MIREX 2012 Submissions - Combining Acoustic and Multi-level Visual Features for Music Genre Classification

Ming-Ju Wu

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

Jyh-Shing Roger Jang

Department of Computer Science,
National Taiwan University,
Taipei, Taiwan
roger.jang@gmail.com

ABSTRACT

The system uses two types of effective features for genre classification. The MLVF (multi-level visual feature) can capture the characteristics of a spectrogram's texture from both local and global views. On the other hand, acoustic features are extracted using universal background model and maximum a posteriori adaptation can represent global timbre characteristics. Based on these two types of features, we then employ SVM to perform the final classification task.

1. INTRODUCTION

This submission is an extension of our previous work [1] which combines the acoustic based GSV (Gaussian Super Vector) [2][3] and song-level visual feature. This feature combination is considered to be effective since we won the second place for the MIREX 2011 genre classification task [4]. In the MIREX 2012 submissions, we improve the proposed visual feature which contains the song-level visual feature and the beat-level visual feature. For song-level visual features, we use Gabor filter bank to extract texture information from the octave-based subbands of spectrogram for each music clip. On the other hand, for beat-level visual features, we apply a beat tracking algorithm to obtain beat synchronized features. The rest of this extended abstract is organized as follows: Section 2 introduce the acoustic feature. Section 3 briefly describes the MLVF. 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 MLVF includes the song-level visual feature and the beat-level visual feature. The flowchart of the proposed MLVF (multi-level visual feature) is shown in figure 1. First, 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.

For the song-level visual feature, the mean and standard deviation of the filtering result are used as the features. For the beat-level visual feature, it can be extracted in the similar way, but it resorts to a beat tracking algorithm to obtain beat information of each music clip. We applied the beat tracking algorithm proposed by Dan Ellis [5].

We had two submissions for the train/test task. For the submission WJ1, the second domain tempo (T2) is selected as the tempo parameter in the beat tracking algorithm. On the other hand, for the submission WJ2, the first domain tempo (T1) is selected as the tempo parameter in the beat tracking algorithm.

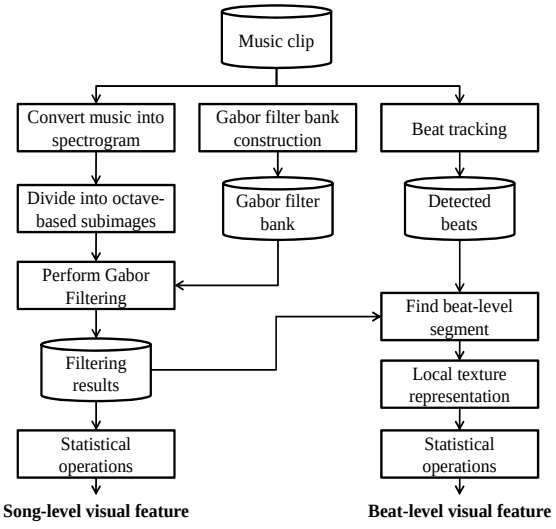

Figure 1. Flowchart of the proposed multi-level visual feature.

4. RESULTS

The results of our submissions for the MIREX 2012 train/test task are shown in figure 2-5. The train/test task includes the genre classification (mixed), genre classification (Latin), mood classification and Classical composer identification. Experimental results show that our submissions achieved the best result for the genre classification (mixed) task. In addition, our submissions also achieved satisfactory results for the mood classification and Classical composer identification.

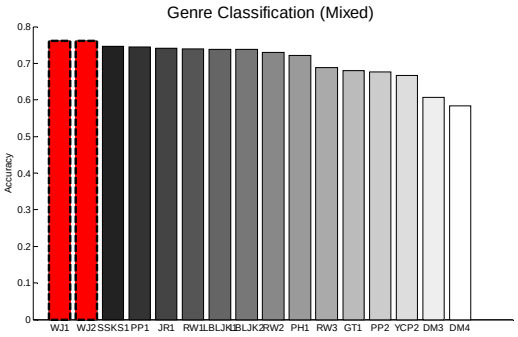

Figure 2. Comparison to other submissions for the MIREX 2012 genre classification (mixed).

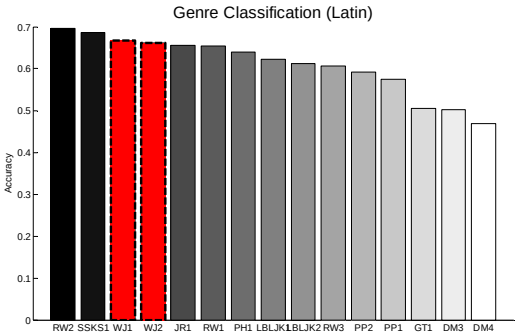

Figure 3. Comparison to other submissions for the MIREX 2012 genre classification (Latin).

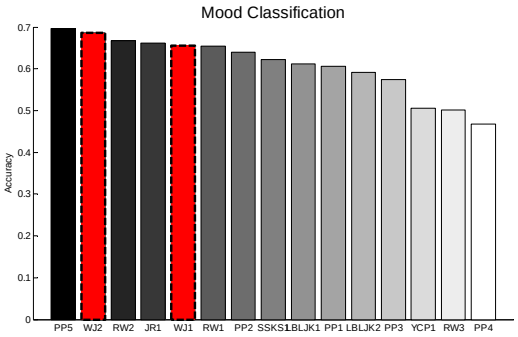

Figure 4. Comparison to other submissions for the MIREX 2012 mood classification.

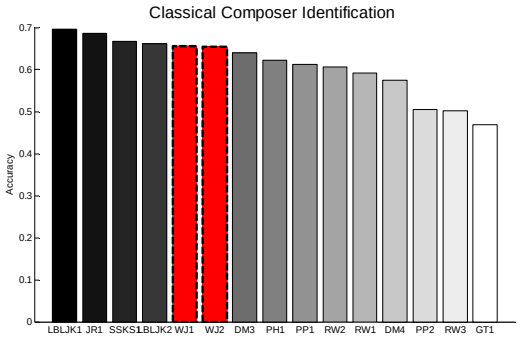

Figure 5. Comparison to other submissions for the MIREX 2012 Classical composer identification.

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.