Automatic Music Video Generating System by Remixing Existing Contents in Video Hosting Service Based on Hidden Markov Model

Hayato Ohya
Waseda University
Shinjuku, Tokyo, Japan
hayato-o@ruri.waseda.jp

Shigeo Morishima
Waseda University
Shinjuku, Tokyo, Japan
shigeo@waseda.jp

1. Introduction

User-generated music video clip called *MAD movie,* which is a derivative (mixture or combination) of some original video clips, are gaining popularity on the web and a lot of them have been uploaded and are available on video hosting web services. Such a MAD music video clip consists of audio signals and video frames taken from other original video clips. In a MAD video clip, good music-to-image synchronization with respect to rhythm, impression, and context is important. Although it is easy to enjoy watching MAD videos, it is not easy to generate them. It is because a creator needs high-performance video editing software and spends a lot of time for editing video. Additionally, a creator is required video editing skill. DanceReProducer (Nakano et al [2011]) is a dance video authoring system that can automatically generate dance video appropriate to music by reusing existing dance video sequences. It trains correspondence relationship between music and video. However, DanceReProducer cannot train video sequence information because it only trains one-bar correspondence relationship. So we improved DanceReProducer to consider video sequence information by using Markov chain of latent variable and Forward Viterbi algorithm.

2. System Overview

Figure 1 shows our system's flow chart. Our system consists of three parts. The first is database construction part. The second is model training part. The last one is video generation part.
In the database construction part, music feature and video feature are extracted from existing video contents. Then each feature is collected per one-bar.
In the model training part, each feature is clustered for Hidden Markov Model (HMM) and HMM parameters are trained by Markov chain.
In the video generation part, music feature and tempo are extracted from music we want to attach video on (input music) and feature is collected by bar-level feature. Then decide video sequence by Forward Viterbi algorithm and using HMM parameter calculated from previous part. Considering state of previous bar video feature in addition to current bar music feature in estimating one-bar feature of video enables video generation training video sequence.

3. Training by Markov Chain

Training model of music and video feature is Markov chain of latent variable. Automaton is separated per one-bar, observation X is bar-level music feature and latent variable Z is bar-level video feature. Additionally, state Y is the cluster of bar-level video feature that is clustered by k-means clustering. Initial state probability π_i and state transition probability a_{ij} from i to j are calculated

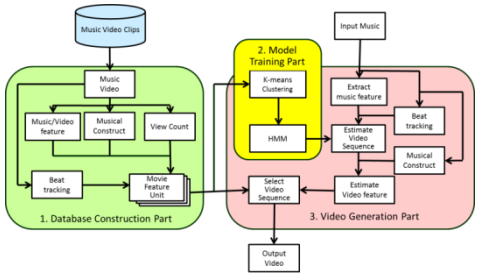

Figure 1. *System's flow chart.*

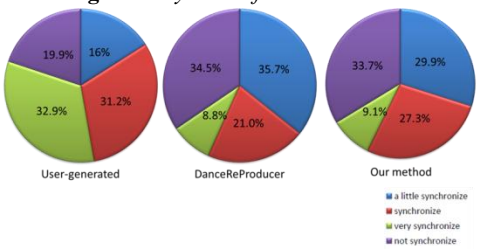

Figure 2. *Percentage of all video time of each assessment time.*

4. Subjective Assessment Experiment

We generated each four videos from four input music by Dance-ReProducer and our method. Number of videos in database is 90. Then we prepared 12 videos in total in addition to four user-generated videos that are generated by using the same music as input music. View count of the user-generated videos is over 50,000. This number can be considered as high-quality video.
The examinees who are 15 men in 20s watched 12 videos and assessed synchronization-level between music and video. Synchronization-level is four step values, in order of increasing level; "not synchronize", "a little synchronize", "synchronize", "very synchronize". Assessment data has time information and its assessment.
Figure 2 shows percentage of all video time of each assessment time in the experiment.

5. Result and Conclusions

Our method enabled video generation considering not only music feature but also video sequence information. At the result of subjective assessment experiment, comparing our method with DanceReProducer, it turns out that "synchronize" has increased and "a little synchronize" has decreased. This verified that considering video sequence information increases synchronization between music and video. Remaining issues, such as feature extraction in detail, will be topics in our future work.

References

T. NAKANO, S. MUROFUSHI, M. GOTO, AND S. MORISHIMA. 2011. DanceReProducer: an Automatic Mashup Music Video Generation System by Reusing Dance Video Clips on the Web. *Proceedings of the SMC 2011*, pp.183–189, July 2011.

SIGGRAPH 2012, Los Angeles, California, August 5 – 9, 2012.
ISBN 978-1-4503-1435-0/12/0008