Automated Classification of Audio Genre using Sequential Forward Selection Method.

Jong-Hak Lee*, Won-Jung Yoon*, Kang-Kyu Lee*, Kyu-Sik Park*
*Dept of Information and Computer Science, Dankook University, Seoul, 140-714, Korea
Tel : +82-2-709-2728 Fax : +82-2-796-2970 E-mail : jhlee@dankook.ac.kr, helloril@dankook.ac.kr, fitz@dankook.ac.kr, kspark@dankook.ac.kr

Abstract: In this paper, we propose a content-based audio genre classification algorithm that automatically classifies the query audio into five genres such as Classic, Hiphop, Jazz, Rock, Speech using digital signal processing approach. From the 20 second query audio file, 54 dimensional feature vectors, including Spectral Centroid, Rolloff, Flux, LPC, MFCC, is extracted from each query audio. For the classification algorithm, k-NN, Gaussian, GMM classifier is used. In order to choose optimum features from the 54 dimension feature vectors, SFS (Sequential Forward Selection) method is applied to draw 10 dimension optimum features and these are used for the genre classification algorithm. From the experimental result, we verify the superior performance of the SFS method that provides near 90% success rate for the genre classification which means 10%~20% improvements over the previous methods

1. INTRODUCTION

As more and more audio data is stored on standalone computers and the Internet, effective automatic audio classification and retrieval systems are required to fully use of audio information. Audio classification is the prerequisites to audio retrieval for helping to reduce the search space and it refers to the process that determines where a given audio signal is speech, Rock music or other types of music.

Musical genre classification based on music content has been a growing area of research in the last few years. Example applications of automatic musical genre classification include search and select music from music digital library (MDL), entertainment industry, virtual reality, and several others on web application. Musical genre information can also be utilized to help music information retrieval (MIR) so that the search can be made to look for data more closely in certain classes than in other classes to improve the search speed and search accuracy.

Most content-based classification methods have three common stages of a pattern recognition problem: feature extraction, training of the classifier based on the sample music, and classification. Depending on different combinations of these methods, several musical genre recognition strategies are employed in these studies. Tzanetakis and Perry [1] who initiated the problem, combined standard timbral features with representations of rhythm and pitch content and they achieved classification performance in the rage of 60% for ten musical genres. The classification accuracy based on just rhythm and pitch content was quite poor such as 23%~28%. In Ref. [2], Li performed extensive comparative study on the selection of features between Daubechies wavelet coefficient and the ones used in [1], and they conclude that the timbral feature is more suitable than rhythmic or pitch content for musical genre classification. Burred et al., [3] suggested hierarchical classification approach and genre dependent feature sets. In 13 musical genres, they can achieve 57.8% classification accuracy. Another interesting approach by Guo and Li [4] was proposed to use support vector machines (SVM) with binary tree recognition. In their work, new metric called distance-from boundary (DFB) with SVM is used to measure music pattern similarity between the classes. On the other hand for the classification strategies, several different methods have been employed in these studies, including • •NN (Nearest Neighbor), Gaussian models, Gaussian mixture models, neural networks, support vector machine and hidden Markov models. Based upon the reviews of these papers, an important observation is that a major improvement could be mainly result from a better description of the music signals, i.e., better music features rather than new efficient classification schemes.

In this paper, we propose an automated classification of audio genre using sequential forward selection method. The proposed system automatically classifies the query audio into five genres such as Classic, Hiphop, Jazz, Rock and Speech. At first, 54 dimensional feature vector is extracted for each audio file and then these features vector is optimized and down-sized into 10 dimensions though Sequential Forward Selection (SFS) method. Finally, the pattern matching using k-NN, Gaussian and GMM between the feature vector of the query and the database is performed and the genre of query music is returned as a query result.

2. FEATURE EXTRACTION AND SFS FEATURE SELECTION

2.1 Feature Extraction

Feature extraction can be thought of as representation conversion, taking low-level representation and identifying higher level features [5]. Before classification, the music signals are normalized to have zero mean and unit variance in order to avoid numerical problems caused by small variances of the feature values. At the sampling rate of 22000 Hz, the music signals are divided into 23ms frames with 25% overlapped hamming window at the two adjacent frames. Two types of features are computed from each frame: One is the timbral features such as spectral centroid, spectral Rolloff, spectral flux and zero crossing rates. The other is coefficient domain features such as Mel-frequency cepstral coefficients (MFCC) and linear predictive coefficients (LPC). The means and standard deviations of these six original features are computed over each frame for each music file to form a total of 54-dimensional feature vector. Table 1 summarizes the feature set used in this paper. These features are well-known in the literature and only the short description of definition is given in the table.

Table 1. Audio feature definition

Feature	Definition
Spectral centroid	It is defined as the center of gravity of STFT magnitude spectrum.
Spectral Rolloff	It is defined as the frequency below when 85% of the magnitude distribution is concentrated.
Spectral Flux	It is the squared difference between the magnitudes of successive spectral distribution.
ZCR	It is the number of time-domain zero-crossings.
MFCC	MFCC is the most widely used feature in speech recognition. It captures short-term perceptual features of human hearing system. Thirteen coefficients are used.
LPC	LPC are a short-time measure of the speech signal with describes the signal as the output of all-pole filter. Ten coefficients are used.

2.2 SFS Feature Selection

Not all the 54-dimensional features are used for musical genre classification purpose. Some features are highly correlated among themselves and some feature dimension reduction can be achieved using the feature redundancy. In order to reduce the computational burden and so speed up the search process, while maintaining a system performance, an efficient feature dimension reduction and selection method is desired. In this paper, we adopt the same SFS method for feature selection to reduce dimensionality of the features and to enhance the classification accuracy. Firstly, the best single feature is selected and then one feature is added at a time which in combination with the previously selected features to maximize the classification success rate. This process continues until all 54 dimensional features are selected. After completing the process, we pick up the best feature lines that maximize the classification success rate. This allows choosing the sub-optimum features for musical genre classification.

3. EXPERIMENTAL RESULTS

The proposed algorithm has been implemented and used to classify music data from a database of about 300 audio files. 60 audio samples were collected for each of the five genres in Speech, Classical, Hiphop, Jazz, and Rock, resulting in 300 files in database. The 300 audio files are partitioned randomly into a training set of 210 (70%) sounds and a test set of 90 (30%) sounds. In order to ensure unbiased classification accuracy because of a particular partitioning of training and testing, this division was iterated one hundred times. The overall classification accuracy was obtained as the arithmetic mean of the success rate of the individual iterations.

For classification purposes, a query to be classified is compared to training feature vectors from different classes. Then the pattern matching using K-NN, Gaussian and GMM between the feature vector of the query and the database is performed and the genre of query music is returned as a query result.

Fig. 1 shows the average classification success rate with all 54 dimensional feature vector described in section 2.1. As seen from the figure, the average success rate is near 70% which has similar results reported in Ref. [1].

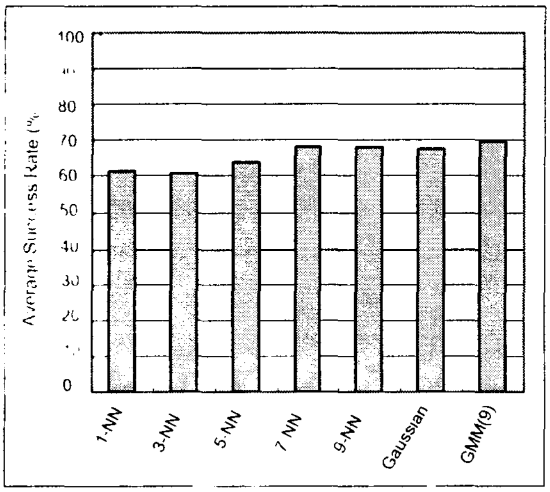

Fig. 1. Average classification success rate using all 54 dimensional feature vectors

Fig. 2 shows SFS feature selection procedure as described in section 2.2. As seen in the figure, the classification performance increases with the increase of features up to 10 with near 90% of accuracy. And it remains constant up to 10~13 features. After 13 features, it even makes the system performance worse.

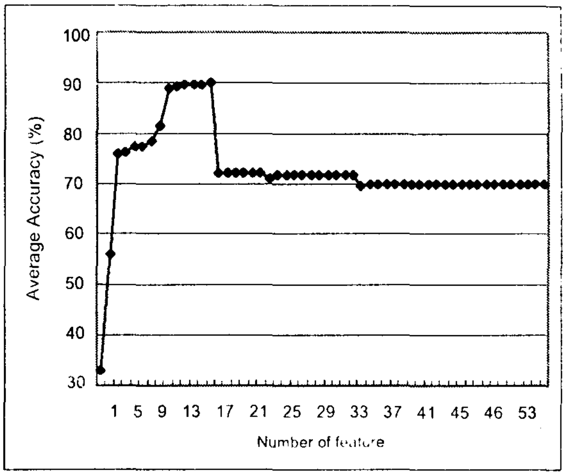

Fig. 2. SFS feature selection procedure

Therefore, we can select only first 10 features to represent each music signals. Table 2. shows detailed SFS performance in musical genre classification in a form of a confusion matrix. As a comparison purpose, the classification results using 54 dimensional feature vectors are included in the table. The numbers of correct classification with SFS lie in the diagonal of the confusion matrix. The numbers shown in parenthesis represent statistics with all 54 dimensional features. From the Table, we see at least 15% improvement of classification performance using only 10 dimensional features derived from SFS feature selection method. The SFS method works fairly well over the genre of Speech, Classical, Hiphop, and Rock while the average rate of correct classification is little lower in Jazz genre. Jazz music is often misclassified as Classical or Hiphop or Rock. This is because the Jazz music contains mixed characteristics over the other types of music and it is very similar to what a human would do. The confusion matrix also shows some notable misclassifications over the Hiphop and Rock music because of the strong beat that they possess in common.

Table 2. Genre confusion matrix with and without SFS

	Classical	Hiphop	Jazz	Rock	Speech
Classical	55(51)	0(0)	10(11)	2(1)	1(1)
Hiphop	1(0)	52(40)	4(12)	6(15)	0(1)
Jazz	2(5)	6(6)	38(27)	12(8)	0(2)
Rock	2(2)	2(14)	8(7)	40(33)	0(2)
Speech	0(2)	0(0)	0(3)	0(3)	59(54)
Average accuracy	92% (85%)	87% (67%)	63% (45%)	67% (55%)	98% (90%)

4. CONCLUSION

In this paper, we propose a content-based audio genre classification algorithm that automatically classifies the query audio into five genres such as Classic, Hiphop, Jazz, Rock, Speech. Two types of features are computed from each frame: One is the timbral features and the other is coefficient domain features such as MFCC and LPC. The means and standard deviations of these six original features are computed over each frame for each music file to form a total of 54-dimensional feature vector. For the classification algorithm, k-NN, Gaussian, GMM classifier is used. In order to choose optimum features from the 54 dimension feature vectors, SFS (Sequential Forward Selection) method is applied to draw 10 dimension optimum features and these are used for the genre classification algorithm. From the experimental result, we verify the superior performance of the SFS method that provides near 90% success rate for the genre classification which means 10%~20% improvements over the previous methods.

References

[1] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, July 2002.

[2] T. Li, M. Ogihara and Q. Li, "A comparative study on content-based music genre classification," in *Proc. of the 26th annual internal ACM SIGIR*, pp. 282-289, ACM Press, July 2003.

[3] J. J. Burred and A. Lerch, "A hierarchical approach to automatic musical genre classification," in *Proc. DAFx03*, 2003, pp. 308-311.

[4] G. Guo and S. Z. Li, "Content-based audio classification and retrieval by support vector machine," *IEEE Trans. on neural networks*, vol. 14, no. 1, pp. 209-215, Jan. 2003.

[5] S. Blackburn, "Content based retrieval and navigation of music, 1999, Mini-thesis, University of Southampton.