A Robust Content-Based Music Retrieval System

Kang-Kyu Lee , Won-Jung Yoon, Kyu-Sik Park
Dankook University
Division of Information and Computer Science
San 8, Hannam-Dong, Yongsan-Ku, Seoul Korea, 140-714
Tel.) 82-2-709-2728, Fax) 82-2-796-2970
E-mail) fitz@dankook.ac.kr, helloril@dankook.ac.kr, kspark@dankook.ac.kr

Abstract:

In this paper, we propose a robust music retrieval system based on the content analysis of music. New feature extraction method called Multi-Feature Clustering (MFC) is proposed for the robust and optimum performance of the music retrieval system. . It is demonstrated that the use of MFC significantly improves the system stability of music retrieval with better classification accuracy.

1. Introduction

The amount of available music data in multimedia databases is in rapid increase with the development of multimedia and network technology. As it continues, an efficient retrieving and browsing given music information becomes more and more important. Currently, the music data are labeled by a file name and a few indexing words, which are far from sufficient for most retrieval tasks. Handling music data using their content becomes a crucial technique for music retrieval.

Music information retrieval based on music content has been a growing area of research in the last few years. Example applications include search and select music from music digital library (MDL), entertainment industry, virtual reality, and several others on web application.

The basic operation of the content-based retrieval system is as follows. First, a suitable music samples must be accumulated and parameterized into so called feature vectors. These features must contain unique characteristics to be discriminated between the music files. Secondly, the pattern matching based on similarity matching between the feature vector of the query and the database is performed and the data entries found then are linked to the proper music file which is retrieved. Several music information retrieval (MIR) strategies have been introduced in these studies depending on different combinations of these methods. Wold *etal* [1] developed a system called "Muscle Fish". There various perceptual features are used to present a sound. A Euclidean distance measure is used to retrieve the music from the database. Foote [2] chooses to use 12 Mel Frequency cepstral coefficients (MFCC) as the audio features. Histograms of sounds are compared and the classification and retrieval is done by using the NN rule. Li [3] used nearest feature line (NFL) method for content-based audio classification and retrieval. In Ref. [4], Guo and Li 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 retrieve audio patterns.

Although many combinations of music features and retrieval strategies have been evaluated in these works, little attention has been paid to the following practical issues. The retrieval results corresponding to different query patterns (or portions) may be much different. A system dependency to the different query portions may cause remarkable uncertainty of the system performance. These clearly calls for a new robust feature extraction method that can characterize a given full-length music signal.

In this paper, we take a new approach that provides a solution to above problem. A new robust feature extraction method called multi-feature clustering (MFC) is proposed to overcome this problem based on K-means clustering algorithm. Basic idea is to extract features over the full-length music signal and then cluster these features in a small number of disjoint subsets. This way allows feature set to characterize whole intervals of music file while maintaining a reasonable size of database.

2. Feature Extraction, Selections and Multi-Feature Clustering (MFC)

2.1 Feature Extraction

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. These features are well-known in the literature and detailed descriptions of these features are omitted.

2.2. 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 Ref. [5], a sequential forward selection (SFS) method is used to meet these needs. 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 not only choosing the sub-optimum features for musical genre classification, but also it helps to keep reasonable size of database in multi-feature clustering (MFC) algorithm in next subsection

2.3. Multi-Feature Clustering (MFC)

As pointed out earlier, the classification results corresponding to different query patterns (or portions) and query lengths within the same music file or same class may be much different. It may cause serious uncertainty of the system performance. In order to overcome these problems, a new robust feature extraction method called multi-feature clustering (MFC) is implemented based on K-means clustering algorithm. Basic idea is to extract features over the full-length music signal in a step of 20 sec large window using SFS method and then cluster these features in four disjoint subsets. This allows feature set to characterize whole intervals of music signal while maintaining a reasonable size of database. The system then compares the query pattern to each one of the four feature sets of music file in trained database, and it retrieves the music similar to query using the Euclidean distance measure. Fig. 1 outlines the proposed MFC method.

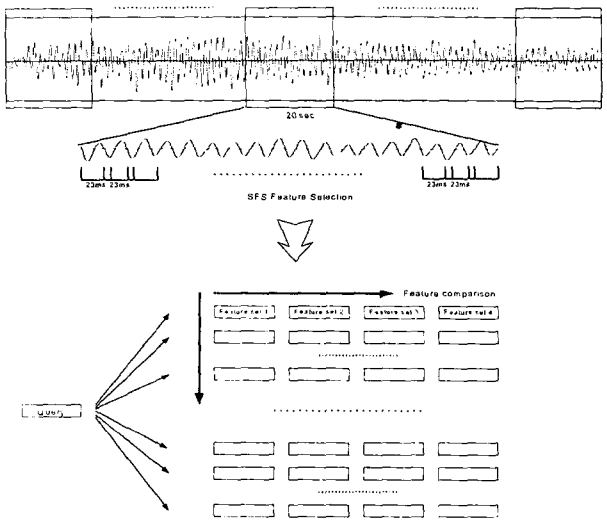

Fig. 1. Multi-feature clustering (MFC) for training and class classification procedure

3. Experiments on Music Retrieval

3.1. Experimental Setup

The proposed algorithm has been implemented and used to retrieve music data from a database of about 240 music files. 60 music samples were collected for each of the four genres in Classical, Hiphop, Jazz, and Rock, resulting in 240 music files in database. The excerpts of the dataset were taken from radio, compact disks, and internet MP3 music files. The 240 music files are partitioned randomly into a training set of 168 (70%) sounds and a test set of 72 (30%) sounds. In order to ensure unbiased retrieval accuracy because of a particular partitioning of training and testing, this division was iterated one hundred times. The overall retrieval accuracy was obtained as the arithmetic mean of the success rate of the individual iterations.

For retrieval purposes, a 20sec query is compared to training feature vectors according to simple Euclidean distance measure. The retrieval accuracy is then measured base on top 10 ranked list rules in query responses. In other words, each of the 300 sounds in the database is used as the query in turn. When a sound is used as the query, it is not used as a prototype, so the prototype set consists of the entire database except the query. This is so called "leave-one-out" tests.

3.2. Results and Analysis

Two sets of experiment have been conducted in this paper.

- Experiment 1: Performance verification of SFS feature selection method
- Experiment 2: Retrieval test using MFC method with different query patterns

Fig. 2 shows average retrieval accuracy using SFS method. From the figure, we can see that the retrieval performance increases with the increase of features up to 9 with near 90% of accuracy. And it remains constant up to 10 ~13 features. After 13 features, it even makes the system performance worse. Therefore, we can select only first 10 features to represent each music signals and it will be used all through out the experiments in this paper.

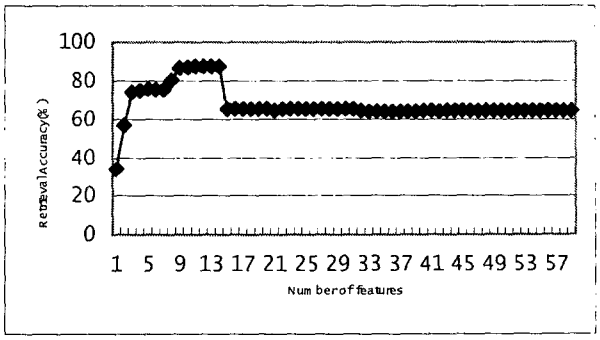

Fig. 2. Average retrieval accuracy using SFS method

As pointed out earlier, the retrieval results corresponding to different query patterns (or portions) may be much different. It may cause serious uncertainty of the system performance. In order to overcome this problem, multi-feature clustering (MFC) is used as explained in Section 2.3. To verify the performance of the proposed method, seven excerpts with fixed duration of 20 sec were extracted from every other position in same query music- at music beginning and 10%, 20%, 30%, 40%, 50%, and 80% position after the beginning of music signal. Fig. 3 shows the retrieving results with seven excerpts at the prescribed query position.

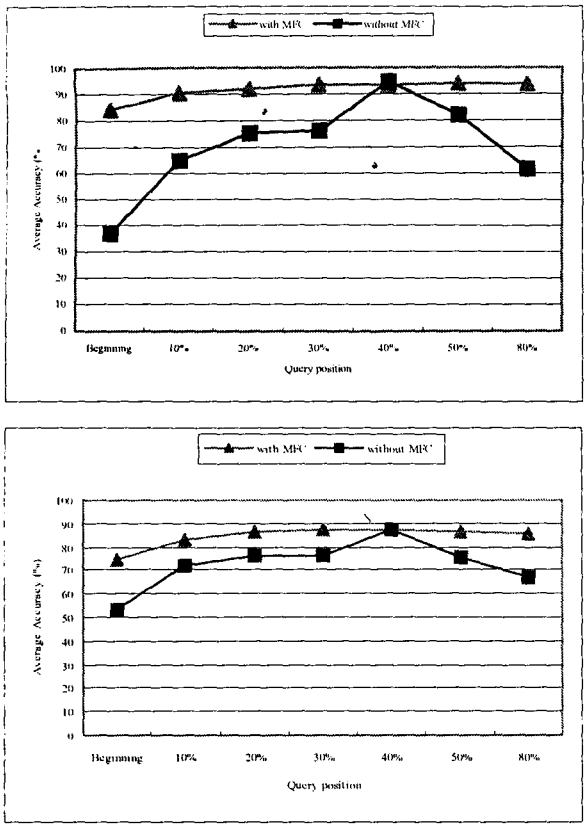

Fig. 3. Retrieval results with different query portions (Top : IR, Bottom : classification)

As we expected, the retrieving results without MFC greatly depends on the query positions and it's performance is getting worse as query portion towards to two extreme cases of beginning and ending position of the music signal. This is no wonder because, in general, the musical characteristics are not rich enough at those extreme intervals of music signal. On the other hand, we can find quite stable classification performance with MFC method and it yields higher accuracy rate in the range of 75% ~ 85%. Even at two extreme cases of beginning and ending position, the system with MFC can achieves classification accuracy as high as 75% which is more than 20% improvement over the system without MFC. This is a consequence of good MFC property which helps the system to build robust musical feature set over the full- length music signal.

4. Conclusion

In this paper, we propose a new robust content-based music retrieval algorithm using multi-feature clustering (MFC) method to overcome the unstable classification problems due to the different query patterns. For the implementation of the proposed algorithm, sequential forward selection (SFS) is first applied to the 54 dimensional feature set to reduce the feature dimension in the order of one-fifth and then multi-feature clustering (MFC) is adopted to build a robust feature vectors. The system then compares the query pattern to each one of the four feature sets of music file in trained database, and it retries the queried music based on Euclidean distance measure. Experimental comparisons for music retrieving with several query excerpts with duration of 15 sec from every other position are presented and it demonstrates the superiority of MFC method in terms of the system stability and accuracy.

References

[1] E. Wold, T. Blum, D. Keislar, and J. Wheaton,"Content-based classification, search, and retrieval of audio," *IEEE Multimedia*, vol.3, no. 2, 1996.

[2] J. Foote, "Content-based retrieval of music and audio," in *Proc. SPIE Multimedia Storage Archiving Systems II*, vol. 3229, C.C.J. Kuo *et al.*, Eds., 1997, pp. 138-147.

[3] Stan Z. Li : "Content-Based Audio Classification and Retrieval Using the Nearest Feature Line Method", 2000, IEEE.

[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] M. Liu and C. Wan, "A study on content-based classification retrieval of audio database," *Proc. of the International Database Engineering & Applications Symposium*, pp. 339 - 345. 2001.