Musical Instrument Recognition using SVM

Prof. Shilpa Sonawane

Assistant Professor
Department of E&TC
JSPM's RSCOE, Tathawade,India

Abstract : **Recently, there is large amount of music on the internet. It is difficult to classify the music based on the music content manually. Instrumental music is often classified or retrieved in terms of instruments played in it. Searching and organizing of music collections requires a mathematical model of music similarity. In this paper, we propose a new method to classify the music automatically based on different features. The classification of musical instrument is done using multi class support vector machine (SVM).We adopt two approaches to the multi-class classification. Finally, computer simulations are done by using real music data in order to prove the effectiveness of the proposed method. The maximum accuracy achieved with svm one vs. one method with wavelet feature is 86% using exponential radial basis function.**

IndexTerms:SVM,Kernel -

I. Introduction

Music data analysis and retrieval has become a very popular research field in recent years. The advance of signal processing and data mining techniques has led to intensive study on content-based music retrieval, music genre classification, duet analysis and, most frequently, on musical instrument detection and classification [5].

Instrument detection techniques can have many potential applications. For instance, detecting and analyzing solo passages can lead to more knowledge about the different musical styles and can be further utilized to provide a basis for lectures in musicology. Various applications for audio editing and audio and video retrieval or transcription can be supported. Other applications include playlist generation, acoustic environment classification and using audio feature extraction to support video scene analysis and annotation.

One of the most crucial aspects of instrument classification is to find the right feature extraction scheme. During the last few decades, research on audio signal processing has focused on speech recognition, but few features can be directly applied to solve the instrument-classification problem. New methods are being investigated for achieving semantic interpretation of low-level features extracted by audio signal processing methods. A framework of low-level and high-level features can be used to create application-specific description schemes. These can be used to annotate music with a minimum of human supervision for the purpose of music retrieval [5],[6].

In this paper, we present a study on feature extraction for instrument classification using machine learning techniques. Four feature schemes were considered: temporal features, spectral features, cepstral features, and wavelet based entropy. The performance of the feature schemes was assessed first individually. Our aim was to find the differences between the different feature schemes and test them with various classifiers, so that a robust classification system could be built. A number of classification algorithms were employed and managed to achieve good accuracies in individual-instrument classification experiment.

The organization of the paper is as follows. In the next section, structure of system is described. The features that are used as discriminating variables are described in section III. The structure of the SVM adopted for the recognition system is discussed in Section IV. Results of experiments are summarized in Section V with concluding remarks presented in Section VI.

II.SYSTEM DESCRIPTION

The samples were collected from The McGill University Master Samples collection, a fabulous set of CDs of instruments playing every note in their range, recorded in studio conditions. We realize that this is a strong constraint and our result may not generalize to a complicated setting such as dealing with live recordings of an orchestra. The purpose of this experiment, however, is to test the effect of the various features and test the performance of different classifiers.

The samples used in our experiment consists of 70% of single instrument files from 10 instruments as a training samples and 30% of testing samples. In this experiment, we used three types of music families. Table 1 shows the category of instrument those are used in our system.

TABLE 1. The musical instrument collection

Instrument Family	Instrument Example
String	Violin,viol,viola,cello,bass,harp,guitar
Brass	Trombone,Trumpet
Keyboard	Piano

Music signals usually contain voiced, unvoiced and many areas of silence or noise. The first step in music analysist is needed to apply a silence removal method to detect "clean" signal. The signal is first divided into frames of 23.2 milliseconds in length. There are 1024 number of samples for each frame. Silence removal algorithm is carried out based on energy feature. To find silence part, we have calculated threshold as a median of energy. If amplitude of frame is less than threshold then it is silence part of the signal [12]. Different feature schemes are applied for feature extraction. These feature vectors are applied to classifier to identify the type of musical instrument. Fig. 1 shows time domain representation of trumpet signal and clean signal after application of silence removal algorithm.

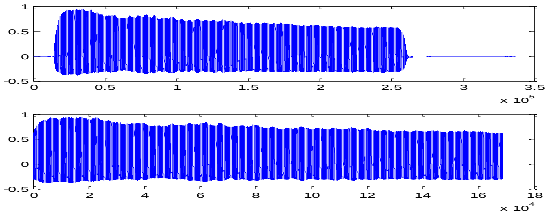

Fig. 1 Silence removed Trumpet Signal

III. FEATURE EXTRACTION

The main objective is to analyze the performance of different features for a robust instrument classifier. The features are the numerical values extracted from a signal that are then fed into the classifier. Here, we use four different extraction methods, namely, temporal features[13], spectral features [13], cepstral features and wavelet based entropy. The audio file is segmented into number of frames . Each frame consists of 1024 samples. After segmentation , each frame is hamming-windowed and 31 features are extracted for each frame. The 31 features from four categories are listed in Table 2.

TABLE 2. Feature Description

Feature Number	Description	Scheme
1	Log attack time	Temporal-based
2	Temporal centroid	
3-5	Mean, std deviation and variance of zero crossing rate	
6	Fundamental frequency	
7-9	Mean, std deviation and variance of autocorrelation	
10-12	Mean, std deviation and variance of spectral centroid	Spectral-based
13-15	Mean, std deviation and variance of spectral flux	
16-18	Mean, std deviation and variance of spectral spread	
19-21	Mean, std deviation and variance of spectral skewness	
22-24	Mean, std deviation and variance of mfcc	Perceptual-based
25-27	Mean, std deviation and variance of delta mfcc	
28-30	Mean, std deviation and variance of double delta mfcc	
31	Wavelet entropy	Wavelet-based

A. Temporal Features

Temporal features are features obtained directly from the time-domain music signal [5], [11].

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

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

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:

$$\text{Zero Crossing Rate} = \frac{1}{N}\sum_{n=1}^{N-1} |\text{sign}(x[n]) - \text{sign}(x[n-1])| \qquad (2)$$

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

Periodicity

The dominant periodicity of a signal is detected using a technique called Autocorrelation. The technique is to multiply the frame by a time-lagged copy of itself, then to measure the amplitude of the new signal. Where the amplitude reaches its peak will be where the peak(s) of the original signal are multiplied by the peak(s) of its copy, i.e. where the first period of the signal has been completed. The value of the time-lag where this peak occurs can then be considered the periodicity of the signal. The autocorrelation function is defined as:

$$\text{Autocorrelation}\,(k) = \sum_{t=1}^{N} x(t)\,x(t-k) \qquad (3)$$

i.e. the signal *x*(*t*) multiplied by a time-lagged copy of itself x(t-*k*).

Log-Attack Time:

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

$$\text{Lat}=\log_{10}(\text{stop\_attack} - \text{start\_attack}) \qquad (4)$$

Temporal centroid :

The temporal centroid is the time averaged over the energy envelop. It allows distinguishing percussive from sustained sounds.

B. Spectral Features:

Spectral features are obtained from the samples in the frequency domain of the musical signal [5],[11].

Spectral Centroid

This is the amplitude-weighted average, or centroid, of the frequency spectrum, which can be related to a human perception of ‘brightness’. 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:

$$\text{Spectral centroid} = \left(\frac{(\sum \text{mag}[i]) \times \text{freq}[i]}{\sum \text{mag}[i]}\right) \qquad (5)$$

where mag= magnitude spectrum and freq=frequency corresponding to each magnitude element

Spectral flux

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

$$\text{spectral flux} = \sum(\text{norm}_f[i] - \text{norm}_f[i])^2 \qquad (6)$$

Spectral spread

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

$$\text{Spectral spread} = \sqrt{\frac{\sum_{k=0}^{N/2}(\text{freq}_k - SC)^2\text{mag}^2}{\sum_{k=0}^{N/2}\text{mag}^2}} \qquad (7)$$

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

Spectral skewness

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

$$\text{Spectral skewness} = \frac{\sum(\text{freq} - SC)^3 \times \text{mag}}{\sum \text{mag}} \qquad (8)$$

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

C. Cepstral feature

Mel frequency cepstral coefficients

In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. They are derived from a type of cepstral representation of the audio clip. The difference between the cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound. The mel scale can be approximated from a Hz value by the formula

$$\mathrm{Melfrequency} = 2595 \times \log_{10}\left(\frac{1+x}{700}\right) \qquad (9)$$

Where x is frequency in Hz

Calculating MFCCs is performed as follows:

1. Calculate the Fourier transform (FFT) of a signal frame;
2. Map the decibel amplitude of the spectrum onto the Mel scale, using overlapping triangular windows; and
3. Calculate the discrete cosine transform (DCT) of this result.

D. Wavelet

In analysis, a discrete wavelet transform is any wavelet transform for which the wavelets are discretely sampled. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution: it captures both frequency and location information i.e. location in time. The wavelet analysis decomposes a signal into “packets” by simultaneously passing the signal through a low pass filter and a high pass filter in a sequential tree like structure. In this experiment we considered fifth level decomposition of Daubencies wavelet .

IV. SVM

The SVM is supervised learning model. A support vector machine constructs a hyperplane in a high or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. A good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class, since in general the larger the margin the lower the generalization error of the classifier. The discrimination function for the nonlinear SVM is described as

$$f(x) = \mathrm{sgn}(w^T K(x_k, x) + b) \qquad (10)$$

Where x_k is the support vectors in the data $\boldsymbol{x}$, the weighting vector $\boldsymbol{w}$ and the threshold b are parameters that decide the discrimination function. In the learning step, the support vectors x_k and the optimal parameters in the discrimination function (the weighting vector $\boldsymbol{w}$ and the threshold b) is decided from the learning data using for the Lagrange’s method of undetermined multipliers.

A The multi-class classification method

The SVM is able to classify into 2 classes. In this experiment, multiclass classification is implemented using SVM. There are two multiclass classification methods. The first is one-against-rest method and the other is one-versus-one method [2],[3].

Fig. 2 shows one-against-rest method when classifying into three different classes such as Class A, Class B, and Class C. If we classify the data into *k* classes, the one-against rest method is used *k* SVM classifiers to classify into the random class and the rest of it. Then, we classify by the outputs for the discrimination function of SVM classifier. Finally, the class is determined from the maximum value in the outputs for the discrimination function. On the other hand, Fig. 3 shows the one-versus-one method by examples for classifying into Class A, Class B, and Class C. If we classify the data into *k* classes, the one-versus-one method is used *kC*2 SVM classifiers. *kC*2 represents the number of the combination selected. Then, we classify by the calculated values. The value is calculated as follows. If the output for the discrimination function of a classifier is the positive, the value is added to the value calculated of the class corresponding to the positive class in the SVM classifier. Otherwise, the absolute value is added to the calculated value of the class corresponding to the negative class in SVM classifier. The class is determined from the maximum value in the value calculated.

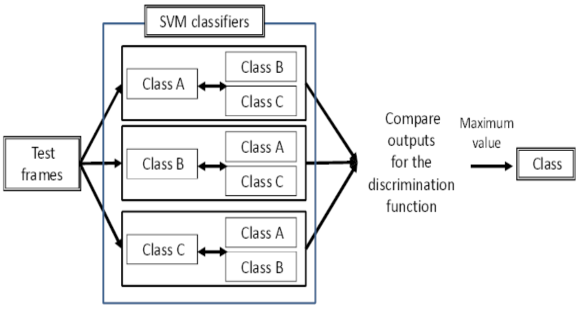

Fig. 2 Flowchart of the one-against-rest method

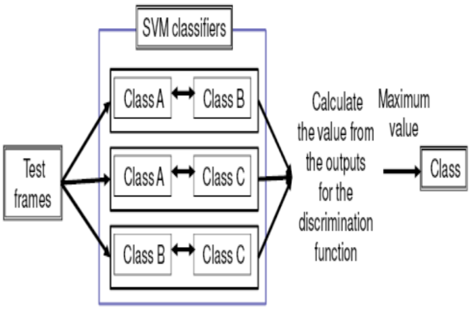

Fig. 3 Flowchart of the one vs. one method

V. SIMULATION RESULTS

The effectiveness of the proposed method is done by calculating accuracy of two multiclass SVM model. We classify much real music data by using the proposed method. In SVM classifier, various basis functions are available. In this experimrnt, the radial basis function, exponential radial basis function and the gaussian kernel are used.Fig. 4 shows the separation of two features using Gaussian kerenel.

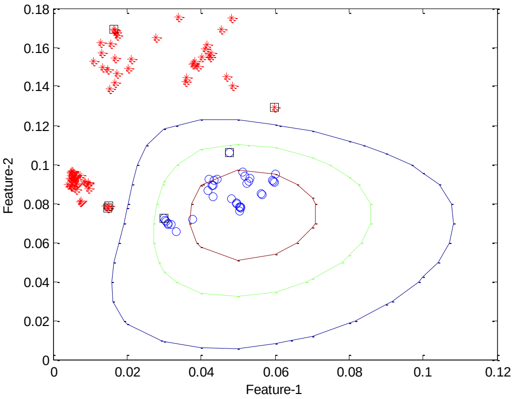

Fig.4 Seperation of two features using Gaussian kernel

Table 3 shows the average accuracy of system using SVM-One against rest and One Vs. One method for different types of kernel. We conclude that Gaussian RBF shows the good result for each feature.

TABLE 3.Average Accuracy of SVM-One against rest and SVM-One vs.One

Feature Scheme	Avg. Accuracy(%) of SVM (One Against Rest)			Avg. Accuracy(%) of SVM (One Vs. One)		
	RBF	ERBF	GRBF	RBF	ERBF	GRBF
Temporal	30	60	75	50	65	70
Spectral	78	77	78	76	77	78
Perceptual	64	78	80	69	74	79
Wavelet	70	79	81	80	86	82

Performance of SVM-One against Rest and SVM-One Vs. One is shown in Fig. 5 and Fig. 6

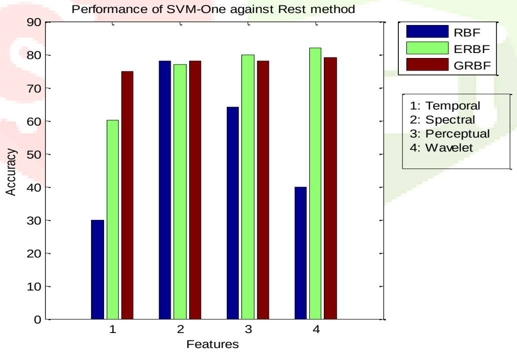

Fig. 5Performance of SVM-One against rest method

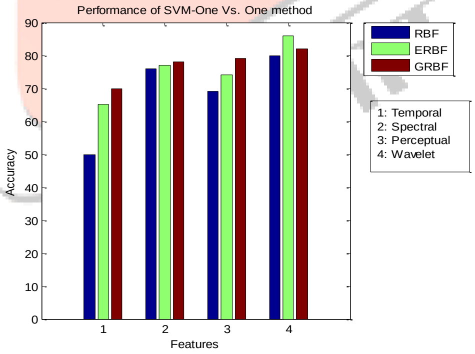

Fig. 6 Performance of SVM-One vs. one method

VI. CONCLUSION

In this paper automatic classification of various instrumental music is discussed by considering four features using SVM. The simulation results of automatic classification proved that SVM-One Vs. One method is best as compare to SVM- One against Rest method. Total 31 features are considered for classification. Therefore selecting correct feature is important task to improve the accuracy. As large number of features are available, there is another issue of dimension of extracted features. In future there is a task to combine wavelet with other features or using some feature selection algorithms to reduce attributes and eliminate similarity to get maximum marginal SVM. Less attributes and higher accuracy are the basis of real-time automatic classification systems.

REFERENCES

[1] Jing Liu, Lingyun Xie, "SVM-Based Automatic Classification of Musical Instruments", International conference 2010

[2] Giovanni Costantini,Massimiliano Todisco,Renzo Perfetti,Roberto Basili,Daniele Casali , "SVM Based Transcription System with Short-Term Memory Oriented to Polyphonic Piano Music" ", IEEE 2010

[3] Tao Li and Mitsunori Ogihara, "Toward Intelligent Music Information Retrieval" , IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 8, NO. 3, JUNE 2006

[4] Harya Wicaksana, Septian Hartono, & Foo Say Wei, "Recognition of Musical Instruments" IEEE TRANSACTIONS 2006

[5] Jeremiah D. Deng, *Member, IEEE*, Christian Simmermacher, and Stephen Cranefield , "A Study on Feature Analysis for Musical Instrument Classification" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 38, NO. 2, APRIL 2008

[6] Changsheng Xu, *Senior Member, IEEE*, Namunu C. Maddage, and Xi Shao , "Automatic Music Classification and Summarization", IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 13, NO. 3, MAY 2005

[7] Kyu-Sik Park, Won-Jung Yoon, Kang-Kue Lee, Sang-Heon Oh and Ki-Man Kim, "MRTB Framework: A Robust Content-Based Music Retrieval and Browsing" IEEE TRANSACTIONS 2006

[8] Qian Ding and Nian Zhang, "Classification of Recorded Musical Instruments Sounds Based on Neural Networks" Proc eedings of the 2007 IEEE Symposium on Computational

[9] Da Deng†, Christian Simmermacher, Stephen Cranefield, "Finding the Right Features for Instrument Classification of Classical Music", Proceedings of the International Workshop on Integrating AI and Data Mining IEEE 2006

[10] Bozena Kostek, "Musical Instrument Classification and Duet Analysis Employing Music Information Retrieval Techniques", Proceedings of IEEE, VOL. 92, NO. 4, APRIL 2004

[11] Qian Ding and Nian Zhang, "Classification of Recorded Musical Instruments Sounds Based on Neural Networks", Proceedings of the 2007 IEEE Symposium on omputational
Intelligence in Image and Signal Processing

[12] Theodoros Giannakopoulos , " A method for silence removal and segmentation of speech signals".

[13] S.Gunasekaran, K.Revathy , "Recognition of Indian Musical Instruments with Multi-Classifier Fusion", 2008 International Conference on Computer and Electrical Engineering

[14] Davyd Madeley, Dr. Roberto Togneri, "Automatic Computer Classification of Solo Musical Instruments"