The research of digital media instrument recognition method based on the distribution overlapping degree of the Gaussian mixture model

Huayun Long

Hunan University of technology Music College of HUT Hunan Zhuzhou 421007, China

Corresponding author's e-mail: yunhua@163.com

Received 10 September 2013, www.cmnt.lv

Abstract

This article mainly studies the musical instrument recognition in digital media based on audio. First of all, the features are studied in this paper. so in this article we choose Gaussian mixture model as models of the instruments, we use K-means algorithm to initialize Gaussian mixture model, and use EM algorithm to train Gaussian mixture mode l. based on this, through the study we find that the convergence of EM algorithm relates to the overlap of Gaussian mixture model, therefore, we use the overlap of Gaussian mixture model to determine the convergence of EM algorithm, and carry out in the traditional Chinese musical instrument recognition system, through the experiments we find that when we use the overlap of Gaussian mixture model to determine the convergence of EM algorithm, the recognition rate can be obviously improved.

Keywords: musical instrument recognition; Gaussian mixture model; K-means algorithm; EM algorithm

1 Introduction

With the rapid development of modern information technology, especially the network technology and the multimedia technology, multimedia data has become a major part of the transmitted data on the Internet. Audio information as well as image information, is a kind of important multimedia information. With larger amount and more types of audio information which people capable of handling, it has become increasingly important to find or retrieve the required information from this flood of information quickly and effectively. Audio information retrieval technology is coming into being in this context, and the instrument identifying both involving the acoustic properties of the sound source and relating to the human perception psychological to the audio, is an important area of audio retrieval and is the basic of depth study on audio retrieval. In recent years, music signal and audio information retrieval technology on computer has become a hot topic. At present, foreign researches' on instrument identifying mainly focus on feature selection and classification study. Eronen et al., applied MFCC features, spectral centroid and spectral flux characteristics to the musical instruments feature recognition system respectively, and they found the MFCC features recognition more satisfactory by comparing the experimental results [1]. Brown et al. make identification to the four classes of the woodwind instruments [2]. Essid used MFCC and its first-order differential as instrument features, and applied the support vector machine (SVM) to classify the 10 kinds of Western musical instrument of solo [3]. Recently Kaminskyj et al. achieved a higher recognition rate by applying the k-NN classifier to the recognition system [4].

2 The pretreatment and feature extraction of the musical instrument recognition

For the characteristic of musical instrument recognition, the researchers' main choice is based on the characteristics of perception, MPEG-7 characteristics and MFCC feature and so on at present. But considering that the LPCC parameter is a very important kind of characteristic parameters and its main advantage is that it can thoroughly remove the motivation information in the process of instruments pronunciation. And its calculation process is very simple. So in this paper we choose LPCC feature as the characteristics of Chinese classical instrument.

2.1THE DEFINITION OF LPCC

LPCC (Linear Prediction Cepstrum Coefficient) is an assumption based on the music signals are autoregressive. Using the linear prediction analysis can obtain a cepstrum feature of cestrum coefficient. Generally, 8~32 d LPCC feature will be a good characterization of channel characteristics [5].

2.2THE EXTRACTION OF LPCC

Music signals Cepstrum can be obtained through Fourier transform, logarithmic modulus and Inverse Fourier Transform of the signal. Because the frequency response $H(e^{jw})$ reflects the channel frequency response and the spectrum envelope of the signal analyzed. Therefore, $\log\left|H(e^{jw})\right|$ is as the LPC cepstrum coefficients of IFT. The system function of the synthesis filter is obtained by linear prediction analysis. It is as shown below.

$$H(z)=\frac{1}{(1-\sum_{i=1}^{p}a_i z^{-i})}, \tag{1}$$

where, $h(n)$ is the impulse response, we will make a deduction of the cepstrum $\hat{h}(n)$. First of all, according to the homomorphic processing method, there is

$$\hat{H}(z)=\log H(z). \tag{2}$$

As $H(z)$ is the minimum phase, that's, it is resolved in the unit circle. So $\hat{H}(z)$ must can expand into series forms, that is

$$\hat{H}(z)=\sum_{n=1}^{\infty}\hat{h}(n)z^{-n}. \tag{3}$$

That's, the inverse transformation $\hat{h}(n)$ of $\hat{H}(z)$ is existing. Assume that $\hat{h}(0)=0$, we take the derivative of both sides of z^{-1}, then

$$\frac{\partial}{\partial z^{-1}}\log\left[\frac{1}{1-\sum_{i=1}^{p}a_i z^{-i}}\right]=\frac{\partial}{\partial z^{-1}}\sum_{n=1}^{\infty}\hat{h}(n)z^{-n}, \tag{4}$$

$$\sum_{n=1}^{\infty}n\hat{h}(n)z^{-n+1}=\frac{\sum_{i=1}^{p}ia_i z^{-i+1}}{1-\sum_{i=1}^{p}a_i z^{-i}}, \tag{5}$$

$$(1-\sum_{i=1}^{p}(a_i z^{-i})\sum_{n=1}^{\infty}n\hat{h}(n)z^{-n+1}=\sum_{i=1}^{p}ia_i z^{-i+1}. \tag{6}$$

We make the constant term on the left and right sides of the type (6) to be equal to the coefficient of the ascending powers of z^{-1}, then the recursive relations between $\hat{h}(n)$ and a_i can be expressed as

$$\begin{cases}\hat{h}(1)=a_1\\ \hat{h}(n)=a_n+\sum_{i=1}^{n-1}(1-\frac{i}{n})a_i\hat{h}(n-k),1\le n\le p\\ \hat{h}(n)=\sum_{i=1}^{p}(1-\frac{i}{n})a_i\hat{h}(n-i),n>p\end{cases}. \tag{7}$$

According to the type (7) it can directly be obtained the cepstrum $\hat{h}(n)$ from the prediction coefficients $\{a_i\}$.

According to the above derivation, The figure 1 shows the calculation process of the LPCC coefficients extraction.

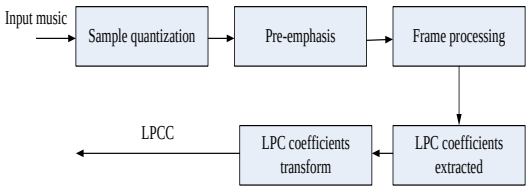

FIGURE 1 The calculation process of LPCC coefficients

3 The description of the GGaussian mixture model[6]

M-GMM probability density function is described as follows

$$P(o,i|\lambda)=\sum_{i=1}^{M}P(o,i|\lambda)=\sum_{i=1}^{M}c_i P(o|i,\lambda), \tag{8}$$

where λ the parameter sets of the GMM mode is, o is a K-acoustic feature vector; i is the number of hidden states. There are M hidden states in M-GMM. c_i is the blend weights of ith component. The value is corresponding to the prior probability of the hidden states i, then

$$\sum_{i}^{M}c_i=1. \tag{9}$$

$P((o,i|\lambda)$ is the Gaussian mixture component in type (8), The corresponding to the observation probability density function of the hidden states i is usually used K-single Gaussian distribution function, It's as shown in the following type (10)

$$P(o|i\mu\lambda)\sum=N(o,i,i)=\frac{1}{(2\pi)^{K/2}\left|\sum_i\right|^{1/2}}\exp\left\{-\frac{o\mu\sum_i^{T-1}o(\mu_i)}{2}\right\}, \tag{10}$$

where μ_i is the mean vector, $\sum_i$ is the covariance matrix, i=1, 2, 3..., M.

There the formula (8) can be understood as that the M-GMM is described by a linear combination of the single Gaussian. That's, the parameter sets λ of GMM can be made up of the weight of the mean vector, covariance matrix and mixed component. That's,

$$\lambda=\{c_i,\mu_i,\sum_i;(i=1,2,...,M)\}. \tag{11}$$

4 The EM algorithm to estimate the parameters of the GMM[7]

When we use the EM algorithm to estimate GMM parameters, the Q function can be defined as

$$Q(\lambda,\hat{\lambda})=\sum_{t=1}^{T}Q_t(\lambda,\bar{\lambda})=\sum_{t=1}^{T}\sum_{i=1}^{M}\frac{P(o_t,i|\lambda)}{P(o_t|\lambda)}\log P(o_t,i|\bar{\lambda}). \tag{12}$$

$P(o,i|\lambda)=c_i P(o_t|i,\lambda)$ has been known, we plug it into (12), then

$$O(\lambda,\bar{\lambda})=\sum_{i=1}^{M}\sum_{t=1}^{T}\frac{c_i P(o_t|i,\lambda)}{P(o_t|\lambda)}\log\bar{c}_i+\sum_{i=1}^{M}\sum_{t=1}^{T}\frac{c_i P(o_t|i,\lambda)}{P(o_t|\lambda)}\log P(o_t|i,\bar{\lambda}). \tag{13}$$

To estimate $\bar{c}_i$, assume $\frac{\partial Q(\lambda,\bar{\lambda})}{\partial \bar{c}_i}=0$. Then we can get

$$\bar{c}_i=\sum_{t=1}^{T}\frac{c_i P(o_t|i,\lambda)}{P(o_t|\lambda)}\Bigg/\left\{\sum_{t=1}^{T}\frac{\sum_{i=1}^{M}c_i P(o_t|i,\lambda)}{P(o_t|\lambda)}\right\}=\frac{1}{T}\sum_{t=1}^{T}\frac{c_i P(o_t|i,\lambda)}{P(o_t|\lambda)}. \tag{14}$$

The probability $P(q_t=i|o_t,\lambda)$ of training data on the assumed hidden states of i can be expressed as

$$P(q_t=i|o_t,\lambda)=\frac{c_i P(o_t|i,\lambda)}{P(o_t|\lambda)}. \tag{15}$$

Therefore the formula (14) can be written as the following form, that's,

$$\bar{c}_i=\frac{1}{T}\sum_{i=1}^{T}P(q_t=i|o_t,\lambda)\cdot \tag{16}$$

Similarly, the mean vector and covariance matrix can be estimated by the formula below, that's,

$$\bar{\mu}_i=\frac{\sum_{t=1}^{T}P(q_t=i|o_t,\lambda)o_t}{\sum_{t=1}^{T}P(q_t=i|o_t,\lambda)} \tag{17}$$

$$\sigma_{ik}^2=\frac{\sum_{t=1}^{T}P(q_t=i|o_t,\lambda)(o_{tk}-\mu_{tk})^2}{\sum_{t=1}^{T}P(q_t=i|o_t,\lambda)}\quad k=1,\ldots,K-1 \tag{18}$$

The revaluation of GMM model parameters by the formula (16), (17) and (18) can guarantee the likelihood function is monotone increasing.

To sum up, the flow chart that GMM model parameters is estimated by using EM algorithm has been shown in Fig. 2.

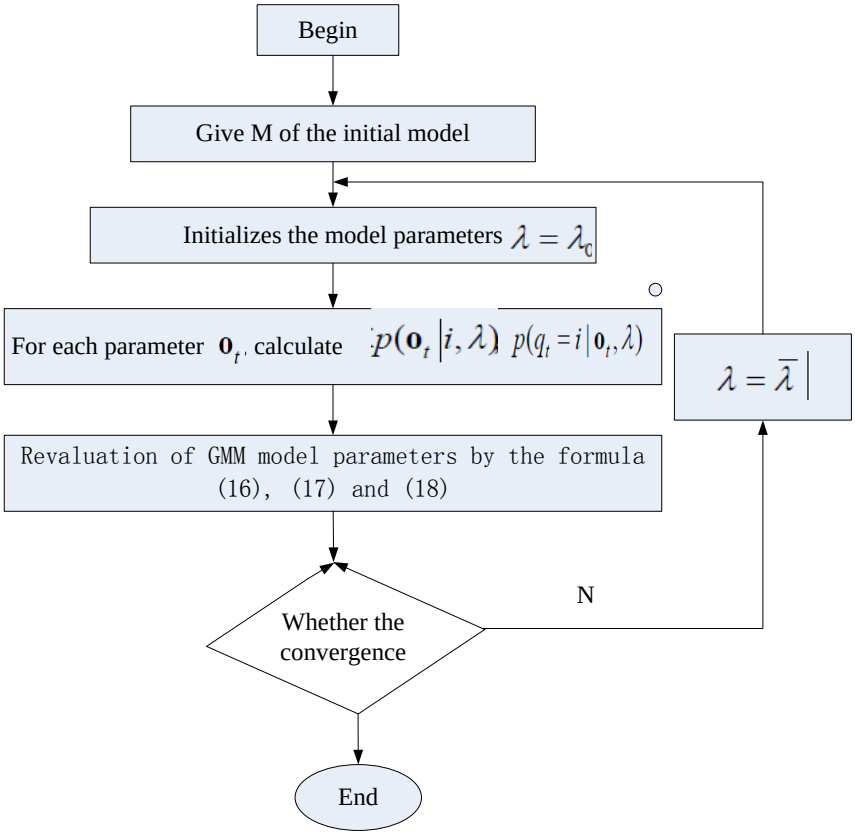

FIGURE 2 The chart of GMM model parameters which is estimated by using EM algorithm

5 K - average clustering algorithm[8]

The k-means clustering method is used to determine the initial weight, the mean and variance. The specific process is as follows.

(i) We freely choose M vectors as the initial clustering center. Generally we can select M vectors in the front of the sample, that's, $(c_1^{(1)}, c_2^{(1)}, ..., c_M^{(1)})$.

(ii) Making a classification of input samples according to the rule of minimum distance. If

$$\left|x_k - c_i^{(m)}\right| \le \left|x_k - c_j^{(m)}\right|, \forall i \ne j \text{ and } i, j = 1, 2, ..., M, \tag{19}$$

where x_k belongs to the ith class.

(iii) Calculating the new clustering center as follow,

$$c_i^{(m+1)} = \frac{1}{N_i} \sum_{x_k \in C_i^{(m)}} x_k \; i = 1, 2..., M, \tag{20}$$

where N_i expresses the number of samples in the ith.

(iv) If $\left|c_i^{(m+1)} - c_i^{(m)}\right| \ge \delta$, go to (2) and continue; Otherwise, go to (5).

(v) The calculation of the initial GMM parameters

$$\alpha_i = \frac{N_i}{T}, \tag{21}$$

$$\mu_i = \frac{1}{N_i} \sum_{x_k \in C_i} x_k, \tag{22}$$

$$\sigma_{ik}^2 = \frac{1}{N_i} \sum_{x_i \in C_i} (x_{ik} - \mu_{ik})^2 \; k = 0, 1, ..., D-1, \tag{23}$$

where T is the total number of samples, D is the dimensions of the covariance matrix.

6 The criterion of musical instrument recognition system

For the musical instrument recognition system with N kinds of musical Instruments, where each of these instruments used a GMM model to represent, it's denoted by $\lambda_1, \lambda_2, ..., \lambda_n$. In the recognition phase, assuming the observation characteristic vector sequence of the testing music is that $O = \{o_1, o_2, ..., o_T\}$, the posterior probability that the instrumen is the nth instrument is

$$p(\lambda_n | O) = \frac{p(O|\lambda_n) p(\lambda_n)}{p(O)} = \frac{p(O|\lambda_n) p(\lambda_n)}{\sum_{m=1}^{N} p(O|\lambda_m) p(\lambda_m)}, \tag{24}$$

where $p(\lambda_n)$ is the prior probability of the nth instruments pronunciation, $p(O)$ is the probability of the characteristic vector sets O under the all instrument condition. $p(O|\lambda_n)$ is the conditional probability that the nth kind of musical instrument produces characteristic vector sets.

The identify results is given by the criterion of the maximum posteriori probability, that's,

$$n^* = \arg \max_{1 \le n \le N} P(\lambda_n | O), \tag{25}$$

where n^* expresses the identify results. In generally, assuming that the prior probability of each instrument pronunciation is equal, that's,

$$P(\lambda_n) = \frac{1}{N}, n = 1, 2, ..., N. \tag{26}$$

In addition, for each instrument, $P(O)$ in the formula (24) is equal.

In order to make a simplified calculation, we usually use logarithmic likelihood function, that's

$$L(O|\lambda_n) = \ln P(O|\lambda_n). \tag{27}$$

The verdict is given by the formula below.

$$n^* = \arg \max_{1 \le n \le N} \ln P(O|\lambda_n) = \arg \max_{1 \le n \le N} \sum_{t=1} \ln P(O_t|\lambda_n). \tag{28}$$

7 Conclusion

In the paper, we give a part of results as shown following.

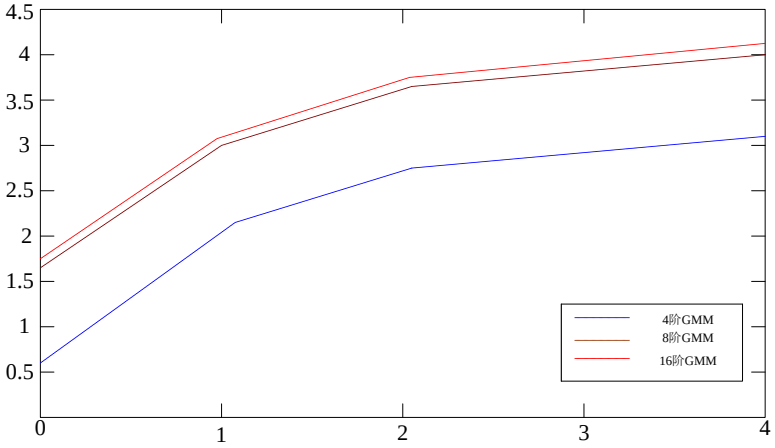

FIGURE 3 The recognition results of different order Gaussian mixture model

It can be seen from the FIGURE 3, the selection of GMM order is very important for the performance of the whole system. Combined with the actual situation of the experiment, it holds that the order of GMM should be elected to be 16 in this paper.

References

[1] Eronen A 2001 Comparison of features for musical instrument recognition *Proc IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz NY* 19-22

[2] Brown J C, Houix O, McAdams S 2001 Feature dependence in the automatic identification of mus ical woodwind instruments *Journal of the Acoustical Society of America* **109**(3) 1064-72

[3] Essid S, Richard G, David B 2004 Efficient musical instrument recognition on solo performance music using basic features *Proceedings of the Audio Engineering Society 25t h International Conference London UK* 89-93

[4] Kaminskyj I, Czaszejko T 2005 Automatic recognition of isolated monophonic musical instrument sounds using knn *Journal of Intelligent Information Systems* **24**(2) 199-221

[5] Furui S 1991 Cepstral analysis technique for automatic speaker verification *IEEE Trans on Signal Processing* **25**(2) 254-72

[6] Reynolds D 1995 Speaker identification and verification using Guassian mixture speaker models *Speech Communication* **17**(3) 91-108

[7] Cardoso J F 1997 Infomax and maximum likelihood for source separation *IEEE Letters on Signal Processing* **4**(5) 112-4

[8] Gao Zhengyan, Zhang Yushuang, Wang Mukun. The nuclear combination method of speaker recognition based on k-means clustering and support vector machine (SVM)

Authors

Huayun Long, born in June,1972, Zhijiang County, Hunan Province

Current position grades: Associate Professor, Hunan University of Technology
University studies: Theory of Music
Scientific interest: Theory Composition
Experience: Graduated from Conservatory of Music of Hunan Normal University in 2000; engaged in advanced studies in Composition Department of the Central Institute of Music from 2003 to 2004; graduated with a master degree from Hunan Normal University in 2009; taught in Conservatory of Music of Hunan University of Technology since 1999, involving the courses of compositing theory, music production and the band training, etc.