# Key Frame Extraction from MPEG Video Stream

Guozhu Liu
College of Information Science & Technology
Qingdao Univ. of Science & Technology
Qingdao 266061,P. R. China
Email: lgz_0228@163.com

Junming Zhao
College of Information Science & Technology
Qingdao Univ. of Science & Technology
Qingdao 266061,P. R. China
Email:zhjm530@sina.com

***Abstract*—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. 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.**

***Keywords-MPEG coding; key frame; shot segmentation; histogram***

## I. INTRODUCTION

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.

Video segmentation and key frame extraction are the bases of video analysis and content-based video retrieval. Key frame extraction [1-2], is an essential part in video analysis and management, providing a suitable video summarization for video indexing, browsing and retrieval. The use of key frames reduces the amount of data required in video indexing and provides the framework for dealing with the video content [3].

In recent years, many algorithms of key frame extraction focused on original video stream. It can introduce processing inefficiency and computational complexity when decompression is required before video processing.

Key frame is the frame which can represent the salient content and information of the shot. 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. Furthermore, the content of the video can be recognized. A basic rule of key frame extraction is that key frame extraction would rather be wrong than not enough. So it is necessary to discard the frames with repetitive or redundant information during the extraction.

A new algorithm of key frame extraction from compressed video data is presented in this paper. We analyze the features of compressed data and finally obtain the key frames.

For video, a common first step is to segment the videos into temporal "shots," each representing an event or continuous sequence of actions. A shot represents a sequence of frames captured from a unique and continuous record from a camera. Then key frames are to be extracted. Video segmentation is the premise of key frame extraction, and key frames are the salient content of the video (key factors to describe the video contents). Figure 1 illustrates the basic framework of our algorithm.

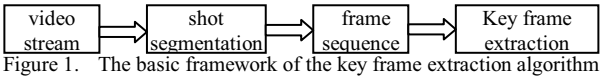


Figure 1. The basic framework of the key frame extraction algorithm from MPEG video stream

## II. ALGORITHM REALIZATIONS

### A. Shot segmentation

Shot segmentation is the first step of the key frame extraction, which mainly refers to detecting the transition between successive shots. The domain of video shot segmentation falls into two categories: uncompressed and compressed. The detection methods can be broadly classified into abrupt transition detection and gradual transition detection [4].

Nowadays, the common used methods of shot transition detection are as follows: pixel-based comparison, template matching and histogram-based method [5-6]. The pixel-based methods are highly sensitive to motion of objects. So it is suitable to detect obvious segmentation transition of the camera and object movement. Template matching is apt to result in error detection if you only simply use this method. In contrast to pixel-based methods, the Histogram-based methods completely lose the location information of pixels. Consequently, two images with similar histograms may have completely different content. In addition, there exist other methods of shot detection, such as boundary detection and dual threshold method.

When camera switches, the video image data will undergo a series of significant changes, such as content change, color difference increase and trajectory

978-0-7695-4261-4/10 $26.00 © 2010 IEEE
DOI

discontinuities. Thus there is a peak between consecutive frames. A color histogram method is adopted to segment the shots according to the frame difference.

The Histogram-based method is the most common 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, $X^2$ histogram matching method is adopted. The color histogram difference $d(I_i, I_j)$ between two consecutive frames $I_i$ and $I_j$ can be calculated as follows:

$$d\left(I_i, I_j\right) = \sum_{k=1}^{n} \frac{(H_{ik} - H_{jk})^2}{H_{jk} + H_{ik}}, (H_{jk} \neq 0) \quad (1)$$

Where $H_i$ and $H_j$ stand for the histogram of $I_i$ and $I_j$, respectively.

A shot transition occurs when $d(I_i, I_j)$ is bigger than a given threshold. The experiment result illustrates that good effect can be achieved. Selecting an appropriate threshold is the key to the method.

*B. Key frame extraction*

Since key frame extraction plays an important role in video retrieval and video indexing, a lot of research has been done on the techniques. The widely used key frame extraction techniques are as follows:

*1)* Key frame extraction based on shot activity. Gresle and Huang [7] computed the intra and reference histograms and then compute an activity indicator. Based on the activity curve, the local minima are selected as the key frames.

*2)* Key frame extraction based on macro-block statistical characteristics of MPEG video stream. Janko and Ebroul [8] generate 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.

*3)* Key frame extraction based on motion analysis. Wolf [9] computed the optical flow for each frame and then used a simple motion metric to evaluate the changes in the optical flow along the sequence. Key frames are then found at places where the metric as a function of time has its local minima.

Nowadays, most of the video are stored in the compressed form of MPEG. The MPEG video compression algorithm has two main advantages: macro block-based motion compensation for the reduction of the temporal redundancy and transform domain based compression for the reduction of spatial redundancy [10]. 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. Figure 2 shows how different types of frames can compose a GOP.

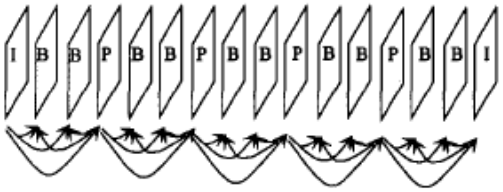


Figure 2. The frame sequence of a MPEG GOP

Within a GOP, an I frame is the first frame and I frames and P frames are 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 and B 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. At the same time, one or two motion vectors are transferred.

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.

In video stream, 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. So encoder can not utilize the macro blocks in previous reference frames to compensate the effect. Therefore, many of the macro blocks should have been coded as "intra" without motion compensation.

An equation is designed to calculate the ratio of macro blocks without motion compensation, which is used to detect whether the P frame is selected as a key frame. The equation is given below:

$$R_p = \frac{no\_com}{com} \quad (2)$$

Where $no\_com$ denotes the number of macro blocks without motion compensation, and $com$ stands for the number of macro blocks after motion compensation.

When $R_p$ peak appears, the P frame can be selected as a key frame.

Shot transition in video stream can also occur at B frame. In the circumstance, great change may take place on the content of B frame compared with the preceding frame. Therefore, the motion vectors come from the reference

frames after the B frame instead of the former ones when B frames are coded by the encoder.

A ratio of backward motion vectors and forward motion vectors is calculated to detect whether the B frame is a key frame. The equation is given as follows:

$$R_B = \frac{back}{forw} \qquad (3)$$

where $back$ is the number of the backward motion vectors and $forw$ denotes the number of the forward motion vectors.

## III. VALIDITY MEASURES

Key frame extraction aims to reduce the amount of video data, and the frame sequence must preserve the overall contents of the original video. Whether the key frames can be accurately detected and extraction is the fundamental rule to measure the validity of the algorithm. Current measurement mainly relies on eye observation. If we simply use this method, it is time-consuming in the case of large-scale video data.

Compression ratio and fidelity[11-12] are chosen to measure the validity of the algorithm in our work. Compression ratio measure is used to evaluate the compactness of the key frame sequence, while fidelity is to measure the correlation degrees of the sets in the image classifications. Fidelity is defined as a Semi-Hausdorff distance between the key frame set and the shot frame set.

Suppose that the key frame set $R$ consists of $K$ frames, $R = \{KF_j \mid j = 1,2,\cdots,k\}$, while the shot frame set $S$ consists of $N$ frames, $S = \{F_i \mid i = 1,2,\cdots,k\}$. Let the distance between any two frames $KF_j$ and $F_i$ be $d(KF_j, F_i)$. Define $d_i$ for each frame $F_i$ as:

$$d_i = min(d(KF_j, F_i)), j = 1,2,\cdots,k \qquad (4)$$

Then the Semi-Hausdorff distance between $S$ and $R$ is given as:

$$d_{sh} = max(d_i),\ i = 1,2,\cdots,N. \qquad (5)$$

The fidelity measure is defined as:

$$fidelity = 1 - \frac{d_{sh}}{\max_i(\max_j(d_{ij}))} \qquad (6)$$

where $d_{ij}$ denotes the dissimilarity matrix of the shot set $S$.

The bigger the fidelity is, the more accurate the global scan of key frames over the original video is.

## IV. EXPERIMENT TEST

A segment of video is selected to do the experiment. Firstly, let the first frame as a key frame, and the ratios are calculated according to equation (2),(3). The frame where a ratio peak occurs is extracted as a key frame. If there are no transitions in a shot, the frames in the shot have high similarity, and there is no significant change among the characteristic curves. Then the first frame can be extracted as a key frame, and finally the key frames of the video can be obtained. The experimental result is as follows:

Figure 3. The segment of the video

Figure 4. Key frames extracted (the first frame, the 10th frame and the 31st frame

The algorithm is based on shot. There is only one shot in the tested video sequence, and there are no gradual transitions and abrupt transitions. So only a shot is obtained after the shot segmentation. The first frame is extracted as a key frame, and other key frames are calculated according to macro block motion when abrupt transitions occur.

## V. RESULTS AND COMPARATIVE ANALYSIS

To verify the validity of the algorithm, it is realized by VC++ in Window XP environment on Intel PD (2.80GHz) with 1G storage. Four segments of video with different characteristics are selected from The Open Video Project as training samples. Experimental test is processed using the method. The results are shown in Table 1.

TABLE I. THE RESULTS OF THE EXPERIMENT

| video sequences | *number of the shot* | *total frames* | *key frames* | *compression ratio (%)* | *fidelity* |
|---|---|---|---|---|---|
| A:BOR10_013 | 8 | 1026 | 12 | 99.3 | 0.7424 |
| B:HURR003 | 12 | 1653 | 22 | 99.5 | 0.7823 |
| C:Indi009 | 17 | 2896 | 30 | 99.1 | 0.7732 |
| D:aircrash | 10 | 1952 | 15 | 99.6 | 0.7534 |

From Table 1, the average compression ratio of the new algorithm is 99.48%, and the average fidelity is 0.76. From the experiment, we can see that the representative key frames can be extracted accurately and semantically from long video sequences or videos with more transitions, reflecting the video content objectively.

Figure 5 shows the results of key frame extraction with the proposed algorithm. The video depicts the process of plane take-off, failure, pilot parachuting. Ten shots are segmented and 15 key frames are extracted from the video with the algorithm. From the key frames, the video content can be clearly acknowledged, including the people, the process of the accident, etc. The result illustrates that the algorithm is valid to segment the shot and extract the key frames and it is of good feasibility and strong robustness.

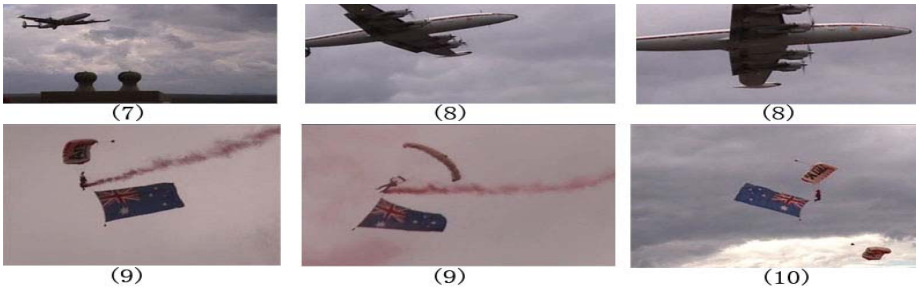


Figure 5. The results of the key frame extraction from the video sequence of air crash

In a second test, we examine the validity of our algorithm comparing the fidelity with two newer algorithms. One algorithm is the key frame extraction combining global and local information[13], and the other is information theory based shot cut/ fade detection and video Summarization[14]. The fidelity values are calculated with the 4 videos above respectively, as illustrated in Figure 6.

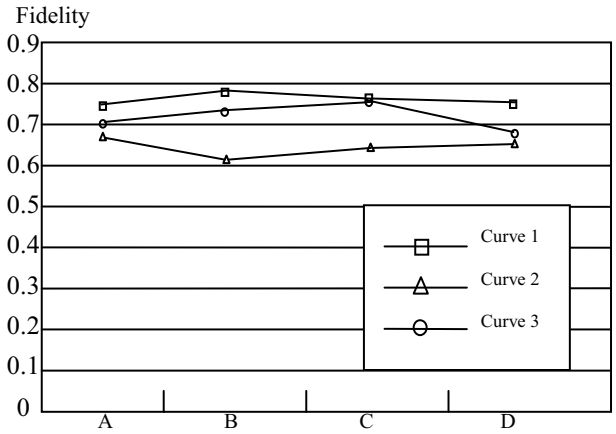


Figure 6. Fidelity values of the three algorithms

Curve 1 indicates the results of our method.

Curve 2 shows the results of the algorithm combining the global and local information, where the parameter values are as follows: $\delta_1 = 0.4$, $\delta_2 = 0.04$, $m = 3$, $\delta = 0.8$, $dis = 13$ and $times = 5$. The algorithm adopts similarity model base on color characteristics. The fidelity value is low for the video B with simple color. At the same time, the algorithm can not achieve good effect for the longer video C. Therefore, it depends on the accuracy of the shot segmentation. Moreover, it also relies on the threshold. When the parameters change, the effects also change accordingly.

Curve 3 is the result of the last algorithm. Good results can be achieved by selecting different thresholds for different video. However, it also depends on the thresholds. Satisfactory results only can be achieved through many experiments.

From the above analysis, the algorithm we proposed is of high accuracy, good feasibility and generality.

## VI. CONCLUSIONS

We have proposed a new algorithm for key frame extraction. It compensates for the shortcomings of other algorithm and improves the techniques of key frame extraction based on MPEG video stream. The experimental results 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.


## ACKNOWLEDGMENT

This research was supported by A Project of Shandong Province Higher Educational Science and Technology Program(J08LJ21). Also, supported by Natural Science

Foundation of Shandong Province of China(ZR2009GL006).

## REFERENCE

[1] D.Feng, W.Siu and H. Zhang, "Multimedia information retrieval and management: Technological Fundamentals and Applications, " *Springer*, pp.44, 2003.

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

[3] T. Liu, H. Zhang, and F. Qi, "A novel video key-frame-extraction algorithm based on perceived motion energy model," *IEEE Transactions On Circuits And Systems For Video Technology*, vol. 13, no. 10, pp. 1006-1013, 2003.

[4] Irena Koprinska, Sergio Carrato, "Temporal video segmentation: A survey," *Signal Processing: Image Communication*, vol. 16, no. 5, pp. 477-500,2001.

[5] C. F. Lam, M. C. Lee, "Video segmentation using color difference histogram," *Lecture Notes in Computer Science*, New York: Springer Press, pp. 159–174., 1998.

[6] A. Hampapur, R. Jain, and T. Weymouth, "Production model based digital video segmentation," *Multimedia Tools Application*, vol. 1, no. 1, pp.9–46, 1995.

[7]P. Gresle, T. S. Huang, "Gisting of video documents: a key frames selection algorithm using relative activity measure," *The 2nd International Conference On Visual Information System*, 1997.

[8]J. Calic, E. Izquierdo, "Efficient key-frame extraction and video analysis information technology: coding and computing," *international symposium on information technology*, pp. 28-33, 2002.

[9]W. Wolf, "Key frame selection by motion analysis," *Proc. IEEE Int. Conf. Acoust., Speech Signal Proc*., vol. 2, pp. 1228–1231, 1996.

[10] D. Le Gall, "MPEG: a video compression standard for multimedia applications," *Communications of the ACM*, vol. 34, pp. 46-58, 1991.

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

[12] 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.

[13]Z. Zhan, W. Yu, "A method of key frame extraction combing global and local information," *Application Research of Computers*, vol.24, no. 11, pp. 1-4, 2007. (In Chinese)

[14] Z. Cernekova, I. Pitas, and C. Nikou, "Information theory-based shot cut/fade detection and video summarization," IEEE Trans. Circuits Syst. Video Technol., vol. 16, no. 1, pp. 82–91, Jan. 2006