Generative Methods for Automated Music Video Editing

Julia Stefan

Generative Methods for Automated Music Video Editing

Julia Stefan

Griffith University, Queensland College of Art
julia.stefan@griffithuni.edu.au

Abstract. In order to enhance viewer experience when watching music videos, a set of methods for generative music video editing has been developed. These algorithms use stochastic selection based on user-defined shot characteristics, as well as temporal constraints to manage narrative structures. The prototype application is introduced and demonstrated by showing two examples of software-generated music videos, and the basic concepts behind the algorithms will be explained.

Keywords. Music videos, generative composition, online videos, non-linear video, adaptability

1 Introduction

Music videos are a popular format for audio-visual experiments. In recent years, they have been used to explore the possibilities presented by online technologies and interactive applications through the use of moving images. Popular examples include The Wilderness Downtown[1], Like a Rolling Stone[2] and Do Not Touch[3]. In comparison to feature films for instance, the shorter running time of music videos makes them a popular format for experiments. This popularity is further increased due to the genre's aesthetic flexibility and its focus on the creation of rhythm and emotional engagement rather than narrative coherence [1].

By implementing options for user interaction, video sequences may become more personal and therefore more engaging for viewers. Interactive online music videos have been used to explore a number of different concepts for interactivity. These include game-like videos and tree-structured, choose-your-own-adventure type narratives. Interactive online music videos also allow for new possibilities of personalisation by embedding content from social networking websites and search engines. These "hypervideos" [2] generally either depend on viewer participation during playback or preliminary input in order to generate variations. Additionally,

1 http://www.thewildernessdowntown.com
2 http://video.bobdylan.com
3 http://donottouch.org

many of these videos are based on comparatively rigid structures in order to maintain coherence and to ensure that the director's concepts are still being realised.

In contrast, the research concerning generative video composition presented in this paper is based on the idea of providing variable music videos, which can be generated in real-time without the requirement of viewer interaction and which therefore depend on processes that are to some extent beyond the director's control [3]. Thus, the methods presented here aim to provide a new creative tool for video directors, as well as an engaging viewing experience for audiences.

2 Music Video Generation

This demonstration presents a prototype application of Dividation, a software for generative music video editing utilising real-time assembling of video sequences from individually prepared pools of video footage using algorithmic decision-making processes based on the creator's clip classifications.

Sequence generation is achieved using stochastic processes based on Markov chains and probabilistic structures to describe the editing dynamics. A database of sufficiently annotated original video footage is used. Assuming the individual shot to be the basic building block of these videos, a sequence is assembled shot by shot according to the video's principal key characteristics as defined by the video director.

Additional structuring methods, such as underlying ordered lists, help to ensure narrative coherence through the specification of forced shots and the definition of temporal windows within which the stochastic selections are constrained.

3 Demonstration

Two music videos are used to demonstrate the video generation method described above: *Majestic* and *Eloise*.

The music video *Majestic* consists of a simple narrative, in which the protagonist is drawn into the televised action of the musicians' performances. Six parameters are used to categorise the video content and are mapped to temporal probabilities. The mapped temporal probabilities describe the amount of visual features that should be seen during particular sections of the song. This music video uses a pool of roughly two hundred clips with a length varying from less than one second up to three minutes. An algorithmic decision-making process is then employed for both the automatic sequencing of the video footage and the cutting of the footage into shots of appropriate lengths.

In comparison, the music video *Eloise* uses a slightly more detailed narrative. Here, the protagonist, Eloise, is seen inside her home, while the musicians are performing the song inside her living room. A secondary protagonist singer urges the woman to leave her house and quit her bad habits, which she finally does. For this video, the editing dynamics are based on the use of three visual parameters and a pre-defined narrative structure. This provides an additional guide to the order of the shots. The video uses approx. 550 pre-cut clips with a length varying from less than one second up to ten seconds. In this example, the algorithmic decision-making process is only applied to the sequencing of the video content. The selected shots are played in their entirety, unless a temporal alteration is required to ensure an accurate synchronisation.

Both videos demonstrate (1) the potential of using generative composition methods for online music video creation, (2) the resulting editing aesthetics, as well as (3) some of the challenges of this process. With the use of randomness as a driving factor for sequence creation, the resulting music videos may not always be coherent to the audience and accurate enough to the editor's intentions. An appropriate management of this problem requires correct and well-balanced definitions of the probabilistic and narrative structures and constitutes a central aspect of this research.

4 Discussion and Future Outlook

Further research will focus on refining the quality of the applied editing algorithms in greater detail and extending the repertoire of specific methods by taking into account the requirements introduced by different progressive and narrative structures. Furthermore, it must deal with the creation of adequate interfaces to involve the editors in the generative composition process. Compared to traditional methods of manual video editing, the creation and editing of music videos by using generative composition methods is a rather abstract creative process because it requires a stronger conceptual involvement in order to properly communicate to the composition software how to edit a specific sequence, rather than manually executing this task oneself. Ongoing research as part of this project will reveal the effectiveness of the editing algorithms in terms of viewer engagement, and its practicality for editors.

References

1. Vernallis, C.: Unruly Media: YouTube, Music Video, and the New Digital Cinema. Oxford University Press, New York (2013).
2. Hammoud, R.I.: Interactive Video: Algorithms and Technologies. Springer, Berlin (2006).
3. Boden, M.A., Edmonds, E.A.: What is generative art? Digital Creativity 20/1-2, 21-46 (2009).