New Approach In Video Summarization Based On Color Feature

Saeideh Sarmadi

Department of engineering, Persian Gulf University, Boushehr, Iran

email address: s.sarmadi@pgu.ac.ir

Abstract

Video summarization is a process to create a meaningful summary of the recorded video within a short period of time. In recent years due to increase in generating videos, we need to summarize videos. General video summarization methods are divided into static and dynamic. Static summary is done by using key frames. However, the dynamic video summarization includes a sequence of small video shots that have been put together in chronological order. In this paper, video summarization is done by using static method and color-based key frame extraction. Then by applying the super resolution algorithm on selected key frames, we obtained frames with higher visual quality. The summarization was performed on several videos. Simulation results show our method produces summaries with high visual quality and computational speed.

Keywords: video summarization, static method, dynamic method, key frame extraction, super resolution.

1. Introduction

In the recent century, the use of digital data including text, image, animation and video is dramatically increasing, while due to time constraints and other reason, watching the whole content of video might not be possible. Regarding efficient storage and quick search, using video summarization is a useful solution since large data processing needs more resources. Video summarization as an active research topic enables users to manage and browse massive videos effectively [1], causing it to be of a great importance and interest. Generally, video summarization is a summary representing an abstract view of the original video sequence that can be used either as a video browser or retrieval systems. Another description of summarization can be either a highlight of the original video sequence which is the concatenation of a user-defined and selected video segments a collection of key frames [2]. Considering the numerous applications of video processing, video summarization is one of the most important things in the area of multimedia technology advances that aims creating summary of video to quick browsing of a collection of large volumes of video database and also useful for video indexing and surveillance.

Video summarization is done in various ways. The two common methods are dynamic and static. In dynamic methods, motioned pictures (skims) are used. Video skimming consists of a collection of image sequences along with the related audios from an original video [12]. Static methods are done by selecting a set of interesting frames of video division. The selected frames which called key frames are used to represent all content of video. So, selecting key frames is very important. For small devices such as

embedded surveillance systems, cellular phones, PDAs and IP cameras, where memory and computing resources are limited, key frame extraction can provide better managing and browsing results, because it is relatively easy to implement and enables users to review video contents in a shorter time [1]. Key frame extraction is done in several ways. For example, in content-based, features such as motion, sound, color, etc. are used for key frame extraction, while the main idea of cluster-based approaches is to select the frame closest to each cluster centroid as a key frame [4].

In 1996, Wolf used motion-based key frame selection [5] and the modified method was presented by Peyrad in [6]. In the method that presented by Zhang [3], the color feature is used for video summarization. In [7] automatic video summarization is done in consumer domain by analyzing the visual and audio content of the video stream. The summarizing rushes videos by motion, object and event understanding has been done by Wang [8]. Widiarto has performed video summarization using key frame selection based on shot segmentation in [9]. Clustering method is presented by Tommy Chheng in [10].

In this paper, we propose a two-phase approach, video summarization method based on color histogram aimed at extracting key frame from video, which is done in the first phase. Since video summarization is used in many issues of security, surveillance, medical and health care, the resolution of selected frames is critical. So, In the second phase, we apply a super resolution algorithm on key frames.

The remaining of this paper is organized as follows. In section II, Problem description and video summarization methods are discussed. In section III, we present our approach in video summarization. Section IV presents our simulation results and section V concludes this paper.

2. Problem description and methods

2.1 Defination of Video Summarization

Video summarization is the condensation of a video into a storyboard that contains key frames or several video segments, the goal of which is to make the video as small as possible, while keeping the flow of the story and essential parts of frames and/or video clips [11]. Video summarization provides a summarized version of video by analyzing the content of video. Summarization aims to create a summary of a digital video, which must satisfy the following five principles:

- The video summary must have precedence entities and actions from the video.
- The summary itself should exhibit reasonable degrees of continuity.
- The summary should be free of repetition
- The summary should present all the content of video.
- The summary should have short time in comparison with original video.

2.2 Applications of Video Summarization

Rapid and continuous growth of video content has caused the need to process videos. On the other hand, large volumes of data in a video are an obstacle for many practical applications such as search and retrieving videos. So we need a mechanism to understand the concept of a video without watching the whole content of it. This mechanism is video summarization. Video summarization is elimination of visual redundancy and duplicate frames from a video and get a short summary of the it. So the obtained summary is visually identical in comparison to the original video content. Video summarization is used

for review the important events, scenes or objects to provide interpretive summary of video. Factors such as time-consuming to watch video records and low quality video data in recent years are sufficient reasons to the need of this scientific field. Providing certain conditions, offering a preview of a television movie, sensitive section of a football game, interesting events in film are some applications of summarization. Other applications are video database management includes different areas such as search engine, digital video library, object indexing and retrieval and object classification, surveillance and in another fields like medical issues, care and supervision.

2.3 Video summarization methods

In order to summarize a video, most of the methods have consist on visual features computed from video frames. Also, there are methods that consider the semantic meaning implied in the video. Thera are two main video techniques in the literature: static video summarization and dynamic video summarization (video skimming) [12].

Static video summary represents a video sequence in a static imagery form (one or more selected frames from the original video, or a synthesized image generated from the selected key frames). Key frame extraction is fundamental process in video content management. It involves selecting one or multiple frames that will represent the content of video and used for generating video summaries. According to different sampling mechanisms, a set of key frames are extracted from shots of original video. Then, the selected key frames are arranged. One of the techniques in this category is clustering. The basic idea is clustering together similar frames/shots and then extraction some frames (generally one frame) per cluster as key frames. These methods are different in features (color histogram, luminance and motion vector) and clustering algorithms.

On the other hand, dynamic summarization consists on selecting the most relevant small dynamic portions (video skims) of audio and video in order to generate the video summary. video skim is a short video composed of informative scenes from the original video presented to the user to be able to receive an abstract of the video story, but in video format. In dynamic summarization, most mechanisms extract and segment video clips from the original video. Techniques for dynamic video skimming include applying SVD, motion model and semantic analysis [12].
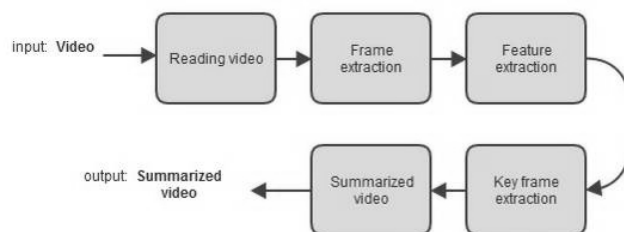

Fig. 1: Block diagram of key frame extraction

3. Proposed method

This section presents our proposed method in video summarization. It is a two phase method in which first phase produces video summary based on color histogram key frame extraction. Video data is usually used in compressed form, since the quality of video frames are decreased during compression, in second phase, we apply a super resolution algorithm to generate video frame with higher visual quality.

3.1 Key Frame Extraction

Block diagram of our video summarization is shown in Fig. 1. since color is an important feature for video frames, we used color feature in key frame extraction. The first key frame is selected manually. Color histogram is a common method to describe the color feature in a frame due to its simplicity and accuracy [3]. We compute color histogram for each frame. Then, by using the Euclidean distance formula, the histogram difference between two consecutive frame is calculated. The distance ($D_{his}(I,Q)$) between two color histograms, I and Q each consisting of N bins, is calculated by the following metric:

$$D_{his}(I, Q) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} a_{ij}(I_i - Q_i)(I_j - Q_j) \tag{1}$$

Where the matrix a_{ij} represents the similarity between the colors corresponding to bins i and j, respectively. If the E-distance is larger than a predefined threshold, the current frames is selected as the last key frame. The threshold value is changeable to control the number of key frames. A larger threshold will produce less key frames. On the contrary, a lower threshold will produce more key frames [3]. After calculation of histogram difference, we define threshold according to the following equation:

$$\text{Threshold} = \text{STD} + \text{Mean} * 4 \tag{2}$$

Where Mean is mean of histogram difference between two frames and STD is standard deviation of histogram difference. We determine threshold by using mean and standard deviation of histogram difference instead of using the histogram differences of two frames in each repetition of key frame extraction. So, as a result, key frames have little redundancy and the execution time has been reduced. Color histogram-based key frame extraction algorithm is described in Algorithm 1.

Algorithm 1: Key Frame Extraction

Input: Video
Output: Summarized Video
begin
1. **Reading a video**
2. **Extracting frames of video one by one**
3. **Compute color histogram of each frame**
4. **Calculate histogram difference between two consecutive frames (hd) according to (1)**
5. **Calculating mean and standard deviation of histogram difference**
6. **Compute threshold according to (2)**

7. **Compare threshold and the mean of histogram difference**

8.	**If (mean(hd) > threshold)**
9.	**Select it as a key frame**
10.	**Else**
11.	**Go to step 2**
12.	**Continue till end of video**
	End

3.2 Super Resolution Algorithm

In the second phase of our proposed method, we apply interpolation as a super resolution algorithm, to enhance the resolution and visual quality of selected key frames. This work is useful, specially, in surveillance applications of different areas like medical supervision.

4. Simulation results

To evaluate the performance of proposed method, Compression Ratio (CR) is computed. CR is used to study the compactness of frames in video shots. Higher value of CR of a method indicate that the method is good. The CR is calculated using the following equation:

$$CR = \frac{Total\ number\ of\ frames\ in\ video\ shot}{number\ of\ selected\ key\ frame} \tag{3}$$

In our experiment, we compare the performance analysis of our proposed method with three other video summarization techniques including VSUHCM (video summarization using higher order color moments) [13], VSUMM1 [14] and VSUMM2 [14] which have used the same video. Table I, shows the comparison of the proposed method with three other video summarization techniques in terms of Compression Ratio according to (3). Our proposed method gives better results for video V28, V31 and V32 [15]. The results of our proposed method and VSUHCM on HD quality video Wildlife.wmv having 901 frames are shown in Fig. 2. As the qualitative analysis shows, our proposed method has good results.

TABLE 1: Comparison of Different Methods (CR)

Video	Duration	Total Frames	VSUHCM	VSUMM1	VSUMM2	Proposed Method
V21	1:48	3266	233.3	217.7	272.2	251.2
V28	1:57	3523	293.6	185.4	220.2	**352.3**
V29	1:03	1917	273.9	213	239.6	239.6
V31	1:23	2467	224.3	224.3	308.4	**352.4**
V32	1:29	2653	221.1	241.2	241.2	**294.8**
V33	1:47	3213	321.3	160.7	247.2	229.5
V34	2:19	4205	300.4	467.2	600.7	420.5
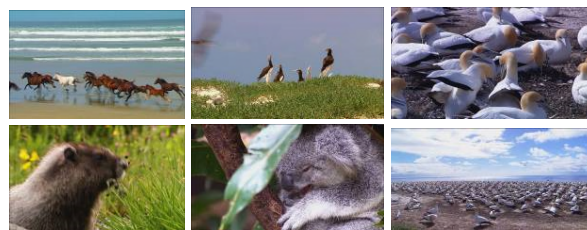
(a)
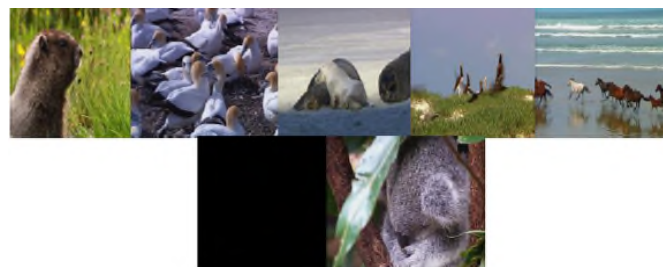
(b)

Fig. 1: Results On HD quality Video Wildlife.wmv (a) our proposed method and (b) VSUHCM

5. Conclusion

In recent years, the volumes of video data have been increasing. To handle this amount of data, different techniques have been developed. Video summarization is an efficient technique in this field. In this paper, we have proposed a new video summarization method which consists of color histogram-based key frame extraction phase and super resolution phase. Simulation results show our method has better operation in comparison to some other methods in terms of visual quality and execution time.

References

[1] Y. Zhang; R. Tao; Y. Wang, "Motion-State-Adaptive Video Summarization via Spatio-Temporal Analysis," in IEEE Transactions on Circuits and Systems for Video Technology , vol.PP, no.99, pp.1-1

[2] R. Kansagara, D. Thakore and M. Joshi, "A Study on Video Summarization techniques," International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, Issue 2, Feb. 2014, pp. 2962-2969, doi:10.15680/ijircce.

[3] J. Oh, Q. Wen, J. Lee and S. Hwang, "Video Abstraction," Video Data Management and Information Retrieval, S. Deb, ed., pp. 321-346, Idea Group Inc. and IRM Press, 2004.

[4] J. Chang, R. M. Hu and H. Zhang, "Summarization Extraction Method for Surveillance Video Based on Color Spatial Distribution Characteristic," *Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on*, Wuhan, 2009, pp. 1-5.

[5] W. Wolf, "Key frame selection by motion analysis," *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, Atlanta, GA, 1996, pp. 1228-1231 vol. 2.

[6] N. Peyrard and P. Bouthemy, "Motion-based selection of relevant video segments for video summarization", Multimedia Tools and Applications, vol. 26, pp. 259-276, 2005.

[7] W. Jiang, C. Cotton and A. C. Loui, "Automatic consumer video summarization by audio and visual analysis," *2011 IEEE International Conference on Multimedia and Expo*, Barcelona, Spain, 2011, pp. 1-6.

[8] F. Wang and C. W. Ngo, "Summarizing Rushes Videos by Motion, Object, and Event Understanding," in *IEEE Transactions on Multimedia*, vol. 14, no. 1, pp. 76-87, Feb. 2012.

[9] W. Widiarto, E. M. Yuniarno and M. Hariadi, "Video summarization using a key frame selection based on shot segmentation," *2015 International Conference on Science in Information Technology (ICSITech)*, Yogyakarta, 2015, pp. 207-212.

[10] Tommy Chheng, "Video summarization using clustering", Department of Computer Science, University of California, Irvine.

[11] D. j. Jeong, H. J. Yoo and N. I. Cho, "Consumer video summarization based on image quality and representativeness measure," *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, 2015, pp. 572-576.

[12] Z. Elkhattabi, Y. Tabii and A. Benkaddour, "Video Summarization: Techniques and Applications", International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 100, pp. 784-789, 2015.

[13] Jadhav, P.S., Jadhav, D.S.: Video summarization using higher order color moments (vsuhcm). Procedia Computer Science 45, 275–281 (2015). International Conference on Advanced Computing Technologies and Applications (ICACTA)

[14] S. E. F. de Avila, A. d. Luz, A. de A. Araújo and M. Cord, "VSUMM: An Approach for Automatic Video Summarization and Quantitative Evaluation," *2008 XXI Brazilian Symposium on Computer Graphics and Image Processing*, Campo Grande, 2008, pp. 103-110.

[15]   The Open-Video Project : http://www.open-video.org