COMPRESSING QUANTIZED TONAL CENTROID VECTORS FOR COVER SONG IDENTIFICATION

Teppo E. Ahonen, Kjell Lemström, Simo Linkola
Department of Computer Science
University of Helsinki
{teahonen,klemstro,slinkola}@cs.helsinki.fi

ABSTRACT

We present a system for MIREX 2011 cover song identification task that is based on a method for measuring similarity between polyphonic, symbolic pieces of music [1].

1. INTRODUCTION

Cover song identification is an important and interesting task in music information retrieval, as successful cover song identification reveals important information on how the tonal similarity in music can be automatically extracted and measured.

We present a system that is based on a method we recently developed for measuring similarity between polyphonic pieces of music that are symbolically encoded. As the task in hand is audio-based, several steps of preprocessing audio data into similar symbolic representation is needed, but otherwise, the method is similar to the method described in [1]. We measure the similarity between two sequences of binarized tonal centroid vectors [3] extracted from the audio files. For similarity measuring, metric called normalized compression distance (NCD) [2] is used.

2. SYSTEM DESCRIPTION

2.1 Binary Chromagram

First, the chroma vector data from the audio file is extracted. We use a window length of 0.1858 seconds with a hop factor of 87.5 percent. The chroma vectors are extracted using MIRToolbox [1]. No beat synchronization is applied.

In order to bridge the gap between the method we presented for symbolic music and the audio data used here, we quantize the continuous chroma vectors to binary values. We experimented with several methods, but eventually had the best results by simply setting the n largest values of the chroma vector to 1, and the rest as 0. For n, the best results were obtained with values under 4, and for the submitted method, we chose to use $n = 3$.

[1] http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox

It should be noted that the binarization of the values is not only a step to make our submission closer to the method presented in [1], but it also seems to have a minor impact on the identification accuracy. See Subsection 3.1 for results and comparison between using non-binarized and binarized chromagrams.

2.2 Tonal Centroid Transform

After the chroma vectors are binarized, they are transformed to tonal centroid vectors [3]. The motivation behind this is to reduce the dimensions from the 12 of the chroma vector to 6 using a method that utilizes musical knowledge.

Tonal centroid vector for time frame t is given by formula [3]:

$$\zeta_t(d) = \frac{1}{||c_t||} \sum_{p=0}^{11} \Phi(d,p) c_t(p) \tag{1}$$

where $0 \leq d \leq 5$ and $0 \leq p \leq 11$. $||c_t||$ denotes the L_1-norm of chroma vector c_t, p is the pitch class index in c_t, d represents which of the dimensions of tonal centroid is being calculated and $\Phi = [\phi_0, \phi_1, \ldots, \phi_{11}]$ is transformation matrix where

$$\phi_p = \begin{pmatrix} \Phi(0,p) \\ \Phi(1,p) \\ \Phi(2,p) \\ \Phi(3,p) \\ \Phi(4,p) \\ \Phi(5,p) \end{pmatrix} = \begin{pmatrix} \sin p\frac{7\pi}{6} \\ \cos p\frac{7\pi}{6} \\ \sin p\frac{3\pi}{2} \\ \cos p\frac{3\pi}{2} \\ \frac{1}{2}\sin p\frac{2\pi}{3} \\ \frac{1}{2}\cos p\frac{2\pi}{3} \end{pmatrix}, 0 \leq p \leq 11. \tag{2}$$

2.3 Tonal Centroid Vector Labeling

The continuous-valued tonal centroid vectors are again binarized. We give the tonal centroid vector values of 0 and 1 according to a threshold vector obtained by taking the median of all possible tonal centroid vectors (i.e. all tonal centroid vectors of the 2^{12} different binary chroma vectors). In practice, the values of the threshold vector are close to zero.

The binarized six-dimensional tonal centroid vectors are then mapped to characters of alphabet size of 64. Thus, each piece of music in the collection is represented as a sequence of characters. These sequences are written to files for the similarity measuring.

2.4 Similarity Measuring

After all the files in the collection have been transformed into sequence representations and the sequences have been written into files, we measure the similarity between each pair of query and target files using normalized compression distance. For files x and y, NCD is denoted [2]

$$NCD(x, y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}, \quad (3)$$

where $C(x)$ is the length of the file x when compressed using compression algorithm C, and xy is the concatenation of x and y. The pairwise distances are finally written to a distance matrix that is the output of the system.

For data compression, we use PPMZ (Prediction by Partial Matching, variant Z) algorithm.

2.5 Key Invariance

To ensure that the possible transposition of the cover version would have minimum impact in the similarity measuring, optimal transposition index (OTI) [4] values between two non-binarized chroma profiles are calculated. For each song, we write the sequences for each 12 transpositions, but measure the pairwise similarity between sequences selected according to the OTI.

3. EVALUATION

3.1 Our Collection

For development purposes, we collected a small dataset of 10 original compositions, with each having four different cover versions. In addition, 150 non-related pieces of music were added to make the evaluation more difficult, making the total size of our data set 200.

TODO

3.2 MIREX 2011 Results

TODO

3.2.1 Mixed Collection

3.2.2 Mazurka Collection

4. CONCLUSIONS

TODO

5. ACKNOWLEDGEMENTS

The work of Teppo E. Ahonen was supported by the Helsinki Doctoral Programme in Computer Science - Advanced Computing and Intelligent Systems (Hecse).

6. REFERENCES

[1] T.E. Ahonen, K. Lemström, and S. Linkola: "Compression-based Similarity Measures in Symbolic, Polyphonic Music," To appear in *Proc. ISMIR'11*, 2011.

[2] R. Cilibrasi and P.M.B. Vitanyi: "Clustering by Compression," *IEEE Transactions on Information Theory*, Vol. 51, No. 4, pp 1523–1545, 2005.

[3] C. Harte, M. Sandler, and M. Gasser: "Detecting Harmonic Change in Musical Audio," Proceedings of the 1st ACM workshop on Audio and music computing multimedia, 2006.

[4] J. Serra, E. Gómez, and P. Herrera: "Transposing Chroma Representations to a Common Key," Proceedings of IEEE CS Conference on The Use of Symbols to Represent Music and Multimedia Objects, 2008.