A Web-Based Video Editing System that Synchronizes Multiple Video Sources

Satoshi Ichimura, Yutaka Matsushita

School of Computer Science
Tokyo University of Technology
Katakura 1404-1, Hachioji, Tokyo 192-0982 JAPAN.
Email: ichimura@cs.teu.ac.jp ; FAX: +81-426-37-2613

Abstract

Home movies edited by amateurs are likely to be a sequence of monotonous scenes, where people record video from a single direction, and use video from a single camcorder when editing video. This is one of the significant reasons why a nonprofessionals' video can easily bore an audience. We developed MediaBlocks, a web-based video editing system. In our system, video clips are collected from multi-users' camcorders via the Internet and shared among the users. The system allows users to easily use other users' video clips, each of which is spontaneously videotaped using different digital video camcorders. MediaBlocks provides the capability to extract the date and time information of the video-recording from the digital video cameras, and automatically synchronize them. Through evaluations of the system, we determined the effectiveness of the system.

1. Introduction

Recent advances in portable digital video camera technology have enabled us to easily record various family events, and PC-based digital video editing has also become quite popular. Although videotaping and video editing is increasing in popularity, the quality of the movie created by these nonprofessionals has not significantly been improved. For instance, home movies edited by amateurs are likely to be a sequence of monotonous scenes, where people record video from a single direction, and use video from a single camcorder when editing video. This is one of the significant reasons why a nonprofessionals' video can easily bore an audience.

Suppose, for example, you attend your children's athletic festival and record a footrace. You might want to simultaneously videotape your children at the start line and at the finish line. However, since you cannot quickly change your camera position, you would have to give up either scene and continuously zoom in and out during the footrace, which is distracting for the audience. As a result, you are most likely to end up recording a typical amateurish video in which your kids are recorded from a single direction with a single camcorder.

On the other hand, professionals are able to create videos in a very different way. For example, pictures on the TV are a sequence of several video shots taken from many different directions and by a number of cameramen. A basic video recording technique is to approach the subject from many angles and not to record for a long period of time. Actually, it is quite unusual that a single cut from a single camera lasts more than ten seconds in TV programs or Hollywood movies. From the cinematography or directorial skills viewpoint, frequent camera switching is a key technique to prevent an audience from becoming bored.

We developed MediaBlocks, a web-based video editing system for sharing video clips among people who attended the same event. In a conventional manner, even when many people record video at the same time in the same place, they don't coordinate with each other when recording or creating a video. In our system, in contrast, video clips are collected from multi-users' camcorders via the Internet and shared among the users. The system allows users to easily use other users' video clips, each of which is spontaneously videotaped using different digital video camcorders. Thereby it allows you to get some video clips of your children, which you failed in recording in the athletic festival, from other parents' camcorders via the Internet.

MediaBlocks provides the capability to extract the date and time information of the video-recording from the digital video cameras, and automatically synchronize them. The system also enables users to create home movies using some professionals' video recording techniques. We describe the system's design, implementation, and evaluation. Through evaluations of the system, we determined the effectiveness of the system.

2. Related Work

2.1 Professionals' Video Recording

When editing a home movie, amateurs normally fail to collect sufficient video clips for editing. As such we are unable to create a rich movie from a small set of video clips extracted from a single camcorder. This makes our home movie an amateurish monotonous video.

On the other hand, professionals use a variety of video clips collected from multiple camcorders. For example, in cinema productions, there is always a small crew called the second unit. They usually record background shots that do not involve the main actors. Professional video editors make it a rule to insert the collected background shots between the main scenes to break up the monotony.

There is also a professionals' video editing rule that video editors don't concatenate video clips of similar size or similar angle. If such similar video clips were played back continuously, the audience would have a sense of discomfort by feeling a time-gap difference between those clips. For this reason, a professional camera crew will always record several video clips with different angles or different zoom scales. Comparing home movies of athletic festival to sports news on TV, it is obvious that using professionals' video recording techniques is helpful in producing a high quality video. However, nonprofessionals are supposed to create home movies having no background shots, having video clips that are similar in size and angle concatenated. As such nonprofessionals cannot prevent an audience from becoming bored to death.

MediaBlocks allows people to share their video clips on the Internet and produce their own movie having other users' video clips inserted as second-unit. We examined several home movies spontaneously videotaped at athletic festivals, and found that each video clip from each camcorder had a different angle and subject size even though those camcorders simultaneously recorded the same subject. This way, we developed the idea that mixing video clips recorded by multiple cameramen could break up the monotony of an amateurish movie.

2.2 Synchronizing Multi-Camera Sources

As mentioned above, using multiple video clips from multiple camcorders for video editing is clearly effective for creating interesting video. It is, however, well-known that synchronizing video scenes from multiple camcorders is terribly time-consuming. Even though camcorders simultaneously record the same subject, the video data extracted from each camcorder holds no clue on how to synchronize. To synchronize multiple video clips extracted from different camcorders, video editors are required to repeatedly press the fast-forward, fast-rewind and slow buttons on the camcorder.

For example, when you synchronize two video clips recorded from two camcorders, first you would have to find a moving object that appeared in both camera sources recorded, and then precisely synchronize the movement of the object. However, images recorded from different directions might not contain the same object you are looking for. You could also use sound to synchronize the two camera sources. However, you might not be able to clearly hear the sound from either camcorder.

Some professional video editing software provides users a way to synchronize multiple camera sources using industry standard SMPTE (The Society of Motion Picture and Television Engineers) time codes. Such software allow users to see SMPTE time codes extract from video camcorders, and semi-automatically synchronize camera sources. However, since SMPTE time codes are sequential serial numbers starting from the beginning of a videotape to the end of the videotape, to use the SMPTE codes for synchronization, you must simultaneously start recording all video camcorders, and you cannot stop or pause recording until the end of the videotape once you start videotaping. You may also need expensive camera-controlling devices to simultaneously control multiple camcorders.

We developed the idea to utilize the date and time information of video-recording, which is embedded in standard digital video format, to synchronize multi-camera sources, and to develop software based on this idea. The international standard (ISO/IEC 61834) specifications[1] for consumer-use digital video camcorder (DV camcorder) describes that DV camcorders must record the recording date and time information in the subcode area of every picture frame (i.e., 30 frames/sec) on a digital video tape. Furthermore, movie files digitally extracted from DV camcorders with DV-capture software (like Windows Movie Maker) keep the date and time information inside the files. We have tested several DV camcorders manufactured by SONY, Panasonic, Canon, and Victor, and confirmed that all the DV camcorders tested had the capability to record the recording date and time information. Furthermore our software was able to extract information from the AVI files extracted from the DV camcorders using DV-capture software on a Windows OS.

When using our software, as long as the internal clock of all DV camcorders are adjusted to the standard

time, multi-camera sources from multi camcorders can automatically synchronize even under the following conditions:

1. Cameraman can start or stop video-recording at anytime.
2. Cameramen need only to use consumer-use DV camcorders and do not need any additional equipment for synchronization.
3. Cameramen can freely change camera position or camera angle.
4. Videotaping can be started from the any point of a partially-used videotape.
5. Video-recording can be extended to another videotape.

Hitchcock [2] and Silver [3] allow users on the Internet to remotely edit movies through web browsers. For example, Hitchcock has capabilities to automatically group similar scenes and remove unstable/blur scenes. They, however, don't provide a mechanism to synchronize multi-camera sources, so that video editors have to spend considerable time and effort using multiple-camera sources.

3. Prototype

MediaBlocks consists of collecting, sharing, and editing video component. Each component is described here.

3.1 Collecting Video Clips

To collect video, MediaBlocks provides a DV-capture tool to create a movie file from a digital video tape using digital video capture. MediaBlocks also extracts the recording date and time from the captured movie file, and converts a high-quality DV-captured movie file into a low-quality streaming-video.

The DV-capture tool runs on a users' PC. Users can control a DV camcorder connected to the user's PC via an IEEE1384 cable by operating the PC software. The software collects each video frame individually from a DV-captured movie file, and detects the recording date and time information embedded in the video frame. This mechanism is based on the DirectShow multimedia architecture implemented on a Windows OS. While the software is capturing a movie, it automatically generates a list of the video clips segmented from the beginning of video-recording (when the record button is pressed) until the end of video-recording (when the stop button is pressed), and stores the list as a file in a CSV format.

To edit videos, it is essential to frequently preview movie clips in the making. In MediaBlocks, users edit video clips over the Internet, previewing of video clips must be performed through a network. However, a DV-captured movie file is too large to preview over the Internet, so we added a function to the DV-capture tool to convert a high-quality DV-captured movie file into a low-quality streaming-video file. For example, a 2GB 10-minute DV movie file (nearly 25 Mbps) is converted into 20 MB of streaming video in RealMedia format (approximately 300 Kbps). The streaming video is small enough to be transmitted over the Internet, and also users do not have to wait to download an entire video file before seeing it. Instead, the video is sent in a continuous stream and is played as it arrives.

The generated streaming video files and the recording date and time information list in the CSV format are uploaded using FTP to the MediaBlocks server. In addition, the original DV-captured movie files are also uploaded to the MediaBlocks server for the purpose of creating a final high-quality video at the end of the video editing process.

Furthermore, the DV-capture tool has the capability to correct the wrong recording date and time information stored in a DV-captured movie file. As MediaBlocks synchronizes video clips based on the recording date and time extracted from the DV camcorders, the internal clock of all DV camcorders must be precisely synchronized. However, in actuality, amateurs' DV camcorders are rarely adjusted to standard time. DV camcorders having a wave clock would solve this problem, but unfortunately none of consumer-use DV camcorders currently on the market have built-in wave clock. To overcome this problem, we developed a mechanism to calculate and correct the time-gap difference between the internal timer of DV camcorders and standard time, and incorporated this mechanism into our DV-capture tool.

The DV-capture tool has the function to obtain standard time from the Internet using NTP (Network Time Protocol), and to obtain internal clock of the DV camcorder connected to the PC via the IEEE 1394 cable. NTP is widely used for synchronizing computer clocks in the Internet; PCs can extract via the IEEE 1394 cable the internal clock of the DV camcorder in camera mode. Then, the DV-capture tool calculates the time-gap difference between standard time from the Internet and internal clock from the DV camcorder, and automatically corrects the recording date and time stored in the DV-captured movie file. Using this mechanism, multiple video clips

from multiple DV camcorders can be automatically and precisely synchronized in the system.

3.2 Sharing and Editing Video Clips

MediaBlocks, which was implemented on a Tomcat server, provides functions to share and edit video clips uploaded from users using a DV-capture tool. Uploaded streaming video clips and the recording date and time information are associated with each other, and are maintained in the database on the server. Video editing can be performed by users on the Internet through web browsers. MediaBlocks provides users with a web-based user interface to cut and paste video data stored in the database. It also provides a web-based video streaming viewer to play the preview video sent from the server to the users' PC.

```
<?xml version="1.0" encoding="UTF-8" ?>
<smil><head><layout>
    <root-layout height="242" width="352" />
<region fit="meet" height="240" id="image0" />
</layout></head><body><par>
    <video begin="0.0" clipBegin="537.7" clipEnd="547.7"
    fill="remove" region="image0"
    src="http://localhost:8080/MovieBlocks/AVI/nakamura0
    2.rm" />
    <video begin="10.0" clipBegin="253.7"
    clipEnd="263.7" fill="remove" region="image0"
    src="http://localhost:8080/MovieBlocks/AVI/watanabe0
    2.rm" />
    <video begin="20.0" clipBegin="220.4"
    clipEnd="230.4" fill="remove" region="image0"
    src="http://localhost:8080/MovieBlocks/AVI/ane01.rm"
    />
</par></body></smil>
```

Figure 1 SMIL description generated after video editing

All the video-editing operations performed by users are recorded into the database on the sever as Synchronized Multimedia Integration Language (SMIL) scripts, where SMIL is a markup language designed to represent multiple media files, and to synchronize a set of independent multimedia objects. Video-editing scripts written in SMIL format, movie clips controlled by the SMIL script are played back as if they were modified according to the script. This feature allows users to edit video data without actually having to modify the video file. MediaBlocks generates a web page embedding a RealOne Player, which supports SMIL2.0 standard, where the embedded player can be controlled using JavaScript associated with graphical user interface objects. For example, users can play or seek a preview-movie described in SMIL by pressing the play button or moving the slider bar on the same web page. Figure 1 shows an example of a SMIL script generated using MediaBlocks server.

SMIL scripts are stored in the database as virtual video clips. A virtual video clip can be a sequence of several camera shots or a part of single shot. The SMIL scripts stored in the database are basically available to other users, and other users can import one or more virtual video clips into their own video clips. Text memos can be attached to any video clips stored in the database, where these memos can be used to search for video clips. Users can also browse a list of video clips stored in the database on their web browser, where each item in the list includes a snapshot image of a movie clip, filename, attached text memo, and play-back button. When the play-back button is pressed, the corresponding movie clip is played on the web browser.

One of the unique features of MediaBlocks is the butterfly view (Figure 2). Butterfly view allows users who are editing their own video clip to know of other video clips that were simultaneously videotaped by other users that are stored on the server.

Suppose you are editing a video clip you videotaped. Your clip (RealMedia streaming video) is displayed in the center of the butterfly view. You may select a necessary portion of the clip by clicking the begin and end button on the web page while playing the streaming video. When you click the begin button, MediaBlocks calculates the recording date and time corresponding to the beginning of the selected portion, searches the database for other video clips that were recorded at the same date and time, and displays a list of all the detected clips on the butterfly view.

Given that the recording date and time of the beginning of the selected portion is TS and the end of the selected portion is TE, the butterfly view displays the video snapshot at TS on the left side of the window, and ones at TE on the right side. Those snapshot images are dynamically extracted from the TS or TE position of the movie clips. If you click either snapshot image on the butterfly view, the selected movie clip will move to the center of the butterfly view, so that you would be able to immediately select the necessary portion from the new clip. This allows you to easily browse the preceding and proceeding movie clips, which may be inserted into the video clip you are working on, and quickly import necessary portions recorded by others into your own movie clip.

Finally, when you complete video editing, the original DV-captured movie files stored in the MediaBlocks server are automatically mixed according to

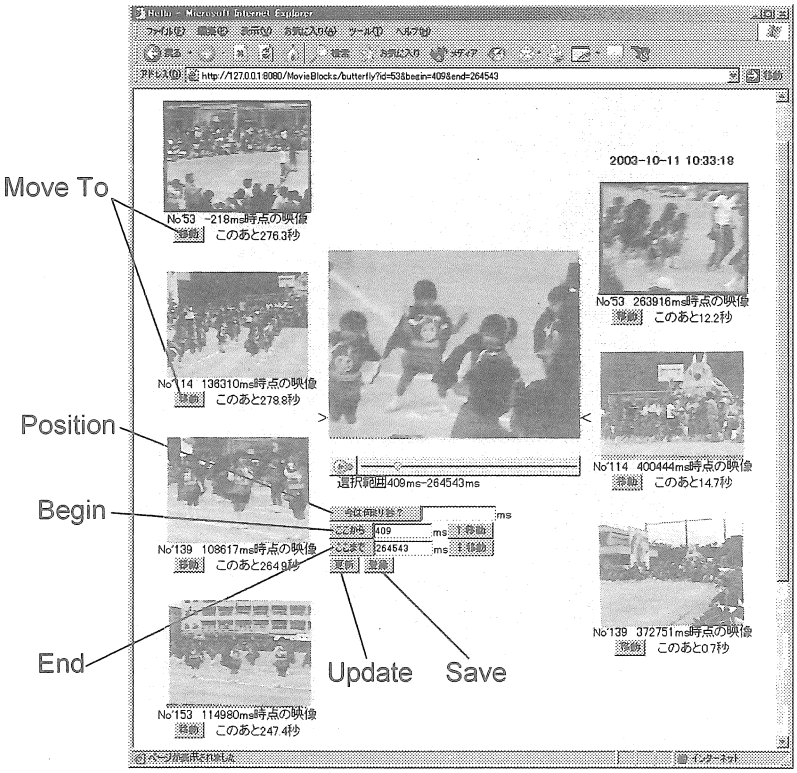

Figure 2 Butterfly view

the SMIL script, and the system creates a new high-quality movie file as a final output. This automatic video-editing is achieved with the DirectShow Editing Service on a Windows OS. For this reason, MediaBlocks server can run on a Tomcat server working on a Windows OS.

4. Evaluations

To evaluate the system, we have collected video clips that were videotaped at an athletic festival at a local kindergarten, and created a movie using MediaBlocks. A number of video clips recorded by five people whose children belonged to the same class in the kindergarten were collected. We didn't give them any directions or restrictions regarding how to record or what to record. They seemed to have recorded their children as usual.

Videotaped scenes included a footrace, a dance, an entrance procession, where children in the same class ran, danced, and walked together. Since we collected video clips that recorded children in the same class, we were able to collect many video clips in which multiple DV camcorders had simultaneously videotaped the same subject from different angles.

To evaluate multi-angle video contents created with MediaBlocks system, we tested subjects by comparing a movie clip produced with nonlinear video editing software to one created with MediaBlocks. Movie A was created from movie clips extracted from a single DV camcorder and edited with nonlinear video editing software in a conventional manner, whereas Movie B was created from movie clips recorded using five DV camcorders and edited with MediaBlocks.

We first concatenated several scenes (in a single camcorder) that people are likely to see through the use of nonlinear video editing software, and created Movie A. This allowed a 6-minute original DV-captured video clip to be summarized to 2-minute video with 10 scenes. Movie A was a typical amateurs' home video having zoomed up on children all the time. Then, we created Movie B by partially modifying Movie A, where we randomly selected five scenes out of Movie A and replaced them with other five scenes extracted from other DV camcorders. The recording date and time of both the old and new scenes were identical.

We surveyed 12 subjects after showing both Movie A and B to them. Six subjects saw movies Movie A and B; other six subjects saw Movie B and A. Each question was evaluated using 5-point scale, and tested using the Wilcoxon's signed rank sum test.

The results as shown in Table 1 indicate Q1 through Q5, Q9, and Q10 were significantly different, and indicate the preference of Movie B. Especially, the results of Q1, Q2, and Q10 express the vast superiority of Movie B, and indicate the advantages of using MediaBlocks. On the other hand, Q6 through Q8 show no significant differences. As mentioned above, we randomly replaced five scenes of Movie A and created Movie B regardless of the risk of breaking context. This might have partially reduced the score of naturalness and easy viewing.

In the next experiment, we had three subjects edit video clips with MediaBlocks, and interviewed them after editing. Before subjects started the video editing, we advised to them not to concatenate video clips of similar size or similar angle and to insert another user's clip between such similar clips.

As a result of the interviews, on the whole, examinees gave us favorable comments like "I was able to

Table 1 A comparison of videos created from single and multiple camcorders.

Q#	Question	Ave. Score of Movie A	Ave. Score of Movie B	Probability (two-side test)
Q1	Is the movie funny?	2.42	3.92	**0.00195
Q2	Is the movie boring?	3.58	1.92	**0.00195
Q3	Is the movie uplifting.	2.25	3.92	**0.00586
Q4	Could you follow the situation?	2.25	4.25	**0.00391
Q5	Could you see the whole picture?	2.58	4.25	**0.00586
Q6	Does the movie appear natural?	3.42	3.42	1.00000
Q7	Is the movie easy to view?	3.67	3.50	0.94531
Q8	Could you see the children enjoying themselves?	3.25	3.58	0.38281
Q9	Did you want to watch the movie again?	1.83	3.25	*0.01172
Q10	What was your overall impression?	2.17	3.92	***0.00098
(n=12, ***: p<0.001, **: p<0.01, *: p<0.05)				

create funnier videos than I had expected with a minimum of fuss.", "It was fun to choose clips from a wide variety of video clips from multi-camera sources.", "It was very good for me to get excellent scenes of my daughter I could not shoot myself at the festival. Thanks!" However, we found that almost all of the video clips uploaded were children-centered video shots, and there were only a small number of clips uploaded from the second unit. In this experiment, since all the cameramen were parents or relatives of the children, they were more likely to concentrate on watching their children through the camera-viewer than recording better video. This implies that it would be better to arrange a small camera-crew, maybe school teachers, who mainly record background shots.

5. Conclusion

Recently, portable digital video cameras are widely used for corporate and sports events, wedding party, kid's party, and so on. We think MediaBlocks is suitable for a wide variety uses to create high quality video.

Wiki is now a popular tool that allows users to freely create and edit Web page, and is used by many online communities. Based on this idea, we are interested in developing MediaBlocks as a tool for creating an online community of people who have videotaped in the same event. However, privacy issues need to be considered if private video files are available to the public. As far as uploaded video files are available to people who know one another, privacy issues may be ignored, but if third person can view the uploaded the movies, portrait rights and copyright of the recorded video need to be managed. These issues are the basis of our future work.

References

[1] http://www.iec.ch/, ISO/IEC 61834: Recording - Helical-scan digital video cassette recording system using 6,35 mm magnetic tape for consumer use (525-60, 625-50, 1125-60 and 1250-50 systems).

[2] Girgensohn, A., Shipman, F., Wilcox, L.: Hyper-Hitchcock: Authoring Interactive Videos and Generating Interactive Summaries, Proc. ACM Multimedia 2003. pp. 92-93 (2003).

[3] Casares, J., Long, A. C, Myers, B. A., Stevens, S. M., Corbett, A.: Simplifying Video Editing with Silver., Proc of ACM CHI'2002 (Interactive Poster Abstract) pp. 672-673 (2002).