Key Frame Extraction Based on Improved Hierarchical Clustering Algorithm

Huayong Liu, Huifen Hao
Department of Computer Science, Central China Normal University
Hubei Wuhan, China
ayhaohuifen@163.com

Abstract—**Key frame greatly reduces the amount of data required in video indexing and provides a suitable abstract for video browsing and retrieval. Key frame extraction plays an important role in content-based video stream analysis, retrieval and inquiry. In order to extract key frame efficiently from different type of videos, in this paper we propose an improved hierarchical clustering algorithm that combing K-means algorithm. The improved hierarchical clustering algorithm is used to obtain an initial clustering result. And K-means is conducted to optimize the initial clustering result and obtain the final clustering result. Finally, the center frame of each clustering is extracted as key frame. Experimental results show that compared with other existing methods, the representations of key frame extracted by our algorithm are better in expressing the primary content of video.**

Keywords-key frame; feature extraction; hierarchical clustering; K-means algorithm

I. INTRODUCTION

With the rapid development of multimedia technology, a large amount of audios and videos are emerging on the Internet and sharing of multimedia videos data has become increasingly popular. Because of the complexity of videos, how to search the interested video quickly and effectively becomes a critical problem to be solved. Key frame is the very representative frame in the series of video frames, which can be used to describe the key image of a video shot. It reflects the main content of a video shot or even a whole video well and truly. In recent years, key frame extraction technology has attracted a wide spread attention by domestic and foreign scholars, and some key frame extraction methods are proposed.

The existing approaches can be categorized into four classes. (1) The approach based on video shot. This approach divides the video stream into several shots, and then it extracts the first frame, the middle frame and the last frame as key frames [1-3]. This kind of method has the advantage of simplicity and low computation complexity. However, in these methods, the disadvantages are 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 motion content in the video shot cannot be efficiently described. (2) The approach based on motion analysis. Wolf[4] calculates the movement of a shot by analyzing the optical flow, and selects the local minimum in the movement as key frames. This approach can express the motion of the video, but the calculation is expensive. Moreover, it does not pay enough attention to the content changes brought by the dynamic accumulation. Consequently, the robustness of this algorithm is not good. (3) The approach based on video content. This method extract key frames based on the change of color, texture and other visual information of each frame [5-6]. When the information changes significantly, the current frame is key frame. This method is very simple and can select corresponding key frames according to the change degree of content in the video shot. But its disadvantage is that the key frames extracted are not always the most representative meaning one and cannot indicate the changes of movement information quantitatively , thus which will cause unstable key frame extraction, for example, extract too much frame when meet frequent camera or mass of content motion. (4) The approach based on clustering. The video frame sequences are classified by clustering into several clusters, and then key frames are selected from every cluster [7-10]. Key frames extracted by this approach can reflect the content of video, so it has become the mainstream method for key frame extraction. However, these methods need to predefine the number of cluster before clustering [11], and the computing time is long. These greatly limit its further development.

Hierarchical clustering algorithm need not designate the number of cluster in advance. However, the speed of convergence is relatively slow. The K-means algorithm is simple, and the convergence speed is fast. But it is sensitive to the initial parameters and is easy to fall into local optimization.

In this paper, based on the methods mentioned above, a key frame extraction method is proposed based on improved hierarchical clustering algorithm. This method use the characteristics of image information entropy to measure the similarity degree of two frames, if the similarity reaches a certain value, they will be merged into the same cluster. Then we extract the clustering center as the key frames.

II. ALGORITHM OF KEY FRAME EXTRACTION BASED ON IMPROVED HIERARCHICAL CLUSTERING

Assuming that the video shot boundary detection has already been done, and some video shots have been distinguished. The algorithm of key frame detection is described as follows.

Starting from the video data, extract the feature of each frame, and calculate the inter-frame similarity by using the Euclidean distance formula.

Then conduct the hierarchical clustering algorithm to obtain an initial clustering result. Conduct the K-means algorithm to optimize initial clustering result.

Finally, output the key frames which are eligible. According to the proposed algorithm, a detection framework is described as Figure 1.

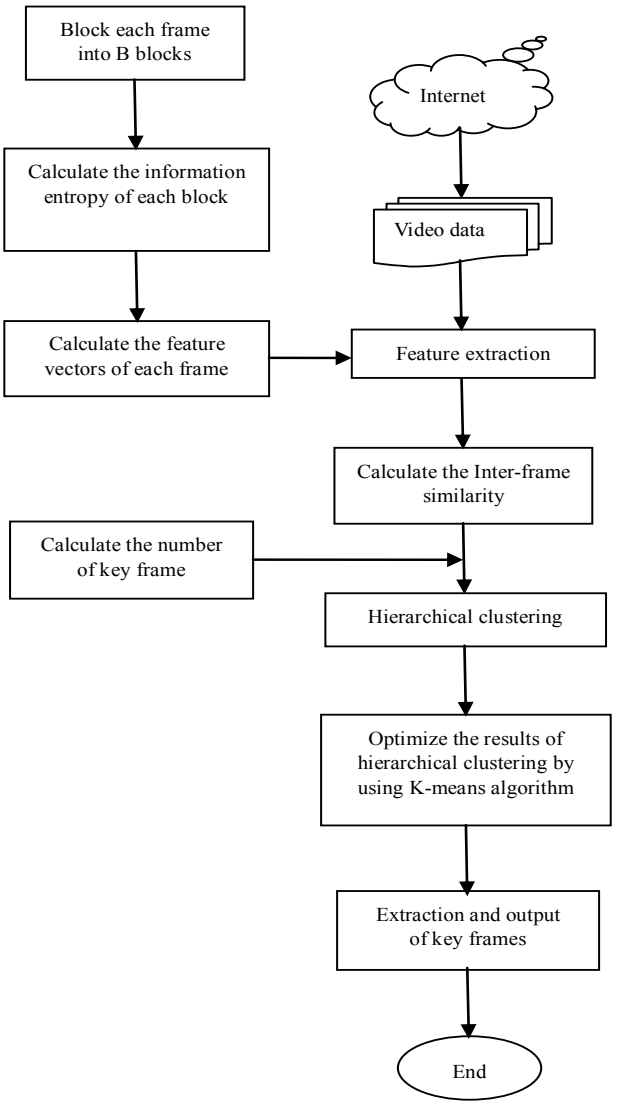

Figure 1. Detection framework of key frame extraction based on improved hierarchical clustering algorithm

A. Feature extraction

In 1948, American mathematician Shannon published his very famous thesis "mathematical theory of communication", which established a fairly comprehensive information theory. Image information entropy reflects the information contained in a certain image. If the information entropy of a certain image is larger, the image will contain more information [12-13]. Using the image information entropy for measuring similarity degree among images has already obtained great success in application.

Before execution of hierarchical clustering, we need to extract the information entropy of each image. Assuming that each image has been divided into B blocks, for the 256 grey-scale image, we can calculate the information entropy of each block, according to the formula (1):

$$H = -\sum_{i=0}^{255} p(x_i) \log_2 p(x_i) \tag{1}$$

Where i denotes the grey scale of each image, $X = \{x_i | i = 1,2,3,4,...,n\}$ denotes the pixel count on grey-scale of i, and $p(x_i)$ denotes the probability of x_i. Consequently the feature vector of each image can be defined as formula (2):

$$F = \{H_j | j = 1,2,...,B\} \tag{2}$$

Where H_j denotes the information entropy of block j-th, F denotes the feature vector of the image.

B. Algorithm Description

Step1. Calculate the Inter-frame similarity. Set each sample (frame) as a cluster, and then calculate the distance between every two clusters according to the Euclidean distance formula (3):

$$d(F_a, F_b) = \left(\sum_{r=0}^{N} (F_a[r] - F_b[r])^2 \right)^{1/2} \tag{3}$$

Where $N = \frac{n(n-1)}{2}$, n denotes the frame number of the video. The above F_a denotes the feature vector of image a, and $d(F_a, F_b)$ denotes the distance between image a and image b.

Step2. Assuming that the numbers of key frames are K, we can calculate the average M and variance V of distance vector according to the formula (4) and (5):

$$M = \frac{1}{N} \sum_{i=0}^{N-1} D_i ; \tag{4}$$

$$V = \frac{1}{N} \sum_{i=0}^{N-1} (M - D_i)^2 ; \tag{5}$$

Where $D=\{d_i|i=1,2,...N\}$ denotes distance vector. In this paper, the value of K is equal to the number of frames whose distance $d(F_a,F_b)>M+2*\sqrt{V}$.

Step3. Merge the nearest two clusters into a new one, and then recalculate the distance between the new cluster and the old ones.

Step4. Repeat step 3, until the number of cluster is K. That is the termination conditions of hierarchical clustering algorithm, as well as the initial clustering center of k-means algorithm.

Step5. Calculate the mean value of each cluster. Assign each frame to the nearest cluster, according to the minimum distance. Recalculate the clustering center.

Step6. Repeat step 5, until the cluster objects do not change. Select the clustering center as key frames.

C. *Analysis of Algorithms*

In the process of key frame extraction, each image frame is divided into B blocks, and then we calculate the information entropy vector of each block respectively. Compared with the algorithm proposed by Angadi [13], the image information especially the features have been fully considered. Using this hierarchical clustering algorithm, we can easily obtain the initial clustering centers for k-means algorithm, which is more accurate than predefined parameter. Moreover, the convergence speed of k-means algorithm is fast.

III. EXPERIMENTAL RESULTS AND ANALYSIS

We have made a lot of experiments on videos which have different characteristics, the length of video sequences ranging from several hundred to several thousand. We use Matlab7.14 for data analysis and get better results. Five representative videos are selected to verify the effectiveness of the algorithm. In order to test the proposed algorithm, we use Precision Ratio (PR) and Recall Ratio (RR) which can be found in the following formulas.

$$PR=\frac{N_c}{N_c+N_f}*100\% \tag{6}$$

$$RR=\frac{N_c}{N_c+N_m}*100\% \tag{7}$$

Where N_c denotes the number of the correct key frames extracted with the proposed algorithm, N_f denotes the number of the false of key frames, and N_m denotes the number of missing key frames.

As the traditional Sequence Difference Histogram algorithm (SDIF) has been widely applied in the key frame extraction [14], and it has good detection performance and fast computing speed, so we applied the SDIF to these videos as comparison in this paper. In this paper, we use artificial method to extract key frames as standard.

TABLE I. RESULTS FOR USING SDIF

Video name	Total frames	Key frames	PR (%)	RR (%)
ad	1235	21	62.1	65.5
cartoon	720	27	68.6	70.4
movie	1087	31	63.4	64.1
news	1802	21	77.4	74.8

Table 1 shows the results that uses the SDIF. The PR and RR is less than 80%, and this shows that traditional method is not effective to extract the key frames entirely.

TABLE II. RESULTS FOR USING IMPROVED HIERARCHICAL

Video name	Total frames	Key frames	PR (%)	RR (%)
ad	1235	21	90.5	89.2
cartoon	720	27	89.7	88.9
movie	1087	31	86.5	87.8
news	1802	21	86.9	90.4

Table 2 shows the results that uses the Improved Hierarchical Clustering Algorithm (IHCA). As shown in the table, the IHCA improves the PR to 86%, and the RR to 87%, showing that the IHCA has more advantages.

In order to illustrate the efficiency of the method and the representative of the extracted key frame, we take a news video for example. The content of the experimental video changes not very quickly, but there is lens movement in it. The key frame sequence extracted from the video display in the Figure 1.


1.jpg


380.jpg


404.jpg


447.jpg

Figure 1. Key frame extracted from news video

From the above results, we found that the content of the last fourteen frames changes greatly, so they are selected as key frames. In order to reflect the change of the subtitle and the action of the announcer (raise her head and bow her head), the first four key frames are chosen, although the content has no big change.

Experimental results show that the extracted key frames which use the improved hierarchical clustering algorithm can describe the main video content effectively. At the same time, the precision ratio and recall ratio of the improved hierarchical clustering algorithm are better than other algorithms. What's more, the redundancy of this algorithm is relatively low.

IV. CONCLUSIONS

According to the shortcomings in the conventional key frame extraction process, in this paper, an optimization method of hierarchical clustering algorithm is proposed to make appropriate improvements, so that the final extracted key frames can be a better description of the video content. Experimental results show that the key frame extraction algorithm can summarize the video content effectively. However, the calculation of the image information entropy is a little expensive. And evaluation criteria of key frame extraction are not perfect. In the future, we will try to solve these issues.

ACKNOWLEDGMENTS

This work is financially supported by self-determined research funds of CCNU from the colleges' basic research and operation of MOE (No. CCNU10A01012).

REFERENCES

[1] Mendi, Engin, and Coskun Bayrak, "Shot boundary detection and key frame extraction using salient region detection and structural similarity," Proceedings of the 48th Annual Southeast Regional Conference, 2010: 66-67.

[2] Abd-Almageed, Wael, "Online, simultaneous shot boundary detection and key frame extraction for sports videos using rank tracing," Proceedings of International Conference on Image Processing (ICIP'08), 2008: 3200-3203.

[3] Feng, Huamin, et al, "A new general framework for shot boundary detection and key-frame extraction," Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, 2005: 121-126.

[4] Wolf, Wayne, "Key frame selection by motion analysis," IEEE International Conference on Acoustics, Speech, and Signal Processing, 1996, 2: 1228-1231.

[5] Sun, Zhonghua, Kebin Jia, and Hexin Chen, "Video key frame extraction based on spatial-temporal color distribution," Proceedings of the Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP'08), 2008: 196-199.

[6] Ding Hong-li, Chen Huai-xin, "Key frame extraction altorithm based on shot content change ratio," Computer Engineering, 2009, 13: 225-231.

[7] Y. Y. Zhu, D. R. Zhou, "An Approach of Key Frame Extraction Based on Video Clustering," Computer Engineering, 2004, 30(4): 12-14.

[8] Y. Yin, H. N. Jiang, "Key frame extraction based on clustering of optimizing initial centers," Computer Engineering and Applications, 2007, 43(21): 165-167.

[9] Ejaz N, Tariq T B, Baik S W, "Adaptive key frame extraction for video summarization using an aggregation mechanism," J. Journal of Visual Communication and Image Representation, 2012, 23(7): 1031-1040.

[10] Chan E Y, Ching W K, Ng M K, et al, "An optimization algorithm for clustering using weighted dissimilarity measures," J. Pattern recognition, 2004, 37(5): 943-952.

[11] Lo C C, Wang S J, "Video segmentation using a histogram-based fuzzy

c-means clustering algorithm," J. Computer Standards & Interfaces, 2001, 23(5): 429-438.

[12] Sun, Lina, and Yihua Zhou, "A key frame extraction method based on mutual information and image entropy," IEEE conference, 2011 International Conference on Multimedia Technology (ICMT'11), 2011: 35-38.

[13] Angadi, Shanmukhappa, and Vilas Naik, "Entropy Based Fuzzy C Means Clustering and Key Frame Extraction for Sports Video Summarization," IEEE conference, 2014 Fifth International Conference on Signal and Image Processing (ICSIP'14), 2014: 271-279.

[14] Lo C C, Wang S J, "A histogram-based moment-preserving clustering algorithm for video segmentation," J. Pattern recognition, 2003, 24(14): 2209-2218.