

Movie Emotional Event Detection based on Music Mood and Video Tempo

Yu-Hao Chen, Jin-Hau Kuo, Wei-Ta Chu, and Ja-Ling Wu

*The Communications and Multimedia Laboratory,
Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan,*

Summary

Movie is one of the most important entertainments in the history of mankind. Being one of the hugest industries in this world, movie industry releases around 4500 movies in the world every year, containing approximately 9000 hours of digital movie contents. As movies have made a lot effect to life and thinking of people, there are many great movies worthy to be seen repeatedly. However, as movies are released continuously, it's more and more difficulty to see every movie from the beginning to the end again. And with the development of content digitalization, it is easier to access the digital content on the Internet now. In addition, more and more movie viewers would like to choose parts of movies according to their favorites. For example, they may want to find their favorite actors, some sad scenes, the best fighting scenes in the movie, and so on. One system that automatically analyzes the movie related data and provides tools for detecting events is of great interest.

In this paper, we present our works about movie analysis by detecting video tempos and music moods. Our work consists of three parts: music mood detection, tempo detection, and integrated applications. The block diagram of the proposed framework is illustrated in Figure 1.

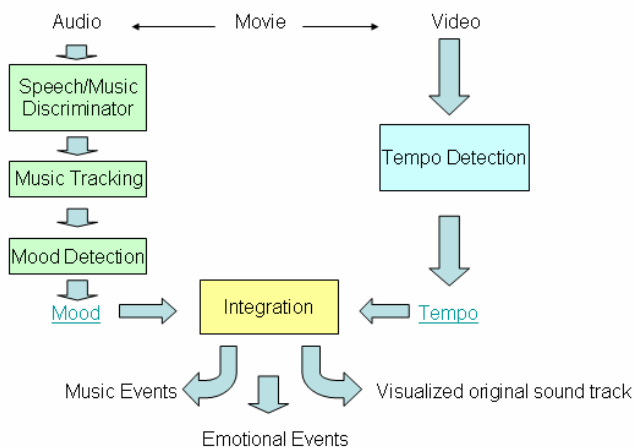


Fig.1. The framework of this paper.

In music mood detection part, we develop three models: speech/music discriminator, music tracking, and mood detection (c.f. Fig.1). In speech/music discrimination, we use the characteristics that there are pauses between syllables and words in speech to classify speech and music. We implement and modify speech/music discriminator proposed by C. Panagiotakis and G. Tziritas [1]. Two audio features, Root Mean Square (RMS) value and Zero Crossing (ZC) rate, are extracted to characterize speech and music. Because music is the main concern of our work, we recognize segments mixed with music and speech as music segments. Since segments whose amplitudes of speech much louder than music are still detected as speech, we developed a Hidden Music Segment Detection (HMSD) method to deal with this situation. Experimental results [4] show that precisions and recalls are improved after using HMSD, and our experimental results show much better detection performance than other proposed methods.

Because music segments with different compositions may be merged together in speech/music discrimination, the music tracking module (c.f. Fig.1.) is used to find the boundaries of different compositions. Timbre and intensity features are adopted to detect changes in audio amplitude and timbre. We modify the mood tracking in [2] to fit the usage of our movie analysis.

After music tracking, we detect music mood in each music segment. Timbre and rhythm features are used, and Gaussian Mixture Model (GMM) is chosen as our training algorithm. Music mood detection is not easy, and two issues make it more difficult. First, classes of music in movies are unlimited. Second, there are many sound effects and speeches in movies. Precisions and recalls in our experimental results are about 40%~60% [4]. We, therefore, modify our mood detection module with user intervention.

In tempo detection part, we incorporate our previous work [3] into our framework. Shot length and motion activity are used to detect tempos of a video. These two video features are selected based on montage editing and motion activity. After tempo detection, each shot has a tempo value to represent its tempo activity.

Applications:

On the basis of music mood and video tempo, we construct three applications: music event detection, emotional event

detection, and Original Sound Tracks (OST) visualization. Music event detection is used to detect events with music. One functionality of music is building up atmospheres. Therefore, in some kinds of movies such as romance ones, viewers will pay more attention to segments with music.

Emotional event detection:

Emotional event detection provides attractive and representative events in movies. We combine tempo and music mood together to find shots which are believed to greatly affect our mood. Because music is our main concern, the emotional event detection is based on music events. There are two principles when we design the weighting functions. First, we focus on shots that their music mood and video tempo have the same trend. Both high tension music and high tempo video can make people nervous or exciting. If a shot with high tempo is accompanied with high tension music, we can consider it as a shot that make people extremely high. A shot with extremely emotion is considered as an attractive shot, which is assigned with higher priority than other shots. The second principle is the notice-ability of music. Because one of music's functions is to build up atmosphere, it may be used to "notice" the appearance of shots. For example, if one person walked happily but accompanied with tragic music playing in the background, it may hint viewers that some disaster will happen. Music can tell us that some event with specified mood will happen. According to these two principles, we detect some attractive shots and reorganize them into emotional events. We sort these emotional events by their weighting values. Then several events with best weighting values are selected and presented to users.

Original Sound Tracks (OST) visualization:

OST visualization is an application developed on the basis of music of movie and emotional events. Users can choose a composition from OST, and our system will combine it with some of the most attractive events. The purpose of this application is to let people recall important contents of movies. Because OST contains the best compositions of a movie, we use it to present music of movie. There are two issues to construct OST visualization. First, we need to detect attractive events in movies. We have proposed a detection method to deal with this issue. Second, we have to trim events to fit the length of OST. After merging and composing trimmed events and the compositions selected from OST, we have visualized OST. The MTV-like clips let people recall their memories about the movie while seeing it.

The experimental results of our work are evaluated by subjective tests. One can visit the link <http://www.cmlab.csie.ntu.edu.tw/~ulyness/ICCE2006/> to viewing the resultant video clips. Eleven people are invited to judge the detected events and give subjective grades. Mean opinion scores from 1 to 5 are used to quantify the performance. Higher scores indicate better performance. The grades scored by viewers who have seen the specified movies are listed in Tables 1 to 4. The "1st sequence" in these tables means the most emotional event after sorting detected events

by mean opinion scores in descending order. Tables 1 to 3 are the experimental results of emotional event detection, and Table 4 is the experimental results of OST visualization.

TABLE 1. SUBJECTIVE TEST RESULTS OF THE EMOTIONAL EVENT DETECTION FOR THE MOVIE "TITANIC"

Movie "Titanic"		
1 st sequence	2 nd sequence	3 rd sequence
4.9	4.5	4.4

TABLE 2. SUBJECTIVE TEST RESULTS OF THE EMOTIONAL EVENT DETECTION FOR THE MOVIE "GHOST".

Movie "Ghost"			
1 st sequence	2 nd sequence	3 rd sequence	4 th sequence
4.0	3.5	2.0	4.5

TABLE 3. SUBJECTIVE TEST RESULTS OF THE EMOTIONAL EVENT DETECTION FOR THE MOVIE "MY SASSY GIRL".

Movie "My Sassy Girl"			
1 st sequence	2 nd sequence	3 rd sequence	4 th sequence
4.6	3.1	4.1	4.6

TABLE 4. SUBJECTIVE TEST RESULTS OF THE OST VISUALIZATION FOR THE MOVIE "MY SASSY GIRL" AND MOVIE "TITANIC".

Name	Movie "My Sassy Girl"	Movie "Titanic"
Average Score	4.2	4.7

The subjective test results of emotional event detection show that though there are some false alarms, the proposed method effectively finds attractive events from movies. The experimental results of OST visualization are also scored well. In the future, some issues or works will be studied to enhance the system. We need to improve the mood detection module such that it detects music mood fully automatically, and cooperate with other modules which can detect moods in other modalities such as mood detection of speech and facial expression detection.

REFERENCE

- [1] C. Panagiotakis and G. Tziritas, "A speech/music discriminator based on RMS and Zero-Crossings", *IEEE Trans. Multimedia*, vol.7, no.1, February 2005
- [2] Dan Liu, Lie Liu, and Hong-Jiang Zhang, "Automatic Mood Detection from Acoustic Music Data", *ISMIR-03*.
- [3] Hsuan-Wei Chen, Jin-Hau Kuo, Wei-Ta Chu, and Ja-Ling Wu, "Action Movies Segmentation and Summarization Based on Tempo Analysis", *MIR'04*, October 2004.
- [4] Yu-Hao Chen and Ja-Ling Wu, "Movie Emotional Event Detection based on Music Mood and Video Tempo", NTU CSIE.