

Musical Genre Classification

Preeti Rao, Sujeet Kini

Abstract— This extended abstract provides overview of the genre classification process in our submission for MIREX 2009. First, the scheme of extraction of features and use of them in the process of Musical Genre Classification is explained. In later part dataset and results obtained are described.

Keywords—ISMIR, Genre, GMM, Timbre

I. INTRODUCTION

Features are extracted from each audio piece and stores in scratch folder provided. These features can later be used for training and testing of Musical Genre Classification System. Features are extracted over each texture window [1]. Features from each texture window are separately used for training the system with same label and classification is also performed in similar fashion i.e. each feature vector is separately classified [2]. Based on appropriate weighting of these classification votes, complete piece of music is classified to particular genre.

II. FEATURES

In this submission texture window of length 1 s is used, over which features are calculated to train and classify the Genre Classification System. Each texture window is divided into smaller analysis windows of duration ~23 ms, over which feature parameters are calculated. Summary of these parameters finally form a feature vector for a texture window. For feature vector of length L , each music clip is expressed as 30-by- L matrix.

Only Timbral features are considered for this submission [1]. MFCC, Spectral Centroid, Bandwidth, Spectral Roll-Off, ZCR, Short Time Average Energy and Spectral Flux features are calculated over each analysis window. Their statistics over texture window forms a feature vector. Low Energy Features are calculated over each texture window.

Mathematical expressions for above features can be found in [1].

III. TRAINING AND CLASSIFICATION

Cluster [3] package with GMM classifier is used for classification. GMM classifier is trained with the

features extracted from the music pieces. Number of mixtures in the Gaussian Mixture Density estimated by the classifier package itself.

Classification is performed for each texture window separately as discussed before. These votes are collected for whole musical piece. By suitable weighting for each vote, a Genre decision is taken for that musical piece.

IV. DATASET

ISMIR2004 genre classification dataset is used for experiments. 30 s non overlapping clips are extracted from each song. Each clip is saved at 22050 Hz sampling rate, with mono channel and 16 bits per samples Final dataset consists of 1845 clips of 7 genres.

V. RESULTS

Three fold cross validation is performed on created dataset. Following table shows various parameters obtained from those experiments.

Parameter	Value	
Maximum Accuracy	67.21%	
Run Time	Feature Extraction	~4.5 ms/clip
	Training + Testing	~ 3 Hrs
Memory	Feature Extraction	~ 10 MB
	Training + Testing	~ 20 MB
CPU	2.4GHz	

VI. REFERENCES

- [1] G. Tzanetakis, "Musical Genre Classification of Audio Classification," in *IEEE Transactions on Speech and Audio Processing*, vol. 10(5), 2002, pp. 293-302.
- [2] J. J. Burred and A. Lerch, "Hierarchical Automatic Audio Signal Classification," in *Journal of the Audio Engineering Society*, vol. 52, 2004, pp. 724-739.
- [3] C. A. Bouman. (1997, Apr.) Cluster: An unsupervised algorithm for modeling Gaussian mixtures. [Online]. <http://www.ece.purdue.edu/~bouman>