MIREX 2011 SUBMISSION – COMBINING VISUAL AND ACOUSTIC FEATURES FOR MUSIC GENRE CLASSIFICATION

Ming-Ju Wu, Jia-Min Ren,

Department of Computer Science National
Tsing Hua University
Hsinchu, Taiwan
{brian.wu, jmren}@mirlab.org

ABSTRACT

The system uses two types of effective features for genre classification. The visual features that can capture the characteristics of a spectrogram's texture patterns. On the other hand, acoustic features are extracted using universal background model and maximum a posteriori adaptation. Based on these two types of features, we then employ SVM to perform the final classification task

1. INTRODUCTION

Since the effectiveness of GSV (Gaussian Super Vector) has been proven in MIREX 2009 [1,2], here we incorporates visual features and GSV for genre classification. This system was described in [3]. For detail explanation, please see the original paper.

2. ACOUSTIC FEATURES

Here we follow the method in [2]. First of all, a universal background model (UBM) is trained from a huge music dataset by using a Gaussian mixture model (GMM) to represent the common distribution of short term features (e.g. MFCCs). The music collection consists of nearly 2000 music clips over different genres. The number of Gaussian mixture component is set to be 30.Next, for a particular music clip, we take the UBM as a prior distribution and use maximum a posterior (MAP) adaptation to establish the corresponding GMM. Thus each music clip can be represented by a set of GMM parameters called GSV.

3. VISUAL FEATURES

We convert each music clip into spectrogram via STFT and perform Gabor filtering to extract visual features. The spectrogram is first divided into the following octave-based subbands: 0~200Hz, 200~400Hz, 400~800Hz, 800~1600Hz, 1600~3200Hz, 3200~8000Hz, and 8000~11025Hz. That is, the original spectrogram image is divided into 7 sub-images. Second, we construct a Gabor filter bank with 6 orientations and 5 scales. Then, each sub-image is filtered with Gabor filter bank. Finally, the mean and standard deviation of the filtering result are used as the visual features. Figure 1 shows an example of filtering results.

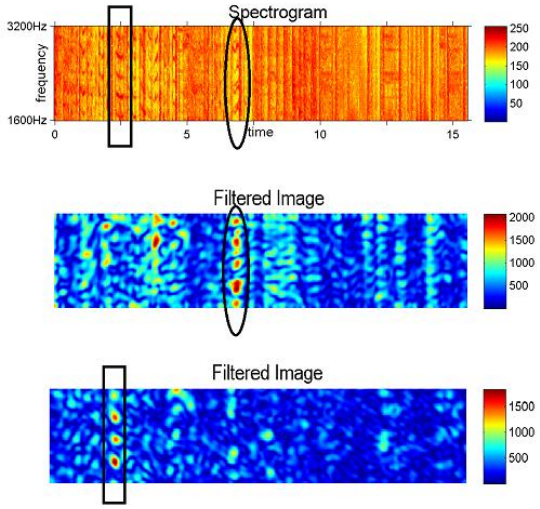

Figure 1. An example showing the effectiveness of Gabor filtering. The top figure is the original spectrogram within the 1600~3200Hz range; the middle and bottom figures are the filtered results with 30 and 120 degree orientations respectively.

4. REFERENCES

[1] 2009: Audio Genre Classification (Mixed Set) Results, Available: http://www.music-ir.org/mirex/wiki/2009:Audio_Genre_Classification_(Mixed_Set)_Results

[2] C. Cao and M. Li, "Thinkit's submission for MIREX 2009 audio music classification and similarity tasks," Available: http://www.music-ir.org/mirex/abstracts/2009/CL.pdf

[3] M. Wu, Z. Chen, J.R. Jang, J. Ren, "Combining visual and acoustic features for music genre classification," in *10th International Conference on Machine Learning and Applications*, 2011, pp. 124–129.