Development of example-based interactive video remixing tool

Naoko NITTA
Osaka University
naoko@comm.eng.osaka-u.ac.jp

1. Project Goal

Video camcorders are becoming increasingly popular for recording personal events. However, people rarely watch the recorded video clips afterward since interesting clips are often intermixed with many longer, less interesting clips. Video remixing, which is to create a new video by mixing a sequence of selected video clips with other media streams such as music clips and video transition effects, is one of the ways to re-purpose such video clips to make them more enjoyable to viewers. Especially, in order to create a video remix with good expressive quality, what kinds of video clips are sequentially arranged and what kinds of music clips and video transition effects are synchronized with the video clips need to be considered, which is not an easy task for average users without any special technical or artistic skills. Thus, the aim of this project is to develop a tool which provides suggestions to users for selecting and arranging video clips, music clips, and transition effects. Existing related work has been trying to explicitly incorporate heuristically determined rules into automatic/semi-automatic video remixing methods. On the other hand, this project aims at developing a video remixing tool which gives suggestions to users about the selection and arrangement of video/music clips and video transition effects based on the temporal and spatial structural patterns which are automatically learned from professionally created video remix examples. Toward this goal, we firstly formulate the problem of video remixing by examining the structures of professionally created video remixes and then propose a video remixing approach composed of 5 processes, each of which is solved based on video remix examples. Finally, we develop a user interface where a user can interactively create a video remix.

2. Technical breakthrough

1) Problem Formulation

A professionally created video remix can be considered as a sequence of L video shots, where each video shot is an excerpt from an original video clip. Generally, a series of shots unified by their content constructs a video scene, so that a video remix can also be considered as a sequence of D video scenes. A music clip is often mixed to a video scene so as to maintain the feeling of continuity in the scene. Moreover, a video transition effect is inserted between each pair of consecutive video shots. Therefore, the target of the video remixing is, given a set of M video clips, N music clips, and K video transition effects, to create a sequence of L video shots, which is a sequence of D video scenes, to mix D music clips to the video scenes, and to select L-1 video transition effects for the L-1 video shot boundaries.

2) Example-based Video Remixing

The target of our proposed example-based video remixing tool is to generate a new video remix based on the patterns learned from H video remix examples, where the h th video remix example can be considered as a sequence of I_h video shots, C_h music clips, and I_h-1 video transition effects. We adopt a video-driven approach, that is, a video shot sequence is firstly created, and then the music clips and the video transition effects are mixed to the generated video shot sequence. Figure 1 shows the overview of the proposed approach composed of the following 5 processes.

I) Video Clip Sequence Creation

Hidden Markov Model (HMM) is used to learn the common temporal structure of H sequences of I_h video shots based on their aesthetic attributed features. Then, a video remix

template which represents a sequence of L video shots constructing D video scenes can be generated from the learned HMM. Candidate video clips can be selected from a given set of M video clips for each of the L video shots based on their similarities of aesthetic attributed features. The user can select a video clip from the selected candidate video clips.

II) Music Clip Selection

Assuming that the compatibility of video scenes and music clips is related to the correlations between their visual and audio features, we firstly learn a non-linear mapping of the music feature space by Relevance Learning based on the three-layered perceptrons, so that the distances among the music clips in the mapped music feature space would correlate to the distances among their mixed video scenes in the video scene feature space for the $H \times C_h$ pairs of video scenes and music clips of video remix examples. Then, after mapping the given set of N music clips to the mapped music feature space, candidate music clips are selected for each video scene created in Step *I)* based on their similarities to the example video scenes and music clips. The user can select a music clip from the selected candidate music clips.

III) Music Boundary Extraction

Inhomogeneous Markov Chain Model (MCM) is used to learn the characteristic auditory changes

around the $H \times (I_h - 1)$ boundaries between consecutive video shots of the video remix examples. Then, the total of L-1 music boundaries which fit the learned patterns are extracted from the D music clips selected in Step *II)*.

IV) Video Shot Extraction

HMM is used to learn the common temporal patterns of the $H \times I_h$ video shots of the video remix examples based on their frame-level visual and audio features. Then, a video shot which fits the learned patterns is extracted from each video clip selected in Step *I)* so that the extracted video shot would fit between the consecutive music boundaries extracted in Step *III)*.

V) Video Transition Effect Selection

Similarly to music boundary extraction, Inhomogeneous MCM is used to learn the characteristic auditory changes and the types of video shots around each type of transition effects in the video remix examples. Then, suitable video transition effects are selected for the L-1 video shot boundaries according to the auditory changes around the L-1 music boundaries extracted in Step *III)* and the types of the L video shots extracted in Step *IV)*.

3) Evaluation

The proposed video remixing approach was evaluated using 61 trailers of action movies as a set of the video remix examples. We also prepared a set of 180 music clips of various genres such as movie soundtracks, classical music, and popular music (Japanese-pop and Western-pop) and a set of 265 home video clips recording a sports field day held by a kindergarten. After creating a video clip sequence composed of $L=32$ video clips by Step *I)* of our approach (*Video A-I*) and by considering the perceived quality in a similar way to existing approaches (*Video B-I*), we applied each process *II)*--*V)* to generate *Videos A-II*--*A-V* and *B-II*--*B-V*. Since only cuts were selected in the process *V)*, 8 video remixes in total excluding Videos *A-V* and *B-V* were presented to 8 subjects, who evaluated the expressive quality of each video remix on a scale of 0-10 with 0 being very bad and 10 being very good. From the results shown in Table 1, we can see that, although the video clip sequence created by our approach ranked the lowest 3 in the scale of 10, the ranking of the video remixes after mixing music clips improved greatly from 3 to the highest 6.2 because the created video clip sequence and the selected music clips were mutually beneficial in increasing the impression of action. This result indicates that the example-based approach can successfully support users in creating video remixes of better expressive quality even with a limited number of video remix examples and the video clip sequence creation and music clip selection are the two most important processes in video remixing.

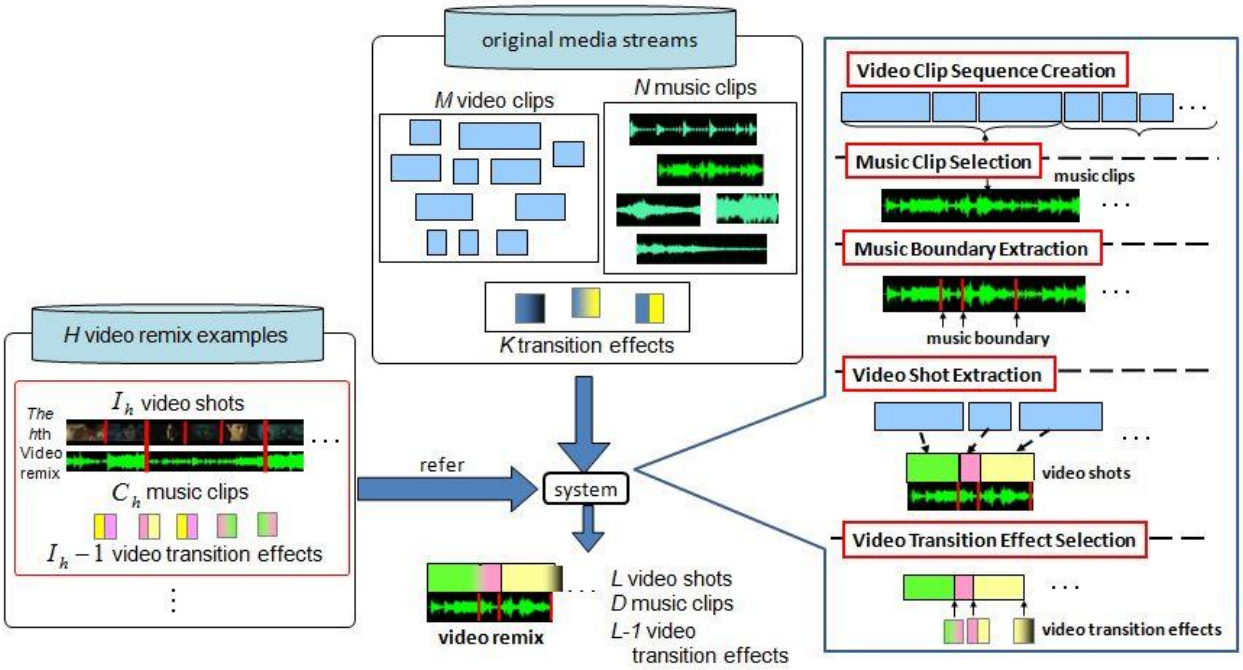

Figure 1: Overview of Proposed Approach

Video Remixes	A				B			
	I	II	III	IV	I	II	III	IV
Length (min:sec)	0:36		0:43		10:56		10:59	
scores	3	5.3	5.8	6.2	3.5	3.8	3.7	3.9

Table 1: Results of Subjective Evaluations

3. Innovative Applications

In order to support users and simultaneously to give them certain amount of control in the two most important processes: video clip creation and music clip selection, a user interface as shown in Fig. 2 was developed. The video remix template created in Step *I)* is presented so that a user can select a video shot in the template to select a suitable video clip for the shot. In order to present video clips in a limited display space by considering their temporal order, a 3-dimensional book-style video clip presentation panel is developed, where each page presents a set of consecutive video clips recording the same event or scene. The pages are then assembled in their temporal order. A video clip is represented by a key frame and is arranged on a page according to its suitability to the video shot selected by the user and its perceived quality related to camera motions and lighting conditions as shown in Fig.3. More specifically, the video clips with the high perceived quality and suitability are placed around the fore edge of the page while those with the low perceived quality and suitability are placed around the spine of the page. Therefore, a user can select video clips from the candidate video clips presented in the fore edges of the pages. After selecting all video clips for a video scene, the user can select the corresponding video scene in the video remix template to obtain a list of candidate music clips suitable for the selected video clips as a result of Step *II)*. By using this user interface, average users without any special technical or artistic skills can be supported in efficiently and effectively create a video remix from their sets of video clips and music clips.

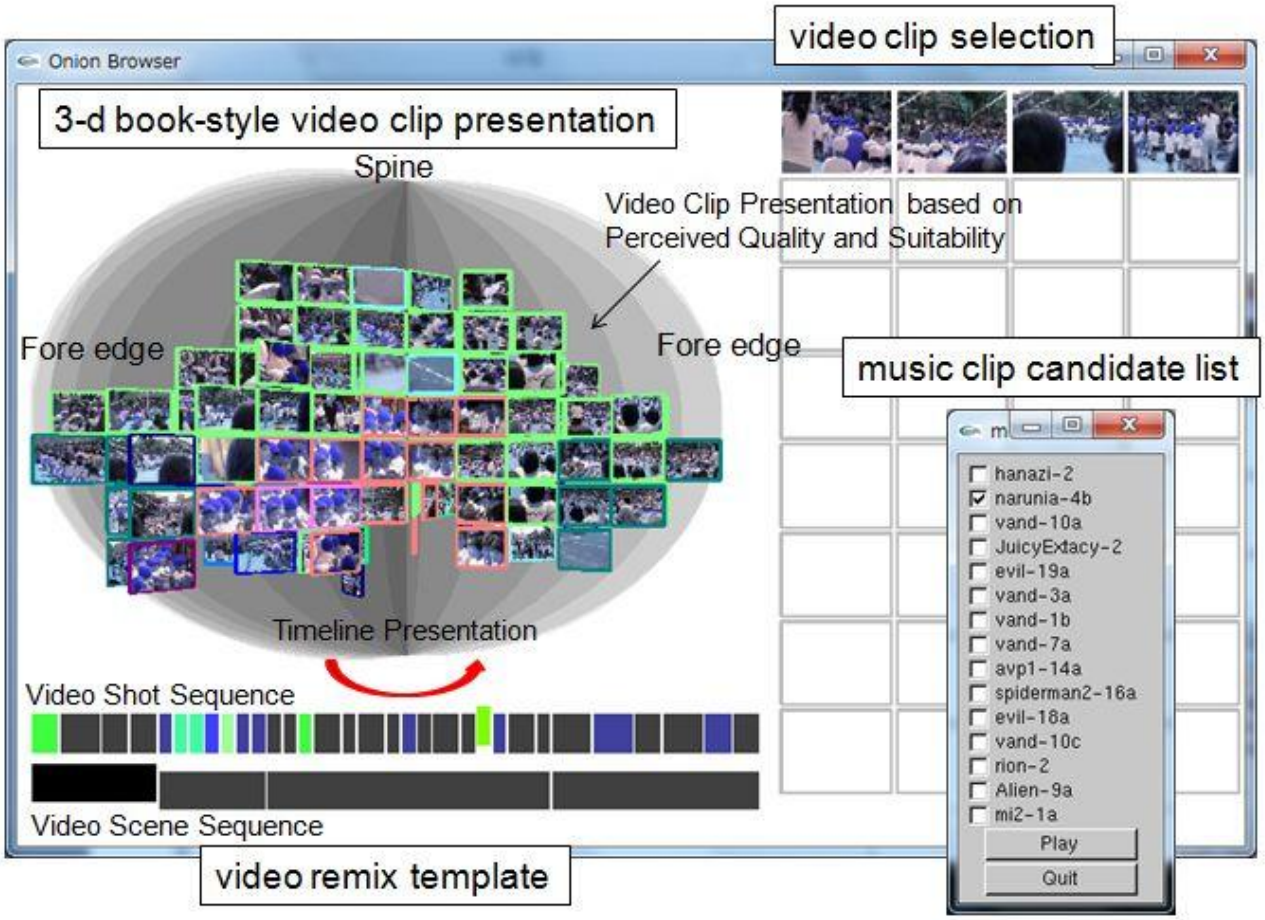

Figure 2: User Interface

4. Academic Achievement

The papers which propose the video remixing approach were published in 1 international journal and 1 domestic conference. The paper which proposes the user interface for video remixing support was accepted for International Conference on Multimedia & Expo (ICME2011) as one of the 59 oral presentations selected from 744 submitted papers.

Figure 3: How to arrange video clips on a page

5. Achievement in Talent Fostering

Jang-il Kim, received his master degree from Graduate School of Engineering, Osaka University for his work on the music clip selection. Yusuke Kumihashi also worked on slide-show generation by using images collected from the Web as an extension of this project and received his bachelor degree from Osaka University School of Engineering.

6. Project Development

All processes will be integrated into the user interface and more evaluations using various types of video remix examples will be conducted.

7. Publications

Paper publication

1) **Naoko Nitta and Noboru Babaguchi, Example-based Video Remixing, Multimedia Tools and Applications, vol.51, no.2, pp.649-673, 2011**
2) **Jang-il Kim, Naoko Nitta, and Noboru Babaguchi, Mixing Music with Video Shot Sequence based on Exemplars, IEICE Technical Report, PRMU2010-294, pp.335-340, 2011**
3) Yusuke Kumihashi, Naoko Nitta, and Noboru Babaguchi, "Example-based Image Selection for Generating Slideshows," IEICE General Conference, D-12-90, p.193, 2011
4) **Naoko Nitta and Noboru Babaguchi, Example-based Video Remixing for Home Videos, International Conference on Multimedia & Expo, 2011 (accepted)**
5) Naoko Nitta and Noboru Babaguchi, Example-based Home Video Remixing Support System, ACM Multimedia 2011 (submitted)