Journal of Computer Science 1 (2): 221-224, 2005
ISSN 1549-3636

Web-based Digital Video Sequencing System

Khalid A. Kaabneh
Department of Computing Studies, Amman Arab University for Graduate Studies
P.O. Box 2234, Amman, 11953, Jordan

Abstract: In order to access and retrieve Internet shared videos, there is a clear need for structuring to decrease search and loading time. Video segmentation is the process of dividing a video clip into segments of similar characteristics. Based on our segmentation research, this study proposes a complete Digital Video Sequencing and retrieval tool, aimed for the Internet users. This ongoing research aspires to provide a fast and efficient visual summarization and searching of video content such that users can efficiently determine whether they are interested in the video before they have to download it from the Internet. This system consists of three major modules. The first module is responsible for automatic segmentation of video scenes. The second module structures the video as a Contents Structure (VCS) ready for lookup. Finally, the last module provides the users with video display Interface Program. Our experiments on a number of various videos demonstrate the efficiency of this system.

Key words: Segmentation, Video Content Structuring, XML, Internet Videos

INTRODUCTION

Video is the most effective medium for capturing the events in the real world around us. It is also the most dramatic medium as it combines both photo-realistic images and sounds. Combining the advantages of video and computers will broaden the scope of information that a computer can process, enhance some existing applications and open the doors to others.

Since the Internet became public, researchers have been looking for better ways of digital videos storage, retrieval and distribution over networks. In the early days, manual classification was used to arrange video databases for future access and retrieval. This manual classification may work for limited video files but not for large databases found on the Internet worldwide. A better approach must be used instead.

A digital content-based information query has been a very hot research topic. In specific, digital video content-based retrieval is gaining more attention because there is a large volume of video content available on the net. Many researchers have investigated video content retrieval problems [1, 2], but there are still many interesting issues to be done. A main problem is the video content matching [3-5]. This is still a challenging problem since it is difficult to extract video features that represent the content for matching and there are many possible algorithms to match the temporal ordering of video features. This complete web-tool will help the Internet users to efficiently browse the contents of shared videos in two ways. First, we generate a Video Contents Structure (VCS) to describe the shared video. Representative images extracted from the video according to the structure are embedded in the VCS, such that they can demonstrate what have been shown in the video and how they synthesize the video contents. After browsing the VCS, users should then have an abstract idea of the video contents. To further describe a video, the second method provided in this research is the generation of a video summary. Since the image-based VCS is a static presentation, it may not be sufficient to describe a video. Therefore, we generate a summarized video by scanning through the video segments according to the structure of the VCS within a short duration. Several options are allowed for the users to customize the video summary and we have designed our own Interface program to give the summary presentation of shared videos. The system architecture is briefly described in the following section.

System Overview: The Web-Based Digital Sequencing System has been composed of three main modules. They are the Video Structuring module, the VCS presentation module and finally, the Presentation module. Figure 1 shown below describes the structure of the system.

Video Structuring Module: At this module, the digital video is restructured in a top-down approach according to its contents. We start by dividing a video into five levels of components [6]. They are whole video, video scenes, video groups, video shots and the video reference frames. The top-down hierarchy of these components is shown in Fig. 2.

The video scenes are represented using our segmentation technique that takes motion into consideration. The technique works by dividing each frame into blocks of a certain size; starting from a reference frame, each block is compared with all the

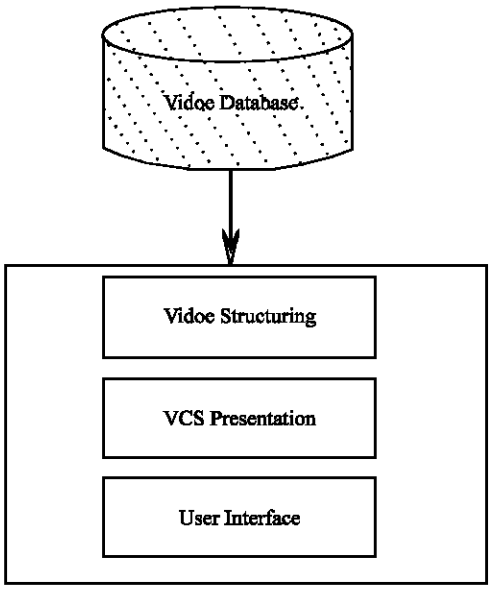

Fig. 1: System Modules

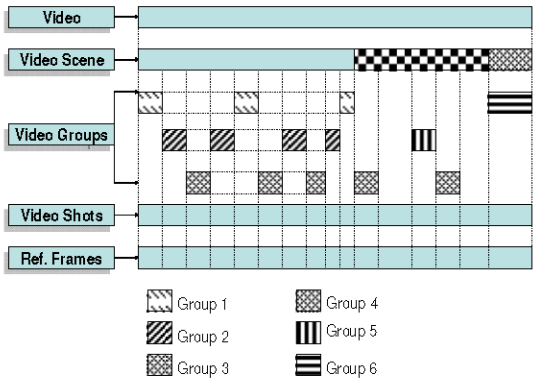

Fig. 2: Hierarchical Representation of a Video

blocks in a region surrounding the corresponding block in the next frame [7]. This technique was selected for its efficiency and effectiveness when using video source.

To further enhance our technique in reference to the frame matching, we have implemented the matching of global color histogram for two frames to determine their visual similarity [8]. The color histogram similarity is calculated by the four raw frame similarities: $FrameColorSim_{bj,ei}$, $FrameColorSim_{ej,ei}$, $FrameColorSim_{bj,ei}$ *and* $FrameColorSim_{bj,bi}$, where, $FrameColorSim_{ej,bi}$, is defined as:

$$FrameColorSim_{x,y} = 1 - Diff_{x,y} \quad (1)$$

and x, y are two arbitrary frames, with $x > y$.

The remaining video components are further organized into a multi-level tree structure as presented in Fig. 3.

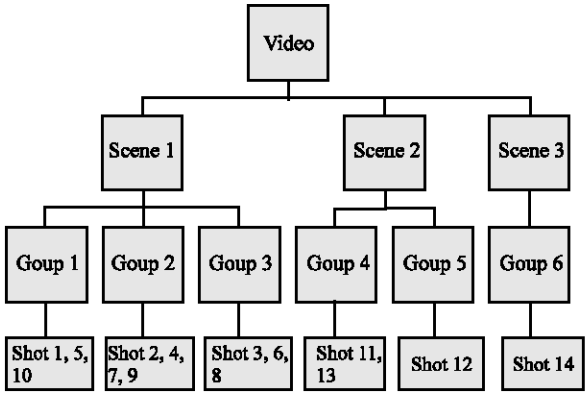

Fig. 3: Video Tree Structure

The tree structure of a video is constructed in a bottom-up manner starting from the shot level. The video tree structure can well demonstrate the organization of the video contents. Different methods on detecting boundary of shots have been used [9, 10]. Most of them can be categorized into several types; however, the histogram-based method is one of the most popular and efficient approaches. By applying these boundary detection methods, a video is divided into shots, which consist of a sequence of similar frames. Some researchers have suggested building up the group of shots by grouping similar shots together [1, 6]. In this way, the number of video fragments can be reduced and grouped analysis would be more effective. At the video scene level, it is composed of content related groups.

VCS Presentation Module: At the VCS module, the VCS video tree structure is generated according to output from the video structuring module. The structure is stored in the XML format [11] for later use. We define a set of elements in Fig. 4 to describe the tree structure. A sample XML video tree structure is shown in Fig. 4.

XML elements	Child nodes	Assoc. attributes
<system>…</system>	video	-
<video> …</video>	scene	String src,
<scene> … </scene>	group	Int id,
<group> …</group>	shot	Int id,
<shot> … </shot>	time, keyframe	Int id,
<time />	-	Int Value,
<keyframe />	-	img

Fig. 4: Set of XML Elements Defined for Referencing

We use XML for storage because of several advantages. First, we can build an organized and compact data structure for using the nested hierarchy of XML. Also, XML is extensible and searchable as it is in a plain-text format. Besides, XML can be used as a standard information protocol for exchanging data between different system modules. It is also convenient

to transform the XML into a web-based presentation by using XSL [12]. XSL provides filtering and sorting functions such that the output web presentation can be organized according to the values of the XML elements. We regard the resulting web-based presentation of the video tree structure as VCS. The video components are ordered along scenes, groups and shots. The key frame image of each shot is shown on VCS together with the corresponding time instance in the video. Therefore, with VCS, we can easily know what and when a video shot shown in the video. This visual video description can give us an abstract idea of the video contents (Fig. 5).

```
<?xml version="1.0"?>
<!DOCTYPE Web-Sequencing System "./toc.dtd">
<system>
<video src="rstp:// source video on server">
<scene id="1">
<group id="1">
<shot id="1">
<time value="2"/>
<keyframe img="./ID-1.jpg"/>
</shot>
<shot id="3">
<time value="10"/>
<keyframe img="./ID-2.jpg"/>
</shot>
</group>
<group id="2">
<shot id="2">
<time value="4"/>
<keyframe img="./ID-3.jpg"/>
</shot>
</group>
</scene>
<scene id="2">
<group id="3">
<shot id="4">
<time value="24"/>
<keyframe img="./ID_4.jpg"/>
</shot>
</group>
</scene>
</video>
</system>
```

Fig. 5: XML Video Tree Structure

User-Interface Module: To facilitate the user's access to a video, we have developed a user interface module, which will incorporate the scene structure information (Fig. 3) together with the representation of frames into a visual tool to grasp the content of a video clip as in Fig. 6. Such interface can be incorporated in a web page for Internet display.

The user can expand the video scenes into more detailed levels such as groups and shots (Fig. 7).

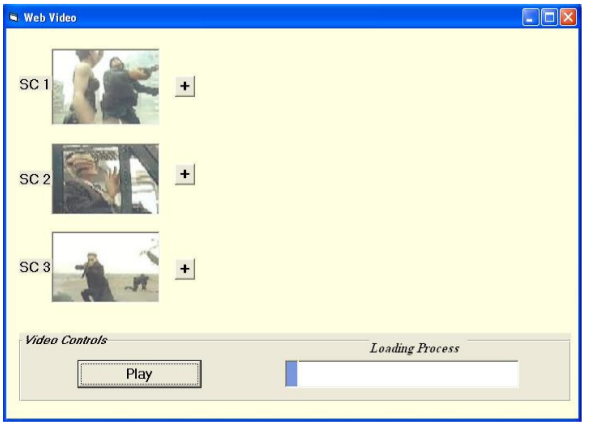

Fig. 6: Display the Video Content (Scene Level)

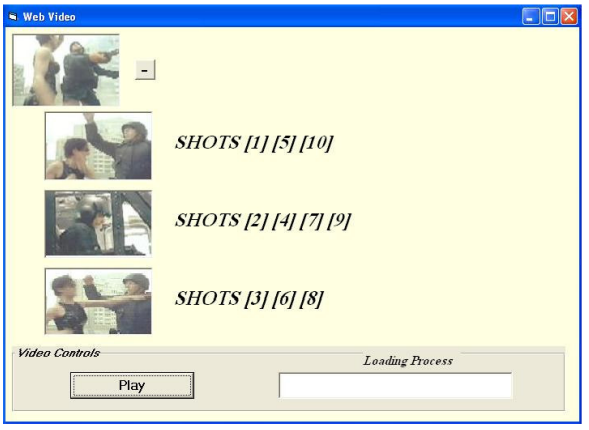

Fig. 7: Detailed Description of Video Content (Group Level)

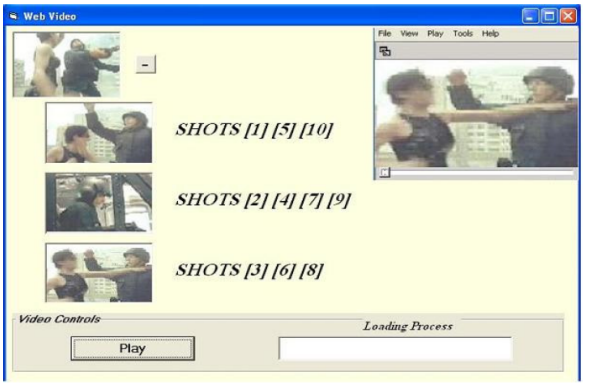

Fig. 8: Digital Video System

The shots are displayed according to their reference number. After browsing the content of a video clip (scenes, groups and shots), the user can display any part interested in and display it by the embedded video player area without doing tedious fast-forward and rewind as in Fig. 8.

EXPERIMENTAL RESULTS

In order to test the effectiveness of this proposed system, we have conducted a test-bed on a various types of MPEG video files with digitization rate of 30 frames/sec. The test set included a number of video types such as romantic-slow, romantic-fast, music video clips, comedy, science fiction, TV sitcoms and some action movies. The video clips were 15-25 min long and of total length of about 180,000 frames. Table 1 gives the experimental results.

Table 1: Experimental Results

Video name	Frames	Shots	Groups	Detected scenes
Romantic-S	20781	135	15	5
Romantic-F	28145	201	27	7
Music	13496	76	15	7
Comedy	34106	186	28	12
Sci-fiction	19855	81	11	5
TV sitcom	25446	413	86	26
Action	36099	345	51	12

To further test our system, the output results were subjected to human subjective evaluation. We have used the same ground truth test-bed as conducted in our segmentation technique [7]. This is shown in the chart in Fig. 9.

From the results in Table 1 and Fig. 9, some observations can be made:

* The sequencing system achieves reasonably good results in most of the movie types in which those results were supported by the human subjective results.
* Segmentation system achieves better performance and output results in slaw movies than fast movies. This is because in fast movies the visual content is normally more complex and difficult to capture.

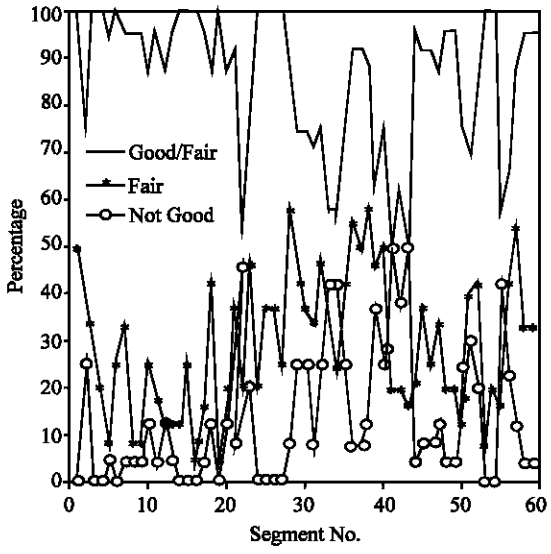

Fig. 9: Output Results of Subjective Evaluation

CONCLUSION

The complete web-based sequencing system presented in this study is a vital tool for video browsing and retrieval aimed for the Internet users. This research provided a fast and efficient visual summarization and searching of video content. This system consists of three major modules. The first module is responsible for automatic sequencing of video scenes based on our newly developed sequencing technique integrated with global color histogram for two frames which further enhanced the frame matching. The second module structures the video as a Contents Structure (VCS) ready for lookup. Finally, the last module provides the users with an option to personalize a video into its summary using an Interface Program, in addition to help the users view the contents of a video quickly before they spend much time to download the whole video. Our experiments on a number of various videos demonstrate the efficiency of this system in most of the movie types.

REFERENCES

1. Koh, J., C. Lee and A. Chen, 1999. Semantic video model for content-based retrieval. In IEEE International Conference on Multimedia Computing and Systems, 2: 472-478.
2. Zhou, J., E. Ong and C. Ko, 2000. Video object segmentation and tracking for content-based video coding. In IEEE International Conference on Multimedia and Expo., 3: 1555-1558.
3. Adjeroh, D., M. Lee and I. King, 1998. A distance measure for video sequence similarity matching. In International Workshop on Multi-Media Database Management Systems, pp: 72-79.
4. Lienhart, R., W. Eielsberg and R. Visualgrep, 1998. A systematic method to compare and retrieve video sequences. In Storage and Retrieval for Image and Video Databases VI, SPIE. 3312: 271.
5. Mohan, R., 1998. Video sequence matching. In Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing. 6: 3697-3700.
6. Rui, Y., T. Huang and S. Mehrotra, 1999. Constructing table-of-content for videos. ACM Multimedia Systems J., Special Issue Multimedia Systems on Video Libraries, 7: 359-368.
7. Kaabneh, K. and H. Al-Bdour, 2004. A new segmentation technique and evaluation. Mu'tah Lil-Buhuth Wad-Dirasat J., 19: 39-51.
8. Yeung, M., B. Yeo and W. Wolf, 1996. Extracting story units from long programs for video browsing and navigation. In: Porc. IEEE Conference on Multimedia Computing and Systems.
9. Boreczky, J. and L. Rowe, 1996. Comparison of video shot boundary detection techniques. J. Electronic Imaging, 5: 122-128.
10. Browne, P., A. Smeaton, N. Murphy, S. Marlow and C. Berrut, 2000. Evaluating and combining digital video shot boundary detection algorithms. In Irish Machine Vision and Image Processing Conference (IMVIP 2000), Belfast, Northern Ireland.
11. Anonymous, 2000. W3C Recommendation, Extensible Makeup Language (XML) 1.0 Specification (2nd Edn.). http://www.w3.org/TR/2000/REC-xml-20001006, 6 Oct.
12. Anonymous, 2001. W3C Recommendation, Extensible Style-Sheet Language (XSL) 1.0 Specification. http://www.w3.org/TR/2001/REC-xsl-20011015, 15 Oct.