CULTURAL STYLE BASED MUSIC CLASSIFICATION OF AUDIO SIGNALS

Yuxiang Liu[1,2]*, Qiaoliang Xiang*[2]*, Ye Wang*[2]*, Lianhong Cai*[1]

[1] Department of Computer Science & Technology, Tsinghua University, Beijing, China
[2] School of Computing, National University of Singapore, Singapore

liuyuxiang06@mails.tsinghua.edu.cn, qiaoliangxiang@comp.nus.edu.sg,
wangye@comp.nus.edu.sg, clh-dcs@tsinghua.edu.cn

ABSTRACT

Music classification based on cultural style is useful for music analysis and has potential applications in retrieval and recommendation systems. In this paper, we present the first attempt to classify audio signals automatically according to their cultural styles, which are characterized by timbre, rhythm, wavelet coefficients and musicology-based features. Machine learning algorithms are employed to investigate the effectiveness of various features on a data set of 1300 music pieces, collected for this research. Experimental results show that the proposed method can achieve an overall accuracy of 86% for six cultural styles, which shows the feasibility of integrating cultural style classification into music analysis and retrieval systems.

Index Terms — Music information retrieval, cultural style, audio classification, feature extraction

1. INTRODUCTION

Culture has a significant influence on music in term of creation, performance and interpretation. People from a certain cultural background often prefer music with a particular cultural style. Therefore cultural style information is useful for music browsing, retrieval and recommendation. This is important in today's Internet enabled global village where people can perform cross-culture music exploration. It can be observed that music pieces with the same cultural style share similar attributes such as the tuning system, musical scale and instrumentation. Based on this observation, we hypothesize that music audio signals can be classified according to their cultural styles using machine learning methods.

In the domain of music analysis, most existing works focus on western music. For instance, existing genre classification systems label all types of non-western music as world music (or traditional music, folk music etc.), despite the fact that music in this category, such as Chinese traditional music and Indian classical music, can be significantly different. Therefore, cultural style classification, which divides world music further into several categories in accordance with Ethnomusicology, complements existing works on genre classification. This paper presents a pioneer work on cultural style based music classification of audio signals. Four sets of features (i.e., timbre, rhythm, wavelet coefficients and musicology-based features) are investigated together with supervised classifiers to determine the cultural style of a music piece. A comparative study shows that timbre features combined with the SVM are most effective.

2. RELATED WORKS

Content-based music classification has been an active research topic recently. Tzanetakis and Cook [1] attempted to classify musical genre via timbre, rhythm and pitch related features, and compared the performance of global features with that of instantaneous features. Octave-based spectral contrast [2] and wavelet coefficients [3] based methods were also investigated. In [4], Lu *et. al*. proposed a hierarchical mood detection framework which had the advantage of utilizing proper features in different steps of classification. Most of the techniques discussed above managed to use acoustic features to represent timbral texture, melody and rhythm of musical signals. Statistical pattern recognition classifiers were trained and used to examine performances of various feature sets.

As non-western music is concerned, Pandey *et. al*.[5] implemented a system to identify Raags, a traditional form in Indian classical music, using heuristic note transcription and HMMs. In order to avoid using unreliable note transcription, Chordia and Rae [6] utilize pitch-class distribution rather than estimated notes for the same task. Norowi *et. al.* [7] used features similar to [1] in Malay music classification and has shown that those features are suitable for Malay music as well. Characteristics of non-western music explored by the above works indicated the possibility of distinguishing non-western music from western music automatically.

To our knowledge, the only work addressing cultural style classification, which is most relevant to the research described in this paper, were presented in [8]. Pitch values as well as their intervals and contours were extracted from monophonic symbolic data and used to track music melody. HMMs were trained and used to classify music pieces via

This work was performed in National University of Singapore. The authors were supported by Singaporean Ministry of Education grant (No. R-252-000-267-112) and National Science Foundation of China grant (No. 60433030).

their melody sequences. The experimental data set contained five hundred pieces of Irish, German and Austrian folk music. The accuracy achieved by this method is 70% for binary classification and 66% for tri-style classification. However, the paper did not present a solution of cultural style classification for audio data. Given that performances of existing music transcription systems are not satisfactory and it is a time-consuming work to manually convert audio data into symbolic format, it is necessary to classify cultural styles of music by analyzing acoustic signals directly.

3. FEATURE EXTRACTION

3.1 Timbre features

Unique instruments are typically employed in music of different cultures. Based on this observation, we employ timbral texture, which is characterized by spectral shape and spectral contrast, for our task. More specifically, signals are segmented into 100ms long 50% overlapped frames. The following features are then extracted from each frame:

- *Spectral Centroid*: amplitude weighted mean of the spectrum;
- *Bandwidth*: amplitude weighted mean of differences between frequency components and the spectral centroid;
- *Spectral Flux*: Euclidean distance between normalized spectrum distributions of two successive frames;
- *Spectral Rolloff*: the frequency boundary, which 85% energy in the spectrum distribution is below;
- *Subband Energy*: the average energy in each subband;
- *Subband Peaks*: the average amplitude of peaks (local maxima of the amplitude envelope) in each subband;
- *Subband Contrast*: the average difference between peaks and successive valleys in each subband.

In our experiments, the number of subbands is set to 8. Thus we obtain a 28 dimensional feature vector for each frame. Then statistics (mean and standard deviation) of the above features are calculated through all frames and used as the input of classifiers.

3.2 Rhythm features

Rhythm strength, rhythm regularity and tempo are important attributes of musical signals. This inspires us to investigate their relevance to cultural style classification.

We segment music pieces into non-overlapping segments of duration about 2 seconds, which is determined empirically. For each segment, signals are separated into six subbands in an octave manner. *Onset curve* is then defined and calculated as the variance of the amplitude envelope in each subband. After summing up six subband onset curves, the following five features are extracted from the overall onset curve, according to the three aspects mentioned above:

- the average strength of a onset curve;
- the average strength of peaks in a onset curve, as well as the ratio between peaks and successive valleys;
- the frequency of peaks as well as the maximum common divisor of peak positions in a onset curve.

Mean and standard deviation of the above features are then computed over the entire piece and concatenated to form a 10 dimensional feature vector.

3.3 Wavelet-based features

Wavelet coefficients have the advantage of containing both global and local information, which was shown to be useful in audio signal analysis [3].

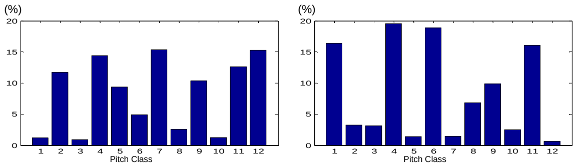
In our implementation, we decompose musical signals via a seven-level Daubechies wavelet filter bank. The normalized histogram of wavelet coefficients is constructed in each level and interpreted as the probability distribution of wavelet coefficients. The first three moments (mean, variance and skewness) are then used to characterize the shape of a distribution. Finally, a 3×8 dimensional feature vector is obtained.

3.4 Musicology-based features

While most existing works only employ low level audio features in their analysis framework, we seek to investigate the effectiveness of musicology-based features for the cultural style classification task.

3.4.1 Chroma distribution and chroma contrast

Chromagram (or pitch class profile) [9] represents the spectrum of musical signals in a chromatic scale, where two pitches separated by octaves have the same chroma value. Energies of pitches with the same chroma value are summed up and mapped into a 12 dimensional chroma vector. In our approach, we average energies of each element in chroma vectors over time and then normalize the total energies into one. This normalized chroma histogram is interpreted as the distribution of 12 pitch classes.



(a) Chroma distribution of J.S.Bach's Sonata No.6 in G major, BWV530 - I. Vivace

(b) Chroma distribution of Chinese traditional music “For the Ox-Cart”

Fig. 1. Chroma distribution of music from different culture

Fig. 1 shows chroma distributions of two pieces of Western classical music and Chinese traditional music respectively. It is easy to observe that the chroma distribution in (a) is much more dispersed than that in (b). The musicological explanation to this phenomenon is Chinese traditional music scales are pentatonic, which has five pitches per octave, compared to western diatonic scales which are made up of seven distinct notes. These differences can potentially help us to construct statistical model for each cultural style.

Suppose the distribution vector of 12 pitch classes is $\{e_1, e_2, \ldots, e_{12}\}$. After sorting the elements in a descending order, we represent the sorted vector as $\{e'_1, e'_2, \ldots, e'_{12}\}$, where $e'_1 > e'_2 > \ldots > e'_{12}$. Then *chroma contrast* is defined as:

$$\textit{chroma contrast} = \sum_{i=1}^{N} e'_i \Big/ \sum_{j=N+1}^{12} e'_j$$

It is shown from our experiments that the average *chroma contrast* (N=6) of Western classical music is around 3.0, while it is as high as 8.8 for Chinese traditional music.

3.4.2 Chord distribution and chord contrast

Chord plays an important role in music expression and the usages of chords are dissimilar in different cultural styles. Although automatic chord recognition has been investigated by other researchers before, the performance is not yet satisfactory. Hence, we estimate the probability of one segment being a certain chord rather than detect and indentify chords from audio signals. In our approach, the confidence score of such probability is obtained as below:

a) 24 chord templates are synthesized via a MIDI device;
b) chroma distributions are extracted from these templates;
c) the dissimilarity measure between audio segments and standard templates is evaluated by symmetrical Kullback-Leibler (SKL) divergences of the chroma distributions. The reciprocal of SKL divergence is considered as the confidence score.

We average the scores through the entire music piece and obtain a 24 dimensional vector which is considered as the chord distribution of a music piece. The ratio of the summation of the N chords with the highest score and that of the rest chords is defined as the *chord contrast.*

3.4.3 Pitch interval histogram

In addition to chroma and chord, the profile of pitch intervals may be another useful characteristic of cultural styles. Again, taking Western classical music and Chinese traditional music for example, the absence of the notes F and B in the Chinese pentatonic scale makes intervals between successive notes either major seconds or minor thirds, while minor seconds may occur in western scales.

Based on this observation, pitch interval histograms are calculated to distinguish musical scales in different cultural styles. For the pitch estimation, we use the well-known YIN algorithm [10] and quantize estimated pitches to semitone levels. In order to avoid octave error which often occurs in a note transcription system, pitch values are mapped to 12 pitch classes. Then intervals between successive pitches are computed and a histogram with 12 bins is constructed as the feature vector.

4. CLASSIFICATION

For classification purposes, three supervised classifiers (decision tree, KNN and SVM) are employed.

In our experiments, a decision tree is generated by the C4.5 algorithm [11] and the optimal tree level is determined automatically through cross-validation among training data.

The K-nearest neighbor (KNN) algorithm assigns labels based on the closest training examples. Here, Euclidean distance is adopted as the distance metric and the value of K is also determined in a cross-validation fashion.

The Multi-class SVM decomposes a multi-category classification problem to a series of binary classification. Both one-against-one (OVO) and one-against-all (OVA) strategies [12] are investigated in our experiments. For each binary SVM, the radius bias kernel function is used and parameter estimation for the kernel function is carried out in a grid search manner.

5. EVALUATION

5.1 Dataset

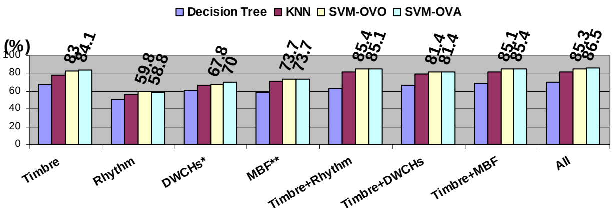
Our collection contains about 1300 music recordings of six cultural styles[1], Western classical music, Chinese traditional music, Japanese traditional music, Indian classical music, Arabic folk music and African folk music. Each style has roughly the same amount of audio samples. Each piece lasts about two to eight minutes and the total duration of the audio data exceeds 100 hours. Audio files, collected from commercial music CDs, are converted to 44.1 kHz, 16 bits/sample, mono WAV format.

5.2 Experiment Setup

The feature set consists of 56 timbre features, 10 rhythm features, 24 wavelet-based features, 13 chroma features, 25 chord features and 12 pitch interval features. Performances of classifiers combined with different kinds of features are evaluated via a random sub-sampling validation strategy, where seventy percent of data, 900 pieces, are randomly selected and used for training while the rest of around 400 pieces serve as testing data. To eliminate the bias caused by a particular partition of the training and testing data, we repeat the evaluation procedure several times and average the results from each trail.

5.3 Result

The overall classification accuracy for six styles is illustrated in Fig. 2. As demonstrated, SVM and KNN outperform decision tree significantly. The highest accuracy of 86.50% is achieved by SVM-OVA using all features.



*** : DWCHs stands for Daubechies Wavelet Coefficient Histograms**
****: MBF stands for Musicology Based Features**

Fig. 2. Accuracy of six-style classification

In term of effectiveness of individual feature sets, timbral texture is most effective. Timbre features alone

[1] Currently we only consider classical and traditional music. Due to the cultural interaction all over the world, modern music, especially popular music, becomes more and more cross-cultural and is hard to distinguish cultural styles among them.

together with SVM can already achieve 84.06% accuracy. We believe that this is because timbre features can effectively reflect the difference of instruments which are specific to cultures.

Experiments show that musicology-based features are not as discriminative as timbre features. Nevertheless, combining musicology-based features with timbre features can improve the classification accuracy by 1% to 2%, compared with that of using timbre features alone. The ineffectiveness of the musicology-based features may be due to the simplistic definition of the features. According to music theory, music styles relate to not only pitches and intervals but also the sequence and the organization of them. Thus, the technologies using only histograms and contrasts are unsatisfactory to model high level musical knowledge. Similar results are also observed in [1].

No improvement is obtained by including rhythm and wavelet related features. A possible explanation is that the rhythm and wavelet coefficients do not contain information related to cultural styles.

	Western	Chinese	Japanese	Indian	Arabic	African
Western	**97.33**	0.27	0.74	0.88	1.36	0.01
Chinese	1.06	**98.36**	0.74	0.29	0.67	0.95
Japanese	0.27	0.27	**94.81**	1.18	0.01	0.63
Indian	0.27	0.27	2.59	**79.41**	6.10	7.93
Arabic	0.80	0.01	0.01	6.76	**77.63**	21.90
African	0.27	0.82	1.11	11.47	14.24	**68.57**

Tab. 1. Confusion matrix for six-style classification based on SVM-OVA using all features (in percentage)

Tab. 1 shows the confusion matrix of the classification performance using SVM with the entire feature set. In the matrix, columns represent the actual styles, while rows correspond to the predicted styles. For western and oriental (Chinese and Japanese) music classification, the accuracy is above 92%. However, Indian classical music, Arabic folk music and African folk music are difficult to classify. (See the bottom right corner of Tab. 1). On one hand, this is mainly due to the diversity existing inside these cultural styles. For example, there are two major traditions of Indian classical music, Hindustani sangeet in north Indian and Carnatic sangeet in south Indian. The two traditions are fundamentally similar but still differ in performance. Such diversity makes it hard to model their characteristics by common features. On the other hand this observation illustrates that standard classifiers which treat features equally without discrimination of different styles, is unsatisfactory for cultural style classification. Inspired by Lu [4], we believe that a hierarchical framework may make reasonable improvement by allowing us to model different subtasks independently. This could be future work.

6. CONCLUSION AND FUTURE WORK

Despite the fuzzy nature of boundaries between different cultural styles, automatic music classification based on cultural characteristics is possible. We have presented in this paper our initial work to address this challenge.

Timbre, rhythm, wavelet coefficients and musicology-based features are extracted from audio signals and used to capture characteristics of cultural styles. A combination of these features and SVM yields an overall classification accuracy of 86%. For some styles, such as Western and oriental (Chinese and Japanese) music, the classification accuracy can reach 95%. A comparative study shows that timbre features are most effective among four feature sets and the addition of musicology-based feature can gain a moderate improvement over the model using timbre features along.

Nevertheless, there is plenty of room for improvement in the proposed approach. Firstly, new features containing sequence and pattern information could be explored. Secondly, a further study on effectiveness of existing features especially in a pair-wise styles classification would be useful. Thirdly, it would be interesting to integrate music knowledge into machine learning framework. Finally, a hierarchical framework which explores different feature sets for each binary classification task, as opposed to multi-category classifiers, can be investigated for better performance.

REFERENCES

[1] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. on Speech and Audio Processing, Vol. 10, Issue 5*, 2002

[2] D. Jiang, L. Lu, H. Zhang, J. Tao, and L. Cai, "Music type classification by spectral contrast feature," *ICME*, Lusanne, Switzerland, 2002

[3] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification," *SIGIR*, Toronto, Canada, 2003

[4] L. Lu, D. Liu and H. Zhang, "Automatic mood detection and Tracking of Music audio Signals," *IEEE Trans. on Audio, Speech and Language Processing, Vol. 14, No.1*, 2006

[5] G. Pandey, C. Mishra, and P. Ipe, "TANSEN: A System For Automatic Raga Identification," *Indian International Conference on Artificial Intelligence*, Hyderabad,India, 2003

[6] P. Chordia and A. Rae, "Automatic Raag Classification Using Pitch-class and Pitch-class Dyad Distributions," *ISMIR*, Vienna, Austria, 2007

[7] N. M. Norowi, S. Doraisamy, and R. Wirza, "Factors Affecting Automatic Genre Classification: An Investigation Incorporating Non-Western Musical Forms," *ISMIR*, London, UK, 2005

[8] W. Chai and B. Vercoe, "Folk Music Classification Using Hidden Markov Models," *IC-AI*, Las Vegas, NV, 2001

[9] T. Fujishima, "Realtime chord recognition of musical sound: a system using common lisp music," *ICMC*, Beijing, China, 1999

[10] A de Cheveigné and H Kawahara, "YIN, a fundamental frequency estimator for speech and music," *Journal of the Acoustical Society of America*, Vol. 111, No. 4, 2002

[11] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993

[12] J. Weston and C. Watkins, "Multi-class support vector machines," *ESANN*, Brussels, Belgium, 1999