韓國 컴퓨터情報學會 論文誌
第11卷 第4號, 2006. 9.

2006-11-4-1-17

텔레매틱스에서 효율적인 장면전환 검출기법을 이용한 비디오 브라우징

신 성 윤*, 표 성 배**

Video Browsing Using An Efficient Scene Change Detection in Telematics

Shin Seong Yoon*, Pyo Seong Bae**

요 약

다중 비디오 프레임들에서 컬러 특징의 효과적이고 효율적인 표현은 시각적 정보관리 시스템에서 매우 중요하다. 본 논문에서는 웹상에서 실시간 사용자 인터페이스에 의해 비디오 내용검색과 비디오 브라우징을 모두 수행하는 비디오 브라우징 서비스(VBS)를 제시한다. 비디오 시퀀스의 장면전환과 키 프레임 추출을 위하여 RGB 컬러 히스토그램과 X2(카이 스퀘어) 히스토그램을 합성한 효율적인 장면전환검출 기법을 제시한다. 장면전환검출에 의해 생성된 키 프레임들은 물리적 그리고 논리적으로 색인화 된다. 본 시스템은 VCR의 비디오 편집과 검색 기능을 포함한다. 비디오 브라우징을 위해서 날자, 분야, 그리고 주제의 세 가지 요소가 이용된다. 비디오 브라우징 서비스는 Apache 웹 서버에서 MySQL, PHP, JMF로 구현되었다.

Abstract

Effective and efficient representation of color features of multiple video frames is an important yet challenging task for visual information management systems. This paper proposes a Video Browsing Service(VBS) that provides both the video content retrieval and the video browsing by the real-time user interface on Web. For the scene segmentation and key frame extraction of video sequence, we proposes an efficient scene change detection method that combine the RGB color histogram with the X2(Chi Square) histogram. Resulting key frames are linked by both physical and logical indexing. This system involves the video editing and retrieval function of a VCR's. Three elements that are the date, the field and the subject are used for video browsing. A Video Browsing Service is implemented with MySQL, PHP and JMF under Apache Web Server.

▸ Keyword : 장면전환검출(Scene Change Detection), 비디오 브라우징(Video Browsing), χ^2 히스토그램(χ^2 Histogram), 비디오 인덱싱(Video Indexing)

• 제1저자 : 신성윤
• 접수일 : 2006.09.11, 심사일 : 2006.09.15, 심사완료일 : 2006.09.20
* 군산대학교 컴퓨터정보과학과 교수 ** 인덕대학 소프트웨어개발과 교수
※ This work was supported in part by a grant from the Regional Innovation System of the Ministry of Commerce, Industry and Energy of Korea.

Ⅰ. 서 론

With rapid increase in digital video, designing systems for efficient browsing and content based retrieval become crucial[1]. In recent years, the enhancement of data compression and computer network technology has enabled the authorization of a large amount of digital video content. However, a large amount of digital video that consisting of a great many frames is not easy to handle. We must have an attractive means to handle our large amount of digital video. It is often difficult to find both the appropriate video file and the portion of the video that is of interest[2]. The users may just want to see some interesting sequence of frames when they watch a video.

In general, for the efficient handling of video, such as retrieval, editing and browsing of video contents, a video stream is divided into a number of scenes or shots.

Scene change detection method divides the original video into a number of scenes or shots, and extract the key frames[3,4]. Key frames can be used to distinguish videos from each other, summarize videos, and provide access points into them[4]. Key frames provide a suitable abstraction and framework for video indexing, browsing and retrieval[5]. A key frame is representative of a scene or a shot, so some researchers call it representative frame[6] as well. However, it is difficult to determine the best representative frame that present the best information of a scene or a shot.

As a result of the great development in video compression and transmission, the use of digital video in multimedia systems and on the Internet is becoming universal. In order to make good use of the valuable multimedia resource, there is a need for content-based indexing, retrieval and browsing of video.

This paper proposes a Video Browsing Service(VBS) based-on scene change detection. A new scene change detection method that combines the RGB color histogram with the X^2(Chi Square) histogram proposed. Extracted key frames are indexed both physically and logically. Video retrieval system involves the function of a VCR(Video Cassette Recorder). Video browsing consists of three functions; retrieval, browsing and playing of the video scenes.

Ⅱ. Related Work

In video retrieval and browsing, key frame extraction is a very important and basic technique. Many researchers have developed key frame extraction methods for efficient video handling and management.

Shot boundary detection segments the video stream into shots, then the first frame of each shot is the shot's key frame[5]. There are some of the significant shot boundary detection techniques, such as the pixel differences, the statistical scene change detection, the compression differences, and the edge tracking etc. Visual contents of the video stream are used for key frame extraction based on shot detection. These visual contents are color and motion features for content-based video browsing[7]. In color feature based criteria, the first frame will be selected as the first key frame. As a reference frame, the distance between this frame and current frames are compared. If the distance exceeds a threshold, the current frame is used as a new key frame. In motion feature criteria, the first and last frame will be selected in a zooming,

and frames depending on overlapped scale in panning are extracted as key frames. Wolf has proposed the key frame extraction method based on motion analysis[8]. This method analyzes the optic flow and the motion metric function. The motion metric is used as a function of time to select key frames at the local minima of motion. Audio and image are used for key frame selection[3]. Key frames are selected using the keywords of audio, the image information and the camera motion. However, this technique is difficult to implement automatically, since processing area is widely spread. Other key frame extraction methods are following. There are frame difference[9, 10], color histogram method[11,12], video object segmentation and tracking[13,14], edge distance[15] and wavelet measurement[16]. Recently, key frame is extracted by structural features of caption[17], video segmentation using the automated threshold decision[18], location-, size-, and color-based key frame extraction[19].

In this paper, proposes a Video Browsing Service (VBS) based on efficient scene change detection method using both the color information and the X^2 histogram difference. 〈Fig. 1〉 shows VBS overview.

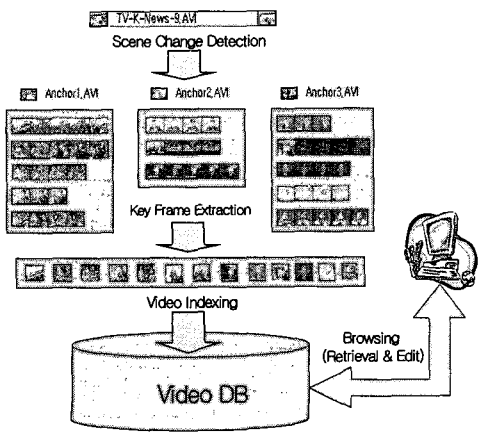


Fig. 1 VBS Overview

III. Scene Change Detection

3.1 Color Histogram and χ^2 Histogram

This section show scene change detection(or shot detection) methods that have used generally. Among the many methods, we will review the color histogram and X^2 histogram that have used by many researchers of multimedia modeling, since this method were of the basis for many histogram differences.

In general, color histogram difference is widely used for the detection of the abrupt scene change [11,12]. Given a frame Ii and a subsequent frame Ii+1, Hi(k) is the k-th histogram value of the i-th frame, and histogram difference can be described as following:

$$d(I_i, I_{i+1}) = \sum_{k=0}^{N-1} |H_i(k) - H_{i+1}(k)| \quad \cdots\cdots (1)$$

In Equation (1), if we use 256-color level, 256 histogram values of frame Ii and Ii+1 and differences of histogram values for each bin are computed. If the sum of all differences exceeds a threshold, scene change is detected. This method is simple, but sensitive to changes of motion and brightness.

And also, color histogram difference can be computed by individually separation of RGB color space. This can be described as following:

$$d_{RGB}(I_i, I_j) = \frac{1}{3}\sum_{k=1}^{n}\left(|H_i^r(k) - H_j^r(k)| + |H_i^g(k) - H_j^g(k)| + |H_i^b(k) - H_j^b(k)|\right) \quad \cdots\cdots (2)$$

Because each color histogram value is computed individually, this method is more complex than the method of Equation (1), but more flexible than it.

A X^2 Histogram gives considerably better results than the color histogram and template matching. Therefore, many researchers apply this method to his study and projects. The X^2 histogram difference can be described as following, where Hi(k) means the k-th bin value of the i-th histogram :

$$d(I_i, I_j) = \sum_{k=1}^{n} \frac{(H_{i(k)} - H_{j(k)})^2}{H_{j(k)}} \quad \cdots\cdots (3)$$

Above equation can be transformed as following:

$$d(I_i, I_j) = \sum_{k=1}^{n} \sqrt{H_{i(k)}^{\ 2} - H_{j(k)}^{\ 2}} \quad \cdots\cdots (4)$$

This method is tolerant to local motion of object and more efficient than other methods of feature extraction. However, this method is also sensitive to brightness.

3.2 A Proposed Histogram

This paper proposes a new scene change detection method that combines the RGB color histogram with the X^2(Chi Square) histogram. Histogram difference will be computed by converting the RGB color space into the YIQ space. This can be described as following:

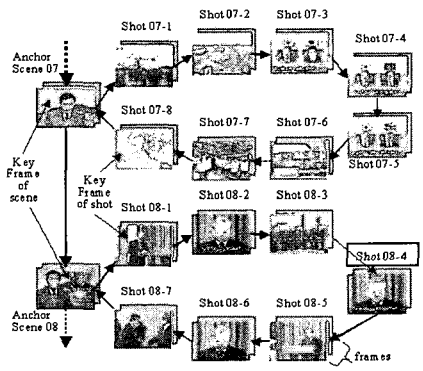
$$d(I_i, I_j) = \sum_{k=1}^{n} \left(\frac{(H_i^r(k) - H_j^r(k))^2}{H_i^r(k)} \times 0.299 + \frac{(H_i^g(k) - H_j^g(k))^2}{H_i^g(k)} \times 0.587 + \frac{(H_i^b(k) - H_j^b(k))^2}{H_i^b(k)} \times 0.114 \right) \Big/ 3 \quad \cdots\cdots (5)$$

This method is more efficient and flexible than other methods, since it is tolerant to the motion of camera and objects, and good for detection of abrupt and gradual scene change.

As a result of this scene change detection method, key frames are extracted at the point when the great frame difference occurred. That is, when the histogram difference exceeds a given threshold, a first frame at just point selected as a new key frame and these key frames are key frames of shots.

3.3 Scene and Shot Boundary

Scene change detection extracts key frames and separates shots from video streams. That is. These key frames are key frames of shots, and shot is consisting of subsequent frames. In news video, an anchor frame is representative of an item of news. Therefore, each anchor frame is the key frame of each news scene that consists of several shots. 〈Fig. 2〉 shows this scene and shot structure.



〈Fig. 2〉 Scene and shot structure

Consequently, the key frame extracted by scene change detection is the key frame of the shot, and shot consist of subsequent frames between current and next key frame. Anchor frame is the key frame of the scene, and scene

consist of subsequent shots between current anchor shot and next anchor shot.

III. Video Browsing Service

Key frame indexing is a fundamental process for the retrieval and browsing of video. Key frame indexing has performed by both physically and logically as 〈Fig. 3〉.

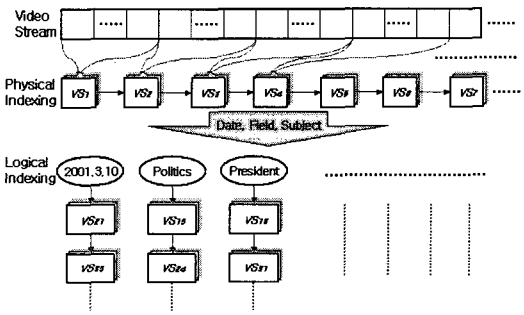


Fig. 3 Physical and logical indexing

Physical indexing links together key frames by the temporal sequence of video scenes(VSi). And logical indexing links together key frames by the date, the field, or the subject, for the efficient retrieval and browsing of the user's interest.

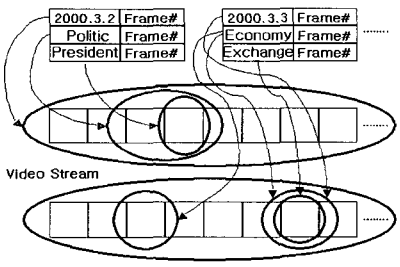


Fig. 4 Logical indexing for retrieval

〈Fig.4〉 shows detailed logical indexing structure for retrieval. Therefore, users can select the required date, field and/or subject, then the proper scene is accessed real-timely and user can view it immediately.

Most users may want to use fast forward and rewind function as in the VCR's. When users will play and watch a video, they may just want to skip over some uninteresting sequence of frames. Therefore, video editing system involves the function of a VCR and provides the video editing services, such as the frame segmentation and merge, the concatenation of frames and scenes, and the deletion of frames and scenes.

In Web, the Video Browsing Service consists of three parts; they are retrieval, browsing and playing of the video scenes. For the retrieval of the video scene, the three retrieval elements that are the date, the field and the subject are used. Each of three elements is used separately or compositively for the retrieval. In the browsing of the video scene, retrieved scenes are viewed and listed. Also, the information of the selected scene and frame is described in detail. The video playing part can play the retrieved or the selected video scene and present all information of the currently playing video scene.

IV. Experimental Result

A study has been carried out using three different News videos of three different TV Broadcasting Stations. Three Broadcasting Stations are KBS, MBC and SBS. In the experiment, the original video streams are AVI format and exactly 300 second long. And video streams are captured by 5 frames per second, then frames are normalized on size of 400 x 300 pixel.

〈Fig. 5〉 shows key frame extraction and editing system. When a captured input video is selected and threshold is given from user, key

frames are extracted. Among the extracted key frames, some of frames that is unnecessary are can be deleted.

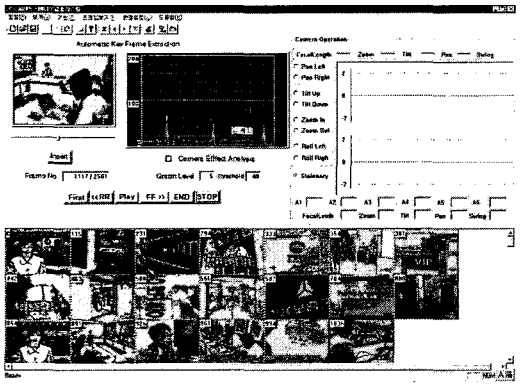


Fig. 5 Example of key frames extraction & editing

Physical indexing is a temporal sequence itself of key frames. The edited key frames are manually indexed on dates, fields and subjects. This is a logical indexing.

〈Table 1〉 shows the results of key frame extraction using three scene change detection methods, where TKF is a total number of extracted key frames and EKF is the number of error key frames. As showing in the results, a proposed method that combined the RGB color histogram with the X^2(Chi Square) histogram is more efficient and optimal, since both the total number of key frames and the number of error key frames are less than other methods.

Table 1. The results of key frame extraction

TV	Scene Detection Method	TKF	EKF
KBS	Color Histogram	46	15
	x^2 Histogram	44	13
	A Proposed Method	38	7
MBC	Color Histogram	46	17
	x^2 Histogram	42	13
	A Proposed Method	34	5
SBS	Color Histogram	58	25
	x^2 Histogram	53	20
	A Proposed Method	40	7

〈Fig. 6〉 shows example of video retrieval. Video scene has been retrieved according to date, field and subject. The area of retrieved video scene can be increased or decreased by user's manipulation. Users can select their interested issue and play and watch the selected video scene.

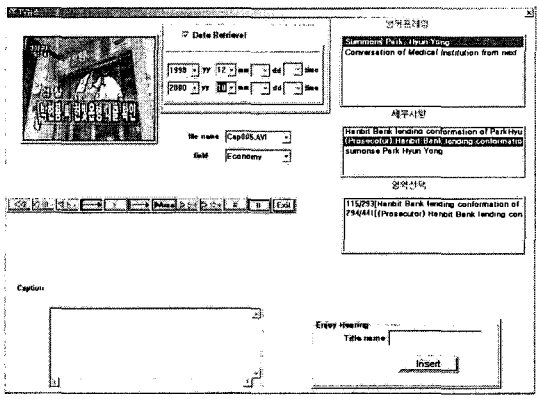


Fig. 6 Example of video retrieval

A video browsing service has implemented with MySQL, PHP and JMF under Apache Web Server. We have developed the medium size of database in MySQL and the database access interface in PHP. And also, we have developed video playing service in JMF (Java Media Framework)

〈Fig.7〉 show video retrieval and list of retrieved video scenes on Web. Date and field are selectable and subject can be retrieved by keyword. The results of this are listed below according to the access frequency. Users can select one of listed results, and then selected item will be browsed as in 〈Fig. 8〉.

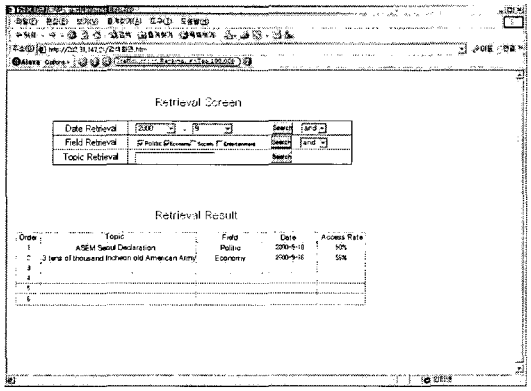


Fig. 7 Example of video retrieval on Web

〈Fig. 8〉 shows video browsing and brief information. The suitable video scenes for selected item of lists are browsed on Web. When users select one of browsed scenes, the brief information of scene is presented.

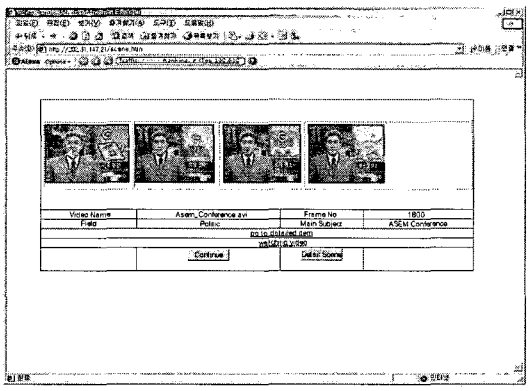


Fig. 8 Example of video browsing on Web

〈Fig. 9〉 shows video playing and detailed information. In 〈Fig. 8〉, if users select the detailed item, detailed information of the scene will be presented, and if users select the video watch, video scene will be played and user can watching the video, and detailed information will be presented too.

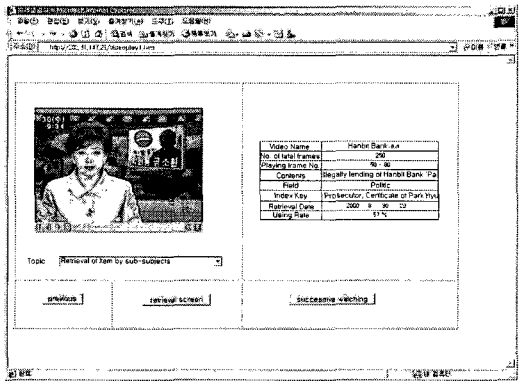


Fig. 9 Example of video playing on Web

V. Conclusion

We proposed a Video Browsing Service(VBS) that provided both the video content retrieval and the video browsing by the real-time user interface on Web. We proposed a new scene change detection algorithm that combined the RGB color histogram with the X^2(Chi Square) histogram. This new method gave better result than the result from individually using of the color histogram or the X^2 histogram. Key frame indexing has been performed by both physically and logically. Also, this system provided the function of a VCR's and the video editing services. On Web, the remote video browsing consisted of three parts: retrieval, browsing and playing of the video scenes. In the future, more efficient key frame extraction method based on semantic features should be developed. And semantic feature-based video retrieval and browsing framework on Web should be more enhanced than current many simple methods.

참고문헌

[1] M.R. Naphade, R. Mehrotra, A.M. Ferman, J. Warnick, T.S. Huang and A.M. Tekalp, "A high

performance algorithm for shot boundary detection using multiple cues,." In IEEE ICIP, vol. 1, p. 884-887, 1998.

[2] Andreas Girgensohn, John S. Boreczky.: Time-Constrained Keyframe Selection Technique, Multimedia Tools and Applications, 11(3), pp. 347-358, 2000

[3] Smith M. A & Kanade T.: Video Skimming for Quick Browsing based on Audio and Image Characterization, TR No. CMU-CS-95-186, School of Computer Science, Carnegie Mellon University ,1995

[4] Zhang H. J., Low C. Y., Smoliar S. W.: Video Parsing and Browsing Using Compressed Data, Multimedia Tools and Applications 1, pp. 89-111, 1995

[5] Boreczky J. S., Rowe L. A.: Comparison of video shot boundary detection techniques, Storage and Retrieval for Image and Video Databases (SPIE), pp. 170-179, 1996

[6] Zhang H. J., Kankanhalli A. & Smoliar S. W.:Automatic partitioning of full-motion video, Multimedia Systems 1, pp. 10-28, 1993

[7] Zhang H. J., Wu J. H., Zhong D., Smoliar S. W.: An integrated system for content-based video retrieval and browsing, Pattern Recognition, Vol. 30, No. 4, pp. 643-658, 1997

[8] Wolf W.:Key frame selection by motion analysis, In Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc, 1996

[9] Hampapur A., Jain R., Weymouth T.: Digital video indexing in multimedia systems, In Proc. Of AAAI-94 Workshop on Indexing and Reuse in Multimedia systems, 1994

[10] Hampapur A., Jain R., Weymouth T.: Production model based digital video segmentation, Multimedia Tools and Apps., Vol. 1, No. 1, pp. 9-46, 1995

[11] Ardizzone ECascia, M. L.: Automatic Video Database Indexing and Retrieval,Multimedia Tools and Applications, Vol. 4, No. 1, pp. 29-56, 1997

[12] Furht B., Smoliar S. W., Zhang H. J.: Video and Image Processing in Multimedia System, Kluwer Academic Publishers, pp. 335-356, 1995

[13] Zhang H. J., Wang J. Y. A., Altunbasak Y.: Retrieval and Compression : A Unified Solution, Proc. ICIP '97, Int. Conf. on Image Processing, pp. I13-16, 1997

[14] Zhong D., Chang S. F.: Spatio-Temporal Video Search Using the Object Based Representation, Proc. ICIP '97, Int. Conf. on Image Processing, pp. I21-24, 1997

[15] Zabih R., Miller J., Mai K.: Feature-based Algorithms for Detecting and Classifying Scene Breaks, Proc. ACM Int. Conf. on Multimedia, pp. 189-200, 1995

[16] Armen F., Hsu A., Chiu M. Y.: Feature Management for Large Video Databases. In Storage and Retrieval for Image and Video Databases, Proc. SPIE, 1908, pp. 2-12, 1993

[17] S. S. Shin, I. K. Kang and Y. W. Rhee "Creation of Soccer Video Highlights Using Caption Information," Journal of the Korea Society of Computer and Information, Vol. 10, No. 5, pp. 65-76, 2005

[18] K. C. Ko and Y. W. Rhee, "Video Segmentation using the Automated Threshold Decision Algorithm," Journal of the Korea Society of Computer and Information, Vol. 10, No. 6, pp. 65-74, 2005

[19] S. S. Shin, I. K. Kang and Y. W. Rhee, "Implementation of Web Based Video Learning Evaluation System Using User Profiles," Journal of the Korea Society of Computer and Information, Vol. 10, No. 6, pp. 137-152, 2005

저 자 소 개

신 성 윤

2003년 2월 군산대학교
컴퓨터과학과 이학박사
2006년~현재 군산대학교
컴퓨터정보과학과 전임강사
〈관심분야〉 비디오 인덱싱, 비디오 요약, 멀티미디어

표 성 배

2004년 2월 숭실대학교
컴퓨터학부 공학박사
1992년~현재 인덕대학
소프트웨어개발과 교수
〈관심분야〉 멀티미디어, 비디오 색인화, 비디오 요약