STRIVING FOR AN IMPROVED AUDIO SIMILARITY MEASURE

Tim Pohle[1], **Dominik Schnitzer**[1,2]
[1)] Dept. of Computational Perception
Johannes Kepler University, Linz, Austria
[2)] Austrian Research Institute for Artificial Intelligence (OFAI)
Vienna, Austria

ABSTRACT

In this submission to MIREX'07, we implement various modifications to the Algorithm G1C by Elias Pampalk which ranked first in last year's MIREX AudioSim task. Although each of the modifications showed only minor effects in our experiments, their combination constantly outperformed the original algorithm in our automated tests. Therefore, we consider it worth submitting the resulting algorithm to MIREX'07.

1 INTRODUCTION

This submission to the MIREX'07 AudioSim evaluation task presents several modification to the G1C algorithm. In the MIREX'06 AudioSim task, this algorithm was ranked first (although there was no statistical significance found between most of the participating algorithms). This algorithm combines a MFCC-Based audio similarity measure [1] with Fluctuation-Pattern based similarity measures [2]. Details of the G1C algorithm can be found in [3]. [1]

2 IMPLEMENTED MODIFICATIONS

2.1 Feature Extraction

In the feature extraction phase, no changes to the original algorithm were implemented, but we changed the following parameters:

- *Framesize* We use a framesize of 1024 samples, and a hopsize of 512 samples instead of the original settings (512/512). In our experiments, this modification led to an improvement in classification accuracy of about 1% on the ISMIR'04 genre classification contest training set.

- *Number of MFCCs* As some of our experiments indicated an improvement of classification performance, we increased the number of MFCCs from 20 to 25.

[1] Note that G1C is not optimized for feature extraction based on 30 sec from the middle of each file.

2.2 Distance Computation

The main modifications are made in the distance computation functions. These changes are outlined in the next sections. The basic computation of distances based on the Fluctuation Pattern descriptors Bass (FPB) and Gravity (FPG) was not changed.

2.2.1 Comparison of Fluctuation Patterns

G1C applies the Euclidean distance measure to determine the distance between Fluctuation Patterns. However, we found the Cosine Similarity measure to be preferable (cf. [4]).

2.2.2 Comparison of Gaussians

G1C uses the KL-Distance to compare the Gaussian component of the song models (cf. [1]). We used an alternative distance measure related to the cosine similarity measure that showed to be a true metric and performed slightly better in our experiments by means of genre classification accuracy (cf. also [5]).

2.2.3 Normalization of Feature Values

The G1C algorithm uses static normalization factors precomputed from various collections. We think that it may be preferable to adapt the normalization factors to the collection. So, we propose to determine the normalization factors by using the individual distance components (Gaussian, FPB, FPG, FP) for each song.

In particular, for each song, the means and standard deviations of the several distance measures are computed over all appearing distances to all other songs in the collection. These values are normalized to zero mean and unit variance.

For example, to normalize the FPB distances of song a to the other songs in the collection, mean and standard deviation of all FPB distances of song a to the other songs in the collection are computed and used for normalization of the FPB distances associated with song a. All components (Gaussian, FPB, FPG, FP) are normalized independently of each other. The normalized distances are combined as in G1C (Equation 1).

$$D = 0.7 \cdot D^{norm}_{Gauss} + 0.1 \cdot D^{norm}_{FP} \quad (1)$$
$$+0.1 \cdot D^{norm}_{FPB} + 0.1 \cdot D^{norm}_{FPG}$$

As the normalization factors are dependent on the seed song, this distance measure is not symmetric. Symmetry is accomplished by

$$D_{final}(a, b) = D(a, b) + D(b, a) \quad (2)$$

In our experiments, these modifications yield better results, as discussed in the next section.

3 PRELIMINARY EVALUATION

As we assume a high correlation between Nearest Neighbor (NN) genre classification accuracy and human judgments, we evaluated the algorithm by this technique. In Tables 2 and 1, the resulting accuracies are given.

NN	1	3	5	10
$G1C$	64.5%	62.7%	60.6%	59.2%
$G1C_{mod}$	65.3%	64.2%	63.0%	61.6%

Table 1. Probabilistic genre classification accuracies after artist filtering on the ISMIR'04 Genre Classification Contest training set. Computations are based on the middle 30 sec of each piece. $G1C$ as published in the MA Toolbox [6]

NN	1	3	5	10
$G1C$	32.2%	31.2%	31.1%	29.9%
$G1C_{mod}$	34.7%	34.1%	33.5%	32.7%

Table 2. Analogue evaluation on the in-house collection consisting of 103 artists / 2445 tracks divided into 13 genres.

It can be seen that on both collections, $G1C_{mod}$ performs slightly but constantly better than the original algorithm. Therefore, it is of particular interest how human evaluators will rate the output of this algorithm compared to G1C.

4 ACKNOWLEDGEMENTS

This research is supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (FWF) under project number L112-N04, and by the EU 6th FP project S2S2 ("Sound to Sense, Sense to Sound", IST-2004-03773). The Austrian Research Institute for Artificial Intelligence also acknowledges financial support by the Austrian Federal Ministries BMBWK and BMVIT.

5 REFERENCES

[1] Michael Mandel and Dan Ellis, "Song-level features and support vector machines for music classification.," in *Proc. ISMIR*, 2005, pp. 594–599.

[2] E. Pampalk, "Islands of Music: Analysis, Organization, and Visualization of Music Archives," M.S. thesis, Vienna University of Technology, December 2001.

[3] E. Pampalk, *Computational Models of Music Similarity and their Application in Music Information Retrieval*, Docteral dissertation, Vienna University of Technology, Austria, March 2006.

[4] Klaas Bosteels and Etienne E. Kerre, "Fuzzy Audio Similarity Measures Based on Spectrum Histograms and Fluctuation Patterns," in *Proc. 2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*.

[5] Jesper Hojvang Jensen, Daniel P.W. Ellis, Mads G. Christensen, and Soren Holdt Jensen, "Evaluation of Distance Measures Between Gaussian Mixture Models of MFCCs," in *Proc. ISMIR*, 2007.

[6] E. Pampalk, "A Matlab Toolbox to compute music similarity from audio," in *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR'04), Universitat Pompeu Fabra*, Barcelona, Spain, 2004, pp. 254–257.