COVER SONG RETRIEVAL BY CROSS RECURRENCE QUANTIFICATION AND UNSUPERVISED SET DETECTION

Joan Serrà[1,2]**, Massimiliano Zanin**[3]**, and Ralph G. Andrzejak**[2]
[1] Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain.
[2] Dept. of Inf. and Com. Technologies, Universitat Pompeu Fabra, Barcelona, Spain.
[3] Universidad Autónoma de Madrid, Madrid, Spain.
joan.serraj@upf.edu, massimiliano.zanin@hotmail.com, ralph.andrzejak@upf.edu

ABSTRACT

This extended abstract briefly outlines our submission to the audio cover song identification task run inside the 2009 music information retrieval evaluation exchange (MIREX). We very briefly introduce cover song identification, give an overview of the submitted system, and comment on the evaluation procedure (specially on the used music collections) and the achieved results. As the submission is completely based on our previous work, we refer to it for further detail.

1. INTRODUCTION

Audio cover song identification is a task that has been receiving considerable attention in the last few years and, therefore, many algorithms for that specific purpose have been developed [1]. In line with this growing interest, since 2006, the music information retrieval evaluation exchange (MIREX) is running the audio cover song identification task [2], which allows for an objective assessment of the accuracy of different such algorithms. In the 2007 edition, our group submitted an algorithm that we subsequently described in [3]. This algorithm, which used a specifically designed chroma similarity measure and a dynamic programming subsequence matching method, yielded the highest accuracy of all algorithms submitted in 2007 and in earlier editions. For the 2008 edition, we used a qualitatively novel approach. The cover song identification measure that we derived from this approach [4] and a composition of this measure with a subsequent cover set detection layer [5] yielded the two highest accuracies of all algorithms submitted in 2008 and in earlier editions. In particular, the accuracy of the 2008 approach clearly surpassed our earlier algorithm proposed in [3]. The system submitted in the present edition [4, 5] is the same system as in 2008, except that for the present edition we do some minor parameter adjustments.

2. SYSTEM OVERVIEW

Given two songs X and Y, we first extract their chroma descriptor time series and transpose one song to the main tonality of the other. As chroma features we use harmonic pitch class profiles (HPCPs) [6]. We employ the same extraction procedure and parameters as in [3], except that (a) the number of HPCP bins is set to 12, (b) we use 464 ms frames with no overlap, and (c) we limit the number of spectral peaks considered to a small integer below 50. Transposition is done via the procedure explained in [7], which includes considering a fixed number of optimal transposition indices N_{OTI}.

From this pair of multivariate time series x and y (i.e. the transposed HPCPs), we form state space representations of the two songs using delay coordinates involving an embedding dimension m and time delay τ. From this state space representation, we construct a cross recurrence plot (CRP) using a fixed maximum percentage of nearest neighbors κ. Subsequently, we use the recurrence quantification measure $Q_{\max}(x, y)$ to extract features that are sensitive to cover song CRP characteristics, which requires the setting of two additional parameters γ_o and γ_e. These steps are detailed in [4]. We tuned the aforementioned parameters to in-house cover song music collection truth and finally set them to $N_{\mathrm{OTI}} = 2$, $m = 9$, $\tau = 1$, $\kappa = 0.1$, and $\gamma_o = \gamma_e = 0.5$.

Because document length is often inversely correlated with relevance, we use it as a normalization factor. This is a common strategy in many information retrieval systems [8]. The final dissimilarity measure between query song X and candidate song Y is obtained by

$$d(X, Y) = \frac{\sqrt{|y|}}{Q_{\max}(x, y)}, \tag{1}$$

where $|y|$ represents the length of the candidate song (in frames).

The final step of the system considers the full dissimilarity matrix obtained with all possible pairwise comparisons and detects cover sets (or clusters). We use “method 1” in [5] with no threshold and just considering the first nearest neighbor for each query song. The usage of this particular method is justified by its computational speed and by the asymmetry of $d(X, Y)$.

3. EVALUATION METHODOLOGY

This year, the task was run with two music collections separately: the so-called “mixed collection” and the “mazurka collection”. The “mixed collection” corresponds to the same music collection used in previous editions. It consists of 1000 pieces containing 30 different cover songs, each represented by 11 different versions for a total of 330 audio files which are complemented by 770 additional songs. The cover songs are meant to represent a variety of genres (e.g. classical, jazz, gospel, rock, folk-rock, etc.) and the variations span a variety of styles and orchestrations. The “mazurka collection” consists of 539 pieces corresponding to 11 selected versions from 49 Chopin mazurkas from the Mazurka Project [1] .

The evaluation is performed in a query/answer framework [8]: using each of the cover song files in turn as the seed/query, the returned list of items is examined for the presence of the other versions of the seed/query [9]. The main evaluation measure is the mean of average precisions (MAP), a common evaluation measure in information retrieval [8, 9]. More details about music collections and evaluations are available at the MIREX wiki [2] .

4. RESULTS AND DISCUSSION

Results for the “mixed collection” and the “mazurka collection” are shown in Tables 1 and 2, respectively. We here just report the summary results. For a more detailed assessment of the results we refer to the MIREX wiki [3] .

Evaluation measures	Submissions		
	TA	RE	SZA
Total number of covers identified in top 10 [3300..0]	646	2046	2426
Mean number of covers identified in top 10 [10..0]	1.96	6.20	7.35
Mean of average precisions [1..0]	0.20	0.66	0.75
Mean rank of first correctly identified cover [1..999]	29.90	2.28	6.15

Table 1. Summary results for the “mixed collection” (submitted system in bold). Range of the evaluation measures in square brackets (from best to worse).

In the present edition there were 3 submitted systems: TA, SE, and SZA. We see that our submitted system (SZA) achieves the best scores in the two music collections with all evaluation measures considered. The only exception is in the mean rank of the first correctly identified cover with the “mixed collection”. This result indicates that, with this concrete collection, the first correct answer obtained with our system might have a higher rank than the one obtained with the RE system. However, this evaluation measure just relates to the first correctly identified item (neglecting the rest). To obtain a general view and to consider more items in the answer, we should look at other measures such as the mean number of covers identified in top 10 or the mean of average precisions.

Evaluation measures	Submissions		
	TA	RE	SZA
Total number of covers identified in top 10 [5390..0]	2843	4757	5165
Mean number of covers identified in top 10 [10..0]	5.27	8.83	9.58
Mean of average precisions [1..0]	0.56	0.91	0.96
Mean rank of first correctly identified cover [1..538]	5.49	1.68	1.61

Table 2. Summary results for the “mazurka collection” (submitted system in bold). Range of the evaluation measures in square brackets (from best to worse).

All the submitted systems achieved particularly high accuracies with the “mazurka collection” (e.g. our system achieved a mean average precision of 0.96). We hypothesize three reasons for this. First, the high accuracies indicate that the “mazurka collection” might be less varied than the “mixed collection”. As we discuss in [1], the more variation there is between covers, the more difficult it is to identify them. In particular, some important musical aspects such as overall timbre, tempo, and structure (which could constitute the main dificulties for a cover song identification system) may not vary within this collection. Second, the “mazurka collection” is more than two times smaller than the “mixed collection”, and it is clear that searching items in smaller collections can significantly increase reported accuracies [1]. Third, the fact that there is a high number of versions of the same song can be intentionally exploited to increase accuracy as we do with the cover set detection module [5]. Overall, one might consider the “mazurka collection” as more of a music identification collection rather than a representative cover song collection with different types of covers [1]. Under this view, the high accuracies obtained with the “mazurka collection” highlight the good performance that cover song identification algorithms might have in tasks such as music identification or audio fingerprinting. Apart, the suitability of the “mazurka collection” could be further questioned because the music metadata is known [4] and, therefore, one could train a system with it. This training couldn’t be with the exact songs used in the MIREX (because they are randomly selected), but could influence algorithms performace on the aforementioned collection. Regarding the “mixed collection”, one should note that the second and third arguments presented above may also hold. This also raises some questions as to its suitability for the evaluation of real-world cover song identification systems.

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

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

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

[4] `http://www.mazurka.org.uk/info/discography`

Apart from the aforementioned results, we want to highlight the efforts we have put in improving the generality and the speed of the system. The first fact resulted in the $Q_{\max}(x, y)$ measure [4], which can be used with any time series by just re-adjusting the parameters outlined in Sec. 2.

5. CONCLUSION

We overview our submission to the MIREX 2009 audio cover song identification task. As it is entirely based on previous work we refer to it for a detailed explanation [4, 5]. We also report the summary results obtained in the task, where our submission outperformed the rest of presented algorithms. We finally discuss on the suitability of the music collections used for evaluation.

6. ACKNOWLEDGMENTS

The authors would like to thank E. Aylon, J. Funollet, and N. Wack for technical support. They also want to thank all the IMIRSEL team for the organization and running of this evaluation. This research has been partially funded by the EU-IP project PHAROS IST-2006-045035 and by the BFU2007-61710 Spanish Ministry of Education and Science grant.

7. REFERENCES

[1] J. Serrà, E. Gómez, and P. Herrera. *Audio cover song identification and similarity: background, approaches, evaluation, and beyond*. Studies in Computational Intelligence. Springer. In press.

[2] J. S. Downie. The music information retrieval evaluation exchange (2005–2007): a window into music information retrieval research. *Acoustical Science and Technology*, 29(4):247–255, 2008.

[3] J. Serrà, E. Gómez, P. Herrera, and X. Serra. Chroma binary similarity and local alignment applied to cover song identification. *IEEE Trans. on Audio, Speech, and Language Processing*, 16(6):1138–1152, August 2008.

[4] J. Serrà, X. Serra, and R. G. Andrzejak. Cross recurrence quantification for cover song identification. *New Journal of Physics*, 11:093017, September 2009.

[5] J. Serrà, M. Zanin, C. Laurier, and M. Sordo. Unsupervised detection of cover song sets: accuracy increase and original detection. *Conf. of the Int. Society for Music Information Research (ISMIR)*, October 2009.

[6] E. Gómez. *Tonal description of music audio signals*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2006. Available online: `http://mtg.upf.edu/node/472`.

[7] J. Serrà, E. Gómez, and P. Herrera. Transposing chroma representations to a common key. *IEEE CS Conference on The Use of Symbols to Represent Music and Multimedia Objects*, pages 45–48, October 2008.

[8] C. D. Manning, R. Prabhakar, and H. Schutze. *An introduction to Information Retrieval*. Cambridge University Press, 2008. Available online: `http://www.informationretrieval.org`.

[9] J. S. Downie, M. Bay, A. F. Ehmann, and M. C. Jones. Audio cover song identification: Mirex 2006-2007 results and analyses. *Int. Symp. on Music Information Retrieval (ISMIR)*, pages 468–473, September 2008.