ISSN:2249-5789
Azra Nasreen et al , International Journal of Computer Science & Communication Networks,Vol 3(3),194-198

Key Frame Extraction from Videos - A Survey

Azra Nasreen
Assistant Professor
Dept of Computer Science & Engineering
RVCE, Bangalore.
Azranasreen@rvce.edu.in

Dr. Shobha G
Professor & Dean
Dept of Computer Science & Engineering
RVCE, Bangalore.
shobhag@rvce.edu.in

Abstract

With the advent of social networking, sharing of multimedia has gained tremendous amount of importance and is the widely used form of communication worldwide. In the process of discovering knowledge from videos, challenge is to process huge amount of information which is resource-intensive. One way to minimize the cost of computations is to reduce the amount of information that undergoes processing. The redundant information can be greatly reduced by extracting key frames that represent the entire video. It is an essential task in any video analysis and indexing applications. This paper carries out an extensive survey of different key frame extraction techniques along with their merits and demerits used in current video retrieval applications.

1. Introduction

Video contains huge amount of information at different levels in terms of scenes, shots and frames. To discover knowledge from videos the issue that needs to be addressed is the elimination of redundant information. The Objective is to remove the redundant data which will significantly reduce the amount of information that needs to be processed. So, key frame extraction is the fundamental step in any of the video retrieval applications. It is necessary to discard the frames with repetitive or redundant information during the extraction. In recent years, many algorithms of key frame extraction focused on original video stream have been proposed. This paper provides an extensive survey in this area to bring out the advantages, drawbacks, suitability to an application, and precision of each method for video retrieval systems. Key frame is the frame which can represent the salient content of the shot. The key frames extracted must summarize the characteristics of the video, all the key frames on the time sequence gives visual summary of the video to the user. There are great redundancies among the frames in the same shot, so only those frames that best reflect the shot contents are selected as key frames to represent the shot. The extracted key frames should contain as much salient content of the shot as possible and avoid as much redundancy as possible. The features used for key frame extraction can include colors (particularly the color histogram), edges, shapes, optical flow, MPEG motion descriptors, MPEG discrete cosine coefficient, motion vectors, camera activity etc. The key frames [1] can be extracted utilizing the features of I-frame, P-frame and B-frame for each sub-lens. Fidelity and compression ratio are used to measure the validity of the method. Experimental results show that the by this method extracted key frames can summarize the salient content of the video and is of good feasibility, high efficiency, and high robustness. A framework [2] has been provided to assess the quality of the video against a given reference summary using both subjective and objective measures. A framework for automatic evaluation is needed based on both subjective and objective measures without the reference summary. Motion is the more salient feature in presenting actions or events in video and, thus, should be the feature to determine key frames. A triangle model of perceived motion energy (PME) is proposed to model motion patterns in video [3] and a scheme to extract key frames based on this model. The frames at the turning point of the motion acceleration and motion deceleration are selected as key frames. The key-frame selection process is threshold free and fast and the extracted key frames are representative. By focusing the analysis on the compressed video features, paper [4] introduces a real-time algorithm for scene change detection and key-frame extraction that generates the frame difference metrics by analyzing statistics of the macro-block features extracted from the MPEG compressed stream. The key-frame extraction method is implemented using difference metrics curve simplification by discrete contour evolution algorithm. This approach resulted in a fast and robust algorithm. Key frames are extracted utilizing the features of I-frame, P-frame and B-frame for each sub-lens. Key frames can also be extracted based on macro-block statistical characteristics of MPEG video stream[5] . The frame difference metrics

are generated by analyzing statistics of the macro-block features extracted from the MPEG compressed stream. The key-frame extraction method is implemented using difference metrics curve simplification by discrete contour evolution algorithm. The MPEG video compression algorithm has two main advantages- (1) Macro block-based motion compensation for the reduction of the temporal redundancy, (2) Transform domain based compression for the reduction of spatial redundancy.

In the compression of the video stream, frames can be grouped into sequences called a group of pictures (GOP). The types of frames can be classified into I frames, P frames and B frame. They are regularly arranged in the video stream and compose the GOPs. Within a GOP, an I frame is the first frame. I frames and P frames act as reference frames. I frames are intra-coded. The frames are processed with discrete cosine transform (DCT) using 8*8 blocks, and DC coefficients contain the main information. P frames are inter-frame coded. P frames refer to the preceding I frame or P frame, and are predicatively coded with only forward motion compensation based on macro blocks. The forward motion vectors for forward motion predication and DCT coefficients of residual error after motion compensation are obtained. B frames are inter-frame coded for forward motion prediction, backward motion prediction and bi-directional motion prediction. Each Macro-block of 16*16 pixels in P frames and B frames search for the optimal matching macro block in corresponding reference frames, then reduce predictive error of motion compensation with DCT coding. Key frames are extracted using the characteristics of I frames, P frames and B frames in the MPEG video stream after shot segmentation. If a scene cut occurs, the first I frame is chosen as a key frame. P frames are coded with forward motion compensation. When a shot transition occurs at a P frame, great change can take place in the P frame corresponding to the previous reference frames. The advantages of this method are: It compensates for the shortcomings of other algorithm and improves the techniques of key frame extraction based on MPEG video stream.

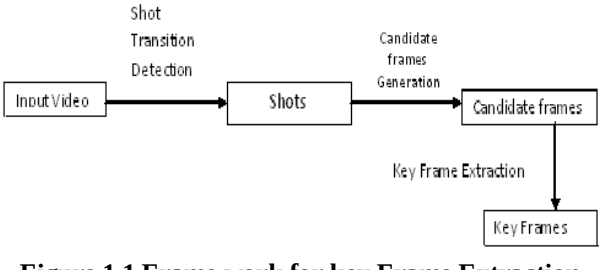

Figure 1.1 Frame work for key Frame Extraction

As shown in figure 1.1, the Input video is segmented into shots using the shot change detection techniques and then once the shot is identified, the key frames can be extracted from the candidate frames to represent each shot. All the key frames can be combined together to create a video summary which represent the video as a whole.

2. Shot Change Detection

The first step in the extraction of key frames is the shot change detection. It mainly refers to detecting the transition between successive shots. Each shot consist of several frames and can be represented by one or more frames based on their relative temporal differences. The detection methods can be broadly classified into abrupt transition detection and gradual transition detection. Shot transitions can be abrupt [5] [6] such as cuts or hard cuts or gradual transitions such as fade dissolve etc. The cut detection can be done in numerous methods such as pair wise pixel comparison, likelihood ratio, histogram difference etc. In Pair wise pixel comparison, the percentage of pixels that have changed in two consecutive frames is measured. If this difference exceeds the preset threshold value a cut is declared. Edge Change Ratio (ECR) algorithm is applied to detect the edge changes in the frames of each shot.

The commonly used methods of shot transition detection are pixel-based comparison, template matching and histogram-based method.

- The pixel-based methods are highly sensitive to motion of objects.
- Template matching is apt to result in error detection.
- A color histogram method is adopted to segment the shots according to the frame difference.

The Histogram-based method is the most commonly used method to calculate frame difference. Since color histograms do not relate spatial information with the pixels of a given color, and only records the amount of color information, images with similar color histograms can have dramatically different appearances. To solve the problem, an improved histogram algorithm X2 histogram matching method is adopted. The color histogram difference d (Ii, Ij) between two consecutive frames Ii and I j can be calculated and an appropriate threshold is selected. A shot transition occurs when d(Ii, Ij) is bigger than a given threshold. The basic idea of ECR is to detect the edges in two contiguous frames, count the number of edge pixels in two frames and define the entering and exiting edge pixels. The entering edge pixels En+1in are the fraction of edge pixels in Fn+1 which are more than a fixed distance r

from the closest edge pixel in Fn. Similarly the exiting edge pixels are the fraction of edge pixels in Fn+1 which are farther than r away from the closest edge pixel in Fn. If the ECR is larger than the pre-specified threshold then the two are considered as a cut. After going through the whole video the hard cuts can be detected and the video is broken down into shots. Once the shot transition detection is done the result will be a video segmented into a number of shots. Now these shots contain a number of frames. The next step is to extract the key frames and remove the redundant frames.

3. Key Frame Extraction

Once the shots are detected the key frames can then be extracted. As suggested previously there are many algorithms proposed for key frame extraction. All of these falls into one of the six categories listed below.

3.1 Sequential Comparison between Frames

In these algorithms, frames subsequent to a previously extracted key frame are sequentially compared with the key frame until a frame which is very different from the key frame is obtained. This frame is selected as the next key frame. The merits of the sequential comparison based algorithms include their simplicity, intuitiveness, low computational complexity, and adaptation of the number of key frames to the length of the shot. However, these algorithms suffer from problems such as the key frames represent local properties of the shot rather than the global properties, the irregular distribution and uncontrolled number of key frames make these algorithms unsuitable for applications that need an even distribution or a fixed number of key frames and also redundancy can occur when there are contents appearing repeatedly in the same shot.

3.2 Global Comparison between Frames

The algorithms are based on minimizing a predefined objective function based on the application.

3.2.1 Even temporal variance. These algorithms select key frames in a shot such that the shot segments, each of which is represented by a key frame, have equal temporal variance. The objective function can be chosen as the sum of differences between temporal variances of all the segments. The temporal variance in a segment can be approximated by the cumulative change of contents across consecutive frames in the segment or by the difference between the first and last frames in the segment.

3.2.2 Maximum coverage. These algorithms extract key frames by maximizing their representation coverage, which is the number of frames that the key frames can represent. If the number of key frames is not fixed, then these algorithms minimize the number of key frames subject to a predefined fidelity criterion. On the other hand, if the number of key frames is fixed, the algorithms maximize the number of frames that the key frames can represent. A greedy algorithm is used iteratively to find key frames.

3.2.3 Minimum correlation. These algorithms extract key frames to minimize the sum of correlations between key frames (especially successive key frames), making key frames as uncorrelated with each other as possible. For instance, represent frames their correlations using a directed weighted graph. The shortest path in the graph is found and the vertices in the shortest path which corresponds to minimum correlation between frames designate the key frames.

3.2.4 Minimum reconstruction error. These algorithms extract key frames to minimize the sum of the differences between each frame and its corresponding predicted frame reconstructed from the set of key frames using interpolation. These algorithms are useful for certain applications, such as animation. The merits of the aforesaid global comparison based algorithms include the following. The key frames reflect the global characteristics of the shot. The number of key frames is controllable. The set of key frames is more concise and less redundant than that produced by the sequential comparison based algorithms. The limitation of the global comparison based algorithms is that they are more computationally expensive than the sequential comparison based algorithms.

3.3 Reference Frame

These algorithms generate a reference frame and then extract key frames by comparing the frames in the shot with the reference frame. The merit of the reference frame based algorithms is that they are easy to understand and implement. The limitation of these algorithms is that they depend on the reference frame: If the reference frame does not adequately represent the shot, some salient contents in the shot may be missing from the key frames.

3.4 Clustering

These algorithms cluster frames and then choose frames closest to the cluster centers as the key frames. The merits of the clustering based algorithms are that they can use generic clustering algorithms, and the global characteristics of a video can be reflected in the extracted key frames. The limitations of these algorithms are as follows: First, they are dependent on the clustering results, but successful acquisition of semantic meaningful clusters is very difficult, especially for large data, and second, the sequential nature of the video cannot be naturally utilized. Usually, clumsy tricks are used to ensure that adjacent frames are likely to be assigned to the same cluster.

3.5 Curve Simplification

These algorithms represent each frame in a shot as a point in the feature space. The points are linked in the sequential order to form a trajectory curve and then searched to find a set of points which best represent the shape of the curve. The key frame extraction method is implemented using difference metrics curve simplification by the discrete contour evolution algorithm. The merit of the curve simplification based algorithms is that the sequential information is kept during the key frame extraction. Their limitation is that optimization of the best representation of the curve has a high computational complexity.

3.6 Objects/Events

These algorithms jointly consider key frame extraction and object/event detection in order to ensure that the extracted key frames contain information about objects or events. The key frames provide temporal interest points for classification of video events. The merit of the object/event based algorithms is that the extracted key frames are semantically important, reflecting objects or the motion patterns of objects. The limitation of these algorithms is that object/event detection strongly relies on heuristic rules specified according to the application. As a result, these algorithms are efficient only when the experimental settings are carefully chosen.

4. Evaluation methods

Because of the subjectivity of the key frame definition, there is no uniform evaluation method for key frame extraction. In general, the error rate and the video compression ratio are used as measures to evaluate the result of key frame extraction. Key frames giving low error rates and high compression rates are preferred. In general, a low error rate is associated with a low compression rate. The error rate depends on the parameters in the key frame extraction algorithms. Examples of these parameters are the thresholds in sequential comparison based, global comparison based, reference-frame based, and clustering based algorithms, as well as the parameters to fit the curve in the curve simplification based algorithms. Users choose the parameters according to the error rate that can be tolerated. In case of extracting key frames from MPEG stream fidelity and compression ratio are used to measure the validity of the method.

5. Conclusion

The extracted key frames should summarize the salient content of the video and the method should be of good feasibility, high efficiency, and high robustness. It should avoid processing inefficiency and computational complexity. The results of extracting using P-frame and I frame show that good fidelity and compression ratio can be achieved. It is not only of good feasibility, high efficiency, but also with low error and high robustness. Sequence Comparison between frames after detecting the shot change using Edge change ratio was found to be efficient as the key frames can be extracted based on global adaptive threshold.

5 6. References

[1] Guozhu Liu, and Junming Zhao, “Key Frame Extraction from MPEG Video Stream”, Second Symposium International Computer Science and Computational Technology (ISCSCT ’09), Huangshan, P. R. China, 26-28,Dec. 2009, pp. 007-011.

[2] Sujatha, C. ; Dept. of Comput. Sci. & Eng., BVBCET, Hubli, India ; Mudenagudi, U. “A Study on Keyframe Extraction Methods for Video Summary”, IEEE Conference on Computational Intelligence and Communication Networks (CICN), 7-9 October 2011,Gwalior, India.

[3] Tianming Liu ; Hong-Jiang Zhang ; Feihu Q, “A novel video key-frame-extraction algorithm based on perceived motion energy model “,IEEE Transactions on Circuits and Systems for Video Technology, vol 13, issue 10, Oct 2003.

[4] Janko Calic, Ebroul Izquierdo, “Efficient key-frame extraction and video analysis”, Multimedia and Vision Research Lab, Queen Mary, University of London.

[5] Ullas Gargi, Rangachar Kasturi, and Susan H. Strayer, "Performance Characterization of Video-Shot-Change Detection Methods", IEEE Transactions on Circuits and Systems for Video Technology, VOL. 10, NO. 1, February 2000.

[6] Weiming Hu, Senior Member, IEEE, Nianhua Xie, Li Li, Xianglin Zeng, and Stephen Maybank, "A Survey on Visual Content-Based Video Indexing and Retrieval ", IEEE Transactions on Systems, Man, and Cybernetic: Applications and Reviews, VOL. 41, NO. 6, November 2011. pp no 797-812.

[7] Costas Cotsaces, Nikos Nikolaidis, and Ioannis Pitas, "Video shot detection and condensed representation: a review," IEEE Signal Processing, vol. 23, no. 2, pp. 28-37, 2006.

[8] S. M M Tahgoghi, Hugh E Williams, James A Thom, and Time Walker, "Video cut deection using frame windows", Proceedings of the Twenty-eight Australian Conference on Computer Science, NewCastle, Austraila, Jan 2006, pp 193-199.

[9] D. Besiris, N. Laskaris, F. Fotopoulou, et al., "Key frame extraction in video sequences: a vantage points approach," International Workshop on Multimedia Signal , pp. 434-437, 2007.

[10] G. Ciocca and R. Schettini, "An innovative algorithm for key frame extraction in video summarization", J. Real-Time Image Process, vol.1, no. 1, pp. 69–88, 2006.

[11] J. Luo, C. Papin, and K. Costello, Toward extracting semantically meaningful key frames from personal video clips: From humans to computers, IEEE Trans. Circuits Syst. Video Technol. , vol. 19, no. 2, pp. 289-301, Feb. 2009.

[12] Vasileios Chasanis, Aristidis Likas and Nikolaos Galatsanos,"Video rushes summarization using spectral clustering and sequence alignment", Proceedings of the 2nd ACM TRECVid Video Summarization Workshop, 2008.

[13] Jafarpour S, Cevher V, Schapire R.E, "A game theoretic approach to expander-based compressive sensing" , IEEE International Symposium on Information Theory Proceedings (ISIT) , August 2011, page(s):464-46.