International Journal of Computer Applications® (IJCA) (0975 – 8887)
Proceedings on Emerging Trends in Electronics and Telecommunication Engineering (NCET 2013)

Musical Instrument Recognition using Wavelet Coefficient Histograms

Kothe R.S.
Rajarshi Shahu College of Engg
Pune, India

Bhalke D.G.
Rajarshi Shahu College of Engg
Pune, India

ABSTRACT

In pattern recognition applications, finding compact and efficient feature set is important in overall problem solving. In this paper, feature analysis using wavelet coefficient histogram for the musical instrument recognition has been presented and compared with traditional features. The new proposed wavelet coefficient histograms features found compact and efficient with existing traditional features. With this work it is justified that the musical instrument information is available in particular frequency sub bands and can be easily extracted using wavelet features. The proposed wavelet based features shows better accuracy than existing traditional features. The database used in this work is from Mc Gill university, Canada . The work is carried out with 18 Musical instrument from different musical instrument families .

Keywords

Musical instrument recognition; Wavelet transform, feature extraction .

1. INTRODUCTION

Recently, music data analysis and retrieval has become an emerging research field in signal processing. Due to tremendous applications of musical instruments, musical instrument recognition (MIR) has attracted the attention of various researchers [1],[2] .

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 [2],[3],[4] .

In musical instrument recognition problem the computer listen to musical sounds and classifies based on feature extraction. Many researchers have proposed different musical instrument features for MIR. Most of these features are based on MFCC and common traditional features .Still benchmark is open to efficiently and compactly representation of Musical instrument features [2],[4] . So finding efficient and compact feature is challenging task in Musical Instrument recognition and open problem. So we have undertaken this problem.

In this paper we have analyzed and compared traditional features with wavelet sub band histograms based features . Also experimentation is carried out to find best wavelet type and number of decomposition levels.

In section I the introduction is explained .The rest of the paper is organized as follows . Section II presents the detailed literature review of the relevant work on Musical Instrument recognition and section III presents the feature extraction schemes . Section IV presents the detailed result analysis with the proposed feature scheme and finally conclusion is given in section V.

2. RELETED WORK

Various feature schemes have been proposed and adopted in the literature of musical instrument classification and recognition. In most of the research articles musical instrument classification and recognition employed mel-frequency cepstral coefficients (MFCC) features. Further, features like temporal, spectral , cepstral, LPC based features have been used by different researchers. These features are called as traditional features.Some researchers have also proposed computational models with different classification algorithms [2],[3],[4],[5]

G. Tzanetaki [7] proposed a new technique in music genre classification. Three feature sets for representing timbre l texture, rhythmic and pitch content are proposed by this author. Using the proposed feature sets, musical genre classification of 61% was achieved for ten instruments. The success of proposed features for musical genre classification testifies to their potential as the basis for other types of automatic techniques for music signals such as similarity retrieval and segmentation which are based on extracting features to describe musical content.

I. Kaminskyj and C. Pruysers [8] described the technique to improve the Musical Instrument Recognition by adding Wavelet packet based features. The author claimed that the recognition accuracy is improved by five percentages by adding Morlet and Daubechies features. Using the proposed feature sets classification accuracy of 87.6 was claimed for 19 musical instruments. Specified work is focused on measuring the generalization performance of the system of monophonic instrument.

Popescu, Gavat and Mithai Datcu [15] suggested a technique of Wavelet analysis for audio signals with music classification application. They used multi-resolution (wavelet) analysis and spectral analysis based features. The Proposed approach uses a no. of features like Mel Frequency Cepstral Coefficient (MFCC), Zero Crossing rate and FFT Coefficients combined with wavelet based features. Using the proposed features the accuracy of 75% was claimed.

Cyril Joder and Slim Essid described temporal integration for audio classification with application to musical classification [1]. Author compared the accuracy using different classifiers like HMM, GMM and DTW. Using the proposed methods the accuracy was claimed to 75%. In this paper the author used temporal features, spectral features, MFCC and wavelet features. The paper suggested that the further work can be extended in designing automatic indexing tools which provides meaningful and efficient means to describe the musical audio content and also can be used in music information retrieval systems.

Tao Li and Qi Li proposed features based on wavelet coefficients at various frequency sub bands of daubechies wavelet for music genre classification and emotional content of the music [4] . The author used features like timbral texture, rhythmic content, pitch content, spectral centroid, and roll off, zero crossing, LPC, MFCC, spectral flux. The author claimed the result up to 80%.

From literature survey it is seen that, finding the unique characteristic of musical instruments is a crucial step and poses challenges for the researchers. Peoples are using common features like MFCC, spectral features, MPEG features and applying to different classifiers. The main problem in Musical Instrument Recognition and classification is to find optimal and compact feature set which allow the classifier to build it quickly and easily [2].

3. FEATURE EXTRACTION SCHEMES

Feature extraction is important part of this research work. Finding optimized and compact feature set is challenging task. Various researchers have used temporal features, spectral features , cepstral features , perceptual features . All features are briefly described here. These features are extracted and the result is compared with our proposed Wavelet histogram based features. Temporal features are obtained from time domain characteristics which take less computational time. Spectral features need more computational time but describes the behavior of the instrument to more extent . Wavelet sub band histogram features describes the characteristics of the instrument globally and need less computations .

3.1 Temporal Features

Temporal features are features obtained directly from in time-domain .

Energy:

Energy is the sum of the amplitudes present in frame and is defined as:

$$Energy = \sum_{0}^{N-1} (x(n))^2 \qquad (1)$$

where x[n] is the amplitude of the sample.

Zero-Crossing Rate:

This is the number of times the signal crosses zero amplitude during the frame, and can be used as a measure of the noisiness of the signal. It is defined as:

$$ZCR = \frac{1}{N}\sum_{0}^{N-1} |\operatorname{sgn}[x(n)] - \operatorname{sgn}[x(n-1)]| \qquad (2)$$

where *sign* = 1 for positive arguments and 0 for negative arguments

Log-Attack Time:

The log-attack time is the logarithm of time duration between the times the signal starts to the time it reaches its stable part. It can be estimated by taking the logarithm of the time from the start to the end of the attack.

$$LAT = \log[startattack - stopattack] \qquad (3)$$

3.2 Spectral Features

Spectral features are obtained from the samples in the frequency domain of the musical signals.

Spectral Centroid (SC)

This is the amplitude-weighted average, or centroid, of the frequency spectrum, which can be related to a human perception or brightness of the instrument. It is calculated by multiplying the value of each frequency by its magnitude in the spectrum, then taking the sum of all these. The value is then normalized by dividing it by the sum of all the magnitudes:

$$SC = \left(\frac{\sum_{k=1}^{K} P(f_k) f_k}{\sum_{k=1}^{k} P(f_k)}\right) \qquad (4)$$

where P(fk) is magnitude spectrum of kth sample and fk is frequency corresponding to each magnitude element

Spectral flux (SF)

This is a measure of the amount of local spectral change. This is defined as the squared difference between the normalized magnitude spectra of successive frames,

$$SF = \left(\sum_{k=2}^{k} \left|P(f_k) - P(f_{k-1})\right|\right) \qquad (5)$$

Spectral spread (SS)

The spectral spread is a measure of variance (or spread) of the spectrum around the mean value μ .It is given by

$$SS = \frac{\sqrt{\sum_{K=0}^{N/2} (P(f_k) - SC)^2}}{\sqrt{\sum_{k=0}^{N/2} (P(f_k)^2}} \qquad (6)$$

where P(fk) = magnitude spectrum corresponding to each magnitude element and SC=spectral centroid.

Spectral skewness(SK)

The skewness is a measure of the asymmetry of the distribution around the mean value. The skewness is calculated from the 3rd order moment.

$$SK = \frac{\sum (freq - SC)^2 \times Mag}{\sum Mag} \qquad (7)$$

where mag= magnitude spectrum, freq=frequency corresponding to each magnitude element and SC=spectral centroid.

3.3 Cepstral feature

Mel frequency cepstral coefficients

Mel Frequency Cepstral Coefficients (MFCCs) are cepstral coefficients used for representing audio in a way that mimics the physiological properties of the human auditory system. MFCCs are commonly used in speech recognition and are finding increased use in music information recognition and genre classification systems. The cepstrum of a signal is the Fourier transform of the logarithm (decibel) signal of the Fourier transform of a signal. In the Mel frequency cepstrum, the frequencies are scaled logarithmically using the Mel scale. A mel is a psychoacoustic unit of frequency which relates to human perception, the mel scale can be approximated from a Hz value by the formula,

$$Melfrequency = 2595 \times \log(\frac{X}{700}) \qquad (8)$$

where X is frequency in Hz

3.4 Wavelet features

The wavelet analysis provides spectro-temporal information of the music signal . The wavelet packet analysis decomposes a signal into "packets" by simultaneously passing the signal through a low decomposition filter (LDF) and a high decomposition filter (HDF) in a sequential tree like structure. There are a large number of mother wavelet filters that can be used for this purpose. In this experiment we considered upto fourth level decomposition of various mother wavelet . The LDF and HDF are intimately related as the HDF is calculated by passing the LDF through a quadratic mirror filter (QMF) . The QMF firstly reverses the coefficients of the LDF and then reverses the sign of every second coefficient. This produces a filter that passes high frequency components. Passing a signal through two of these filters produces two packets: (1) the Approximation (from the LDF) and (2) the Detail (from the HDF). This is referred to as level 1 decomposition. The level 1 packets can then be passed through another pair of filters to produce a total of 4 packets (level 2 decomposition). This operation can be continued indefinitely, although after a certain point, which has to do with the signal sample length, the packets from one instrument become less distinguishable from that of other instruments, affecting classification accuracy. Fig.1 shows the wavelet packet tree upto four level.

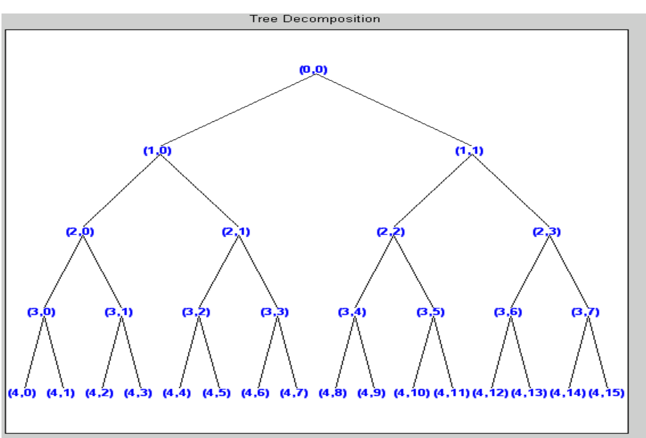

Fig. 1. Wavelet packet tree

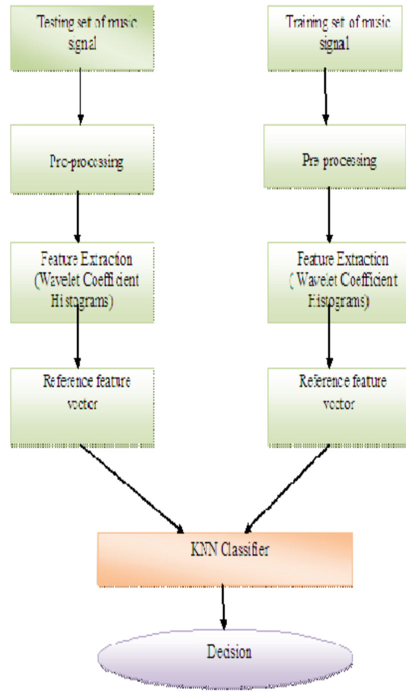

Fig.2 Proposed block diagram for Musical Instrument classification

Different 104 features which are shown in the table 1 are extracted and reference feature set is formed separately using temporal, spectral , cepstral , perceptual and wavelet sub band histogram based features . The results are tested using KNN classifier. The wavelet sub band histogaram based features shows good accuracy than all other features.

Table 1 : Features used

Feature No	Feature
1-39	MFCC, ΔMFCC, ΔΔMFCC
40-42	Mean SC, Std SC, variance SC
43-45	Mean SS, Std SS, variance SS
46-48	Mean SF Std SF, variance SF
49	LAT
50-52	Mean SK , STD SK , variance SK
53-100	Mean , STD , Variance of sub band of each histogram (Four level decomposition)
101	LAT
102-104	Mean ZCR, Std ZCR, variance ZCR

Choice of Mother wavelet and Number of decomposition levels

The choice of mother wavelet would be that which provides best distinguishing features having a short filter length which is computationally efficient. The author performed some classification experiments using the wavelet coefficient histogram based features using Daubechies 1 through 8, Symlet 2 through 8 , and Coiflet 1 through 5 wavelets. Number of decomposition levels are finalized where it gives best result. Through our experiment four decomposition level gives good result.

In the experiment we calculated the sub band histograms of each sub band and three moments of each histograms.

Database

We performed the experiment with separate training and test set database which is used from Mc Gill University database[11].We have used total 18 Musical instruments as follows.70% Notes are used for training and 30 % notes are used for testing .

1.Bass 2. Guitar 3.Violin 4.Viola 5.Cello 6Trumpet 7.Tuba 8.Trombone 9. Coronet 10.Flute 11. Oboe 12.Saxophone 13. Horn 14.Piano 15.Xylophone 16.Accordion 17.French Horn 18. Lute

4. RESULT ANALYSIS

Fig.3 shows the decomposition Flute E2 note. Fig 4 shows the accuracy of musical instrument recognition using different mother wavelet. Db4, Db5 wavelets shows better accuracy . The choice of decomposition level is four because for four level decomposition we achieved better accuracy.

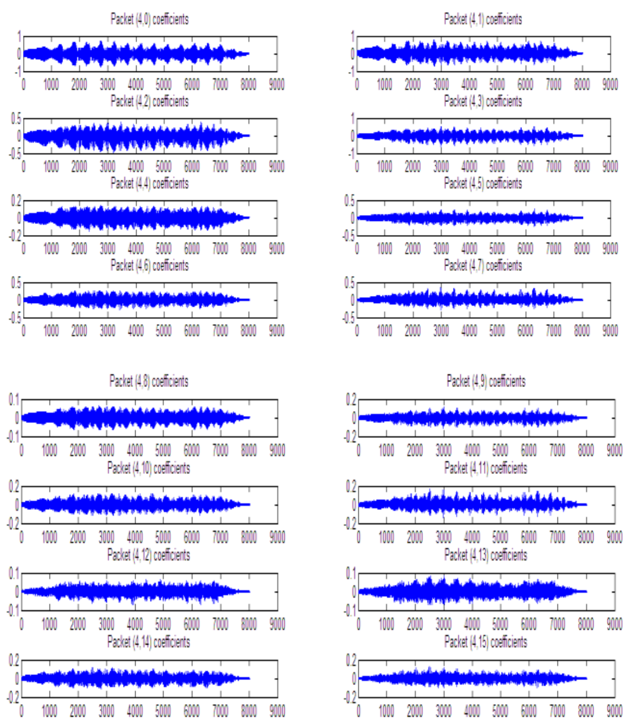

Fig3: four level decomposition of flute E2

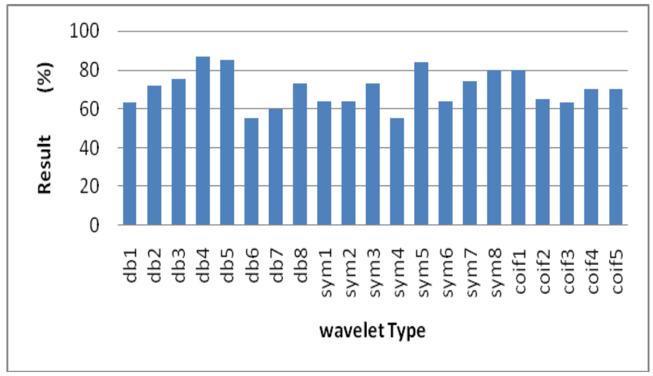

Figure 4 : Result using different wavelet types

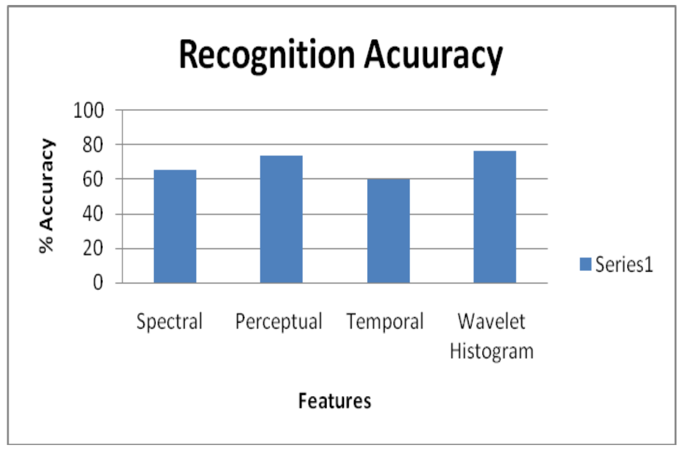

Fig 5: Recognition accuracy using different features

Fig.5 shows the overall accuracy using spectral features (40-48), Perceptual features (1-39), Temporal features(101-104), Wavelet histogram features (53-100)

5. CONCLUSION

We have proposed an innovative feature set for musical instrument recognition based on wavelet coefficient histograms , which is compact and optimized. The result reflects that musical instruments information is available in specific frequency band information. The energy in different sub band can easily retrieved using wavelet transform. The overall accuracy of the system using Wavelet Transform recorded is 76.83% compared to 73.82 % using MFCC and other features. The Db4 and Db5 give good accuracy. An improvement of the results may be possible by adding other timberal features, features selection algorithms and different machine learning algorithms.

6. RFERENCES

[1] C. Joder, S. Essid, G. Richard, and S. Member, “Temporal Integration for Audio Classification With Application to Musical Instrument Classification,” *IEEE Transactions on Speech and Audio Processing*, vol. 17, no. 1, pp. 174–186, 2009.

[2] J. D. Deng, C. Simmermacher, and S. Cranefield, “A study on Feature analysis for Musical Instrument Classification,” *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, vol. 38, no. 2, pp. 429–438, 2008.

[3] J. Garcia, A. Barbedo, and G. Tzanetakis, “Musical Instrument Classification using Individual Partials,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 1, pp. 111–122, 2011.

[4] T. Li, Q. Li, and M. Ogihara, “Music feature extraction using wavelet coefficient histograms,” *US Patent 7,091,409 B2*, 2006.

[5] T. Li, M. Ogihara, and Q. Li, “A comparative study on content-based music genre classification,” *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval - SIGIR ’03*, vol. 2, pp. 113–116, 2003.

[6] A. Holzapfel and Y. Stylianou, “A statistical approach to musical genre classification using non-negative matrix factorization,” in *IEEE international conference on acoutsics , speech and signal processing*, 2007, pp. 692–696.

[7] G. Tzanetakis, S. Member, and P. Cook, “Musical Genre Classification of Audio Signals,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[8] C. Pruisers,I. Kaminskyj and Schnapp, “Wavelet analysis in musical instrument sound classification,” in *international symposium on signal processing and its applications*, pp. 1–4, 2005

[9] M. G. et Al, “A Wavelet Packet representation of audio signals for music genre classification using different ensemble and feature selection techniques,” in *ACM SIGMM international workshop on Multimedia information retrieval*, 2003, pp. 102–108.

[10] G. Tzanetakis, G. Essl, and P. Cook, “Audio analysis using the discrete wavelet transform,” *Proc. Conf. in Acoustics , speech and signal processing* , pp. 1–6, 2001.

[11] “Mc gill University Master Samples www.music.mcgill.ca/resources/mum/.html/mums.html.”

[12] M. Daniels, “Classification of Percussive Sounds Using Wavelet-Based,” ,CCRMA , Stanford University thesis , 2010.

[13] A. Saxena, “Application of The Wavelet Transform In The Processing of Musical Signals,” *Thesis*, vol. 2, no. April, pp. 1–5, 2005.

[14] F. Germain, “The wavelet transform Applications in Music Information Retrieval,” *McGill University, Canada, Thesis*, pp. 1–29. 2009

[15] M. Popescu, A.; Gavat, I.; Datcu, “Wavelet analysis for audio signals with music classification applications,” in *proc. of speech technology and human computer dialogue*, pp. 1–6.,2009.