



REVIEW OF PURPOSED METHOD FOR KEY FRAME EXTRACTION FROM VIDEOS

Riya, Sumandeep kaur

Department of Computer Engineering, Punjabi University
Patiala, India

Abstract: In recent days, vision based surveillance in public spaces has dramatically increased due to its wide applications such as traffic system management, terrorism/crime deterrence, and crowded activity monitoring and so on. These surveillance videos often contain a large amount of frames. Taking an example of a frame of 25fps, there are $250 \times 3600 = 90,000$ frames in total for one hour of video. The huge volume of video data is a barrier to many practical usages. So, work on key frame extraction going on to extract key frames from videos which give an overall summarization of a video.

Keywords: Key frame extraction, video summarization

INTRODUCTION

In recent years, use of video bases information is increasing more and more. Due to this many research is done in the area of video. Key frame is very useful in the area of video abstraction, video summarization, video editing and animation. With the large amount of video data available, it has become increasingly important to have the ability to quickly search through and browse through these videos. For that Key frame is very useful for the users. There are many methods to find out key frame. In general key frame extraction, video is first converted into scenes that are called scene segmentation. Then it is converted into different shots and among them shot boundary is detected using threshold/edge detection method. Then after the detection of shot boundary the key frame is extracted from different shots.

USES OF KEY-FRAME EXTRACTION:

- **Video transmission:** In order to reduce the transfer stress in network and invalid information transmission, the transmission, storage and management techniques of video information become more and more important. When a video is being transmitted, the use of key frames reduces the amount of data required in video indexing and provides the framework for dealing with the video content. Each frame can only choose the latest coded and reconstructed key frame as its reference frame. After coding and packetisation, compressed video packets are transmitted with differentiated service classes. Key frame along with difference values are sent from the source.
- **Video annotation:** Video annotation is the extraction of the information about video, adding this information to the video which can help in browsing, searching, analysis, retrieval, comparison, and categorization. Annotation is to attach data to some other piece of data (i.e. add metadata to data). To fasten the access of video, it is annotated. It is not momentous to analyze each video frame for this, so key

frames are found and only these are analyzed for annotation purpose.

- **Video indexing:** Key frames reduce the amount of data required in video indexing and provides framework for dealing with the video content. Before downloading any video over the internet, if key frames are shown besides it, users can predict the content of the video and decide whether it is pertinent to his search. Other applications such as creating chapter titles in DVDs and prints from video using the key frame picture and the difference values the picture is reconstructed at the destination.
- **Video summarization:** Video summarization is a compact representation of a video sequence. It is useful for various video applications such as video browsing and retrieval systems. A video summarization can be a preview sequence which can be a collection of key frames which is a set of chosen frames of a video. Key-frame-based video summarization may lose the spatio-temporal properties and audio content in the original video sequence; it is the simplest and the most common method. When temporal order is maintained in selecting the key frames, users can locate specific video segments of interest by choosing a particular key frame using a browsing tool. Key frames are also effective in representing visual content of a video sequence for retrieval purposes. Video indexes may be constructed based on visual features of key frames, and queries may be directed at key frames using image retrieval techniques.

LITERATURE SURVEY

In this section key frame extraction from videos and their enhancement algorithms survey is done.

Hannane, et al.[1], In today's digital era, there are large volumes of long-duration videos resulting from movies, documentaries, sports and surveillance cameras floating over internet and video databases (YouTube). Since manual processing of these videos are difficult, time-consuming and expensive, an automatic technique of abstracting these long-duration videos are very much desirable. In this backdrop,

this paper they presents a novel and efficient approach of video shot boundary detection and key frame extraction, which subsequently leads to a summarized and compact video. There proposed method detects video shot boundaries by extracting the SIFTpoint distribution histogram (SIFT-PDH) from the frames as a combination of local and global features. In the subsequent step, using the distance of SIFT-PDH of consecutive frames and an adaptive threshold video shot boundary are detected. Further, the key frames representing the salient content of each segmented shot are extracted using entropy based singular values measure. Thus, the summarized video is then generated by combining the extracted key frames. The experimental results show that our method can efficiently detect shot boundaries under both abrupt and gradual transitions, and even under different levels of illumination, motion effects and camera operations (zoom in, zoom out and camera rotation). With their proposed method, the computational complexity is comparatively less and video summarization is very compact.

Paliwal, *et al.* [2], Text extraction from still images and videos are on high demand these days in the area of multimedia for data retrieval and many other. Text extraction from the videos is a challenging task in image processing because of complex and sometimes highly illuminated background in videos it is quite complicated to extract text from the running video. Video segmentation and key frame extraction play an important role in extraction of text from a video. This paper present a brief study on how different methods and algorithms that have been used till date for text extraction from videos and still images. Through there are numbers of existing literatures to image processing and segmentation, they attempt to give a more elaborate image for a comprehensive review. With some tables and figures, they brief in the content.

Antonis, *et al.*[3], the extraction of representative key-frames from video shots is very important in video processing and analysis, since it constitutes the basis for several important tasks such as video shot summarization, browsing and retrieval as well as high-level video segmentation. The extracted key-frames should capture a great percentage of the information of a shot content, while at the same time they should not present similar visual information. Clustering or segmentation methods are usually employed to extract key-frames. A major difficulty is caused by the large variety in the visual content of videos. Thus, using a single image descriptor (color, texture etc) to extract key-frames is not always effective, since there is no single descriptor surpassing the others in all video cases. To tackle this problem, they proposed an approach for the weighted fusion of several descriptors that automatically estimates the weight of each descriptor. The weights reflect the relevance of each descriptor for the specific video shot. Moreover, they are used to form a composite similarity matrix as the weighted sum of all the similarity matrices corresponding to the individual descriptors. This matrix is then used as input to a spectral clustering algorithm that partitions shot frames into groups. Finally the medoid frame of each group is selected as key-frame. Numerical experiments using a variety of videos demonstrate that our method is capable of efficiently summarizing video shots regardless of the characteristics of the visual content of a video.

K.S. Thakre, *et al.* [4], video partitioning and key frame extraction (KFE) are the key foundations of video analysis and Content based video retrieval. The use of key frames reduces the amount of data that is necessary in video indexing and provides the outline for dealing with the video content. In the last few years, many algorithms of key frame extraction concentrated on the original compressed video stream. This can increase computational complexity when decompression is required before video processing. Key frame is the frame, which can be a prototype of the salient content and information of the shot. The key frames extracted must summarize the significant features of the video in time sequence. Therefore, there is a commensurate need of an efficient and secured key frame selection technique in an efficient CBVR system. They proposed an algorithm for key frame extraction of compressed video shots using adaptive threshold method. Extensive computation on 200 plus video clips is performed and results are accurate and satisfactory.

Chin Yeow Wong, *et al.* [5], color image enhancement technique introduced in this work aims at maximizing the information content within an image, whilst minimizing the presence of viewing artefacts and loss of details. This is achieved by weighting the input image and the interim equalized image recursively until the allowed intensity range is maximally covered. The proper weighting factor is optimally determined using the efficient golden section search algorithm. Experiments had been conducted on a large number of images captured under natural indoor and outdoor environment. Results showed that their proposed method is able to recover the largest amount of information as compared to other current approaches. The developed method also provides satisfactory performances in terms of image contrast, and sharpness.

Sheena C.V and N.K Narayanan [6], 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. Their methods to summarize video data are extraction of key-frame. They proposed 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.

Kuldeep Singh and Rajiv Kapoor [7], this paper presents a robust contrast enhancement algorithm based on histogram equalization methods named Median-Mean Based Sub-Image-Clipped Histogram Equalization (MMSICHE). The proposed algorithm undergoes three steps: (i) The Median and Mean brightness values of the image are calculated. (ii)The histogram is clipped using a plateau limit set as the median of the occupied intensity. (iii) The clipped histogram is first bisected based on median intensity then further divided into four sub images based on individual mean intensity, subsequently performing histogram equalization for each sub image. This method achieves multi objective of preserving brightness as well as image information content (entropy) along with control over enhancement rate, which in turn suits for consumer electronics applications. This method avoids excessive enhancement and produces images with natural enhancement. The simulation results show that

MMSICHE method outperforms other HE methods in terms of various image quality measures, i.e. average luminance, average information content (entropy), absolute mean brightness error (AMBE) and background gray level.

Xian-Mei Liu, *et al.* [8], In this paper, they present a new solution for extracting key frames from motion capture data using an optimization algorithm to obtain compact and sparse key frame data that can represent the original dense human body motion capture animation. The use of the genetic algorithm helps determine the optimal solution with global exploration capability while the use of a probabilistic simplex method helps expedite the speed of convergence. By finding the chromosome that maximizes the fitness function, the algorithm provides the optimal number of key frames as well as the low reconstruction error with an ordinary interpolation technique. The reconstruction error is computed between the original motion and the reconstruction one by the weighted differences of joint positions and velocities. The resulting set of key frames is obtained by iterative application of the algorithm with initial populations generated randomly and intelligently. They also present experiments which demonstrate that the method can effectively extract key frames with a high compression ratio and reconstruct all other non key frames with high quality.

Guoliang Lu, *et al.* [9], recent years have witnessed a dramatic growth of the deployment of vision based surveillance in public spaces. Automatic summarization of surveillance videos (ASOSV) is hence becoming more and more desirable in many real-world applications. For this purpose, a novel frame-selection framework is presented in this paper, which has three properties: 1) *un-supervision*: it can work without requirements of any supervised learning or training; 2) *efficiency*: it can work very fast, with experiments demonstrating efficiency faster than real-timeness and 3) *scalability*: it can achieve a hierarchical analysis/ overview of video content. The performance of proposed framework is systematically evaluated and compared with various state-of-the-art frame selection techniques on some collected video sequences and publicly-available *ViSOR* dataset. The experimental results demonstrate promising performance and good applicability for real-world problems.

PROBLEM FORMULATION

In recent days, vision-based surveillance in public spaces has dramatically increased due to its wide applications such as traffic system management, terrorism/crime deterrence, and crowded activity monitoring and so on. These surveillance videos often contain a large amount of frames. Taking an example of a frame rate of 25 fps, there are $25 \times 3600 = 90,000$ frames in total for one hour of video. The huge volume of video data is a barrier to many practical usages. So, work on key frame extraction going on to extract key frames from videos which give an overall summarization of a video. From the survey, it's found that in the most of the technique the comparative of two consecutive frames histogram is done with a fixed threshold. So, we found following issues on which further work can be done.

1. In the videos, similarity metrics are used for key frame extraction without taking into the consideration the illumination variances in the color intensities posed to the motion object during time.
2. Most of the key frame extraction techniques are used for cyber and crowd activity monitoring. The key frames of videos are low quality so enhancement is required for further processing. (Rachida Hannanel, March 2016)

PURPOSED METHODOLOGY

1. First step is to extract frames.
2. Second step is to calculate histogram difference between two consecutive frames.
3. Third step is calculating threshold value.
4. Next step is to compare the consecutive frames with threshold value for extracting the key frame.
5. Performance parameters are calculated for the purposed technique.

REFERENCES

- [1] Rachida Hannanel, Abdessamad Elboushaki, Karim Afdel, P. Naghabhushan and Mohammed Javed, "An Efficient Method for Video Shot Boundary detection and key frame extraction using SIFT Point Distribution Histogram," International Journal Multimedia Information Retrieval, Vol.5, pp.89-104, March 2016.
- [2] Sonal Paliwal, Rajesh Shyam Singh, and H.L. Mandoria, "A Survey on various text detection and extraction techniques from videos and images," International Journal of Computer Science Engineering, Vol. 6, pp.1-10, June 2016.
- [3] Antonis Ioannidis, Vasileios Chasanis, and Aristidis Likas, "Weighted Multi-view Key Frame Extraction," Elsevier: Pattern Recognition Letter, Vol.72, pp.52-61, February 2016.
- [4] K.S. Thakre, A.M. Rajurkar, and R.R. Manthalkar, "Video Partitioning and Secured Key Frame Extraction of MPEG Video," International Conference on Information Security and Privacy, pp. 790-798, December 2015.
- [5] Chin Yeow Wong, Shilong Liu, Md Arifur Rahman, Stephen Ching-Feng Lin, Guannan Jiang, Ngiaming Kwok and Hariyan Shi, "Image Contrast Enhancement using Histogram Equalization with Maximum Intensity Coverage," Journal of Modern Optics, Vol. 63, No. 16, pp. 1618-1629, 2016.
- [6] Sheena C.V and N.K Narayanan, "Key Frame Extraction by analysis of histogram of video frames using statistical methods," 4th International Conference on Eco-Friendly Computing and Communication Systems, pp. 36-40, 2015.
- [7] Kuldeep Singh and Rajeev Kapoor, "Image Enhancement via Median-Mean Based Sub-Image-Clipped Histogram Equalization," Optik, Vol. 125, pp. 4646-4651, April 2014.
- [8] Xian-Mei Liu, Ai-min Hao, and Dan Zhao, "Optimization based key frame extraction for motion capture animation," Springer: Visual Computation, Vol. 29, pp.85-95, 2013.
- [9] Guoliang Lu, Yiqi Zhou, Xueyong Li, and Peng Yan, "Unsupervised, Efficient and Scalable Key-Frame Selection for Automatic Summarization of Surveillance Videos," Multimedia Tools Application, January 2016.
- [10] Swagat Kumar Behera, Dr. Satyasis Mishra, and Debaraj Rana, "Image Enhancement using Particle Swarm Optimization," International Journal of Engineering Research and Technology, Vol. 4, pp.1049-1055, March 2015.