韓國 컴퓨터情報學會 論文誌
第12卷 第1號, 2007. 3.

2007-12-1-1-7

바이올린 음원을 이용한 스펙트랄 롤오프 포인트의 최적점 검출

김 재 천*

Detection of the Optimum Spectral Roll-off Point using Violin as a Sound Source

Kim Jae Chun *

요 약

음악을 분류하기 위해 특성함수를 사용하여 추출한 특성값 벡터를 사용한다. 본 실험에서는 특성값 벡터를 추출하기 위해 스펙트랄 롤오프, 분산, 평균 피크레벨을 사용하였다. 이중에서 스펙트랄 롤오프는 저음프레임과 고음프레임의 상대적인 비를 나타낸다. 최적의 롤오프 포인트를 찾기 위하여 롤오프 포인트를 0.05에서 0.9까지 0.05간격으로 증가시키며 반복실험 하였다. 롤오프 포인트를 증가시키며 분류성공률을 관찰하였다. 그리고 실험에 사용된 음원데이터는 바로크바이올린과 현대바이올린 연주이다. 두 종류의 악기는 모양과 주파수대역에 있어서 유사하지만 약간의 대역차와 질감의 차이를 가지고 있다. 이러한 특성이 최적의 롤오프 포인트를 찾는데 유용할 것으로 판단하였다. 실험결과 롤오프 포인트 0.85에서 가장 높은 분류성공률 85%를 나타냈다.

Abstract

Feature functions were used for the classification of music. The spectral roll-off, variance, average peak level, and class were chosen to make up a feature function vector. Among these, it is the spectral roll-off function that has a low-frequency to high-frequency ratio. To find the optimal roll-off point, the roll-off points from 0.05 to 0.95 were swept. The classification success rate was monitored as the roll-off point was being changed. The data that were used for the experiments were taken from the sounds made by a modern violin and a baroque one. Their shapes and sounds are similar, but they differ slightly in sound texture. As such, the data obtained from the sounds of these two kinds of violin can be useful in finding an adequate roll-off point. The optimal roll-off point, as determined through the experiment, was 0.85. At this point, the classification success rate was 85%, which was the highest.

▶ Keyword : DSP, Pattern Recognition, Feature Function, Spectral Analysis

---

• 제1저자 : 김재천
• 접수일 : 2007.2.14, 심사일 : 2007.2.23, 심사완료일 : 2007. 3.19.
* ANSAN College, Dept. of Digital Information & Communication, Professor

Ⅰ. 서 론

The present searching systems discriminates against multimedia files by using titles of the files. If the file is incorrectly titled, there is no way to know the contents of the file. It would be helpful in processing music files and greatly increase the speed at which they are found, if there were a system that recognized the contents of music files and classified music files automatically.

The main factor of music automatic classification is finding characteristics which discriminate musical form.The key to automatic music classification is finding the characteristics that would allow one to discriminate between musical forms. To compare different kinds of music, some data were extracted from music and were converted to compact data.

The feature vector results from the pre-processing that is done before the classification of the data. It can be obtained through the conversion of raw data, which can be used for direct comparison and classification. But since the dimensions of raw data are usually very high and the information is just repeated, the feature vector is not commonly used in classifying music [1]. Usually data file(Raw data) is processed by Fourier transform, and then attributes extracted from raw data. Among those data, some attributes are selected to make feature vector.

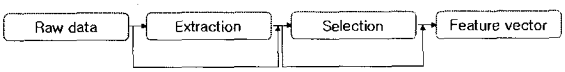

그림 1. 입력된 데이터셋으로부터 특성값 벡터의 생성
Figure 1. Generating a feature vector from a dataset input

The astounding ability of human's music perception abilities are revealed by D. Perrot and R. O. Gjerdige [2]. Classical, blues, country, dance, jazz, latin, pop, r&b, rap and rock, 10 genres used for experiment and each genre is composed of 8 songs. 3000ms, 475ms, 400ms, 325ms, 250ms length samples picked from each of the songs. 52 college students participated in the experiments. They are heard arbitrarily chosen samples, subsequently choosing one genre within the above 10 genres that best described it. The results compared to the CD vendor's classification.

Accuracy of success was 70% using 3000ms samples. Considering the obscurity of the music genre in some cases, the success rate is relatively high. While the success rate of the experiment is 40% when using 250ms samples. This research showed that any high level inferences are not necessary to genre classification. As implied in the short time sample where it was difficult to grasp rhythm, melody and structure, the students in the experiment totally depended on spectral, timbral attributes for their judgments.

Purpose of this paper is to find optimal parameter in the process of feature vector extraction. Feature functions generate feature vector. Feature functions which used in this paper are listed chapter Ⅱ. Among these functions, spectral roll-off function is analysed to obtain exact parameter and the result of experiment is presented in chapter Ⅲ. Conclusion is presented in chapter Ⅳ.

Ⅱ. Feature Functions

To classify musical categories, it is indispensable to extract some characteristics from music that represents a short time frame within the music sample. Characteristics can be called features. Feature values are acquired from raw data. As the dimension of raw data is very high, some converting process is required. The human auditory system transforms the sound into frequency components. Fourier analysis is a powerful tool used in music recognition systems as it can decompose sound into frequency.

There are several kinds of feature functions that can be used to analyze the attributes of music signals. These functions are based on sound processing. To demonstrate how changing the parameters affects the spectral roll-off function, the following four feature vector extraction functions were used in the experiment [3].

■ Variance

The variance of the PSD value, obtained from the FFT transformed data.

■ Spectral roll-off

Roll-off is a measure of spectral shape. It is defined as the frequency R corresponding to r% of the magnitude distribution. It can be seen as a generalization of the spectral centroid, which is the roll-off for r = 50%. Eq (1) shows relation of r and R. When the r is fixed, corresponding R value is used for feature vector.

$$\sum_{n=1}^{R} M_t[n] = r \cdot \sum_{n=1}^{N} M_t[n] \quad \text{(1)}$$

$M_t[n]$ is the magnitude of the FFT at frame t and frequency bin n. To find the optimal r value, an experiment was conducted in which the r value was increased from 0.05 to 0.95 in step 0.05.

■ Average peak level

The average peak level is the mean peak level. It is regarded as a peak because its value goes beyond the critical point. The amplitude and frequency of the peak are counted.

$$AVpeaklevel = \sum \frac{x(t)}{f_p} \quad \text{(2)}$$

$x(t)$ is the amplitude and frequency of the peak.

■ Class

Class is the category to which a specific data is assigned. The classification system that was used in this experiment can be likened to supervised learning. The class is used for training in the supervised learning system.

Some class names that were used in this experiment were modern, and the others were baroque. The classifier also uses class names for training purposes. The result of the classification was compared to the class name input to aid in determining whether the classification was right or wrong.

III. Experiment

Data files were processed iteratively, and their variance, spectral roll-off, average peak level, and class features were obtained. Fifteen-seconds-long modern and baroque violin files were prepared, and each violin's data were contained in 100 files. Each experiment was conducted by increasing the roll-off point from 0.05 to 0.95 in step 0.05. Thus, 19 experiments in all were conducted to be able to extract a feature vector.

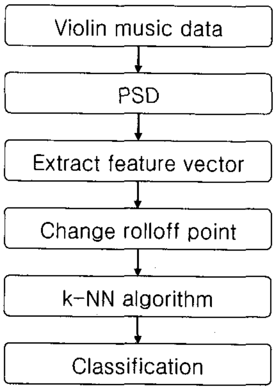

그림 2. 실험순서
Figure 2. Flowchart of the experiment

The k-nearest neighborhood classifier [4] was applied despite the fact that it is computationally intensive. In a music data classification system, it exceeds other kinds of algorithms [5]. k-NN algorithm uses Minkowski metric(3) which estimates the distance between feature vectors,

$$d_p(x_i,x_j) = \sqrt[p]{\sum_{k=1}^{d}(|x_{i,k}-x_{j,k}|)^p} = \| x_i - x_j \|_p \cdots\cdots (3)$$

The estimation of successive distances utilizes Euclidean distance(4) which measures distance of 2 or 3 dimensions. In Minkovsky metric, when p=2, It means Euclidean distance.

$$d_2(x_i,x_j) = \sqrt[2]{\sum_{k=1}^{d}(|x_{i,k}-x_{j,k}|)^2} = \| x_i - x_j \|_2 \cdots\cdots (4)$$

In k-NN algorithm, y value is estimated using k value of x_i which is closest to x. Here k is kind of coefficient, which can be weighted according to importance. The tenfold cross validation method was used to acquire training in the classification system. Cross validation is a general way of training, in which the data are split into n segments and the other (n-1) parts are used to acquire training in the classification system. The last nth segment was used to test the trained system [6].

Table 1 is an example of a feature vector. The numbers from the right side represent the variance, spectral roll-off, average peak level, and class, respectively. Among these features, the class is labeled when the data is inputted for the train of classifiers. As one row is extracted from one violin sound file, 200*4 = 800 features are thus extracted at each roll-off point. As all the 19 roll-off points were used, a total of 200*19*4 = 15,200 features were used to find the optimal roll-off point.

Feature Vectors at the Spectral Roll-off Point 0.40
0.004920, 0.156220, 1378.125000, modern
0.004454, 0.159770, 1378.125000, modern
0.007294, 0.206301, 1378.125000, modern
0.003660, 0.132678, 1378.125000, modern
0.004243, 0.144828, 1378.125000, modern
.
.

그림 3. 추출된 특성값 벡터의 예
Figure 3. Example of an extracted feature vector

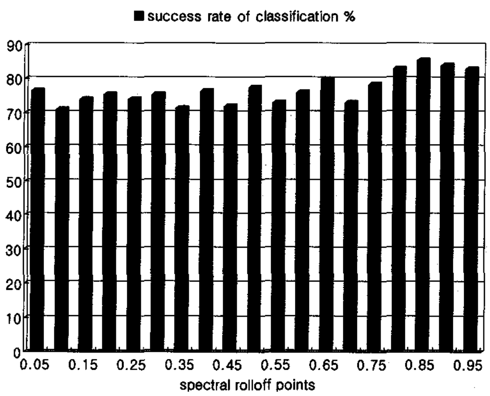

그림 4. 롤오프 포인트 변화에 따른 분류 성공률
Figure 4. Success rates of various roll-off points

As shown in Fig. 3, when the spectral roll-off point was about 0.85, the highest classification success rate of 85.5% was achieved. Table 2 shows the detailed result of the classification (when the roll-off point was 0.85). At the left side of Table 2 is a list of class inputs, the upper part being the experiment result labeled class. From the 200 data inputs, 11 data were classified correctly.

표 1. 롤오프 포인트 0.85에서 분류결과
Table 1. Results of the success rate at the spectral roll-off point 0.85

	Output: modern violin	Output: baroque violin
Input: modern violin	87	13
Input: baroque violin	16	84

표 2. 전체 정확도
Table 2. All-over accuracy

Correctly Classified Instances	171
Incorrectly Classified Instances	29
Kappa statistic	0.71
Mean absolute error	0.145
Root mean squared error	0.3808
Relative absolute error	29%
Root relative squared error	76.1577%
Total Number of Instances	200

The kappa statistics represent the success rate of the classification result compared to the data input. If kappa = 1, the result completely coincides with the class input. If kappa = 0, the result does not coincide with the class input. A negative value of the kappa statistic means that the probability of the classification success is lower than that of arbitrary classification [7]. The classification can be regarded as excellent if the kappa statistic value is higher than 0.9. If it is lower than 0.7, improvement is needed.

Ⅳ. Conclusion

The music recognition field has many areas to explore due to it's youth. A voice recognition tool can be applied to this area because they have a lot in common. Several experiments has been done, although they use similar tool in their experiments. Parameters should be different. In this paper, proper parameter of roll-off point function is found in music classification.

Modern and baroque violins are similar in their shapes and sounds. Their frequency band and sound texture, however, are slightly different. These characteristics can be helpful in determining the optimal parameters of the feature function, but subtle differences may be regarded as the effects of a changing roll-off point.

Although the roll-off point was changed, the overall classification success rate still exceeded 70%. This also proves the effectiveness of the other feature functions. As the roll-off point increased, the success rates fluctuated from 71% to 85.5%. It was earlier mentioned that in the above section, the roll-off point was 0.8 and the success rates turned out to be more than 80%. Thus, the relatively high proportion of the lower part in PSD could make it more successful. At first, the optimal roll-off point was estimated to be at both extremes in value. As expected, at the roll-off point 0.85, the highest classification rate of 85.5% was achieved. In other's research, they use roll-off point around 0.80 without logical explanation. This paper suggests optimal parameter of roll-off point in music classification and the results are expected to be used in this field.

참고문헌

〔1〕 T. Zhang and C. Kuo. Hierarchical system for content-based audio classification and retrieval. Technical Report, University of Southern California, Los Angeles, CA, 1998.

〔2〕 D. Perrot, and R. O. Gjerdigen, Scanning the dial : An exploration of factors in the identification of musical style, In Proceedings of the 1999 Society for Music Perception and Cognition, 1999.

〔3〕 G. Tzanetakis, and P. Cook. Automatic musical genre classification of audio signals in Proc. Int. Symp. Music Information Retrieval (ISMIR), Oct. 2001.

〔4〕 R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification. New York: John Wiley & Sons, Inc., 2001.

〔5〕 J.C. Kim, K.S. Kwak, "Performance Comparison of Classification Algorithms in Music Recognition using Violin and Cello Sound Files." The Journal of KICS, Vol. 30, No. 5C, 2005; pp. 305-312.

〔6〕 A.K. Jain, R.P.W. Duin, and M. Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1999; 22(1):4-37.

〔7〕 B.S. Evritt. Moments of the statistics kappa and the weighted kappa. British J. Math. Statist. Psych. 1968; 21:97-103.

저자소개

김 재 천

1996년 :인하대학교 전자공학과 졸업(공학사)

1999년 :인하대학교 대학원 첨단정밀공학과 졸업(공학석사)

2006년 : 인하대학교 대학원 전자공학과 졸업(공학박사)

2006~현재 :안산1대학 교수

관심분야 :데이터베이스, 패턴인식, DSP, 알고리즘