Audio Genre Classification by Clustering Percussive Patterns *

◎ Emiru Tsunoo†, George Tzanetakis††, Nobutaka Ono† and Shigeki Sagayama†
(† The University of Tokyo, Japan, †† University of Victoria, Canada)

1 Introduction

Due to the increasing size of digital music collections available on computers, interest in music information retrieval (MIR), especially automatic genre classification has recently surged. In this task, not only instrumental but also rhythmic information is thought to be important. If representative unit rhythmic patterns in music can be identified they can be used to genre classification directly from audio signals.

In previous research, timbral, rhythmic and pitch features have been used for audio genre classification [1]. However, the rhythmic features were based on overall statistics of periodicities, not directly on temporal information. More closely related research includes Dixon [2] and which extract a periodical pattern from acoustic signals. These approaches can successfully discriminate styles such samba or tango.

In this paper, we describe an approach for extracting unit percussive patterns from a number of audio tracks and propose a patterns occurrence histogram as a feature for genre classification. Finally, the effectiveness of the proposed feature is verified experimentally.

2 Rhythm Pattern Clustering

2.1 Challenges in Rhythm Pattern Clustering

Bar-long percussive patterns are frequently common and characteristic of a particular genre or style. Automatically detecting these patterns is a "chicken-and-egg" problem in that sets of bar-long unit patterns may be determined only after their corresponding unit boundaries are given, and vice versa. This is complicated by tempo fluctuations and also by harmonic sounds. In the next subsections, we describe our approach to solving these challenges.

2.2 Iterative Update of Percussive Pattern Cluster and Segmentation

Generally, harmonic and percussive sounds are mixed in the observed spectrograms of audio pieces. Therefore in order to perform percussive pattern analysis it is useful to separate these components by using the method described in Ono [3], that is based on the difference of general timbral features.

We have proposed an approach to solved problems above iteratively on the spectrogram whose percussive sounds are emphasized in [4]. Given the initial seed templates, the alignment is calculated using a one-pass dynamic programming algorithm and the optimal segmentation of the input spectral patterns is calculated. Based on the calculated alignment and a similar approach to k-means clustering the input templates are adapted by averaging segments that belong to the same cluster. By iterating these two step, the total summation of distance cost gradually converges.

When used in the context of genre classification, the same set of templates is used for the one-pass DP algorithm calculation for all the pieces. In addition all corresponding segments of all music pieces are collected and averaged in the template update phase. That way a certain number of representative percussive templates common to a particular genre or style can be identified.

3 Rhythm Pattern Feature Extraction

Ideally percussive patterns for a particular genre or style would be fixed and if so then genre classification could be performed simply by looking at which particular patterns are used in a music piece. However in practice there is no guarantee that patterns are fixed for a particular genre/style, i.e., the patterns of a particular music piece will belong to more than one genre in many cases.

One possible way to extract feature vector is count up which percussive pattern templates are contained in a song and calculating the genre pattern occurrence histogram, so that supervised learning classifier can be used.

If M percussive pattern templates are learned from genre g ($g = 1, \ldots, G$), an alignment can be calculated using dynamic programming to calculate the templates $T_{m,g}$ that exist in the song s. Then, the occurrence number of the patterns from genre g can be simply calculated by summation as follows:

$$c_{s,g} = \sum_{m=1}^{M} c_{s,m,g} \tag{1}$$

where $c_{s,m,g}$ is the number of the template $T_{m,g}$ in the song s, and the eventual G dimensional pattern occurrence histogram features vector $\boldsymbol{x}$ of song s can be written as

$$x_g = \frac{c_{s,g}}{N_s} \tag{2}$$

which is normalized by N_s, the number of measure in the song s.

---

* 打楽器パターンのクラスタリングによる音響信号からの自動ジャンル認識、角尾 衣未留 (東京大学)、George Tzanetakis (ヴィクトリア大)、小野 順貴、嵯峨山 茂樹 (東京大学)

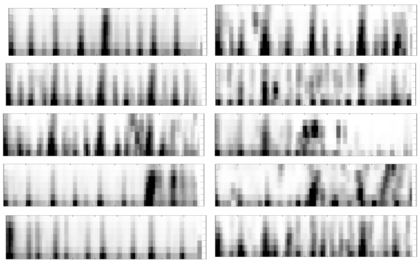

Fig. 1 Example of learned 10 common percussive spectrogram patterns (Blues)

4 Experimental results

4.1 Dataset

Experiments with the proposed algorithms were conducted on both the GTZAN dataset [1] as well as a dataset of ballroom music [5]. The former had 10 genres: blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, and rock. The latter contains 8 dance styles: chacha, foxtrot, quickstep, rumba, samba, tango, viennesewaltz, and waltz. Both of the datasets have 100 songs per genre all of which are single-channel and sampled at 22.05kHz.

4.2 Template learning and feature extraction

First, common percussive pattern templates were learned using the proposed algorithm for each genre. The proposed algorithm was implemented using the audio processing framework, *Marsyas*[1] [6].

To ensure that the template learning didn't become accidentally good for classification, we divided each dataset into two parts and obtained two sets of templates for each genre. In this experiment, 10 templates were learned from each from genre or style, and the number of iterations was fixed to 30 times. The example of learned templates of blues is shown in Fig. 1.

After template learning, using the one-pass DP algorithm all segments are labeled and the pattern occurrence histograms from Eq. 2 are calculated. That way 10 dimensional and 8 dimensional feature vectors are obtained in each case.

4.3 Classification results

To train a classifier in feature space, the "Weka" machine learning toolkit [7] was employed. All the results shown are based on 10-fold cross-validation using a linear SVM as a classifier. The results using only the rhythmic pattern features are shown in Table 1, This shows the proposed features have enough information for genre classification because the accuracy is significantly above the baselines of random classification.

An existing state-of-the-art genre classification system which uses 68 dimensional timbral features like MFCC proposed by Tzanetakis was used for comparison. This system achieved 72.4% on the GTZAN dataset and 57.625% on the ballroom dataset. Merging the timbral features and rhythmic features, the classification accuracies shown in Table 2 were obtained. These results are higher than existing genre classification systems and verify the effectiveness of the proposed features.

[1] http://marsyas.sness.net/

Table 1 Genre classification accuracy using only rhythmic pattern features

Features	GTZAN	Ballroom
Baseline	10.0%	12.5%
Rhythmic (from template set #1)	43.6%	54.0%
Rhythmic (from template set #2)	42.3%	55.125%

Table 2 Genre classification accuracy using merged features with existing timbral features

Features	GTZAN	Ballroom
Existing (Timbre)	72.4%	57.625%
Merged (from template set #1)	76.1%	70.125%
Merged (from template set #2)	76.2%	69.125%

5 Conclusions

We discussed an approach to extracting common percussive patterns for particular genres/styles and using pattern occurrence histograms as features for genre classification. Experiments over music pieces from various genres confirmed that the proposed algorithm can improve the accuracy of classification systems based on timbral information.

Future work includes using n-gram model approach rather than only looking at the uni-gram histogram. In addition more experiments with the parameters of the algorithms such as the number of templates to be learned need to be conducted.

Acknowledgment This research was partly supported by CrestMuse Project under JST.

References

[1] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transaction on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[2] S. Dixon, F. Guyon, and G. Widmer, "Towards characterization of music via rhythmic patterns," in *Proc. of the 5th Int. Conf. on Music Information Retrieval*, 2004, pp. 509–516.

[3] N. Ono, K. Miyamoto, H. Kameoka, and S. Sagayama, "A real-time equalizer of harmonic and percussive componets in music signals," in *Proc. of the 9th Int. Conf. on Music Information Retrieval*, September 2008, pp. 139–144.

[4] E. Tsunoo, N. Ono, and S. Sagayama, "Rhythm map: Extraction of unit rhythmic patterns and analysis of rhythmic structure from music acoustic signals," in *Accepted for ICASSP*, 2009.

[5] "Ballroomdancers.com," http://www.ballroomdancers.com/.

[6] G. Tzanetakis, *Marsyas-0.2: A Case Study in Implementing Music Information Retrieval System*, chapter 2, pp. 31 – 49, Idea Group Reference, 2007, Shen, Shepherd, Cui, Liu (eds).

[7] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2005.