DRAFT: A REFINED BLOCK-LEVEL FEATURE SET FOR CLASSIFICATION, SIMILARITY AND TAG PREDICTION

Klaus Seyerlehner, Markus Schedl, Peter Knees, Reinhard Sonnleitner
Dept. of Computational Perception, Johannes Kepler University, Linz, Austria

ABSTRACT

In our submission we use a set of so-called block-level features (BLF) for three different tasks, namely genre classification, tag classification and music similarity estimation. Compared to the submission in 2010 two additional feature were added to the feature set. This abstract gives an overview on the feature set and presents some specific details of the submitted algorithms.

1. INTRODUCTION

In our system the same set of features is extracted for all three tasks. The feature extraction is implemented in MATLAB. All submitted algorithms also contain a classification part, which is based on the WEKA machine learning toolbox [3]. In the following subsection we first discuss the audio features set used in our submissions especially pointing out the differences to the last year's submission. Then in the subsequent sections we discuss the most important algorithmic details of our submissions.

2. AUDIO FEATURES

In all our submissions we extract the same set of block-level features (BLF), as we did in our last year's submission [18]. Compared to our 2010 submission two additional features (GT, LSG) are extracted within the proposed block-processing framework. Altogether, the extracted feature set consists of the following BLF:

- Spectral Pattern (SP)
- Delta Spectral Pattern (DSP)
- Variance Delta Spectral Pattern (VDSP)
- Logarithmic Fluctuation Pattern (LFP)
- Correlation Pattern (CP)
- Spectral Contrast Pattern (SCP)
- **Local Single Gaussian Model (LSG)**
- **George Tzanetakis Model (GT)**

For a more detailed description of these features and their extraction process we refer to [17, 19, 20]. Here we only briefly present the specific details of the newly introduced patterns in our feature set.

2.1 Local Single Gaussian Model (LSG)

The Local Single Gaussian Model is a timbral feature. Similar to [10] for each block — in this case a block consists of a consecutive set of 100 MFCC frames — the mean and covariance over the block's MFCCs feature vectors are computed. Together mean and covariance form the local feature vector. To come up with a global song-level feature vector mean and variance are computed separately for each dimension of the local feature vectors.

2.2 George Tzanetakis Model (GT)

The GT model is another timbral feature based on MFCCs. In contrast to standard MFCCs 200 Mel filters and 50 MFCC coefficients are used to more precisely model the spectral envelope of an audio frame. Then similarly to [22] together mean and standard deviation over a block's MFCC vectors form the local feature vector and finally the local feature vectors are once more summarized using mean and variance to generate the global song-level feature vector.

Figure 1 visualizes the proposed set (except the LSG and GT model, which do not have an obvious visual representation) of features for two different songs, a Hip-Hop and a Jazz song.

3. GENRE CLASSIFICATION

The genre classification approach itself is rather straight forward. The presented block-level features are combined into a single feature vector that forms the input to the classification stage. To train and predict genre labels the WEKA support vector machine implementation (SMO) is used. Comparing the results to our MIREX 2010 submission we obtain improved classification results on some well-known datasets (see table 1). According to these experiments the newly introduced block-level features seem to further improve the classification accuracy of our system.

4. AUTOMATIC TAG PREDICTION

In general tag prediction can be viewed as a simple extension of the genre classification approach from single to

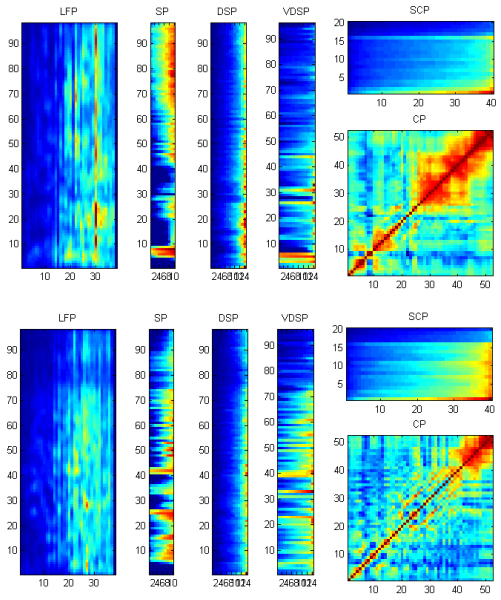

Figure 1. *Visualization of the proposed block-level patterns for a Hip-Hop song (upper) and a Jazz song (lower).*

multi-label classification. In tag classification there is, instead of a single classifier like in genre classification, one classifier per tag. In contrast to our last year's submission, where we made use of a PCA to reduce the high dimensionality of the block-level feature set in order to reduce the runtime, this year we directly use the full feature set. To meet the runtime requirements the SVM classifier was replaced by a random forest classifier. According to our experiments this approach clearly outperformed our 2010 submission. Furthermore, we swapped from using a fixed binarization threshold to also using a dynamic threshold as proposed in [11].

5. MUSIC SIMILARITY ESTIMATION

Our music similarity estimation approach is based on two distinct components: *Block-Level Feature Similarity* and *Tag Affinity Based Similarity*. The following two subsections present the algorithmic details of these two components.

5.1 Block-Level Feature Similarity

To directly estimate music similarity based on the presented block-level features we follow the approach presented in [19]. First, pairwise song similarities are estimated by computing the Manhattan distance for each of the presented block-level features separately (expect for the LSG pattern which is not used in this task). Then in a second step the individual distance matrices resulting from the individual patterns are combined into a single distance matrix. This is realized by by first normalizing the individual distance matrices using a distance space normalization approach (DSN) [14, 16, 19] and then combining the individual matrices by summing up the corresponding pairwise distances over all matrices. The weights for the contribution of the individual patterns to the overall similarity are the same as last year and are defined in [19]. For the GT model, which was not part of the last year's algorithm we set the weight to $w = 1$.

Reference	Dataset	Accuracy
Tazanetakis et al. [23]	GTZAN	61.00%
Holzapfel et al. [4]	GTZAN	74.00%
Lidy et al. [9]	GTZAN	76.80%
Seyerlehner et al. [17]	GTZAN	77.96%
Panagakis et al. [12]	GTZAN	78.20%
Li. et al. [8]	GTZAN	78.50%
Bergstra et. al. [1]	GTZAN	83.00%
MIREX 2010 Submission	GTZAN	85.49%
MIREX 2011 Submission	GTZAN	87.03%
Panagakis et al. [13]	GTZAN	92.40 %
Panagakis et al. [12]	ISMIR2004all	80.95%
Lidy et al. [9]	ISMIR2004all	81.40%
Seyerlehner et al. [17]	ISMIR2004all	83.72%
MIREX 2010 Submission	ISMIR2004all	88.27%
MIREX 2011 Submission	ISMIR2004all	88.52%
Pohle et al. [15]	ISMIR2004all	90.04%
Panagakis et al. [13]	ISMIR2004all	94.38 %
Holzapfel et al. [5]	Ballroom	86.90%
Jensen et al. [6]	Ballroom	89.00%
Pohle et al. [15]	Ballroom	89.20%
MIREX 2010 Submission	Ballroom	92.44%
MIREX 2011 Submission	Ballroom	92.51%
MIREX 2010 Submission	Homburg	60.37%
MIREX 2011 Submission	Homburg	61.74%
MIREX 2010 Submission	1517-Artists	50.92%
MIREX 2011 Submission	1517-Artists	52.79%
MIREX 2010 Submission	Unique	75.41%
MIREX 2011 Submission	Unique	75.86%

Table 1. *Comparison of classification accuracies achieved by music genre classification approaches.*

5.2 Tag Affinity Based Similarity

For the tag affinity based music similarity as proposed in [2,24] we use a set of about 1500 classifiers pretrained on 4 different tag collections yielding a probabilistic tag affinity vector per song. The training data contained the *Magnatagatune* [7] dataset and three additional datasets. Then for each dataset separately a similarity estimate is derived using the Manhattan distance between the auto-tag vectors of each pair of songs. The similarity estimates resulting from each dataset are then once more combined using the DSN approach .

Finally, to generate the overall similarity matrix the matrices of both components (Block-Level Similarity and Tag Affinity Based Similarity) are simply added to combine them. Comparing this slightly modified algorithm to our last year's submission, preliminary experiments indicated an improvement of the artist filtered genre clustering eval-

uation criterion. This is also inline with the results from the MIREX evaluation, where SSPK2 yielded a artist filtered genre neighbourhood clustering of **59.67%**, while SSKS3 achieves **60.12%**. In contrast to our expectations both broad and fine scores decreased slightly. One possible explanation could be the *portfolio effect* [21] and refining this task by portfolio filtering the generated recommendations would be reasonable.

6. ACKNOWLEDGMENTS

This research was supported by the Austrian Research Fund (FWF) under grant L511-N15.

7. REFERENCES

[1] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kégl. Aggregate features and adaboost for music classification. *Machine Learning*, 2006.

[2] T. Bertin-Mahieux, D. Eck, F. Maillet, and P. Lamere. Autotagger: A model for predicting social tags from acoustic features on large music databases. *Journal of New Music Research*, 2008.

[3] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: An update. *SIGKDD Explorations*, 2009.

[4] A. Holzapfel and Y. Stylianou. Musical genre classification using nonnegative matrix factorization-based features. *IEEE Transactions on Audio, Speech, and Language Processing (TASLP-08)*, 2008.

[5] A. Holzapfel and Y. Stylianou. A scale transform based method for rhythmic similarity of music. In *Proc. of the 2009 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-09)*, 2009.

[6] J. H. Jensen, M. G. Christensen, and S.H Jensen. A tempo-insensitve representation of rhythmic patterns. In *Proc. of the 17th European Signal Processing Conf. (EUSIPCO-09)*, 2009.

[7] E. Law and L. Ahn. Input-agreement: A new mechanism for collecting data using human computation games. In *Proc. of the 27th Int. Conf. on Human Factors in Computing Systems (CHI-09)*, 2009.

[8] T. Li, M. Ogihara, and Q. Li. A comparative study on content-based music genre classification. In *Proc. of the 26th ACM SIGIR Conf. on Research and Development in Informaion Retrieval*, 2003.

[9] T. Lidy, A. Rauber, A. Pertusa, and M. Inesta. Improving genre classification by combination of audio and symbolic descriptors using a transcription system. In *Proc. of the 8th International Conference on Music Information Retrieval (ISMIR-07)*, 2007.

[10] M. Mandel. Svm-based audio classification, tagging, and similarity submission. In *online Proc. of the 7th Annual Music Information Retrieval Evaluation eXchange (MIREX-2010)*, 2010.

[11] S. R. Ness, A. Theocharis, G. Tzanetakis, and L.G. Martins. Improving automatic music tag annotation using stacked generalization of probabilistic svm outputs. In *Proc. of the 17th ACM Int. Conf. on Multimedia (MM -09)*, 2009.

[12] I. Panagakis, E. Benetos, and C. Kotropoulos. Music genre classification: A multilinear approach. In *Proc. of the 9th International Conference on Music Information Retrieval (ISMIR-08)*, 2008.

[13] Y. Panagakis, C. Kotropoulos, and G.R. Arce. Music genre classification using locality preserving non-negative tensor factorization and sparse representations. In *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, 2009.

[14] T. Pohle and D. Schnitzer. Striving for an improved audio similarity measure. In *online Proc. of the 4th Annual Music Information Retrieval eXchange (MIREX-07)*, 2007.

[15] T. Pohle, D. Schnitzer, M. Schedl, P. Knees, and G. Widmer. On rhythm and general music similarity. In *Proc. of the 10th International Society for Music Information Retrieval Conference (ISMIR-09)*, 2009.

[16] D. Schnitzer, A. Flexer, M. Schedl, and G. Widmer. Using mutual proximity to improve content-based audio similarity. In *Proc. of the 12th Int. Conf. for Music Information Retrieval (ISMIR-2011)*, 2011.

[17] K. Seyerlehner and M. Schedl. Block-level audio feature for music genre classification. In *online Proc. of the 5th Annual Music Information Retrieval Evaluation eXchange (MIREX-09)*, 2009.

[18] K. Seyerlehner, M. Schedl, T. Pohle, and P. Knees. Using block-level features for genre classification, tag classification and music similarity estimation. In *online Proc. of the 7th Annual Music Information Retrieval Evaluation eXchange (MIREX-2010)*, 2010.

[19] K. Seyerlehner, G. Widmer, and T. Pohle. Fusing block-level features for music similarity estimation. In *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, 2010.

[20] K. Seyerlehner, G. Widmer, M. Schedl, and P. Knees. Automatic music tag classification based on block-level features. In *Proc. of the 7th Sound and Music Computing Conference*, 2010.

[21] Klaus Seyerlehner. *Content-based Music Recommender Systems: Beyond simple Frame-Level Audio Similarity*. PhD thesis, Johannes Kepler University, 2010.

[22] G. Tzanetakis. Marsyas submissions to mirex 2010. In *online Proc. of the 7th Annual Music Information Retrieval Evaluation eXchange (MIREX-2010)*, 2010.

[23] G. Tzanetakis and P. Cook. Musical genre classification of audio signal. *IEEE Transactions on Audio and Speech Processing*, 10(5):293–302, 2002.

[24] K. West, S. Cox, and P. Lamere. Incorporating machine-learning into music similarity estimation. In *Proc. of the 1st ACM Workshop on Audio and Music Computing Multimedia (AMCMM-06)*, 2006.