COVER SONG IDENTIFICATION BASED ON SIMILARITY FUSION

Ning Chen
East China University of Science and Technology

ABSTRACT

We describe a similarity fusion based cover song identification scheme. The Harmonic Pitch Class Profile (HPCP) is chosen as the musical descriptor. First, the similarity between HPCP descriptors of two songs are obtained based on Qmax function and Dmax function, respectively. Then these two similarities are fused via Similarity Network Fusion (SNF) technique, which was originally proposed for combining different kernels for predicting drug-target interactions. Finally, the fused similarity is used to identify the cover versions.

1. INTRODUCTION

Cover Song Identification (CSI) technique tries to identify an alternative version, performance, rendition, or recording of a previously recorded musical composition by measuring and modeling the similarity between two tracks based on their content. It can be used for music searching and organization, music rights management and licenses, and music creation aids, etc. In 2006, CSI task was included by the Music Information Retrieval Evaluation eXchange (MIREX) for the first time.

With the development of CSI techniques, various musical descriptors and similarity functions have been proposed for CSI task. For example, in [1], the Qmax similarity function was proposed for CSI task. When it is combined with Harmonic Pitch Class Profile (HPCP) [2] descriptor, the obtained CSI scheme yielded the highest accuracy of all algorithms submitted in 2009 to MIREX. However, due to the possible alignment constraints in Qmax, it fails to identify the cover versions when the Cross Recurrence Plot (CRP) includes such phenomenon as shown in Figure 1(a), where there is serious short disruption of diagonal. This phenomenon may be resulted from the skip of some chords or part of the melody when performing the cover version. To solve this problem, we modified Qmax by changing the possible alignment constraints from Figure 2 (a) to Figure 2 (b) to obtain a new measure, called Dmax in [3]. As shown in Figure 1 (b) and (c), in the cases shown in Figure (a), Dmax performs better than Qmanx. Further, in this extended abstract, we fuse Qmax and Dmax based similarities between HPCP descriptors via Similarity Network Fusion (SNF) [4] technique to incorporate the advantage of both of these two similarity measures. The advantages of the proposed scheme are that: i) Since it is based on one musical descriptor, the computational complexity is low; ii) It deals with the difference in scale and noise in each similarity measure; iii) It captures both shared and complementary information carried by different similarity measures; iv) it can be used to fuse a large number of similarity measures.

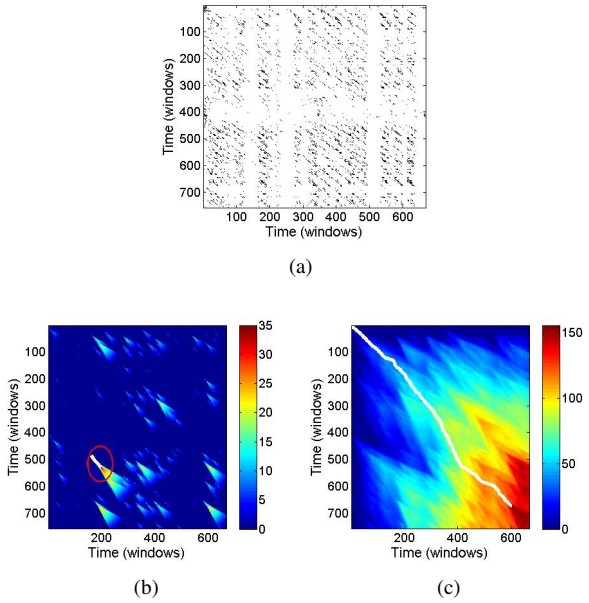

Figure 1. (a) CRPs for the song *Addicted to Love* as performed by Tina Turner and Robert Palmer and the corresponding cumulative matrix obtained by (b) Qmax (Qmax=35) and (c) Dmax (Dmax=155.5).

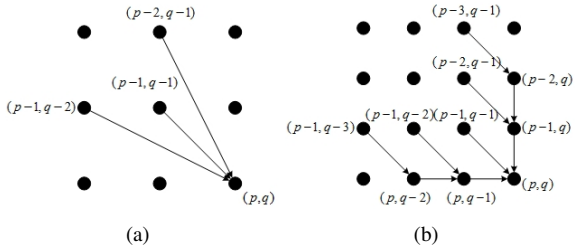

Figure 2. Possible alignment constraints in (a) Qmax and (b) Dmax

2. SYSTEM OVERVIEW

For the given two tracks X and Y, we first extract their H-PCP descriptors with the same extraction procedure and parameters as shown in [1]. Then transposition is performed via the method proposed in [5]. The transposed H-PCP time series are denoted as x and y, respectively. Then a state space representations of these two tracks is formed by delay coordinates involving the embedding dimension m and the time delay τ, and a CRP is constructed using a fixed maximum percentage of nearest neighbors κ. Subsequently, Qmax and Dmax function are used respectively to obtain the similarity between two tracks. The parameters chosen to calculate CRP are $m = 15$, $\tau = 2$, $\kappa = 0.1$. For Qmax and Dmax, the penalty for a disruption onset and that for a disruption extension are set as $\gamma_{o\_Qmax} = 5$, $\gamma_{e\_Qmax} = 0.5$ and $\gamma_{o\_Dmax} = \gamma_{e\_Dmax} = 0.5$, respectively. Finally, these two kinds of similarities are fused via SNF technique [4] with following parameter setting: number of neighbors $K = 12$, hyperparameter $\alpha = 0.36$ and number of iterations $T = 20$.

3. EVALUATION METHODOLOGY

This year, the CSI task was run on two music collections: Mixed Collection and Mazurka Collection. The Mixed collection is composed of 1000 tracks containing 30 different "cover songs", each represented by 11 different "versions" for a total of 330 files which are complemented by 670 additional tracks. The Mazurka Collection consists of 539 pieces corresponding to 11 selected versions from 49 Chopin mazurkas from the Mazurka Projects [1] .

To evaluate the performance of a CSI scheme, the query/answer framework [6] is adopted. Each of the cover song is used as query and the returned list of items is examined for the presence of the other versions of the query. The specific evaluation metrics include: (i) Total number of covers identified in top 10; (ii) Mean number of covers identified in top 10; (iii) Mean of Average Precision (MAP); (iv) Mean rank of first correctly identified cover.

4. ACKNOWLEDGE

This work is supported by the National Natural Science Foundation of China (61271349)

5. REFERENCES

[1] Serra, Joan and Serra, Xavier and Andrzejak, Ralph G: "Cross Recurrence Quantification for Cover Song Identification," *New Journal of Physics*, Vol. 11, No. 9, pp. 093017, 2009.

[2] Gómez, Emilia: *Tonal description of music audio signals*, PhD thesis, UPF Barcelona, 2006.

[3] Yang, Fan and Chen, Ning: "Cover Song Identification Based on Cross Recurrence Plot and Local Alignment," *Journal of East China University of Science and Technology*, Vol. 42, No. 2, pp. 247–253, 2016.

[4] Wang, Bo and Mezlini, Aziz M and Demir, Feyyaz and Fiume, Marc and Tu, Zhuowen and Brudno, Michael and Haibe-Kains, Benjamin and Goldenberg, Anna: "Similarity network fusion for aggregating data types on a genomic scale," *Nature methods*, Vol. 11, No. 3, pp. 333–337, 2014.

[5] Serra, Joan and Gómez, Emilia and Herrera, Perfecto: "Transposing chroma representations to a common key," *Proceedings of IEEE CS Conference on The Use of Symbols to Represent Music and Multimedia Objects*, pp. 45–48, 2008.

[6] Manning, Christopher D and Raghavan, Prabhakar and Schütze, Hinrich and others: *Introduction to information retrieval*, Cambridge university press Cambridge, 2008.

[1] http://www.mazurka.org.uk