Send Orders for Reprints to reprints@benthamscience.ae

Open Access

Design and Achievement of Teaching Video Retrieval System Based on Key Frame

Huang Lanying*, Xiong Zenggang, Zhang Xuemin, Li Zhimin and Ye Conghuan

School of Computer and Information Science, Hubei Engineering University, Hubei, 432000, China

Abstract: The persistent development of large-scale teaching video highlights the insufficiency of the current teaching video retrieval. Now, a teaching video retrieval system based on key frame has been designed and achieved. The system uses mutual information extraction for the video key frame extraction. The feature of key frame uses colour information; the similarity judgment uses simplified histogram for computation. In case of inquiry and retrieval, such a video retrieval method as combines key word and key frame is applied. The system had been tested and verified, and the result proves that the system meets the application requirements.

Keywords: Color feature, key frame retrieval, similarity, teaching video.

1. INTRODUCTION

The rapid development in the field of multimedia and network technique promotes the roaring emergence of the large scale teaching video in colleges and universities. On the other hand, its management and application efficiency is always poor. The popular retrieval for teaching video resource is achieved with key word or text (*e.g.* such a retrieval as based on field including program title, program making time or shooting site, etc.). But, the teaching video which features not being structural and rich content makes the retrieval mode based on key word or text has been not suitable to the video resource retrieval. It would be an urgent issue to establish and organize the database for large scale teaching video resource for the purpose of quick and effective retrieval [1, 2].

The references [3-5] cited by the paper designs a teaching video retrieval system based on key frame. The system uses mutual information extraction for the video key frame extraction. The feature of key frame uses color information; the similarity judgment uses simplified histogram for computation. In case of inquiry and retrieval, such a video retrieval method as combines key word and key frame is applied.

2. GENERAL ARCHITECTURE OF TEACHING VIDEO RETRIEVAL SYSTEM

The teaching video retrieval system based on key frame introduced by the paper mainly consists of video key frame processing submodule, database, video retrieval sub module. Teaching Video Retrieval System is show in Fig. (**1**).

In the video key frame processing submodule, the video data is preprocessed with the techniques of shot detection, extraction of key frame and feature extraction and the establishment of the video feature index is achieved; the video retrieval submodule uses the method based on the combined application of key word and key frame to describe user retrieval. The main task of the submodule is completing the establishment of the video feature index and the similarity computation and feature matching for the characters submitted by user and then sequence the similarity computation result and deliver the result to user; the video feature database in the in the database mainly includes content feature (low-level feature) and video external information feature (high-level feature). Herein, the description of video external information (high-level feature) uses manual tagging and automatic tagging. The whole inquiry course of the system is that:

(1) User raises the inquiry conditions (*e.g.* legend or feature of key word and key frame).

(2) And then the system would use the conditions inputted by the user to launch similarity matching *via* feature in index or feature database.

(3) Original video database (or key frame database) with like feature would be positioned.

(4) Inquiry result would be present.

(5) This inquiry is completed or next one launched.

3. DESIGN AND ACHIEVEMENT OF MAIN PARTS OF SYSTEM

3.1. Extraction of Video Key Frame

A key frame means a key image frame which functions describing a shot. It embodies main content of a shot and is also the base to shape video index. The application of key frame could largely reduce the data volume of the teaching video index, speed up retrieval and transmission and save storage space [4, 8, 10]. In the course of extraction of key frame, the first operation is to apply the shot boundary detection technique to divide a video stream into various shots.

1874-110X/14 2014 Bentham Open

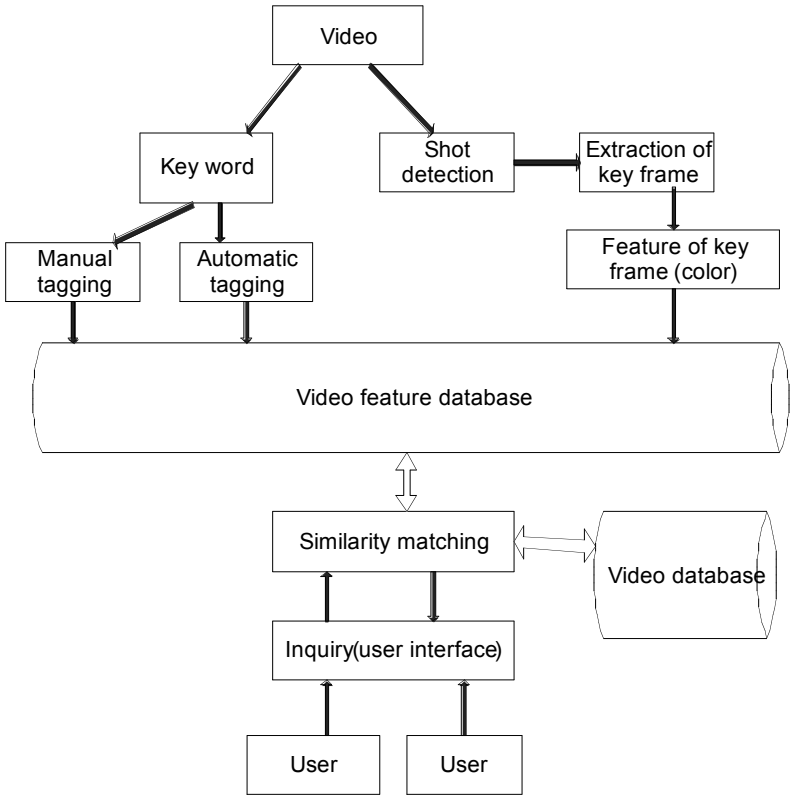

Fig. (1). Teaching video retrieval system.

Secondly, relevant algorithm is used to extract one or more key frame from each shot which would be crucial to describe shot content. In conclusion, the teaching video retrieval based on content feature has its core of the extraction of key frame which would directly decide the performance quality of the retrieval system [4, 8, 10]. The typical extraction method includes: (1) that based on shot boundary. The method features simple achievement because the quantity of the key frame is certain (one frame or two frames). But its effect is not stable [4, 8]. (2) That based on vision content. The method allows user to depend on change degree of shot content to select relevant quantity of key frame. However, the frame selected may present no representative. It means that selecting excessive key frame may be probable [4, 8]. (3) That based on motion analysis. The method allows user to depend on shot content to select relevant quantity of key frame. However, such method features hard difficulty and work for achievement [4, 8, 9]. (4) That based on clustering. When this method is applied, the quantity of the key frame is decided by categories which are also subject to threshold appointed. The selection of threshold has great influence on the result of the extraction of key frame. Moreover, the computation is also somewhat heavy frames [4, 7, 8].

Considering the characteristic of the above extraction of key frame, on the base of the references [6, 9, 10], the paper introduces a method for extraction of key frame based on mutual information, detailed as follows: The mutual information in consecutive two frames is used to measure the similarity between two images. Meanwhile, according to the characteristic of the shot sequence, the clustering method is applied to extract one or more key frame for each shot.

With given image A and B, mutual information is defined as:

$$I(A, B)=H(A)+H(B)-H(A,B) \quad (1)$$

From the Formula (1), it could be known that mutual information I(A, B) means the information volume shared by two frames. Among, H(A) and H(B) are respectively entropy of video A and B. H(A,B) is union entropy of both. When I(A, B)=0, it means A independent of B each other. Larger the I(A, B) value is, higher the similarity between two images would be. For two consecutive neighbouring frames, the mutual information of its three components RGB would be computed independently. As in Equation (1), the mutual information at the component R for neighbouring image frame could be expressed as follows:

$$I^{R}_{(t,t+1)} = \sum_{a=0}^{a=L-1} \sum_{b=0}^{b=L-1} p_{AB}(a,b) * \log \frac{p_{AB}(a,b)}{p_A(a) p_B(b)} \quad (2)$$

General mutual information between image t and image t+1 could be expressed as follows:

$$I_{(t,t+1)} = I^{R}_{(t,t+1)} + I^{G}_{(t,t+1)} + I^{B}_{(t,t+1)} \quad (3)$$

wherein, PA (a) and PB (b) respectively means probability density distribution in video frame A and B. PAB(a,b) means union probability distribution of video frame A and B.

The algorithm for the extraction of key frame detailed as follows:

1) To get mutual information $I_{(t,t+1)}$ of neighbouring frame in video;

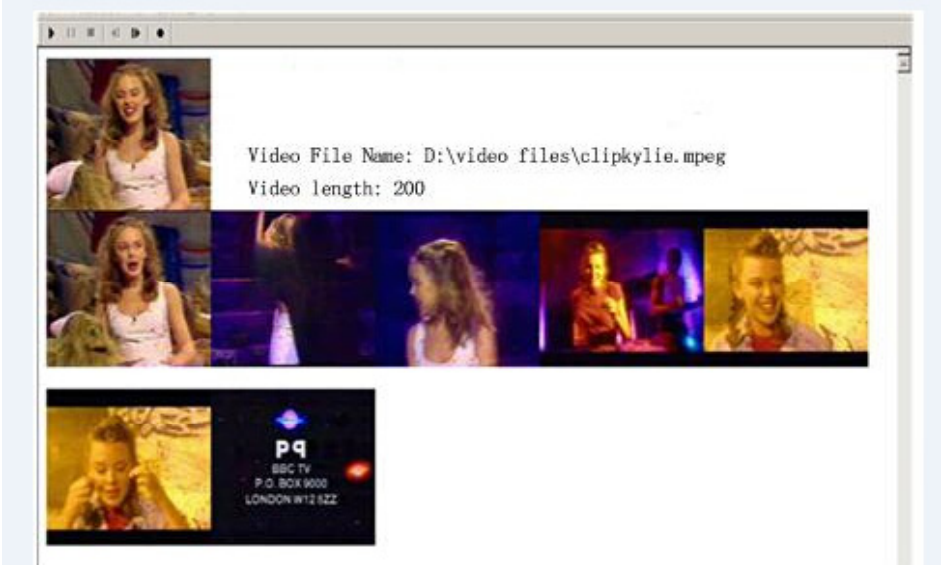

Fig. (2). Extraction of key frame in video clipkylie.

Table 1. Test result.

Video Sequence	Length/Frame	Shots	Key Frames Extracted	Visualized Key Frames
Feature film	4370	30	38	34
News film	204	12	18	13
Popular science film	2280	3	13	6
Cartoon film	5670	51	62	56

2) In the light of initial threshold θ1 define the category, *i.e.* if mutual information between neighbouring frame I (t, t+1)< θ1, new category defining begins;

3) Compute average mutual information of each category;

4) Merge the category in accordance with threshold θ2;

5) Select such a frame as has mutual information most close to average mutual information of the category as key frame candidate.

Refer to Fig. (**2**) for extraction result (θ2=0.4). Fig. (**2**) shows the result of the extraction of key frame by mutual information in the video clipkylie (foreign video retrieval study website http://www.videokeyframes.de)

The paper also introduces the test aiming at the performance of the extraction of key frame by mutual information (see Table **1**). The data in Table **1** comes from the experiment respectively aiming at feature film, news film, popular science film and cartoon film, etc. with video format of MPEG and AVI.

The test result in Table **1** verifies that the key frame extracted could better indicate teaching video by little data processing and computation. Its shortcoming is the selection of θ2 is subject to manual debugging.

3.2 Feature Extraction

The frame image feature of a general video includes two categories, one based on text (*e.g.* key word and note, etc.) and the other based on vision (*e.g.* color, texture, shape and object surface, etc.). The main extraction object introduced in the paper is the color and vision of the key frame. The feature extracted would be stored in the feature database for user inquiry and retrieval; in case of retrieval, the feature of the example video frame would be used to compare with the feature of the video in the database. Within set similarity, all the video which meets the requirement would be locked. The system uses color feature and RGB model. The statistic data of the color histogram could be achieved by Formula (4):

$$H(k) = nk/N \quad = 0,1,\ldots\ldots,L\text{-}1 \tag{4}$$

wherein, k means the value range of the feature of the frame image, L value range of color, nk the quantity of the pixel which is in the frame image with a feature value and N the sum of the frame image pixel. The color histogram achieved by the above computation is a feature vector of dimension *K*, which is the statistic result for the color value of all the pixels in the frame image [3, 4, 8].

3.3. Retrieval Based On Key Frame

When the video is abstracted into key frame, the inquiry of teaching video would become a retrieval which is launched in accordance with similarity of certain feature, to key frame or video, stored in video database, which is similar to inquiry description; in the video retrieval sub module, the system introduced in the paper uses a method which combines key word with key frame to describe the inquiry requirement of the user; the subsystem includes user inquiry model and similarity matching module. The retrieval based

on key word or text uses precise matching of key word or text; the retrieval based on key frame relies on computing and inquiring the similarity matching in terms of visual feature between key frame image (video abstract) and candidate image (video database) [3, 4].

Similarity matching, Euclidean distance function in Formula (5) is used to measure similarity between frame features [3, 4, 8].

$$d^2(x,y)=(x_1-y_1)^2+(x_2-y_2)^2+...+(x_n-y_n)^2=(x-y)^T(x-y) \quad (5)$$

where in, x and y mean image feature vector, and n means dimensions of feature vector.

The user inquiry module in the system could follow the inquiry conditions set by user to execute similarity computation based on video feature and feature provided by user and feature matching. Then, the system would sequence the result in accordance with degree of correlation and deliver the result to user. If a key frame is locked, user may use media player to watch the video clip that the key frame represents.

Fig. (**3**) is a key frame extracted after inquiry of sample video clip or shot. Fig. (**4**) shows post-search partial search result which all contains similar video (a screenshot from front four similar videos, the result shown sequenced down in terms of similarity, information including video similarity, video size and video format, etc.)

Fig. (3). Key frame of animal.

CONCLUSION

The paper designs a teaching video retrieval test system based on key frame. The system functions respectively by foreground and background program. The foreground program is developed respectively with ACCESS and ASP (Active Server Pages), the background program developed with

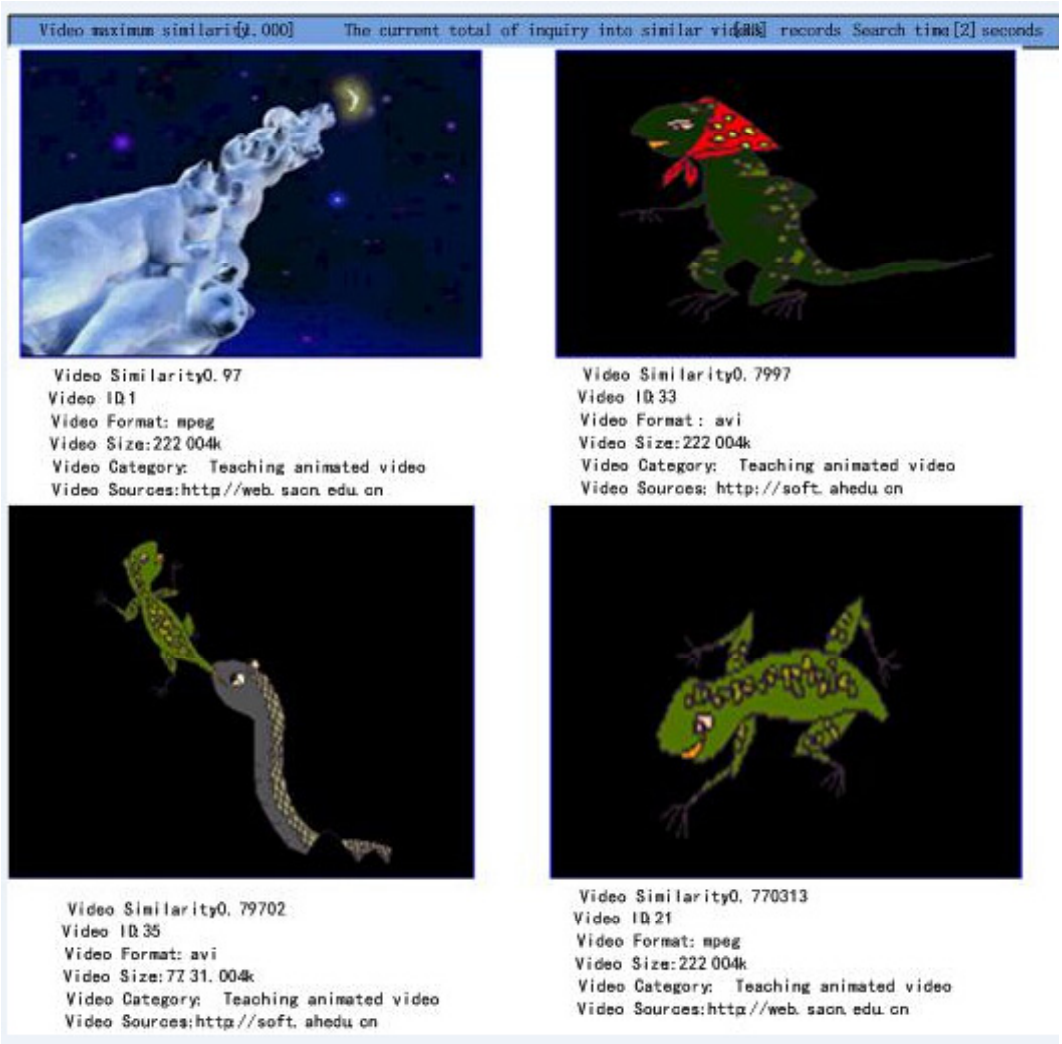

Fig. (4). Result of partial video retrieval based on key frame.

Visual C++. It functions making the video structured and extracting key frame and feature and saving both into database; when the content feature of video is extracted, the system could use some key frames in each video to generate a video summary which would be saved into database. On the retrieval result display interface, the relevant summary video would be opened to user for browse. The test result indicates that the key frame-based retrieval to the teaching video has a relatively right result which is competent to meet the application requirement. However, the shortcoming is objective, including that the test data are limited, the threshold where the key frame is extracted has not verified with massive data, it is impossible to express the dynamic detail of the video, and the system reliability is still subject to further test.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work is partly Supported by National Natural Science Foundation of China (No. 61370092), Natural Science Foundation of Hubei Province of China (No. 2013CFC005, No. 2014CFB188), Hubei Provincial Department of Education Outstanding Youth Scientific Innovation Team Support Foundation (No. T201410) and Hubei Provincial Department of Education Research Foundation (No. 20122604, No. 20122604, No. 2012367, No. XD2014255).

REFERENCES

[1] Y. Feng, "*Video Retrieval Techniques and Teaching Content-Based Applied Research*", Master's Thesis, Henan Normal University, 2008 (in Chinese)

[2] X. F. Meng, and X. Ci, "Big data management: concepts, techniques and challenges," *Journal of Computer Research and Development*, vol. 50, no. 1, pp. 146-169, 2013. (in Chinese)

[3] P. Muneesawang, and L. Guan, "iARM-an interactive video retrieval system," In: *Proceedings of the IEEE International Conference on Multimedia and Expo*, Taibei, Taiwan, 2004, no. 1, pp. 285-288.

[4] C. Ji, "*Video Retrieval Research Keyframe Extraction Method*", Master's thesis, Nanjing University, 2005 (in Chinese)

[5] V. Lavrenko, S. L. Feng and R. Manmatha, "Statistical models for automatic video annotation and retrieval", In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Quebec, Canada, 2004, no. 3, pp. 1044-1047.

[6] C. Chan, Y. Weiying, and Z. Zhang, "Mutual information based keyframe extraction method global optimization", *TV Technology*, vol. 7, pp. 6-28, 2011. (in Chinese)

[7] X. Pan, and H. Wang, "A key frame extraction method based on kernel clustering", *Computer Development*, vol. 3, pp. 29-31, 2005. (in Chinese)

[8] H. Yang, and Z. Ming, "Content-based video key frame extraction technology research", *Journal of Fujian Compute*r, vol. 5, pp. 49-52, 2010. (in Chinese)

[9] L. Huang, "Key frame extraction method based on mutual information", *Journal of Hubei University*, vol. 3, pp. 54-56, 2008. (in Chinese)

[10] L. Huang, and C. Ye, "Application of mutual information key frame extraction method in the teaching video retrieval", In: *Proceedings of International Conference on Multimedia Technology*, Hangzhou, China, 2011, no. 1, pp. 1120-1123.

Received: September 22, 2014 Revised: November 30, 2014 Accepted: December 02, 2014

© Lanying *et al.*; Licensee *Bentham Open.*

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/-licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.