Chapter 69
An Improved Key Frame Extraction Algorithm of Compressed Video

Xiaoping Wang

Abstract The key frame extraction technology is the basis of content-based video retrieval, and a key frame extraction algorithm is based on the lens for compressed video sequence. Its core is based on the compression encoding characteristics of video sequences, only need to be part of the decoding, use the DC component of I-frame information to construct DC thumbnail, and combine the important difference of different areas of the image frame's DC information to make the similarity measure, and thus achieve key frame extraction. The experiments show that the algorithm is significantly improved than the traditional one about the two indicators of the recall and retrieval time, especially for the more dramatic news documentaries, films, and other local sports video sequences.

Keywords Key frame extraction • Compressed domain • The MPEG

69.1 Introduction

Presently, the key frame extraction based on the shot, light, movement descriptor to a variety of methods. Among them, the most common method is lens-based key frame extraction. A video is divided into the lens, the first frame of each scene (or the first and last frames) as the key frame of the lens. This method is relatively simple, regardless of the contents of the lens; the number of key frames is relatively OK (the first frame, last frame, or both are selected), the drawback is less stable, because the first and last frames of each scene are not always able to reflect the main content of the lens. The key frame extraction method based on the lens is studied mainly from two areas: the pixel domain and compressed domain.

X. Wang (✉)
College of Mathematics and Computer Science, Yangtze Normal University, Fuling, Chongqing, China
e-mail: wxp1102@163.com

Z. Zhong (ed.), *Proceedings of the International Conference on Information Engineering and Applications (IEA) 2012*, Lecture Notes in Electrical Engineering 217,
DOI: 10.1007/978-1-4471-4850-0_69,

69.1.1 Lens-Based Key Frame Extractions in the Pixel Domain

In the so-called pixel domain, this refers to the space/time domain compared with transform domain, the video data exists in the form of people's daily scene, people's accustomed features (such as color, texture, shape, and motion vectors). Pixel domain detection is the use of these features to get the clip of a video sequence. The key for the shot segmentation is to find the difference between the different camera images. Currently, we have developed some more mature ways to do key frame extraction, full use of the video data, time/space, global/local, static/dynamic, and other kinds of information. Histogram comparison method is the most traditional and common method. In a continuous video sequence, if there is no special treatment, a small gap is formed between the adjacent two frames. In this way, the characteristics of adjacent frames are also almost the same. There are many algorithms for comparing two frame histogram differences that typically include the Euclidean distance, X the square of detection, dual-threshold comparison method, and the sub-block division method.

69.1.2 Lens-Based Key Frame Extractions in the Compressed Domain

More and more video data are saved in compressed form such as JPEG, MPEG2X; thus it is necessary to study compressed video sequence key frame detection method. This test is carried out usually two ways:

1. First, full-decompression (e.g., Huffman decoding, DPCM decoding, DCT inverse transform and motion compensation) is used to form a video sequence and then used the pixel domain-based approach to realize key frame extraction. The disadvantage of this method is to calculate more and low efficiency.
2. Second, partial decompression, which directly uses the features in the compressed video data to analyze and process, saving decoding time and reducing the computational complexity at the same time.

Currently, image and video compression aspects of international standards, such as JPEG, MPEG, of H. 261 and H. 263, are based on DCT. DCT is converting the pixel values of the two-dimensional space into two-dimensional frequency domain coefficient values; the frequency domain transform coefficients and the pixel domain are closely related and express the contents of the image frame to a certain extent. Early Arman and others used DCT coefficients to detect MPEG; this method was later extended to the MPEG compressed stream for the shot segmentation.

69.2 The Improved Key Frame Extraction Algorithm Based on the Lens

69.2.1 Algorithm Basis

The adjacent image frames of the video sequence have similarity and continuity, which is the theoretical basis of the key frame extraction based on lens. Yang Sheen et al. construct the key frame extraction system accordingly.

Known in the MPEG21/2 international standard (video part), MPEG21/2 video sequence is constituted by a number of image groups (group of picture GOP), and each GOP is composed by a range of the I, P, B frames of mutual interval forecast and generation; in each group, the first frame is always I-frame; I-frame adopts coding information of the image itself, and P, B frames are obtained by the forecast. Each shot must include the I-frame and has been confirmed by experiment (MPEG21/2 video encoding requires an I-frame in 13 frames in every shot. The lens, which is composed of the uninterrupted consecutive frames, and its playing time should be by s unit that can make sense, so calculating the frame rate (24 fps), each lens must include I-frames). Therefore, the key frame established in the lens can completely delete P and B frames and generate video sequence file that is composed only by the I-frame. In addition, considering the image compression in MPEG21/2 standards is based on the DCT, the transformation is the basic unit of 8×8 sub-blocks for the transformation, can decode the I-frame to a certain extent, remove the DC coefficients, and restore DC thumbnail. And then adopt the template matching method; use the difference between the thumbnails as a similarity measure between the two frames to achieve the key frame extraction.

69.2.2 Algorithm Thinking

The analysis found that for the two image frames within the same lens, they are very similar from the statistical sense; two images belong to a different lens, which is very small in similarity. Solutions starting from the sub-block, considering the sub-block in the middle position of each image frame in the video sequence depicts visual information of the scene core, compared with sub-block in the same image frame surrounding the location, and the sub-block information at the center position are more important [1]. As a result, the difference of the sub-block at the center position plays an especially important role in determining the difference between the two adjacent image frames and should be treated specially [2]. This article on the basis of literature introduces the DC coefficient of the weight difference, design, and proposes an improved key frame extraction algorithm based on the lens in the compressed domain [3]. According to the theoretical thinking, the sub-block difference of the image frame at the middle position gets more reference

value than the sub-block difference at the peripheral location and constructs weight difference of the DC coefficient schematic diagram (Fig. 69.1).

In Fig. 69.1, each sub-block corresponds to a DC coefficient; the depth degree of color shows the importance of sub-block information in a different location. The deeper the color shows that it is more important position in the current image frame, it is necessary to give a larger weight value of the corresponding DC coefficient in the similarity measure. Finally, the difference of thumbnail as a similarity measure of the two adjacent image frames is calculated as follows:

$$D(I_i, I_{i+1}) = \sum_{k=0}^{n} \left[H_i(k) - H_{i+1}(k)\right]^2 / \left[H_i(k) + H_{i+1}(k)\right]^2$$

Of which, $I_i\, I_{i+1}$ represents the first I and I +1 frame, respectively; $H_i\, H_{i+1}$ represents the I and I + 1 of the I-frame DC thumbnail histogram information. When D (Ii, Ii + 1) reaches a peak, and then identifies the two I-frames from a different lens, extraction the first frame of the lens as a key frame. The essence of the proposed algorithm can be vividly interpreted as the amplification of the sub-block difference in the image frame.

69.3 The Analysis of Experiments and Results

For the improved algorithm, the selection of the characteristic parameters and decision rules of determining the key frame is the key. Specifically speaking, how to select the so-called core area scope, how to determine the weight values of the DC coefficients in the range of the core region, and how to select frame difference threshold to extract the key frames. These issues directly affect the performance of the merits of the proposed algorithm.

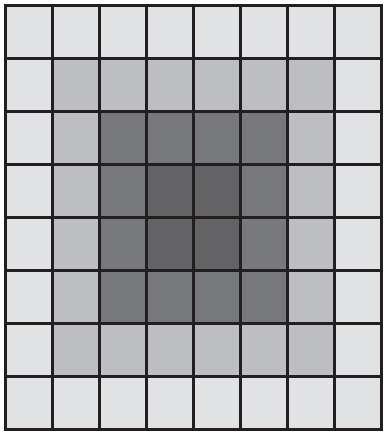

Fig. 69.1 The schematic diagram of 8 × 8 sub-block DC coefficient about weight difference

69.3.1 Experimental Performances

The recall and accuracy of the test model to measure the improved algorithm in this article were adopted. Retrieving the integrated query, the description, matching, and extraction processing have the possibility of success and failure. According to the principle of pattern recognition, you can get four conditions in Table 69.1, corresponding to the four basic parameters.

Using the basic parameters in Table 69.1 can define the commonly used recall and precision in order to characterize the retrieval performance. Defined as follows:

Recall rate = associated with the correct search results
All associated with the results = $[A/(A + C)] \times 100\ \%$
Precision = associated with the correct search results
All retrieved results = $[A/(A + B)] \times 100\ \%$

This paper selected the animation, film, advertising, science, education, and other video clips to test the effectiveness of the key frame extraction algorithm designed in this paper. The properties of the test video clips are shown in Table 69.2.

69.3.2 Analysis of Experimental Data

Followed by template matching method, Euclidean distance is divided into sub-block method, I-frame DC coefficient method; retrieval results of the improved algorithm for key frame extraction in compressed video sequences. The

Table 69.1 The basic parameter of expressed retrieval ability

Result of retrieval	Associated	No association
Retrieved	Retrieved correctly (A)	Retrieved wrong (B)
No retrieved	Missing retrieved (C)	Right refuse (D)

Table 69.2 Test the property of video fragment

Video source	Characteristic	Test frame	Key frame	Screen feature
Animation	Scene change fast more movement	625	24	No color 352*288
Film	Scene change slowly more movement	859	18	No color 352*288
Ad	Scene change fast more switch	527	16	Color 160*120
Science and education	Scene change slowly	385	6	No color 352*240

detection time represents T (unit: s), the number of key frames K (unit: frames), and video sequence length L (unit: frames).

Experimental results show that, after full decoding, key frame extraction results in pixel domain in the precision of this indicator are slightly better than the compressed domain methods. Rich and complete image information was obtained after full decoding, which sub-block partition method is the best.

But the large-scale decoding of the compressed file may result in longer detection time and less effective real-time detection. Although take compressed domain methods, such as I-frame DC coefficient method and the improved algorithm, it is better than the pixel domain methods in the detection time, and the partial decoded image information is limited, so it is slightly worse in the indicators of precision. From the experimental results, the compressed domain methods in the recall rate showed a good performance; unit time of the recall percentage is higher than the detection method of the pixel domain.

It can be seen from the data in Tables 69.3, 69.4, and 69.5; the improved algorithm in this article has improved retrieval time and the recall rate compared with the traditional division of the sub-block detection method. It increases the sensitivity of the image motion of the center of the lens position, making it

Table 69.3 Test result of animation (L = 530, K = 20)

Test method	The number of correlate images	The number of valid images	Recall rate percentage	All retrieved results percentage	Test time
Template matching	28	16	82.1	62.1	31″460
Euclidean distance	20	12	76.9	62.4	36″420
Sub-block divide	24	13	82.6	71.2	46″620
I-frame DC coefficient	18	12	76.5	65.4	25″830
My algorithm	25	15	86.6	70.8	26″250

Table 69.4 Test result of film (L = 820, K = 22)

Test method	The number of correlate images	The number of valid images	Recall rate percentage	All retrieved results percentage	Test time
Template matching	36	24	81.5	61.6	1′16″260
Euclidean distance	30	27	86.2	73.8	1′24″210
Sub-block divide	32	24	84.8	76.1	1′39″510
I-frame DC coefficient	25	25	80.7	68.4	56″280
My algorithm	28	26	96.2	72.5	59″650

Table 69.5 Test result of AD (L = 410, K = 20)

Test method	The number of correlate images	The number of valid images	Recall rate percentage	All retrieved results percentage	Test time
Template matching	27	16	84	61.6	18″25
Euclidean distance	25	15	79	62.7	23″27
Sub-block divide	23	16	84	71.3	27″73
I-frame DC coefficient	24	15	80	63.8	16″14
My algorithm	26	17	81	68.5	18″15

more suitable for more intense news documentaries, films, and other local sports video sequences; the key frame in the lens is not missing, but there will be a small amount of redundancy.

69.4 Conclusions

This paper put forward an improved key frame extraction algorithm based on lens in the compressed domain. Considering the encoding characteristics in the compressed video sequence domain, the proposed algorithm only use the DC component of the I-frame information and in accordance with the theory that sub-block difference of the image frame in the middle position is more valuable than the ones of the peripheral location, the proposed algorithm only use the DC component of the I-frame information. Experimental results show that the extracted key frames using the proposed algorithm can better reflect the contents of the video lens.

References

1. Shi LC (2011) Key frame extraction algorithm based on rough set in compressed domain. Comput Eng 12:190–198
2. Xiao-ge PU (2010) A survey on key techniques of content-based video retrieval. Info Sci 14:80–88
3. Wen H (2012) Content-based video forensics. Comput Sci 9:48–56