MIREX 2019: SALIENT CHROMAGRAM FOR COVER SONG IDENTIFICATION

Jin S. Seo
Gangneung-Wonju National University
Dept. Electrical Eng.

ABSTRACT

This document describes our submission to 2019 MIREX cover song identification task based on the salient chromagram. We apply the proposed salient chromagram to the sequence-alignment based cover song identification. Experiments on a cover song dataset confirm that the proposed salient chromagram improves the cover song identification accuracy.

1. INTRODUCTION

Computing music similarity for cover song identification is interesting and challenging since various types of differences in timbre, rhythm, song structure, main key, and lyrics may occur during cover song generation. To cope with the differences, we have to find invariant properties shared by an original song and its cover version. One commonly used musical property for cover song identification is the tonal contents of music, such as chromagram or pitch class profiles, which is independent of timber and loudness and thus suitable for cover song identification [6]. These features are a representation of the spectral energy in the frequency range of each one of the twelve semitones. In this submission, a salient chromagram extraction method is proposed to improve cover song identification accuracy. Details of extracting the chroma-based features can be found in [5]. The cover song identification can be performed by temporally aligning the chromagram vector sequences of two songs. Sequence-alignment based methods [6] [2] try to find best alignment between feature sequences from two songs by adopting techniques used in speech recognition or DNA sequence identification, such as dynamic time warping [3] or Smith-Waterman (SW) algorithm [7].

2. SALIENT CHROMAGRAM AND SEQUENCE ALIGNMENT

Extracting a salient chromagram for cover song identification is not a trivial task considering various types of differences between the original and its cover versions. It is almost impossible or at least not viable for now to cope with all these variations separately. In this submission, we try to improve the saliency of the chromagram. An interesting previous work is the chroma DCT-reduced log pitch (CRP) proposed in [4] where the upper-frequency coefficients of the discrete cosine transform (DCT) is utilized in obtaining a timber-invariant chromagram by assuming that the lower-frequency DCT components of the spectral energy are closely related to the aspect of timbre. The proposed salient chromagram is an extension of the CRP. The chromagram extraction is based on a pitch-frequency scale used in [5]. The input music signal is decomposed into 88 frequency bands with center frequencies corresponding to the MIDI pitches $p = 21$ to $p = 108$ (which correspond to the keys of a piano). Further details on the frequency band positions and bandwidth are described in [5]. At each of the 88 subbands, the short-time mean-square power (local energy) is calculated. As a result, we obtain a sequence of 88-dimensional feature vectors where the entries correspond to MIDI pitches to $p = 21$ to $p = 108$. As in [5], we add 20 zeros at the beginning and 12 at the end to construct a 120-dimensional feature vector where the entries correspond to MIDI pitches from $p = 1$ to $p = 120$. Then we apply a logarithmic compression on the pitch representation to account for the logarithmic sensation of sound intensity [4]. By chroma binning, which adds up the corresponding values of the pitch representation that belong to the same chroma, we can obtain 12-dimensional chromagram. Muller and Ewert [4] proposed the CRP by taking the DCT on the 120-dimensional pitch representation, keeping the upper coefficients, and applying the inverse DCT before the chroma binning.

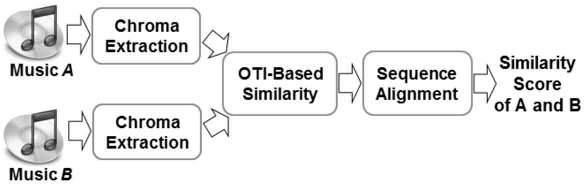

Figure 1. The music-similarity computation for the cover song identification based on the optimal transposition index and sequence alignment.

The baseline cover song identification method, considered in this paper, is presented in [6] where the optimal transposition index (OTI) is used for the chromagram similarity, and the SW algorithm is used for the local sequence alignment. The OTI-SW method has shown one of the best

Attribution: Jin S. Seo. “Mirex 2019: Salient Chromagram for Cover Song Identification”, 20th International Society for Music Information Retrieval Conference, Delft, The Netherlands, 2019.

Table 1. Identification performance of the covers80 dataset. Accuracy measures are the average rank of the first correctly identified cover, $Rank_1$, precision at one, $P@1$, and the mean of average precision, MAP.

Methods	$Rank_1$	$P@1$	MAP
Salient Chromagram	17.23	0.613	0.669
CRP [4]	24.86	0.556	0.605
CLP	24.14	0.588	0.631

performance in MIREX test [1] and studied in depth [6]. Fig. 1 shows a general block diagram of the OTI-SW, which composes of three modules: preprocessing, similarity matrix creation, and sequence alignment. Preprocessing comprises chromagram sequence extraction and a global chromagram averaging for each song. Based on the OTI between two input chromagram sequences, a binary similarity matrix is then computed. The binary similarity matrix is the input to the local sequence alignment by SW algorithm, which gives the highest score to the best aligned subsequence. Finally, the highest score is normalized on the temporal lengths of two input sequences. Details of each step are in [6].

3. EVALUATION

The cover song search performance of the proposed multi-scale chroma n-gram indexing was evaluated on two cover song datasets. The first cover song dataset (abbreviated as covers80) is the one that was used by Dan Ellis in his work [1]. The covers80 consists of 80 original and cover song pairs, which are available online.

For a fair comparison of retrieval performance, the same order chroma features were used for all the considered approaches ($M = 12$). Each song in the datasets was converted to mono at a sampling frequency of 22050 Hz and then divided into frames of 200 ms overlapped by 100 ms where the 12-dimensional chromagram was computed as a low-level feature for each frame. We extracted the chroma log pitch (CLP) and the CRP using the chroma toolbox [5] with the default parameter settings. Table 1 shows that the cover song identification performance of the considered chromagrams for the covers80 dataset when combined with the OTI-SW method. The proposed salient chromagram outperformed the other chromagrams.

4. CONCLUSION

For a reliable cover song identification, improving chromagram saliency is crucial due to the wide range of possible distortions which may occur during cover song generation process. This paper proposes a new salient chromagram, which improves robustness against timber change and additive noise. Experimental results on a dataset show that the proposed salient chromagram is effective in improving cover song retrieval accuracy.

[1] Available at https://www.music-ir.org/mirex/wiki/2007:Audio_Cover_Song_Identification_Results

5. REFERENCES

[1] D. Ellis and G. Poliner. Identifying cover songs with chroma features and dynamic programming beat tracking. In *Proceedings of ICASSP-2007*, pages 1429–1432, 2007.

[2] P. Foster, S. Dixon, and A. Klapuri. Identifying cover songs using information-theoretic measures of similarity. *IEEE Transactions on Audio, Speech, and Language Processing*, 23(6):993 – 1005, June 2015.

[3] Y. Kim and H. Jeong. A systolic FPGA architecture of two-level dynamic programming for connected speech recognition. *IEICE Transactions on Information and Systems*, 90(2):562 – 568, February 2007.

[4] M. Muller and S. Ewert. Towards timbre-invariant audio features for harmony-based music. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3):649 – 662, March 2010.

[5] M. Muller and S. Ewert. Chroma toolbox: MATLAB implementations for extracting variants of chroma-based audio features. In *Proceedings of ISMIR-2011*, pages 215–220, 2011.

[6] J. Serra, E. Gomez, P. Herrera, and X. Serra. Chroma binary similarity and local alignment applied to cover song identification. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(6):1138 – 1151, August 2008.

[7] T.F. Smith and M.S. Waterman. Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1):195 – 197, March 1981.