ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

Music Genre Classification using Chromagram and GMM

R. Thiruvengatanadhan

Assistant Professor, Dept. of Computer Science and Engineering, Annamalai University, Annamalainagar, Tamilnadu, India

ABSTRACT: Automatic music genre classification is very useful in music indexing. Chromagram is one of the feature extraction method uses in classification of musical genre that is based harmonic information in music structure analysis of music signals. Searching and organizing are the main characteristics of the music genre classification system these days. This paper describes a new technique that uses Gaussian Mixture Model (GMM) to classify songs. Gaussian mixture model classify music audio into their respective classes by learning from training data. The proposed feature extraction and classification models results in better accuracy in music genre classification.

KEYWORDS: Music, Feature Extraction, Chromagram, GMM.

I.INTRODUCTION

Musical genres have no strict definitions and boundaries as they arise through a complex interaction between the public, marketing, historical, and cultural factors. This observation has led some researchers to suggest the definition of a new genre classification scheme purely for the purposes of music information retrieval [1]. In addition to this, the advancement in digital signal processing and data mining techniques has led to intensive study on music signal analysis like , content-based music retrieval, music genre classification, duet analysis, Musical transcription, Musical Information retrieval and musical instrument detection and classification. Musical Instrument detection techniques have many potential applications such as detecting and analyzing solo passages, audio and video retrieval, music transcription, playlist generation, acoustic environment classification, video scene analysis and annotation etc.

The techniques for automatic genre classification would be a valuable addition to the development of audio information retrieval systems for music. Advanced music databases are continuously achieving reputation in relations to specialized archives and private sound collections. Due to improvements in internet services and network bandwidth there is also an increase in number of people involving with the audio libraries. But with large music database the warehouses require an exhausting and time consuming work, particularly when categorizing audio genre manually. Music has also been divided into Genres and sub genres not only on the basis on music but also on the lyrics as well [2]. This makes classification harder. To make things more complicate the definition of music genre may have very well changed over time [3]. For instance, rock songs that were made fifty years ago are different from the rock songs we have today.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

II. CHROMAGRAM

Pitch is a property related to perception and sound is ordered on a scale related to frequency [4]. The audio signal is decomposed into bands of varying frequency [5]. Chroma feature representation is an effective and powerful method to describe harmonic information in music structure analysis [6]. Pitch class is a collection of pitches that share the same chroma. Two dimensions characterize music, tone height and chroma [7]. The dimension of tone height is partitioned into the musical octaves. The range of chroma is usually divided into 12 pitch classes, where each pitch class corresponds to one note of the twelve tone equal temperament. The spectral energy of each of the 12 pitch classes is represented by chromogram.

It is based on a logarithmized short time Fourier spectrum. The chromagram represents an octave-invariant (compressed)spectrogram that takes properties of musical perception into account [8]. The chromagram is extracted as follows: In the pre-processing stage, short segmented frames are extracted and windowed from music signal using framing and windowing Following that, shifting the centre frequencies of the subband filters of the multi rate bank is necessary for the global tuning of a recording as shown in Fig 1. An average spectrum vector is calculated and the derivation of an estimate for tuning derivation is done by stimulating the filter bank shifts using weighted binning techniques.

The pitch representation is performed by the decomposition of a given music signal on 88 frequency bands with centre frequencies corresponding to the pitches A0 to C8 (MIDI pitches p=21 to p=108) in order to extract the chroma features.

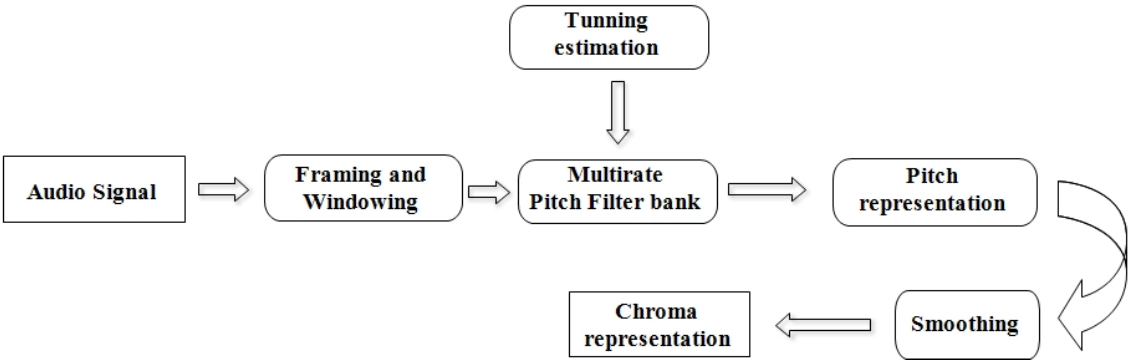

Fig. 1 The Chromagram Computation.

III.GAUSSIAN MIXTURE MODELS (GMM)

The probability distribution of feature vectors is modeled by parametric or non parametric methods. Models which assume the shape of probability density function are termed parametric. In non parametric modeling, minimal or no assumptions are made regarding the probability distribution of feature vectors [9]. In this section, we briefly review Gaussian mixture model (GMM), for audio classification. The basis for using GMM is that the distribution of feature vectors extracted from a class can be modeled by a mixture of Gaussian densities.

The iterative Expectation Maximization (EM) algorithm is used to estimate the parameters of GMM. EM algorithm is one of the most popular clustering algorithms used to estimate the probabilistic models for each Gaussian component. The Expectation step (E-step) and Maximization step (M-step) are iterated till the convergence of the parameter [10]. EM algorithm finds out maximum likelihood estimation of parameters. The E-step computes Expectation of likelihood assuming parameters and M-step computes maximum likelihood estimates of parameters by maximizing the expected likelihood found in E-step.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

IV.EXPERIMENTAL RESULTS

The database

The music data is collected from music channels using a TV tuner card. A total dataset of 100 different songs is recorded, which is sampled at 22 kHz and encoded by 16-bit. In order to make training results statistically significant, training data should be sufficient and cover various genres of music.

Acoustic feature extraction

In this work fixed length frames with duration of 20 ms and 50 percentages overlap (i.e., 10 ms) are used. The objective of overlapping neighbouring frames is to consider the harmonic information characteristic of audio content. An input wav file is given to the feature extraction techniques. Chromagram 12 dimensional feature values will be calculated for the given wav file. The above process is continued for 100 number of wav files.

Classification

When the feature extraction process is done the music should be classified. We select 75 music samples as training data including 25 classic music, 25 pop music and 25 rock music. The rest 25 samples are used as a test set. A mean vector is calculated for the whole audio and it is compared either to results from training data or to predefined thresholds.

Gaussian mixtures for the three classes are modeled for the features extracted. For classification the feature vectors are extracted and each of the feature vectors is given as input to the GMM model. The distribution of the acoustic features is captured using GMM. We have chosen a mixture of 2, 5, 10 mixture models. The class to which the audio sample belongs is decided based on the highest output.

The performance of the system for 2, 5 and 10 Gaussian mixtures is shown in Table.1. The distribution of the acoustic features is captured using GMM. The class to which music sample belongs is decided based on the highest output. Table.1 shows the performance of GMM for music classification based on the number of mixtures. Music classification using GMM gives an accuracy of 92%.

Table 1: Performance of GMM for different mixtures.

GMM	**2**	**5**	**10**
Classic	92%	90%	91%
Pop	83%	80%	88%
Rock	87%	88%	90%

V.CONCLUSION

In this paper, we have proposed an automatic music genre classification system using GMM. Chromagram is calculated as features to characterize music content. GMM learning algorithm has been used for the classification of genre classes of music by learning from training data. The proposed classification method is implemented using EM algorithm approach to fit the GMM parameters for classification between classic, pop and rock by learning from training data. Experimental results show that the proposed audio GMM method has good performance in musical genre classification scheme is very effective and the accuracy rate is 92%.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

REFERENCES

[1] F. Pachet and D. Cazaly, “A classification of musical genre,”inProc.RIAO Content-Based Multimedia Information Access Conf., Paris,France, Mar. 2000.
[2] O.M. Mubarak, E. Ambikai rajah and J. Epps, “Novel Features for Effective Speech and Music Discrimination,” IEEE Engineering on Intelligent Systems, pp. 342-346, 2006.
[3] Dijk, L. Van. Radboud Universiteit Nijmegen Bachelorthesis Information Science Finding musical genre similarity using machine learning techniques, 1–25. 2014
[4] FitzGerald. D and J. Paulus, 2006, Unpitched Percussion Transcription, in Signal Processing Methods for Music Transcription, Springer, pp. 131-162.
[5] Zhe Zuo, 2011, Towards Automated Segmentation of Repetitive Music Recordings, Master's Thesis, Saarland University.
[6] Müller. M, 2007, Information Retrieval for Music and Motion, Database Management & Information Retrieval Springer.
[7] Shepard, 1964, Circularity in Judgements of Relative Pitch, The Journal of the Acoustical Society of America, vol. 36, pp. 2346-2353.
[8] Schroder M. R., B. S. Atal, and J. L. Hall, 1979, Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human Ear, Journal of the Acoustical Society of America, vol. 66, pp. 1647-1652.
[9] Kim H.-G. and Sikora T., “Automatic Segmentation of Speakers in Broadcast Audio Material,” *IS&T/SPIE’s Electronic Imaging 2004*, San Jose, CA, USA, January 2004.
[10] Christel, M., Kanade, T., Mauldin, M., Reddy, R., Sirbu, M., Stevens, S., and Wactlar, H., “Informedia, Digital Video Library,” *Communications of the ACM*, vol. 4, no. 38, pp. 57-58, 1995.