COVER SONG IDENTIFICATION BASED ON DATA COMPRESSION

Teppo E. Ahonen
Department of Computer Science
University of Helsinki
teahonen@cs.helsinki.fi

ABSTRACT

We present a system for cover song identification. Our approach combines chord sequence estimation with a similarity metric called normalized compression distance.

1. INTRODUCTION

Audio cover song identification is a challenging music information retrieval task, and in recent years it has been actively studied by various researchers in the music information retrieval (MIR) community. Here, we present a system for cover song identification that combines chord sequence estimation with a similarity metric called normalized compression distance. The system and several evaluations for it are presented more precisely in [1].

Normalized compression distance (NCD) [2] is a distance metric that uses a compression algorithm to measure the similarity between two objects. For strings x and y, NCD is calculated as

$$NCD(x, y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}, \qquad (1)$$

where $C(x)$ is the length of the string x when compressed with a fixed lossless compression algorithm C and $C(xy)$ the length of the compressed version of the concatenation of strings x and y [2].

2. SYSTEM DESCRIPTION

We use chord estimation technique that is based on hidden Markov model [3] to convert the audio signal to a symbolic harmonic feature representation. After the chord sequences have been estimated, the distances between them are calculated based on Equation (1). The components of the system are depicted in Figure 1.

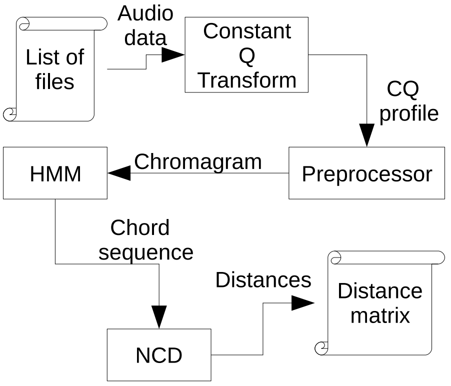

Figure 1. Blueprint of the system components.

2.1 Feature Extraction

To obtain a chromagram, the original audio signal is processed using constant Q transform [4] with a frequency range from 65.4 Hz to 1046.5 Hz. For audio with sampling rate of 44100 Hz we use a window length of 8192 samples and hop size of 1024 samples. No beat detection is used, as we have empirically discovered frame-based approach leading to better results. The constant Q profile is turned into a chromagram by folding all octaves into one and normalizing each vector.

The chromagram is given as observation vectors to a 24-state ergodic hidden Markov model (HMM), with parameters set as suggested in [3]. The model is trained with the EM algorithm, and the Viterbi algorithm is used to decode the most likely state transition path.

2.2 Calculating the NCD Values

After the HMM has been trained and the most likely state transition path has been obtained, the chord sequences are transformed into a string format. In [1], we used a representation that describes the difference between subsequent chords in one character. For MIREX 2009, we tried a different approach where we describe each chord with its relation to the most frequent chord in the sequence.

Our choice of the compression algorithm is the PPMZ algorithm. PPMZ is a statistical compressor, and as such, usually more efficient compression algorithm than for example gzip or bzip2. For the PPMZ implementation, we use the pzip algorithm [1].

[1] http://muq.org/˜cynbe/compression/ppm.html

3. MIREX 2009 EVALUATION

In MIREX 2009, the audio cover song identification was divided into two subtasks: the mixed collection and the mazurka collection. For more information on the task and complete results of all submitted algorithms, we suggest visiting the MIREX 2009 wiki [2] .

3.1 Mixed Collection

The mixed collection is the original audio cover song identification dataset. The dataset consists of 1000 pieces of music from various genres, and includes 30 groups of cover performances, each consisting of 11 versions, making a total of 330 cover versions. Each cover version is used as a query file and the evaluation is based on the distance matrix. The results for our algorithm are presented in Table 1.

Total number of covers identified in top 10	646.00
Mean number of covers identified in top 10	1.96
Mean (arithmetic) of average precisions	0.20
Mean rank of first correctly identified cover	29.90

Table 1. Results for the Mixed Collection subtask

3.2 Mazurka Collection

The mazurka collection is a dataset consisting of a subset from the mazurka.org dataset put together by Craig Sapp. The collection consists of 49 mazurkas with 11 versions from each, making a total of 539 pieces of music. Each file is used as a query and, as with the mixed collection, the evaluation is based on the distance matrix. The results for our algorithm are presented in Table 2.

Total number of covers identified in top 10	2843.00
Mean number of covers identified in top 10	5.27
Mean (arithmetic) of average precisions	0.56
Mean rank of first correctly identified cover	5.49

Table 2. Results for the Mazurka Collection subtask

4. DISCUSSION

Out of the three algorithms submitted to the audio cover song identification task, our approach performed evidently weakest. As we speculated in [1], the weakest link of the system is probably the chord estimation. Keeping in mind that in MIREX 2008 the best-performing pre-trained algorithm for audio chord detection obtained an average overlap score of 0.66 [3] , the chord estimation is still not a solved task. The estimated chord sequences are distorted, and the additional noise in the sequences weakens the compression rate, which in turn causes distortion in the NCD values. Although NCD has been proved to be robust against noise [5], here the sequences are short and constructed from a small alphabet, thus enabling even a small amount of noise to have a major impact on the identification.

Another possible hindrance is the selected mid-level representation itself. An alphabet of 24 chords is probably too reducing for cover song identification: similar chord sequences do appear in different pieces of music that are nevertheless original and not different versions of the same composition. In addition, when estimating chords, melodic information is somewhat lost.

Thus, for future implementations, it seems to be a better idea to concentrate on the chroma vectors themselves instead of attempting to estimate the chord sequences from them.

5. ACKNOWLEDGEMENTS

The author wishes to thank the IMIRSEL group for organizing MIREX 2009 and running the evaluations.

This work was supported by the Academy of Finland (Grant #129909).

6. REFERENCES

[1] Teppo E. Ahonen. Measuring harmonic similarity using PPM-based compression distance. In *Proceedings of Workshop on Exploring Musical Information Spaces*, pages 52–55, Corfu, Greece, October 2009.

[2] Ming Li, Xin Chen, Xin Li, Bin Ma, and Paul Vitányi. The similarity metric. *IEEE Transactions on Information Theory*, 50(12):3250–3264, December 2004.

[3] Juan P. Bello and Jeremy Pickens. A robust mid-level representation for harmonic content in music signals. In *Proceedings of 6th International Conference on Music Information Re trieval*, pages 304–311, London, UK, September 2005.

[4] Judith C. Brown. Calculation of a constant Q spectral transform. *Journal of Acoustic Society of America*, 89(1):425–434, January 1991.

[5] Manuel Cebrián, Manuel Alfonseca, and Alfonso Ortega. The normalized compression distance is resistant to noise. *IEEE Transactions on Information Theory*, 53(5):1895–1900, May 2007.

[2] http://www.music-ir.org/mirex/2009/index.php/Audio_Cover_Song_Identification_Results

[3] http://www.music-ir.org/mirex/2008/index.php/Audio_Chord_Detection_Results