MIREX 2013 Submissions for Train/Test Tasks (Draft)

Ming-Ju Wu

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

ABSTRACT

Our submissions apply two types of features, including visual features and acoustic features. The visual features can capture characteristics of a spectrogram's texture from both local and global views. On the other hand, acoustic features are used to represent global timbre characteristics. The two types of features are concatenated to a single long feature vector. Then SVM is used for classification.

1. INTRODUCTION

Our submissions in MIREX 2011 are based on our previous work [1], which combines the acoustic based GSV (Gaussian Super Vector) [2][3] and song-level-based texture features to form a new feature set. This feature set can be considered as effective since it won the first place for the MIREX 2011 genre classification task [4]. In our MIREX 2012 submissions, the feature set contain GSV, song-level-based and beat-level-based texture features. As for our MIREX 2013 submission JJ1, we add the beat-level-based heterogeneity features on the same feature set that were used in MIREX 2012. The new features provide heterogeneity measure of spectrogram's texture of IBIs (Interbeat Intervals). Besides, our MIREX 2013 submission JJ2 is the same as our MIREX 2012 submission, which is used for comparison. The rest of this extended abstract is organized as follows: Section 2 introduce the acoustic feature. Section 3 briefly describes the proposed visual features. Experimental results are shown in Section 4.

2. ACOUSTIC FEATURES

The GSV is applied as our acoustic feature, since it demonstrated the discriminative power of previous MIREX competition [2]. Here we follow the method in [3]. 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

The proposed visual features include song-level-based texture features, beat-level-based texture features, and beat-level-based heterogeneity features. The flowchart of the proposed visual features is shown in Fig. 1. First, we convert each music clip into spectrogram via STFT and perform Gabor filtering. The spectrogram is 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 the Gabor filter bank.

For song-level-based texture features, the mean and standard deviation of Gabor filtering results are used as features. For beat-level-based texture features, we apply a beat tracker [5] to divide Gabor filtering results of the spectrogram into IBIs. Then we also use mean and standard deviation for these IBIs as features. For beat-level-based heterogeneity features, they can be extracted from the SSM (Self-Similarity Matrix) of the IBIs' texture. This feature can be used to describe the heterogeneity of spectrogram texture of each octave-based subband.

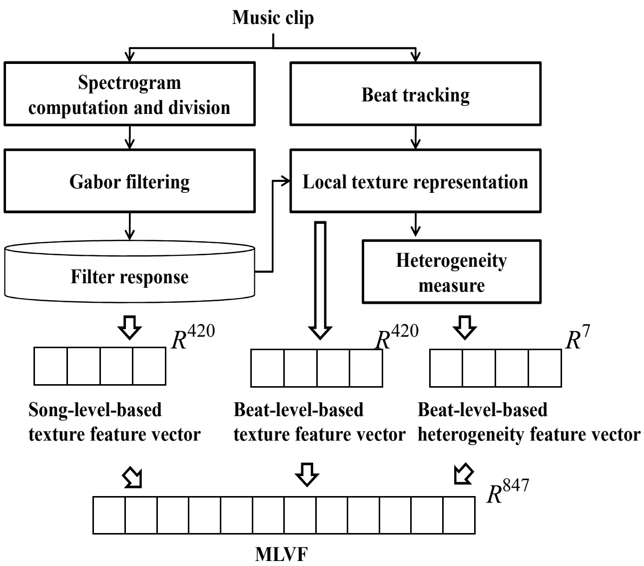

Figure 1. Flowchart of the proposed visual features.

4. RESULTS

The results of our submissions are shown from Fig. 2 to Fig. 4. Experimental results show that our submissions achieve the best accuracy for genre classification (mixed) and mood classification tasks. In both tasks, our JJ1 perform better than JJ1, demonstrating the usefulness of our beat-level-based heterogeneity features. For the genre classification (Latin) task, AP1 and SSKS1 perform better than our submissions.

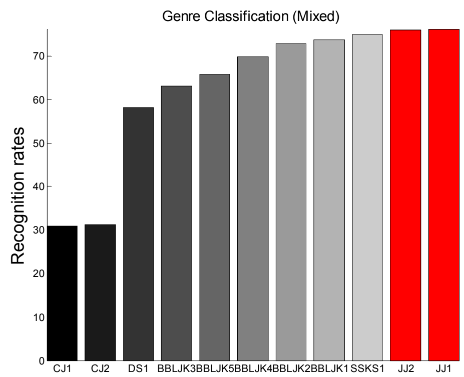

Figure 2. Comparison with other submissions for the MIREX 2013 genre classification (mixed) task.

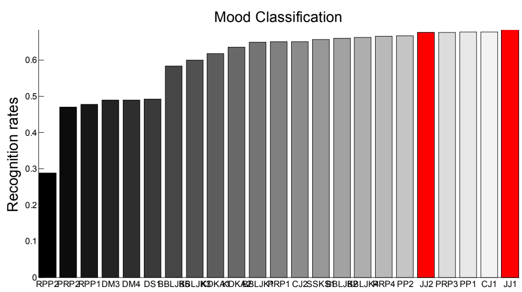

Figure 4. Comparison with other submissions for the MIREX 2013 mood classification task.

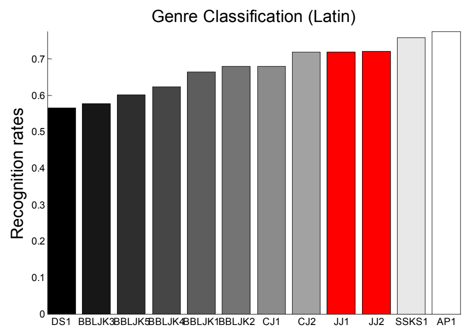

Figure 3. Comparison with other submissions for the MIREX 2013 genre classification (Latin) task.

5. REFERENCES

[1] 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.

[2] 2009: Audio Genre Classification (Mixed Set) Results, Available: http://www.music-ir.org/mirex/wiki/2009:Audio_Genre_Classification_(Mixed_Set)_Results

[3] 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

[4] 2011: Audio Genre Classification (Mixed Set) Results, Available: http://www.music-ir.org/nema_out/mirex2011/results/act/mixed_report/

[5] D.P.W. Ellis, "Beat tracking by dynamic programming," *Journal of New Music Research*, vol. 36, no. 1, pp. 51-60, July 2007.