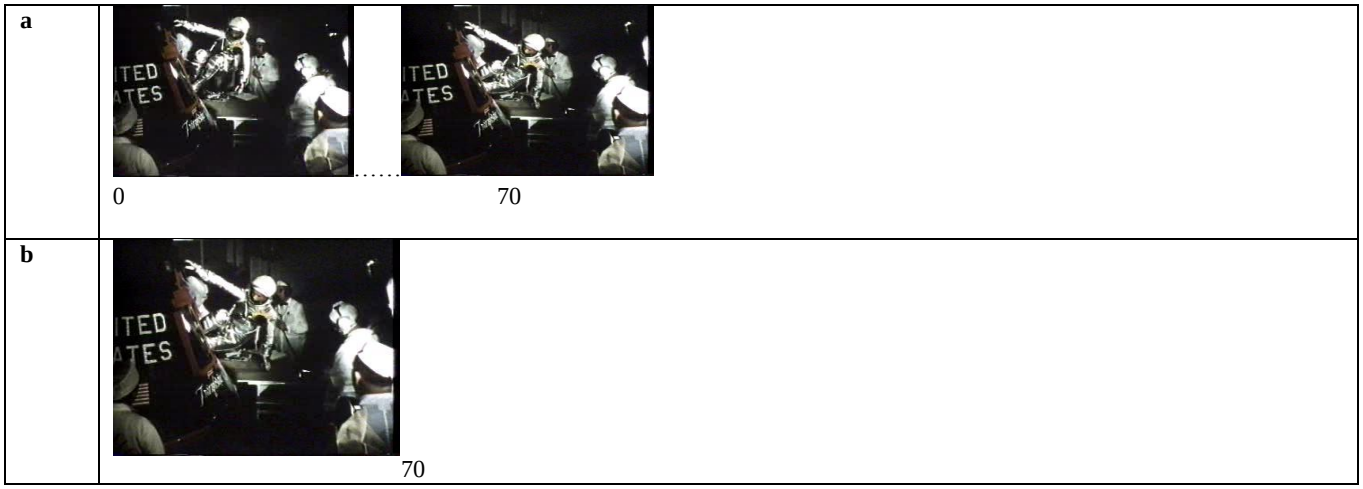

Fig. 4: Result of the proposed system. a) shows the original frames of the shot starting from frame 1 and ending at 72. b) shows the resulting keyframe

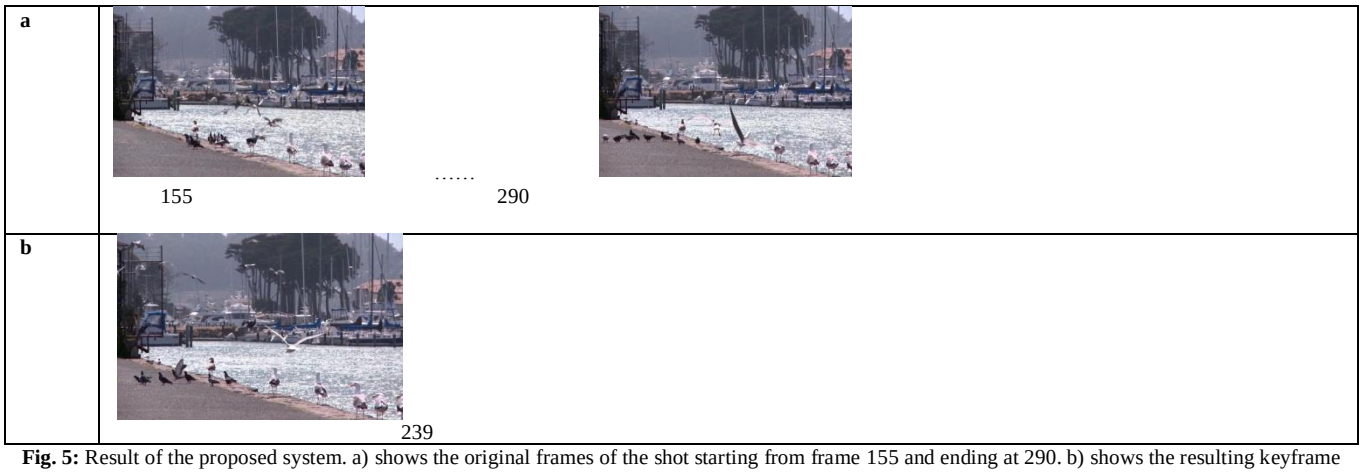

Fig. 5: Result of the proposed system. a) shows the original frames of the shot starting from frame 155 and ending at 290. b) shows the resulting keyframe

References

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

[2] Z. M. Lu and Y. Shi, "Fast video shot boundary detection based on SVD and pattern matching," IEEE Transactions on Image Processing, vol. 22, no. 12, pp. 5136–5145, 2013.

[3] R. Hannane, A. Elboushaki, K. Afdel, P. Naghabhushan, and M. 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, pp. 89–104, 2016.

[4] I. H. Ali and T. T. AL Fatlawi, "Key Frame Extraction Methods," International Journal of Pure and Applied Mathematics, vol. 119, no. 10, pp. 485–490, 2018.

[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. pp. 797–819, 2011.

[6] G. Gao and C. H. Liu, Video Cataloguing: Structure Parsing and Content Extraction. 2015.

[7] B. T. Truong and S. Venkatesh, "Video abstraction," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 3, no. 1, p. 3–es, 2007.

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

[9] S. V. Porter, M. Mirmehdi, and B. T. Thomas, "A shortest path representation for video summarisation," in Proceedings - 12th International Conference on Image Analysis and Processing, ICIAP 2003, 2003, pp. 460–465.

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

[11] T. Liu, X. Zhang, J. Feng, and K. T. Lo, "Shot reconstruction degree: A novel criterion for key frame selection," Pattern Recognition Letters, vol. 25, no. 12, pp. 1451–1457, 2004.

[12] A. M. Ferman and A. M. Tekalp, "Two-stage hierarchical video summary extraction to match low-level user browsing preferences," IEEE Transactions on Multimedia, vol. 5, no. 2, pp. 244–256, 2003.

[13] J. Calic and E. Izquierdo, "Efficient key-frame extraction and video analysis," in Proceedings - International Conference on Information Technology: Coding and Computing, ITCC 2002, 2002, pp. 28–33.

[14] Q. Zhang, S. P. Yu, D. S. Zhou, and X. P. Wei, "An efficient method of key-frame extraction based on a cluster algorithm," Journal of Human Kinetics, vol. 39, no. 1, pp. 5–13, 2013.

[15] Rong Pan, Yumin Tian, and Zhong Wang, "Key-frame extraction based on clustering," in 2010 IEEE International Conference on Progress in Informatics and Computing, 2010, vol. 2, pp. 867–871.

[16] A. Nasreen, K. Roy, K. Roy, and G. Shobha, "Key Frame Extraction and Foreground Modelling Using K-Means Clustering," in Proceedings - 7th International Conference on Computational Intelligence, Communication Systems and Networks, CICSyN 2015, 2015, pp. 141–145.

[17] B. Ghanem, T. Zhang, and A. Narendra, "Robust video registration applied to field-sports video analysis," Computer Engineering, pp. 1473–1476, 2012.