Gabor-LBP Features and Combined Classifiers for Music Genre Classification

Haiqian Wu,Ming Zhang
Department of Information Engineering, Shanghai Maritime University, Shanghai, 201306, China
E-mail: wuhaiqian@163.com

***Abstract*—In this paper, we propose the combination of two visual features with the Gabor filters and LBP for music genre classification. In addition, we use combined classifiers training and testing data to reach a final decision. The experiment shows that we not only reduce the amount of time extracting features, but also achieve a satisfactory result.**

Keywords-genre classification; Gabor filters; LBP; Combined classifiers

I. INTRODUCTION

With the development of the Internet, the amount of music has been increased rapidly in recent years. How to organize such vast amount of music is one of the main tasks of many music record companies and music websites. Classification is a very good strategy. At present most of the music database is organized according to one of music's attributes such as title, artist, album, genre, etc. Among these methods, genre classification is getting more and more attention because music genre is one of the most commonly used descriptors of music. But currently most music genre classification is performed manually, which cost too much time and money. In addition, because of the inherent subjectivity of genres, different people may have different opinions on the labels of a music piece, which would limit the recognition rate. These lead to the development of automatic music genre classification.

The process of music genre classification usually divided into two steps [1]: feature extraction and classification. In the first step, useful information representing the music is extracted from the music signals. In the second step an algorithm or a mathematical model is proposed to identify the labels of music with respect to their features.

Currently most of the state-of-the-art research on this field focuses on the development of new feature sets. One of the groundbreaking works was introduced by Tzanetakis and Cook [2], where they proposed three sets of features for representing timbral texture, rhythmic content and pitch content and achieved an accuracy of 61% on a dataset of 1000 music pieces. Lidy and Rauber [3] propose two new feature representations: Statistical Spectrum Descriptors and Rhythm Histogram features. Holzapfel and Stylianou [4] suggest a new feature set based on Nonnegative Matrix Factorization (NMF) for the description of the vertical structure of music.

A new feature is introduced based on the spectrogram which is computed from each music clip through the short-time Fourier transform (STFT). Deshpande et al. [5] use the texture-of-texture models to pick up features from the spectrogram. Later, Wu et al. [6] propose the use of Gabor filters to extract visual features of texture in the spectrogram. Costa et al. [7, 8] use another two texture features: Gray Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP). Both of them are good descriptors for music genre classification.

Considering the efforts done in the previous work, we make some improvement about the feature extraction and classification based on the contribution of Wu et al. [6]. We extract two visual features with the Gabor filters and LBP from the music clip; then train and test the classifier separately and combine each classifier's output to reach a final decision according to the product rule [9]

II. FEATURE EXTRACTION

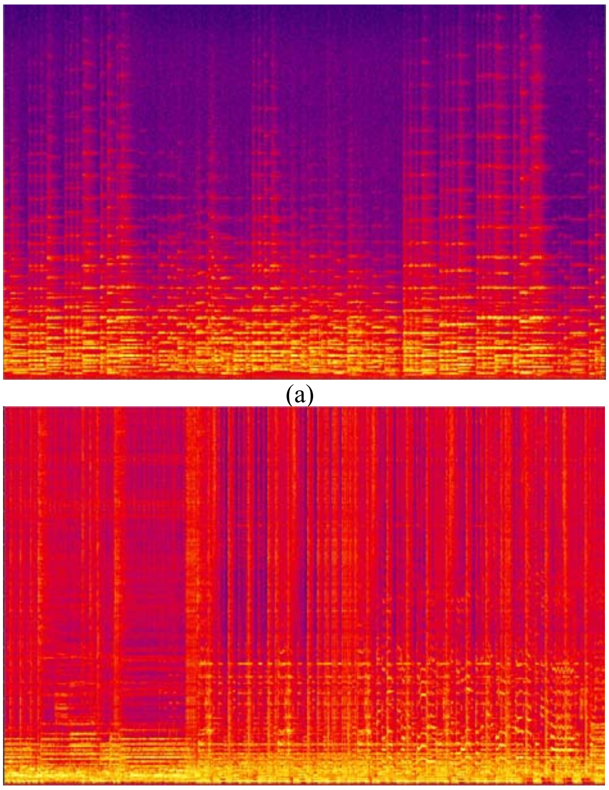
Spectrogram is computed from audio signals through the STFT with a window size of 1024 samples using the Hanning window. Fig.1 shows two spectrograms taken from music clips of different genres. We can see that the spectrogram of classical music has very clear horizontal lines while the spectrogram of disco music has a lot of vertical lines. Even though music clip seems to have no direct relation to its spectrogram, the music in different genres definitely has different spectrogram and the texture features are the most obvious one to distinguish these images.

Because of the large computation cost in the feature extraction via Gabor filters, in this paper we only consider a part of the frequency of the music. In other words, after converting the music into a spectrogram through STFT, we divide the spectrogram into the following sub-bands: 0-200Hz, 200-400Hz, 400-800Hz, 800-1600Hz, discarding the remaining part. The reason why we choose 4 sub-bands will be discussed in the experiment section. The method would definitely reduce the recognition accuracy, so we combine the LBP features [10] derived from the overall spectrograms, which cost less than Gabor filters. Fig.2 shows the flowchart of feature extraction. Some data processing is needed after the extraction.

A. Gabor Filter Features

We introduce the main steps of extracting the Gabor filter features.

Published by Atlantis Press, Paris, France.



(a)

(b)

Figure 1. Spectrogram of different music clips:(a) classical (b)disco

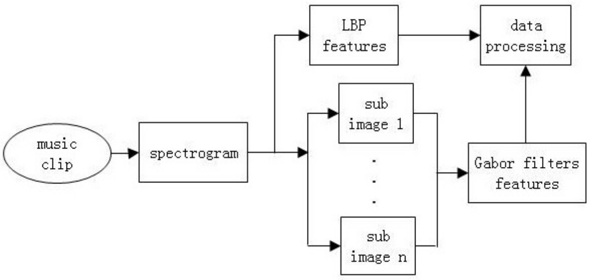


Figure 2. Flowchart of feature extraction

First, the convolution of the sub-image with the Gabor filters is computed:

$$G(m,n) = \sum_x \sum_y I(m-x, n-y)\psi(x,y) \qquad (1)$$

where I is the sub-image and $\psi(x,y)$ is 2-D Gabor filter in the spatial domain:

$$\psi(x,y) = \exp(-(\frac{x^2+y^2}{2\sigma^2}))\exp(\frac{j2\pi x}{\lambda}) \qquad (2)$$

Then the energy is computed as follow:

$$E = \sum_m \sum_n |G(m,n)| \qquad (3)$$

Finally, the mean and standard deviation of energy are used:

$$\mu = \frac{E}{M*N} \qquad (4)$$

$$s = \sqrt{\frac{\sum_m \sum_n (|G(m,n)| - \mu)}{M*N}} \qquad (5)$$

where M and N are the width and height of the sub-image respectively.

All sub-images take the same steps with different scales and orientations. More details can be found in [6]

B. LBP features

We also illustrate briefly how to compute LBP features from the spectrogram [10].

Fig.3 shows texture *T* in a local neighborhood of an image:

$$T \approx t(g_0 - g_C, ..., g_{P-1} - g_C) \qquad (6)$$

where g_C is the gray level intensity of pixel C(the central pixel) and g_0 to g_{P-1} corresponds to the gray level intensities of the P neighbors which equally spaced on a circle of radius R(In Fig.3, P=8, R=1).

Most of the time, it is sufficient just to consider the signs of the difference of gray value between the center pixel and its neighbors:

$$T \approx t(s(g_0 - g_C), ..., s(g_{P-1} - g_C)) \qquad (7)$$

where s(x) is the sign function:

$$s(x) = \begin{cases} 1 & if \quad x >= 0 \\ 0 & if \quad x < 0 \end{cases} \qquad (8)$$

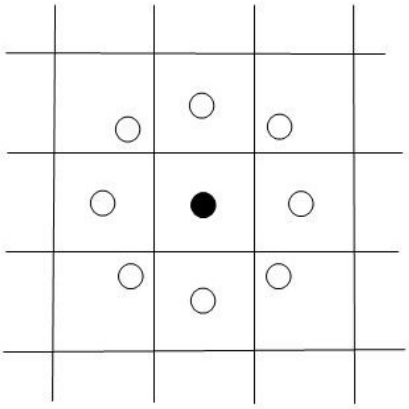


Figure 3. Neighbor sets for texture T

Then the LBP value can be obtained as follows:

$$LBP_{P,R}(x_C,y_C)=\sum_{P=0}^{P-1}s(g_P-g_C)2^P \qquad (9)$$

Instead of using the whole value, 58 possible uniform values and an extra value which is obtained by all the non-uniform patterns are advised in the experiment of [8].

III. CLASSSIFICATION

Support Vector Machine (SVM) is commonly used classifier in music genre classification. But SVM is nonlinear classifier and it takes more time training and testing data than linear classifier. On the other hand, linear classifiers are much ineffective in terms of accuracy to some extent. As a compromise, we use the combination of linear and nonlinear classifiers. To compensate for the loss of accuracy, we didn't use the definite outputs of the classifiers. Instead, the probabilistic outputs are used and they are combined with product rule [9] to reach a final decision. Costa et al. [8] has shown in their experiment that product rule is much better than other three decision rules.

Here we use the PRTools [11] to help conduct our experiment. Three classifiers of different kinds will be used in the training and testing process, namely: Fisher, LDC and KNN. The first two are linear classifiers and the third is nonlinear one.

First, each vector extracted from every sub-image and the overall LBP vector is used to train and test the above three classifiers separately. Then their probabilistic outputs are combined according to the product rule to determine the final label of the song. Fig.4 shows this process.

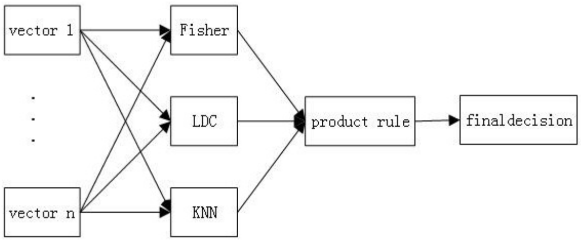


Figure 4. Combined classifier

IV. EXPERIMENT RESULT

In order to be able to compare our work to previous studies, we used the publicly available dataset: GTZAN [2]. This dataset contains 1000 music clips composed of 10 genres: blues, classical, country, disco, hippop, jazz, metal, pop, reggae and rock. They are stored as 22050Hz, 16-bit, mono audio files. The classification results are calculated using a 10-fold cross-validation.

We conduct a series of experiments to explain why we use 4 sub-bands in the feature extraction section.

Fig.5 shows the recognition rate using feature vectors of only one band. It can be seen that the recognition rates of the 7 sub-band range from 49% to 64%, all of which lower than 65%. Obviously only one sub-band's result does not meet our requirement. Furthermore, the bandwidth of the sub-band gets larger in the higher frequency domain(except for the last one), which means the corresponding sub-image becomes bigger and costs much more time computing the Gabor filter features.

Fig.6 and Fig.7 shows the recognition rate using successive sub-band and the corresponding time cost respectively. We can conclude from Fig.6 that the join of higher frequency sub-bands decease the recognition rate vastly. We can also see that the recognition rate gets the highest when using the sub-band of 0-1600Hz and corresponding time cost of feature extraction is very satisfactory. So in our experiment we only use 4 sub-bands.

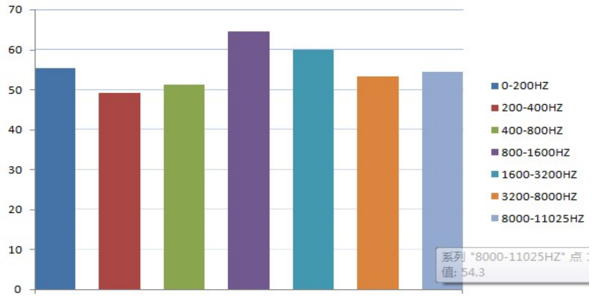


Figure 5. Sub-band recognition rate

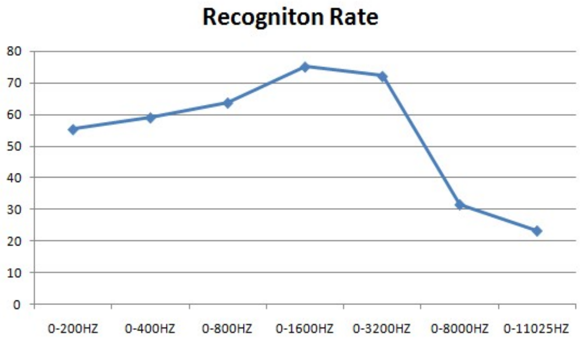


Figure 6. Successive sub-band reconnition rate

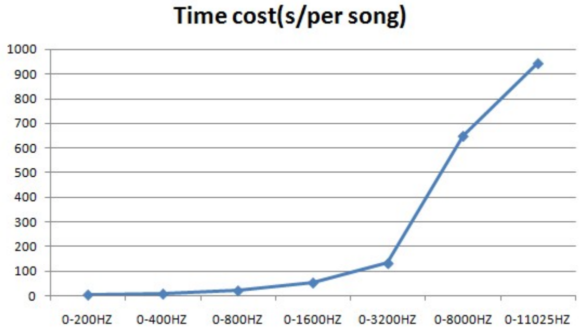


Figure 7. Time cost of computing Gabor filter features

TABLE I. TIME COMPARISION

Features	***Time(s/per song)***
Gabor features (7 sub-image)	943
Gabor features(4 sub-image)+LBP	53

TABLE II. CONFUSION MATRIX

Genre	***a***	***b***	***c***	***d***	***e***	***f***	***g***	***h***	***i***	***j***
blues	77	0	1	6	0	8	8	0	0	0
classical	0	91	4	0	0	4	1	0	0	0
country	5	1	83	1	0	5	3	2	0	0
disco	2	2	2	82	5	0	2	5	0	0
hippop	0	0	2	0	87	0	4	7	0	0
jazz	9	2	4	0	0	82	3	0	0	0
metal	0	0	3	1	0	0	96	0	0	0
pop	1	0	3	8	3	0	5	80	0	0
reggae	4	0	0	5	1	0	0	3	76	11
rock	0	0	0	0	0	0	1	0	4	95

In addition, we record the time cost of the visual feature extraction in [6] and ours. The results are shown in Table 1.One can observe that we decrease the time significantly.

Table 2 shows the confusion matrix of our best classification result. Compared to the performance using only visual features [6], we increase the recognition rate of disco and rock to 82% and 95% respectively. Moreover, our best classification accuracy reaches about 84.9%, which is comparable to [6].

V. CONCLUTION

In this paper we have improved the visual features with Gabor filter features and LBP, proposed the linear and nonlinear combined classifiers for music genre classification. Our experiment shows a preferable result in some aspect. More improvement can be done to achieve higher classification accuracy, such as finding a better descriptor of the spectrogram, which will be the subject of future work.

ACKNOWLEDGMENT

We would like to thank to Dr. Costa and Dr. Wu for their kind hearted help and generous sharing of some algorithm.

REFERENCE

[1] T.Li,M.Ogihara,Music genre classification with taxonomy, International Confence on Acoustics, Speech, and Signal Processing(ICASSP), 2005.

[2] G.Tzanetakis, P. Cook, Musical genre classification of audio signals,IEEE Transactions on Speech and Audio Processing, vol. 10,2002,pp.293–302.

[3] T. Lidy and A. Rauber. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. InProc. ISMIR, 2005.pp.34-41.

[4] A. Holzapfel and Y. Stylianou. Musical genre classification using nonnegative matrix factorization-based features. IEEE Transactions on Audio, Speech, and Language Processing, vol.16, 2008,pp.424–434,

[5] H. Deshpande, R. Singh, U. Nam, Classification of music signals in the visual domain, Proceedings of the COST-G6 Conference on Digital Audio Effects, 2001, pp. 1–4.

[6] M. Wu, Z. Chen, J.R. Jang, J. Ren, Combining visual and acoustic features for music genre classification, 10th International Conference on Machine Learning and Applications, 2011, pp. 124–129.

[7] Y.M.G. Costa, L.S. Oliveira, A.L. Koerich, F. Gouyon, Music genre recognition using spectrograms, 18th International Conference on Systems, Signals and Image Processing, 2011, pp. 151–154.

[8] Y.M.G. Costa. L.S. Oliveira, A.L. Koerich, F. Gouyon, J.G.Martins, “Music genre classification using LBP textural features”. Signal Processing,vol. 92,2012,pp.2723–2737

[9] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers,IEEE Transactions on Pattern Analysis and Machine Intelligence,vol.20,1998,pp.226–239

[10] T. Ojala, M. Pietik□inen, T. M□enp□□, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24,2002,pp.971–987.

[11] R.P.W. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, D.M.J. Tax, S. Verzakov PRTools4.1, A Matlab Toolbox for Pattern Recognition , Delft University of Technology, 2007.