Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 70 (2015) 36 – 40

Procedia Computer Science

4th International Conference on Eco-friendly Computing and Communication Systems

Key-frame extraction by analysis of histograms of video frames using statistical methods

Sheena C V[a], N. K. Narayanan[a]

[a]*Department of Information Technology,Kannur University, Kannur 670 567, India*

Abstract

Summarization of videos for different applications like video object recognition and classification, video retrieval and archival and surveillance is an active research area in computer vision. One of the methods to summarize video data is extraction of key-frame. This paper proposes a method of key-frame extraction using thresholding of absolute difference of histogram of consecutive frames of video data. The experiment is conducted on KTH action database. For evaluation purpose compression ratio and fidelity value is calculated and it is able to achieve reasonably higher accuracy rate.

Peer-review under responsibility of the Organizing Committee of ICECCS 2015

*Keywords:*Key-frame; video summarization; absolute difference; histogram; thresholding.

1. Introduction

Key-frame extraction from video data is an active research problem in video object recognition and information retrieval. Key-frame refers to the image frame in the video sequence which is representative and able to reflect the summary of a video content. By using the key-frame it is able to express the main content of video data clearly and reduce the amount of memory needed for video data processing and complexity greatly. So we could make the storage organization, retrieval and recognition of video information more convenient and efficient [1]. Thus key-frame extraction is an efficient method for video summarization. Depending on the content complexity of the shot one or more key-frames can be extracted from a single shot. A shot is defined as unbroken sequence of frames recorded from a single camera, which forms the building blocks of video. In video data which contains multiple shots it is necessary to identify individual shots for key-frame extraction.

Literature classifies key-frame extraction in to sequential based approaches and cluster based approaches. In sequential based approaches visual features and temporal information are used to determine key-frames. That is the

Peer-review under responsibility of the Organizing Committee of ICECCS 2015
doi:10.1016/j.procs.2015.10.021

variation between the visual content of frames is estimated and the key-frame is selected whenever there is a considerable change. In cluster based approaches the basic idea is to produce the key-frame by clustering frames of a shot. The frames corresponding to representative of each cluster are selected as a key-frame. It should be noted that the clustering should preserve the temporal order of frames in video data. Also, the extracted key-frames may be static or dynamic in nature. The static key-frames are those frames that are extracted from the video which hold the important content of the video. Thus they are representative of video. Dynamic key-frames preserve dynamic nature of video in the sense that they are temporal ordered sequence of key-frames extracted [2].

This paper presents a method of computation of key-frame using thresholding of absolute difference of histogram. It computes thresholding point using mean and standard deviation of absolute difference of histogram for comparative study of feature difference of consecutive video frames. The organization of the paper is as follows. In section 2 a brief overview of related work is presented. In section 3 proposed frame works for key-frame extraction is discussed. The results are discussed in section 4 and finally conclusions are drawn in section 5.

2. Previous Works

In this section some of the approaches of key-frame extraction is discussed. A detailed review of existing techniques is done by Li et al. [3].One of the simplest methods for key-frame extraction is to select first frame of each shot segment as key-frame. But this may not be a proper choice, since the rest of the frames are not inspected to determine whether they are also represented by the selected key- frame [4]. Zhao et al. exploited the concept of curve segmentation to extract key-frames. In their work each frame is represented by the use of color histogram where each frame is represented by using CIE U V color model with 256 bins. The difference between adjacent frames is estimated and plotted versus frames as a point in a 2D plane. The plotted curve is analyzed to estimate the sharp corners and the frame corresponding to the sharp corner are treated as key-frame of the shot [5]. A fuzzy based key-frame selection approach is presented by Doulamis et al. Each frame is represented in the form of a fuzzy color and a motion histogram. Normally features are extracted on the entire frame, where as in this work each frame is segmented using the multi-resolution recursive shortest spanning tree algorithm. For each segment a fuzzy color and motion histogram is extracted and stored as a representative. The frames which are not similar to each other are selected based on the cross-correlation criterion by the use of genetic algorithm [6]. Mukherjee et al. have proposed a model to estimate key-frames from a video shot based on the randomness of representative of the frames. For each frame spatial and Har wavelet based features are extracted. In case of spatial, average intensity and busyness of intensity defined over 3×3 image mask are calculated and in case of Har wavelets, frequency space points at different resolutions are estimated. Individual features are used to estimate the randomness between frames and the frames with high randomness are selected as key-frames. Also to take a decision based on both spatial and wavelet features Dempster – Shafer theory of evidence is used [7].

Li Liu et al. developed a new method of key-frame extraction using correlated pyramidal motion-feature for human action recognition. To select key- frame for each action sequence he used Ada - Boost learning algorithm [8]. A method of key frame extraction based on unsupervised clustering was adopted by Zhuang et al. They used color histogram of each frames of video computed in HSV color space and a threshold to control clustering density. The key frame selection is employed only to the clusters which are big enough to consider as key cluster. From each key cluster a representative is selected as a key-frame [9].Gong and Liu proposed a technique for video abstraction based on singular value decomposition. The method calculated color histogram of video frames in RGB color space. To incorporate spatial information each frame is divided into 3×3 blocks and a 3D histogram is created for each block. The nine histogram are then concatenated together to form a feature vector. These feature vectors are used to form clusters and for each cluster the frame closest to the cluster center is selected as a key frame [10].

The related works of key-frame extraction reveals that they select predefined number of key-frames and should take care of temporal ordering of frames during clustering. Proposed method extract distinct key-frames keeping temporal ordering of video sequence based on threshold obtained from mean and standard deviation of absolute difference of histogram of consecutive frames.

3. Key – frame extraction using absolute difference of histogram of consecutive frames

This section explores a method of key-frame extraction algorithm based on absolute difference of histogram of consecutive image frames. It is a two phase method in which first phase compute threshold using mean and standard deviation of histogram of absolute difference of consecutive image frames. Second phase extract key – frames comparing the threshold against absolute difference of consecutive image frames. The algorithm starts by extracting video frames one by one. After preprocessing each video frames histogram difference between two consecutive frames are calculated. The mean and standard deviation of absolute difference of histogram is calculated to fix a threshold point. The threshold (T) is computed using following equation.

$$T = \mu_{adh} + \sigma_{adh} \quad (1)$$

Where μadhmean of absolute difference and σadhis the standard deviation of absolute difference. Once the threshold is obtained next phase determine the key-frames by comparing the absolute difference of histogram against threshold. The proposed algorithm is given below.

Step.1 *Extract frames one by one*

Step. 2 *Histogram difference between two consecutive frames*

Step. 3 *Calculate mean and standard deviation of absolute difference*

Step. 4 *Compute threshold*

Step. 5 *Compare the difference with T and if it is*

> Tselects it as a key-frame else go to step 2

Step. 6 *Continue till end of video*

The experiment is conducted on KTH action database,a database containing six types of human actions (walking, jogging, running, boxing, hand waving and hand clapping) performed several times by 25 subjects in four different scenarios: outdoors s1, outdoors with scale variation s2, outdoors with different clothes s3 and indoors s4. Currently the database contains 2391 sequences. All sequences were taken over homogeneous backgrounds with a static camera with 25fps frame rate. The sequences were down sampled to the spatial resolution of 160x120 pixels and have a length of four seconds in average [11]. Fig.1 shows illustration of KTH action database. Some frames of KTH action database under the class running is shown in Fig.2

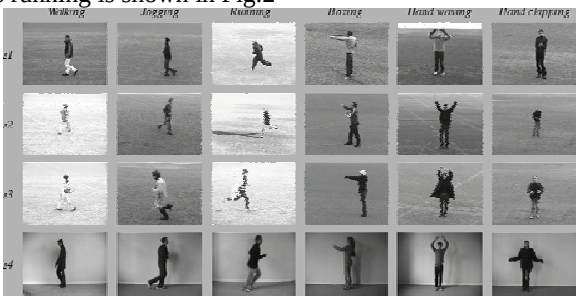

Fig.1 KTH action database illustration

4. Experiments and results

To evaluate the performance of the proposed technique fidelity measure and compression ratio is computed. Fidelity measure is used to study the effectiveness of a method in preserving the global content of a shot. The compression ratio is used to study the compactness of the shot content. Higher values of fidelity and compression ratio of a method indicate that the method is good [12].

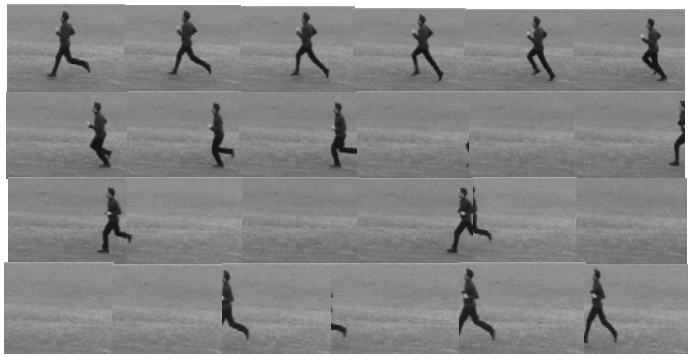

Fig.2 Sample frames of class running

Fidelity is a maximum of minimal distance between the set of key-frames of a shot and frames of correspondingshot or semi Hausdorff distance. The fidelity value is computed as follows. Let SF={Fi/1≤i≤N} be a set of N numbers of frames in a shot S and KF={KFj/1≤j≤l}be set of l number of key-frames obtained for the shot S. The semi Hausdorff distance between the set of frames SF to the set of key-frames KF is the maximum of the minimal distance between individual frames of the shot SF and the key-frame in the set KF. The distance between jthkey-frame KFj to the image frame of a shot is calculated as follows

$d_j=\min\{dis(KF_j,F_i)\}$, $F_i/1\le i\le N, 1\le j\le l$ (2)

Then semi Hausdorff distance between SF and KF is

$d_{sh}=\max\{d_j\};1\le j\le l$ (3)

The compression ratio (CR) is used to study the compactness of shot due to selected key-frames and it depends on the number of key-frames selected. Compression ratio is computed using the equation

$$CR=\frac{\text{total number of frames in video shot}}{\text{number of key - frames selected}} \quad (4)$$

Table. 1 gives the compression ratio (CR) and fidelity value of the selected video data in KTH action database. In KTH action database all video data are represented as single shot of individuals' actions so shot detection is not performed. Key-frame extraction is directly performed on the noise filtered video data. The key-frames obtained for the video data person01_running_d1_uncomp.avi is shown in Fig. 3.

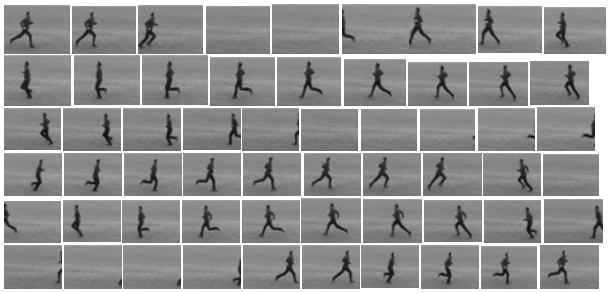

Fig .3 Key- frames of data *person01_running_d1_uncomp.avi*

Table .1 Compression ratio and fidelity criteria for selected videos in database

Video	Compression Ratio (CR)	Fidelity (d_{sh})
person01_running_d1_uncomp.avi	5.6780	1457
person01_running_d2_uncomp.avi	8.9024	1381
person01_running_d3_uncomp.avi	6.4815	1480
person02_running_d1_uncomp.avi	6.8261	1584
person02_running_d2_uncomp.avi	7.5354	1867

5. Conclusion

In this paper a method for automated extraction of key-frames for video summarization is presented. The role of key- frame extraction in Human Action Recognition is to reduce redundant frames that can lead dimensionality reduction of feature vector for classification. Directly representing the video sequence by all the frames, which contain redundant and indiscriminate information, would confuse the classifier in action recognition [13]. In future we have to extract visual features from the extracted key-frame for recognition. Majority of the existing key-frame extraction algorithm consider predefined number of key-frames to be extracted on the other hand those which are automated need high computation and ordering of temporal data. But the proposed algorithm is able to compute the key-frames using simple calculations of histogram of absolute difference of consecutive frames in video data. The values of compression ratio and fidelity criteria show that the results obtained are reasonably accurate.

References

1. Zhonglan Wu and Pin Xu , " Research on the technology of Video key-frame extraction based on clustering", IEEE Fourth international conference on Multimedia Information networking and security, 2012, p. 290-293,
2. Naveed Ejaz et al. , " Adaptive key-frame extraction for video summarization using an aggregation mechanism", Journal of Visual Communication 23 (2012), p.1031-140.
3. Y. Li et al. Techniques for movie content analysis and skimming: tutorial and overview on video abstraction techniques, IEEE signal processing magazine 23(2)2006 p. 27-50
4. Suresh C Raikwar, Charul Bhatnagar and Anand Singh Jalal, "A frame work for key-frame extraction from surveillance Video", 5th International Conference on Computer and Communication Technology", IEEE, 2014, p. 297-300.
5. Zhao et al. "Key-frame extraction and shot retrieval using nearest feature line",Proceedings of ACM Workshop on Multimedia, 2000,p. 217-220.
6. Doulamis et al. "A fuzzy video content representation for video summarization and content based retrieval", Journal of signal processing, 2000,p.1049-1060
7. Mukhargee et al., "Key-frame estimation in video using randomness measure of feature point pattern", IEEE transactions on circuits on systems for video technology, vol.7,no.5,May 2007, p. 612-620.
8. Li Liu, Ling Shao, Peter Rockett, "Boosted key-frame and correlated pyramidal motion feature representation for human action recognition", Pattern Recognition 45 (2013), p. 1810-1818.
9. Zhuang Y, Rui Y, Huang T.S and Mehvotra S, " Adaptive key-frame extraction using unsupervised clustering", Proceedings of International conference on Image Processing, 1998, p 866-870.
10. Gang Y and Liu, "Video summarization using singular value decomposition", Proceedings of Computer Vision and Pattern Recognition, 2000, p 347-358.
11. http://www.nada.kth.se/cvap/actions/.
12. S. Manjunath, "VARS: Video Archival and Retrieval System", Ph.D Thesis, 2012.
13. Li Liu et al., " Learning discriminative key poses for action recognition", IEEE transactions on cybernetics , Vol.43, No.6, Dec 2013, p. 1860-1870.