A MULTI-LAYER VIDEO BROWSING SYSTEM

Herng-Yow Chen and Ja-Ling Wu
Communication & Multimedia Lab.
Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan, R.O.C.

Abstract

This paper proposes an effective and efficient browsing system which is capable of representing the contents of a video sequence in a multi-layer manner. In this system, users can rapidly preview some pre-processed video shots, based on some simple scene change detection algorithm, to find out the particular video segments which are of interest. Because users just preview some video shots instead of examining all the video frames, this shot-based browsing method is superior to the conventional frame-based browsing one from the view points of the network bandwidth, user friendliness, and/or browsing efficiency. Based on the proposed browsing tool, a Video-on-Demand (VOD) prototype system has been implemented in the Communication and Multimedia Laboratory of National Taiwan University.

I. Introduction

Digital video has become the most important medium for multimedia computing and communication, with wide spread applications such as broadcasting, education, publishing, and entertainment. With the expansion of options on low cost mass storage medium, improved compression techniques, less costly and sophisticated input modules, and higher network bandwidth, the availability and the demanding of digital video are rapidly increasing. Many institutions researchers are developing Video-on-Demand (VOD) systems that will allow user to play digital movies on line. These systems are designed to deliver digital movies to a set-top box connected to a TV or multimedia computer in a home. System developers usually have focused on the server architecture and file system performance or network bandwidth problems, and on various resource scheduling problems of servicing many concurrent customers [1-7]. They have not pay enough attention to the problems of how to find a movie efficiently from a large collection of digital videos (movies). In fact, a significant issue arises in any large video library is how to efficiently locate a desired set of video sequences, this explains why content-based video query for large video database is now a hot research topic. Moreover, the facility of program preview should be one of the primary functions in the prospective Video-on-Demand or digital video rental store services, and the services should charge the user based on the rule of : the longer preview time you take the more money you should pay. So, the ability of allowing users to preview the outline of video, such as a set of video shots, efficiently in advance (so as to rapidly choose the most relevant or interested video segments) is highly desirable.

This paper describes the design and implementation of a multi-layer video browsing tool/system for assisting the users to find out the interested video segments easily. In this system, users need just to preview some video shots instead of examining every video frames; therefore, the proposed shot-based browsing method is superior to the conventional frame-based browsing method from the view points of the network bandwidth, user friendliness, and/or browsing efficiency. Based on the proposed browsing tool, a prototype VOD system has been implemented on Ethernet LAN environment.

The remainder of this paper is structured as follows. Firstly, we review the most primary key technique for video segmentation process in automatic

Manuscript received June 12, 1995 0098 3063/95 $04.00 © 1995 IEEE

video parsing or content-based video query research area, the so-called "scene change detection", and then describe the algorithm we adopted. Secondly, the basic idea and the architecture of the proposed multi-layer video browsing system is presented and discussed in detail. Thirdly, a prototype VOD system will be described in which the proposed multi-layer video browsing tool is used to assist the remote customer in the client sites to rapidly find the video material of interest.

II. Video Parsing or Video partition

In VOD or on-line video rental stores services, service provider must maintain a large video database and support query interface for purchasing provided services. Besides, from a user friendliness view point, service system should also provide an efficient and effective browsing interface for remote customers to grasp the outline of selected material as quick as possible. Nevertheless, traditional keyword-based query methods and frame-based browsing techniques have many drawbacks [8], the main reason is that text descriptions of video for keyword-based query can be both inadequate and inaccurate, and then frame-based browsing support just fast-forward and/or fast-backward playbacks. is really deficient specially for a long video.

To solve these problems, many researchers have been dedicated to the subject of automatic *"video parsing"* or *"video partition"* that parses a video sequence into some structured units, such as *video shots* or *clips*, which are more appropriate for further retrieval and index processing. A common technique for the first task in video parsing is *"scene change detection"* or *"cut detection"*. A scene change or cut in a video sequences can be defined as a discontinuity between two adjacent video shots. We consider a video shot, composing of a sequence of continuous frames without *explicit* scene change, as a basic indexing unit for further retrieving. Comparing to the explicit scene change, however, some video special effects such as fade-in, fade-out, cross-fade and dissolve, may also be applied on two distinct video shots for providing a continuous transition in visual perception. This kind of scene change is *implicit* and ambiguous but should be detected by an intelligent scene change detection algorithm.

There are a lot of algorithms have been proposed to perform scene change detection [9-17]. A scene change implies that, at the change instance, a significant change of continuous physical quantities (such as luminance and chrominance) occurs between two consecutive frames on either side of a shot. Therefore, calculation of the "quantitative difference measure" of a pair of frames (usually two adjacent frames are performed), and if the difference exceeds a given threshold, it is considered that a scene change occurs. However, different algorithms, adopt different metrics and/or distance measures for quantitative difference, produce different performance. Hence, finding a suitable metric for video scene change detection is the most important issue for automatic video parsing. From view points of the data processing, scene change detection algorithm can be divided into two categories : one is in the spatial domain and the other is in the transform domain.

For spatial domain scene change detection techniques, two common mechanisms are usually used. The first one is based on comparing corresponding pixels or block regions on the pair of frames, the second one is based on the derived histogram statistics, such as intensity or color histogram [8][10]. These two types of algorithms all require that each video frame be represented as an array of pixels. If the input video is in a compressed form, the compressed input source must be decompressed first and then apply the detection algorithm. On the other hand, for transform domain scene change detection, all computations can be worked directly on the compressed representations. Arman et al. [15] have proposed DCT-based scene change detection algorithm for Motion-JPEG based video, which is based on the fact that the DCT coefficients are mathematically related to the spatial domain and represent the content variations of the videos. Zhang et al. [17] have proposed a hybrid algorithm for transform domain detection, specially for MPEG coded video, this algorithm is based on motion analysis and combined with Arman's algorithm to overcome some scene change false positives introduced by traditional scene change detection algorithm under the circumstances of gradual transition (such as fade-in-out), camera operation (such as camera panning) and fast object motions.

Notice that the browsing function of the proposed system is independent of which scene change detection algorithm is adopted, in other words, the better the scene change detection algorithm is adopted, the better the system can perform.

In the proposed browsing system, the compressed domain (DCT-based) scene change detection technique presented in [15] is adopted because of its high processing speed. It takes advantage of the information already encoded in the compressed data, specifically in the DCT-based compression standards such as JPEG [18] and MPEG [19], so as to save the overhead for computing the inverse discrete cosine transform (IDCT) (which must be considered and will play a dominant role in the scene change detection processes on the spatial-domain).

III. The multi-layer video browsing system

The proposed system is composed of two components : a video partition component (VPC) and a video browsing component (VBC). VPC is usually working at the video providing part, i.e. in the server end, and will partition the input video sequence into shots by any techniques and algorithms for scene change detection or motion tracking, and then generates some abstraction descriptions for each video shot. These abstraction descriptions include a *"representative frame"* (or key frame) which is the first frame of a continuous video shot (a continuous video segment which has no appreciable scene change) and a *"quantitative value"* (the associated scene change value which reflects the level of scene changes). Figure 1. shows the basic concept of VPC.

The VBC provides a video previewing environment in the client part, i.e. in the user end, which make use of the abstraction descriptions (including a lot of shot icons and their corresponding scene change levels) offered by VPC to provide shot browsing and/or shot video playback facilities. Furthermore, a multi-layer browsing facility which can dynamically adjust the scene change threshold values to produce finer or coarser video shots are also implemented in the system . For example, a VOD can get some video shot representations in its coarsest scheme, the user can then rapidly decide which video shot is closest to the desired one. If the user still can not make a precise decision on whether to subscribe this video shot or not, he can press a selected shot icon to display the candidate video segments with finer precision shots for further detail previewing. This browsing mechanism can be implemented easily and provide efficient browsing facilities. Figure 2. shows the basic concept of VBC.

IV. Applications : A Video On Demand Prototype System

Based on the proposed browsing system, a prototype Video On Demand system has been implemented in the Communication and Multimedia Laboratory of National Taiwan University. This prototype system is developed on a SUN Sparcstation coupled with a Parallax board for performing motion JPEG real-time compression. POET OODB [20] is currently used as the platform for video database management system (DBMS) and the network backbone is Ethernet LAN. At present, the video data are stored in an external disk array and it is expected the next release version will port the storage to a jukebox system in the near future.

At the current stage, the video partition engine in the video server can capture video data from external devices, such as LD or VCR, process the video data (including video editing and scene change automatic detection) and save the so-obtained features into the DBMS for later usage. We also provide a multi-layer video browser which will help users to select movies easily.

Figure 3. shows the interactive behavior of subscribing a specific video between the video provider and the customer in a typical VOD applications. First, client connects to a VOD server, and then server will send movie categories which will be represented in icon form on user screen, as shown in the upper window of Figure 4. User can play back the video program directly by double clicking on the selected video icon, or choose the preview mode by single clicking on the video icon. Then, the client will send a preview request to video server for requiring a specific movie program of interested. Once server receives the preview request message, the corresponding preview information specific to one movie which composes of the key frames and its corresponding shot values will be sent to the client. The client browsing system will display those key frames in icon form, as shown in the bottom window of Figure 4. The volume of preview information is much less than the volume of the whole original video, therefore, this approach saves a lot of network bandwidth and transmission time and other overheads. In the preview window, user can rapidly preview a pre-processed video shots by dragging the scroll bar, or pressing a specific shot icon and then adjusting a scene change threshold to get finer/coarser video shots, as shown in Figures 5 to 7. Finally, if client user likes the previewed material he can subscribe the complete video program for normal playback.

V. Conclusion

Due to the growth of requirement in video services, many efforts have been made on providing a more friendly environment for users to obtain the desired information. This paper presents an efficient video browsing system to provide flexible representations of the contents of a video sequence. Abstraction of video shots is presented in a multi-layer manner instead of examining all frames in the conventional browsing approaches. Based on the proposed browsing tool, a prototype VOD system has successfully been implemented in the Communication and Multimedia Laboratory of National Taiwan University.

Acknowledgment

This research was supported by the National Science Council of Republic of China under the contract no. NSC 83-0425-E-002-140.

References

[1] Gerard A. Wall, James G. Hanko, and J. Duane Northcutt, "Bus Bandwidth Management in a High Resolution Video Workstation," *Proceeding of 3rd International Workshop on Network and Operating System Support for Digital Audio and Video*, pp. 274-288, 1992.

[2] Federighi, C. and Rowe, L., "A Distributed Hierarchical Storage Manager for a Video-on-Demand System," *1994 IS&T/ SPIE Symposium on Electronic Imaging: Science and Technology*, San Jose, CA, February, 1994.

[3] Haskins, R., "The Shark Continuous-Media File Server," *Proceeding of IEEE COMPCON '93*, San Francisco, CA, February 1993..

[4] Peter Druschel, Mark B. Abbott, Michael Pagels, and Larry L. Peterson, "Analysis of I/O Subsystem Design for Multimedia Workstations, "*Proceeding of 3rd International. Workshop on Network and Operating System Support for Digital Audio and Video*, pp. 289-301, 1992.

[5] Gerard J. M. Smit, Paul J. M. Havinga, "The Architecture of Rattlesnake: A Real-Time Multimedia Network," *Proceeding of 3rd International Workshop on Network and Operating System Support for Digital Audio and Video*, pp. 15-24, 1992.

[6] Keeton, K. and Katz, R., "The Evaluation of Video Layout Strategies on a High-Bandwidth File Server," *Fourth Int'l. Workshop on Operating Systems and Network Support for Digital Audio and Video*, 1993.

[7] Rangan, P.V., H. M., Ramanathan, S., "Designing an On-Demand Multimedia Service,", *IEEE Communications Magazine*, Vol. 30, No. 7, July 1992, pp. 56-64.

[8] S-K. Chang and A. Hsu, "Image information systems: Where do we go from here ?" *IEEE transactions on Knowledge and Data Engineering*, pp. 431-442, October, 1992.

[9] Andrew W. Davis., "Motion image processing: Striking possibilities," *Advanced imaging, August 1992*, pp. 22-56.

[10] Takeshi Kikukawa and Satomi Kawafuchi., "Development of an automatic summary editing system for the audio visual resources," *Trans. of the Institute of Electronics, Information and Communication Engineers*, Vol. J75-A, No.2, 1992, pp.204-212.

[11] Akio Nagasaka and Yuzuru Tanzka., "Automatic video indexing and full-video search for object appearances," *Trans. of Information Processing Society of Japan*, Vol.33, No.4, 1992, pp.543-550.

[12] Kiyotaka Otsuji, Yoshinobu Tonomura and Yuji Ohba., "Video browsing using brightness data," *Proc. of SPIE. VCIP'91. Vol 1606, 1991*, pp.980-989.

[13] Hirotada Ueda, Takafumi Miyatake and Satoshi Yoshizawa., "IMPACT: An interactive natural-motion picture dedicated multimedia authoring system," *CHI'91 Conference Proceedings, 1991*, pp.343-350.

[14] Kiyotaka Otsuji, Yoshinobu Tonomura., "Projection detecting filter for video cut detection," *ACM Multimedia 93 USA*, pp. 251-257.

[15] Farshid Arman, Arding Hsu, and Ming-Yee Chiu., "Image processing on compressed data for large video databases," *ACM Multimedia 93 USA*, pp. 267-272.

[16] AI Lab, University of Michigan., "Knowledge Guided Parsing in Video Databases", *ACM Multimedia 93 Tutorial.*

[17] H.J. Zhang, A. Kankanhalli, and S.W. Smoliar, "Automatic Partitioning Full-motion Video," *Multimedia Systems*, pp. 10-28, 1993

[18] G. K. Wallace. "The JPEG still picture compression standard," *Communications of ACM*, 34(4):30-44, April 1991.

[19] D. Le Gall. "MPEG: A video compression standard for multimedia applications," Communications of ACM, 34(4):46-58, April 1991.

[20] *P.O.E.T Programmer and Reference Guide*, BKS Software Entwicklungs GmbH, Berlin 1992

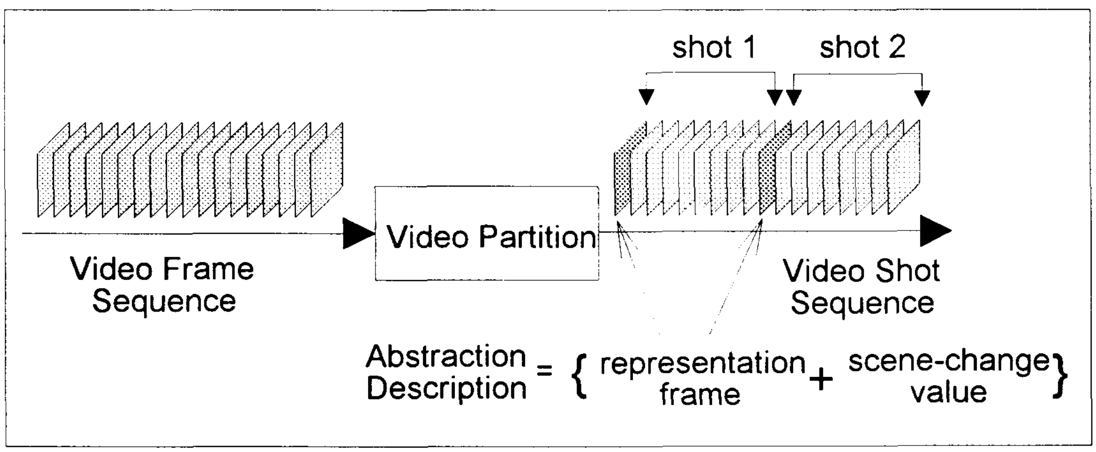

Figure 1. Video Partition Component (VPC) in a VOD Server Site

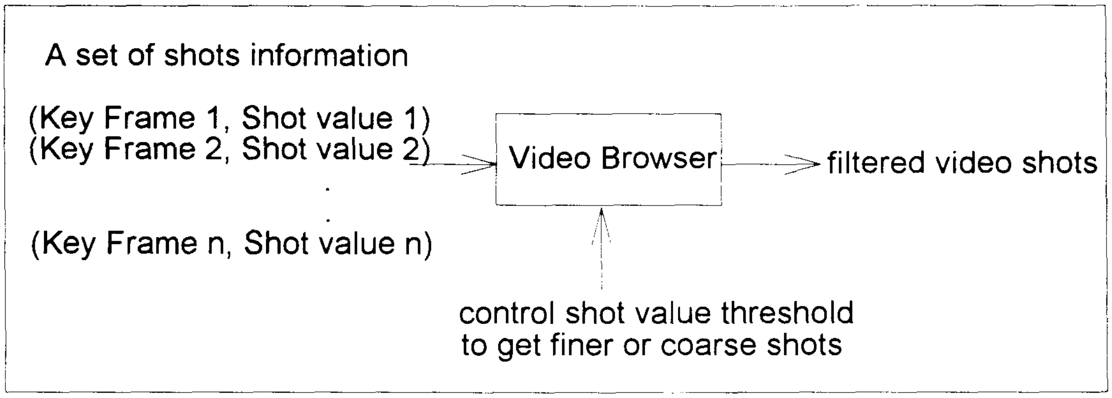

Figure 2. Video Browsing Component (VBC) in a VOD Client Site

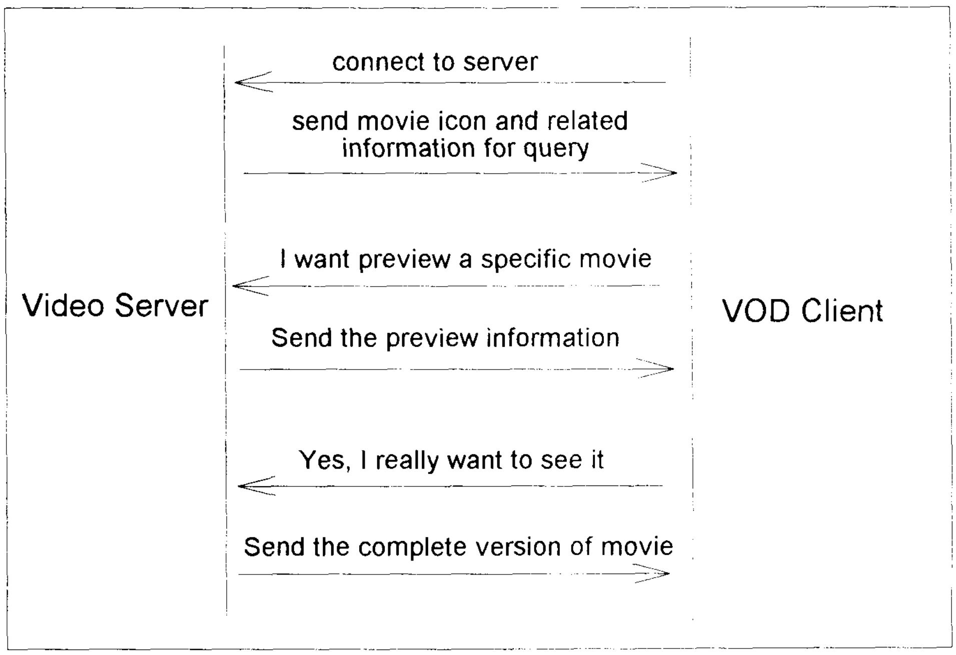

Figure 3. The Interactive Behavior between a VOD Client and a VOD Server

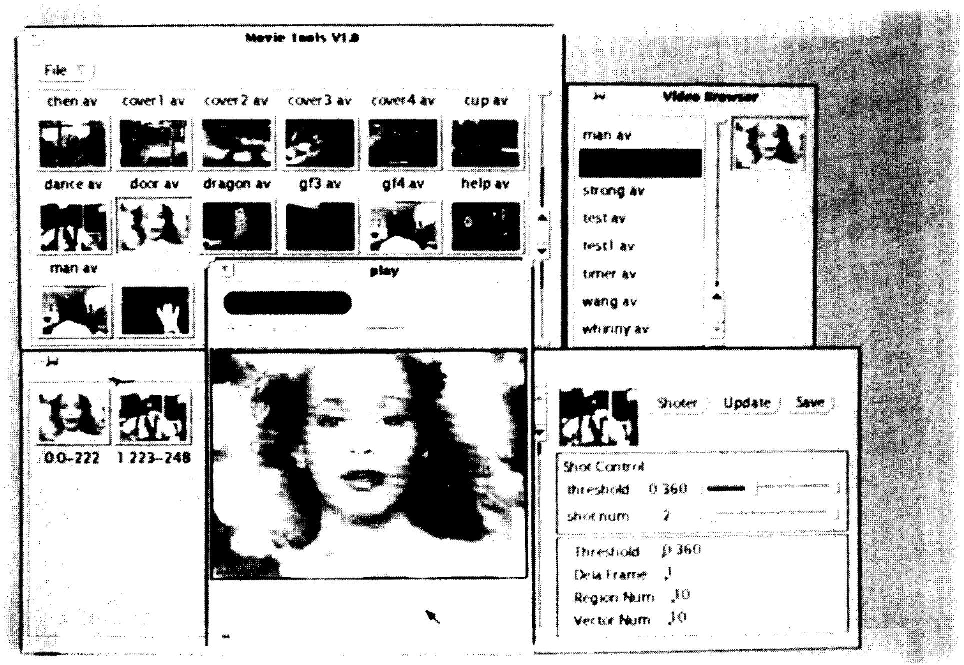

Figure 4. The Photo of prototype Video On Demand System (The movie categories are represented as icon form in the upper window)

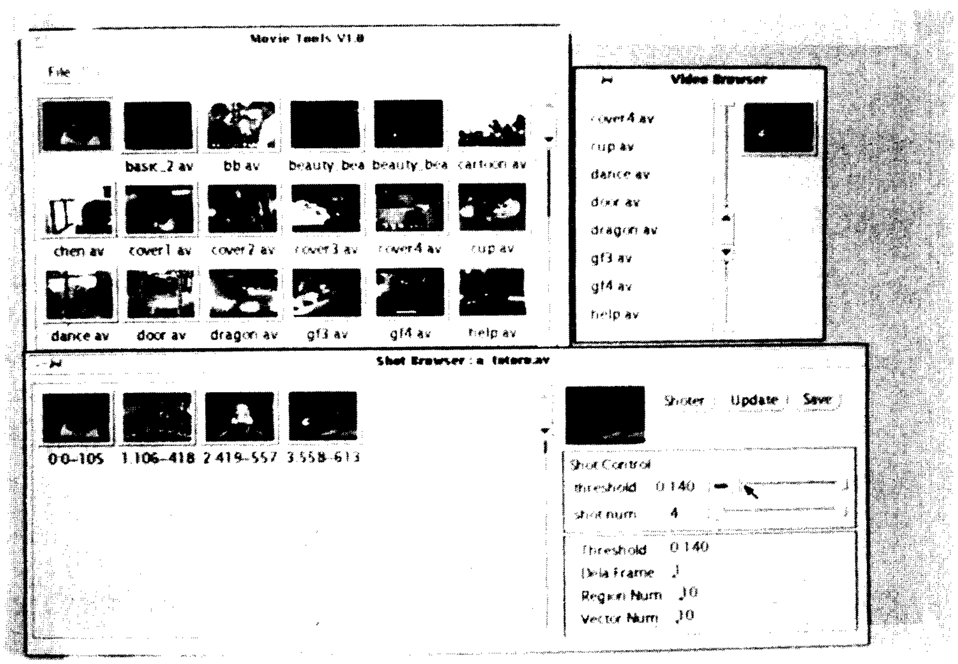

Figure 5. The Largest Scene Change Threshold Produces The Coarsest Video Shots (as shown in the left-bottom window)

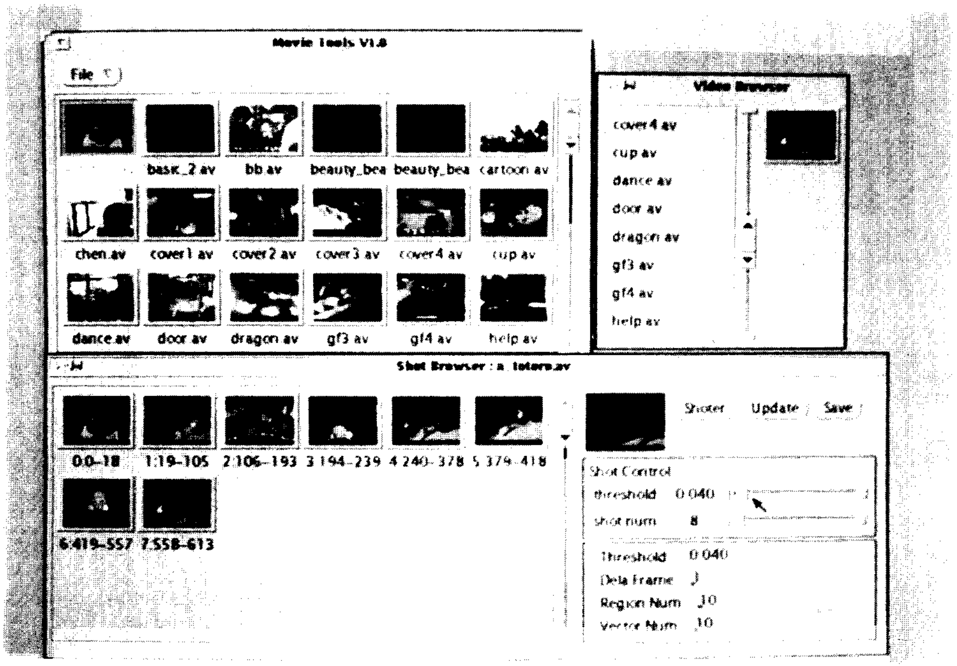

Figure 6. The Smaller Scene Change Threshold Produces The Finer Video Shots (as shown in the left-bottom window)

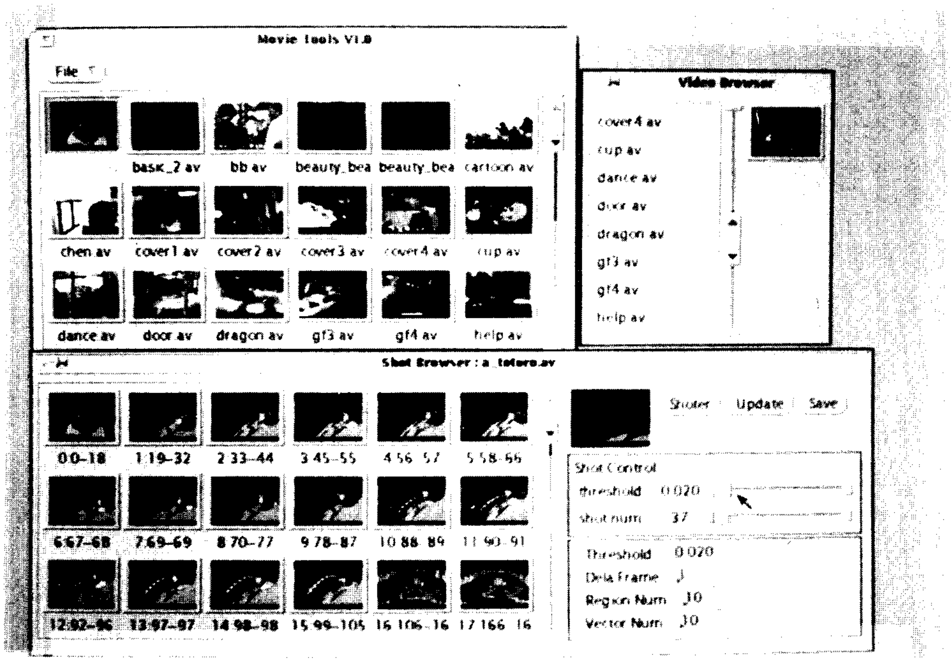

Figure 7. The Smallest Scene Change Threshold Produces The Finest Video Shots (as shown in the left-bottom window)

Biographies

Herng-Yow Chen received the B.S. degree in Computer Science and Information Engineering from Tamkang University, Tamshoei, Taiwan, ROC. He has been a Ph.D. student in the Department of Computer Science and Information Engineering at National Taiwan University, Taipei, Taiwan, ROC, since 1992.

His research interests include Digitial Signal Processing, Image Coding, Data Compression, Multimedia Synchronization Modeling and Multimedia Systems. He is a student member of IEEE.

Ja-Ling Wu was born in Taipei, Taiwan, on November 24, 1956. He received the B.S. degree in Electrical Engineering from the Tamkang University, Tamshoei, Taiwan, in 1979, the M.S. and Ph.D. degree in Electrical Engineering from the Tatung Institute of Technology in 1981 and 1986, respectively.

From 1986 to 1987 he was an Associate Professor of the Electrical Engineering Department at Tatung Institute of Technology, Taipei, Taiwan. Since 1987 he has been with the Department of Computer Science and Information Engineering, National Taiwan University, where he is presently a professor.

Professor Wu was the recipient of the 1989 Outstanding Youth Medal of China and the Outstanding Research Award sponsored by the National Science Council, from 1987 to 1992. He published more than 100 technical and conference papers. His research interests include Neural Networks, VLSI Signal Processing, Parallel Processing, Image Coding, Algorithm Design for DSP, Data Compression, and Multimedia Systems.