A Review Paper on Key Frame Extraction

Jignesh M. Chhaiya[1] Ashish M. Kothari[2]
[1]PG Student, [2]Asst. Professor
[1,2]A.I.T.S., Rajkot, Gujarat, India

Abstract— In the recent years, use of video bases information in increasing more and more. Due to this many research is done in the area of video. Owing to the decreasing cost of storage devices, higher transmission rates, and improved compression techniques, digital video is becoming available at an ever increasing rate. To extract valid information from video, without any loss of information, much attention is being paid to video processing technology. For this key frame is very useful technique. Extracting a small number of key frames that can abstract the content of video is very important for efficient browsing and retrieval in video databases. Consequently, technologies for video segmentation and key-frame extraction have become crucial for the development of advanced digital video systems. In this paper, different methods of key-frame extraction are shown and a comparison is done between the outputs of all these methods.

I. INTRODUCTION

The process of extracting frames which are representative of the video is known as key frame extraction. The world as a living space is shrinking, are we really shrinking or have we found a new horizon to live in. It is true we are expanding leaps and bounds in Giga bytes and terra byte world. Recent advances in technology have made tremendous amounts of multimedia information available to the general population. A video in simplest of words is agglomeration of data. With the ever escalating videos the systems for processing these videos need to be developed. Analyzing these videos as small data packets for the simplicity of human effort is the need of the hour. Extraction of key frames from the video and to analyze only these frames instead of all the frames present in the video can greatly improve the performance of the systems. Key frame is the frame which can represent the salient content and information of the video. The key frames extracted must summarize the characteristics of the video, and the image characteristics of a video can be tracked by all the key frames in time sequence. In shot based method shots of the original video are first detected, and then one or more key frames are extracted from each shot. Methods of shot transition detection are: pixel-based comparison, and histogram-based method. The pixel-based methods are susceptible to motion of objects. So it is suitable to detect segmentation transition of the camera and object movement. But in this method as each pixel is compared the time required is more. The Histogram-based methods entirely lose the location information. For example, two images with similar histograms may have completely different content.

II. REVIEW OF CLASSIC KEY FRAME EXTRACTION METHOD

Following is the brief summary of the works carried out in this domain. Various methods employed for key frame extraction and shot boundary detection are reviewed here.
Varieties of schemes have been employed yet now which are as follows:

A. Paper [1]

In paper [1] authors describe key frame extraction technique based on intuition that higher the motion more the key frames required for summarization.
They obtain the key frame by dividing the shot in parts of equal cumulative motion activity and selecting the frames located at the half way point of each sub segment.
Furthermore they establish empirical relation between the motion activity of a segment and required number of key frames and compute them.

B. Paper [2]

In paper [2] author present an approach to key frame extraction for structuring user generated videos on video sharing websites (e. g. YouTube). Their approach is intended to link existing image search engines to video data. User generated videos are, contrary to professional material, unstructured, do not follow any fixed rule, and their camera work is poor. Furthermore, the coding quality is bad due to low resolution and high compression.

In a first step, they segment video sequences into shots by detecting gradual and abrupt cuts. Further, longer shots are segmented into sub shots based on location and camera motion features. One representative key frame is extracted per sub shot using visual attention features, such as lighting, camera motion, face, and text appearance which key frames are useful for indexing and for searching similar video sequences using MPEG-7 descriptors.
They compare their key frame extraction approach against the key frame extraction of IBM Multimedia Analysis and Retrieval System (IMARS). Their key frame extraction is only based on visual differences. Many extracted key frames by IMARS are shaky, blurred and extracted during gradual transitions due to ignoring motion and visual attention features. Unlike IMARS, our approach extracts the key frame with the highest amount of visual attention and minimal motion intensity to get the steadiest frame. The key frames are evaluated by the mean opinion score (MOS), because it is hard to judge objectively the quality.;

C. Paper [3]

In paper [3] they propose a novel method for key-frame extraction based on dominant-set clustering. Key frames play

an important role in video abstraction. Clustering is a popular approach for key-frame extraction.
Compared with the existing clustering-based methods, the proposed method dynamically decides the number of key frames depending on the complexity of video shots produces key frames in a progressive manner and requires less computation. Experimental results on different types of video shots have verified the effectiveness of the method.

D. Paper [4]

In paper [4] their approach uses shot boundary detection to segment the video into shots and the k-means algorithm to determine cluster representatives for each shot that are used as key frames. Furthermore they performed an additional clustering on the extracted key frames to provide a video summarization.

E. Paper [5]

In paper [5] In the emerging research field of content-based video copy detection, efficient representation of video content at key frame level is crucial, due to the fact that similarity search is mainly performed between content-representative frames. In this paper a sequential search algorithm that bypasses the process of temporal video segmentation is proposed for key frame extraction in MPEG videos.
They aim at providing an efficient, real-time and fully automatic way of extracting key frames in videos, where not only the laborious task of offline video database indexing is avoided, but also query video processing is performed in the same manner as the reference video database. Significant reduction in computational cost is achieved by exploiting DCT coefficients in feature extraction. The effectiveness of the proposed scheme is evaluated in terms of quality and speed on the manually annotated TREC Vid 2007 test video dataset.

F. Paper [6]

In paper [6] Query by key frame or video example is a convenient and often effective way to search in video database. This paper proposes a new approach to support such searches. The main contribution of the proposed approach is the consideration of both feature extraction and distance computation as a whole process. With a video shot represented by key-frames corresponding to feature points in a feature space, a new metric is defined to measure the distance between a query image and a shot based on the concept of Nearest Feature Line (NFL). They use the "breakpoints'' of feature trajectory of a video shot as the key frames and use the lines passing through these points to represent the shot. When combined with the NFL method, it helps to achieve a better performance, as evidenced by experiments.

G. Paper [7]

In Paper [7] In this paper, a new key frame extraction method is presented, which not only is based on the traditional idea of clustering in the feature extraction phase but also effectively reduces redundant frames using the integration of local and global information in videos.
Experimental results on the TRECVid 2007 test video dataset have demonstrated the effectiveness of our proposed key frame extraction method in terms of the compression rate and retrieval precision. Key frame extraction methods aim to obtain a set of frames that can efficiently represent and summarize video contents and be reused in many video retrieval-related applications.
An effective set of key frames, viewed as a high-quality summary of the video, should include the major objects and events of the video, and contain little redundancy and overlapped content.

H. Paper [8]

In paper [8] they propose an innovative approach to the selection of representative (key) frames of a video sequence for video summarization. By analyzing the differences between two consecutive frames of a video sequence, the algorithm determines the complexity of the sequence in terms of changes in the visual content expressed by different frame descriptors.
The algorithm, which escapes the complexity of existing methods based, for example, on clustering or optimization
Strategies, dynamically and rapidly selects a variable number of key frames within each sequence. The key frames are extracted by detecting curvature points within the curve of the cumulative frame differences. Another advantage is that it can extract the key frames on the fly : curvature points can be determined while computing the frame differences and the key frames can be extracted as soon as a second high curvature point has been detected. Video summarization, aimed at reducing the amount of data that must be examined in order to retrieve the information desired from information in a video, is an essential task in video analysis is and indexing applications.

I. Paper [9]

In paper [9] they present a compact representation for animating meshes based on novel key-frames extraction and animating mesh simplification approaches.
Three-dimensional animating meshes have been widely used in the computer graphics and video game industries. Reducing the animating mesh complexity is a common way of overcoming the rendering limitation or network bandwidth.
In contrast to the general simplification and key-frames extraction approaches which are driven by geometry metrics, the proposed methods are based on a deformation analysis of animating mesh to preserve both the geometric features and motion characteristics.
These two approaches can produce a very compact animation representation in spatial and temporal domains, and therefore they can be beneficial in many applications such as progressive animation transmission and animation segmentation and transferring.

J. Paper [10]

In paper [10] they propose an efficient algorithm for video sequence matching using the modified Hausdorff distance and the directed divergence of histograms between successive frames.
To effectively match the video sequences with a low computational load, we use the key frames extracted by the cumulative directed divergence and compare the set of key frames using the modified Hausdorff distance.

To manipulate large video database, effective video indexing and retrieval are required. A large number of video retrieval algorithms have been presented for frame wise user query or video content query, whereas a few video-sequence matching algorithms have been investigated.
Experimental results with color video sequences show that the proposed algorithms for video sequence matching yield better performance than conventional algorithms such as histogram difference, histogram intersection, and Chi-square test methods.

K. Paper [11]

In paper [11] they introduce a new algorithm for key frame extraction based on un-supervised clustering. The algorithm or both computationally simple and able to adapt to the visual content .The efficiency and effectiveness are validated by large amount of real world data

L. Paper [12]

In Paper [12] with the features of MPEG compressed video stream, a new method is presented for extracting key frames. Firstly, an improved histogram matching method is used for video segmentation.
Secondly, the key frames are 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 extracted key frames can summarize the salient content of the video and the method is of good feasibility, high efficiency, and high robustness.
In order to extract valid information from video, process video data efficiently, and reduce the transfer stress of network, more and more attention is being paid to the video processing technology.
The amount of data in video processing is significantly reduced by using video segmentation and key-frame extraction. So, these two technologies have gradually become the focus of research.

III. CLASSIFICATION OF KEY FRAME EXTRACTION METHODS

Commonly, some conservative methods are used to extract key frames such as preferring to extract more key frames to obtain important information of video streams. When extracting many key frames, the irrelevance between two adjacent frames is considered as the principle of key frame extraction.

A. Video shot based Method

In the video shot based method, some classic methods include frame average method and histogram average method. The frame average method is to extract the average value in a specific location from the video shot, then select the frame whose pixel value mostly close to the average value as key frame; histogram average method is to compute the average value of the histogram of all frames in the video shot, then select the frame mostly close to the average histogram as key frame.
These methods have the advantage of low computation complexity and easy computation, and extracted frames have average representative meaning. However, in these methods, the complexity of content in the current video shot is not considered and the number of key frames is limited as a fixed value, and the content in the video shot with many changes cannot be completely described.

B. Content Analysis based Method

This method is to extract key frames based on the change of color, texture and other visual information of each frame, when this information changes significantly, the current frame is key frame.
The basic idea is that the first frame is selected as the new frame, and is viewed as reference frame, then the back frames are compared with the reference frame in order, the k-th frame do not become the new key frame until the distance between the k-th and (k-1)-th frame exceed a specific threshold. The frame
This method can select corresponding key frames according to the change degree of content in the video shot. But its disadvantage is that it is not sensitive to the movement of video camera and cannot quantitatively indicate the changes of movement information, thus which will cause unstable key frame extraction.

C. Motion-based Analysis Method

Movement of a shot can be analyzed by analyzing the optical flow, and select the local minimum in the movement as key frames. The movement of a shot is defined as formula：

$$M(k) = \sum\sum | O_x(i,j,k) | + | O_y(i,j,k) |$$

Where $O_x(i, j, k)$ is the intra-pixel (i , j) X-component of optical flow, and $O_y(i, j, k)$ is the intra-pixel (i , j) Y-component of optical flow .
The disadvantage is that the method did not have strong robustness because it was depended on local information, and it did not pay enough attention to the content changes brought by the cumulative dynamic either.

IV. CLUSTER-BASED METHOD

It can be a good response to the main shot content when the cluster analysis method was applied to the key frame extraction. Algorithm idea is as follows: first initialize a cluster center, and then determine the current frame is classified as this class or as a new class of cluster centers by calculating the distance between the current frame and the center, and finally including all the frames which is nearest from the cluster centers as key frames. It is difficult for cluster algorithm to obtain general cluster parameters.

V. CONCLUSION AND FUTURE WORK

In this paper, I reviewed different methods of key frame extraction with its advantages and disadvantages in direction of obtaining best and reliable method for key frame extraction.

REFERENCES

[1] MOTION ACTIVITY BASED EXTRACTION OF KEY FRAMES FROM VIDEO SHOTS by Ajay divakaran, Regunathan Radhakrishnan, Kadir A Perker published in IEEE 2002

[2] FEATURE-BASED VIDEO KEY FRAME

EXTRACTION FOR LOW QUALITY VIDEO SEQUENCES By Pascal Kelm, Sebastian Schmiedeke, and Thomas Sikora published in IEEE 2009
[3] KEY-FRAME EXTRACTION USING DOMINANT-SET CLUSTERING by Xianglin Zeng, WeimingHu, Wanqing Liy, Xiaoqin Zhang, Bo Xu published at SCSSE, University of Wollongong, Australia
[4] KEY FRAME EXTRACTION FOR VIDEO TAGGING & SUMMARISATION published at German Research Center for Artificial Intelligence (DFKI)
[5] REAL-TIME KEYFRAME EXTRACTION TOWARDS VIDEO CONTENT IDENTIFICATION by Maria Chatzigiorgaki and Athanassios N. Skodras published in IEEE 2009
[6] KEY FRAME EXTRACTION AND SHOT RETRIEVAL USING NEAREST FEATURE LINE by Li Zhao†_, Wei Qi*, Stan Z. Li*, Shi-Qiang Yang†_, H. J. Zhang* published at Microsoft Research China
[7] INTEGRATION OF GLOBAL AND LOCAL INFORMATION IN VIDEOS FOR KEY FRAME EXTRACTION by Dianting Liu, Mei-Ling Shyu, Chao Chen, Shu-Ching Chen published at Florida International University, Miami, FL 33199, USA
[8] AN INNOVATIVE ALGORITHM FOR KEY FRAME EXTRACTION IN VIDEO by G. Ciocca, R. Schettini published at Bicocca degli Arcimboldi 8, 20126 Milano,Italy
[9] ANIMATION KEY FRAME EXTRACTION AND SIMPLIFCATION USING DEFORMATION ANALYSIS by Tong-Yee Lee, Chao-Hung Lin, Yu-Shuen Wang, and Tai-Guang Chen published at IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 18,
[10] AN EFFECTIVE ALGORITHM FOR VIDEO SEQUENCE MATCHING USING THE MODIFIED HAUSDORFF DISTANCE AND THE DIRECTED DIVERGENCE by Sang Hyun Kim, Rae-Hong Park published at IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 12
[11] ADAPTIVE KEY FRAME EXTRACTION USING UNSUPERVISED by Yueting Zhuangy,Yong Rui,Thomas S. Huang and Sharad Mehrotra published at University of Illinois at Urbana-Champaign
[12] KEY FRAME EXTRACTION FROM MPEG VIDEO STREAM by Guozhu Liu, and Junming Zhao published at ISCSCT ’09

