VRAPS: VISUAL RHYTHM-BASED AUDIO PLAYBACK SYSTEM

Trista P. Chen, Ching-Wei Chen, Phillip Popp, Bob Coover

Gracenote, Inc.
{tchen,cwchen,ppopp,bcoover}@gracenote.com

ABSTRACT

VRAPS (Visual Rhythm-based Audio Playback System) is an interactive multimedia application that uses a novel visual rhythm detection technique to allow a user to control the playback speed of an audio signal by moving in front of a video camera. As traditionally defined in the context of music, a beat represents a distinctive musical event such as the hitting of a drum or the start of a new melodic note. Similarly, we define a visual beat to indicate distinctive visual events such as rhythmic gestures and dance motions. In this paper, we propose a method for automatically detecting visual beats and tempo from a video signal. As a demonstration of this technique, we present an interactive, real-time audio playback system that continuously adjusts the playback speed of an audio signal based on the visual rhythm detected from a real-time video capture device.

Keywords—visual beats, visual rhythm, visual tempo, novelty detection, motion analysis, video signal processing, audio tempo, beat detection, time-stretching, audio signal processing

1. INTRODUCTION

In the context of music, rhythm may be defined as the regular or semi-regular pattern of pulses resulting from significant melodic, harmonic, or percussive events. These events are also called beats, and are usually characterized by a pulse, or rapid increase in energy in the audio signal. In addition, a tempo may also be defined as the rate of repetition of these beat events. Much research has been done in the area of musical beat detection and tempo estimation to automatically extract these events and repetition rates from audio signals [1][2].

In this paper, we define visual rhythm as a regular or semi-regular pattern of pulses resulting from significant visual events. Visual rhythm similarly consists of visual beats, which are the significant visual events, and visual tempo, which is the beat repetition rate. One example of visual rhythm is illustrated by observing a dance performance, where we can perceive visual beats and tempo from the rhythmic movements of the dancer.

This paper is organized as follows. In Section 2, we describe our visual rhythm detection algorithm, including video feature extraction, novelty function, and beat detection. In Section 3, we describe a demonstration system named VRAPS (Visual Rhythm-based Audio Playback System), which uses our visual rhythm detection technique to allow a user to control the speed of music playback by making rhythmic motions in front of a video camera. We then summarize the paper in Section 4.

2. VISUAL RHYTHM DETECTION

The key observation guiding our proposed method for visual rhythm detection is that visual beats are perceived when there is a significant and rapid change in some aspect of the visual content. These changes may be a result of different visual events, such as lighting changes, scene transitions, camera motion, or motion of people or objects. An example of visual beats can be seen in the rhythmic movements of a dancer. Dance movements are often characterized by coordinated movement of one or more objects in a limited number of directions. In this context, visual beats are usually perceived at times when this motion changes abruptly, such as when motion starts, stops, or changes direction. Our algorithm attempts to automatically detect where these significant changes occur in a video signal, as summarized in Figure 1.

To do this, we first extract video features that describe the motion characteristics of a video signal. The features used in the current implementation include a 2-dimensional histogram of an optical flow [6] feature, and a foreground mask [7] feature. Use of other visual features is also possible as long as it captures the motion change of the video. We then analyze these 2-dimensional features to derive a 1-dimensional novelty function. In the current implementation, we use a variation of the 2-dimensional moment. The final stage is to analyze how the novelty...
function changes over time, and detect significant fluctuation events as our visual beats. Our algorithm for this stage includes smoothing out very rapid changes in the novelty function, then taking the derivative to find periods of greatest changes, and then using peak-picking techniques to detect local extrema or peaks, which are our visual beats. In addition to detecting the locations of visual beats, we also use the novelty function, as well as the detected beat locations, to estimate the rhythmic repetition rate of the visual beats, or visual tempo.

As mentioned earlier, visual beats may be perceived as a result of significant visual events other than motion, such as scene transitions, sudden camera motion, or lighting changes. Prior work [3][4][5] has focused only on features derived from human motion. Although the VRAPS (Visual Rhythm-based Audio Playback System) demonstration presented here uses a setup with a fixed camera and consistent environmental lighting, the underlying visual rhythm detection algorithm takes into account all the visual events mentioned.

3. VRAPS DEMONSTRATION SYSTEM

In order to demonstrate our visual rhythm detection algorithm, we have developed VRAPS (Visual Rhythm-based Audio Playback System), an interactive audio playback application that allows a user to control the speed of a selected audio file by dancing or performing other rhythmic motions in front of a video camera. The system diagram is presented in Figure 2.

The user first selects an audio file that they wish to listen to. Conventional audio rhythm detection techniques (including beat detection and tempo estimation) are applied to extract the audio rhythm of the chosen audio file. The system then plays the audio file through speakers or headphones, at its original speed.

The user can then dance or make other rhythmic movements in front of a video camera. The system will use the algorithm described in the previous section to detect, in real-time, the visual rhythm present in the user’s motion. This visual rhythm is compared with the detected audio rhythm to determine how well they are synchronized. The audio signal is then either sped up or slowed down using conventional audio time-stretching techniques [8] so that the audio rhythm is synchronous with the detected visual rhythm.

The time-stretching factor is updated on a continuous basis, depending on the instantaneous visual rhythm detected from the user’s movements. Therefore, the user can interactively control the playback speed of the audio by changing the speed of their dancing motion.

The VRAPS system is implemented as a Windows application running on standard PC hardware. A consumer-grade webcam is used as a video capture device. The OpenCV software library is used to access the incoming video signal, as well as to perform some of the video feature extraction. Custom software was developed in C++ to perform the visual rhythm detection, audio rhythm detection, A/V synchronization, audio time-stretching and playback.

4. SUMMARY AND FUTURE WORK

We proposed a new concept of visual rhythm as a pattern of repeating visual events in a video signal. An interactive audio-visual rhythm matching system called VRAPS (Visual Rhythm-based Audio Playback System) was demonstrated, where audio playback speed is adjusted dynamically according to the real-time detected visual rhythm. Future applications include applying the proposed visual rhythm detection algorithm to non-fixed camera video inputs, and to music videos with creative camera and lighting changes.

5. REFERENCES