2016 IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT)
Rajarshi Shahu College of Engineering, Pune India. Dec 2-3, 2016

Musical Instrument Recognition using K-Nearest Neighbour and Support Vector Machine

R.S.Kothe, D.G.Bhalke,P.P.Gutal
RSCOE, Electronics and Telecommunication department.,SPPU Pune, India
rskothe@gmail.com, bhalkedg2000@yahoo.co.in, pratima.gutal@gmail.com

Abstract-. **In this paper, we present a model to detect and distinguish individual musical instrument using different feature schemes. The proposed method considers ten musical instruments. The feature extraction scheme consists of temporal, spectral, cepstral and wavelet features. We developed k-nearest neighbor model and support vector machine model to test the performance of system. Our system achieves the 60.43% of recognition rate using k-nearest neighbor classifier with all features. A two prong approach was taken to the multi-class classification which were SVM-one against rest &SVM-one vs. one. The accuracy of SVM in both cases is 73.73% with all features using radial basis function. Using weight factor method knn shows 73% accuracy while SVM shows 90.3% accuracy using exponential kernel function. Using weight factor method knn shows 73% accuracy while SVM shows 90.3% accuracy using exponential kernel function.**

Keywords-feature extraction, temporal, spectral, perceptual, cepstral, wavelet, KNN, SVM, RBF, EBRF, GRBF.

I. INTRODUCTION

Numerous attempts have been undertaken for automatic construction of recognition of musical instruments. Varied performances have been derived withvaried approach and scopes. As things stand today, much of the work required to tag and catalogue music with search terms is conducted manually. Cataloguing musical recordings by instruments often requires a either trained musician or a priori knowledge about the recording. Due to a lack of a samples database that was both suitable for the project, this project was also concerned with collecting recordings from six musical instruments for training and testing the classifier.

In this paper, we predict the sound of musical instrument based on feature extraction using machine learning techniques. Four feature schemes are considered: temporal features, spectral features, cepstral features, and wavelet based entropy. The performance of the feature scheme was assessed individually.

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

In our work we aim at classifying the musical instrument sounds. The samples have been collected for the Master Samples collection of McGill University which has examples of all the notes in their range recorded in studio conditions.In order to classify musical instruments properly several stages are needed: pre-processing, feature extraction (parameterization), and the actual classification process as shown in Fig.1. An audio file stored in WAV format is passed to a silence removal algorithm which detects music segment. This music segment is passed to a feature extraction function. The feature extraction function calculates 31 numerical features that characterize the sample. When training the system, this feature extraction process is performed on training data to create a matrix of column feature vectors. This matrix is then applied as an input to the classifier. Two classifiers are used to predict the sound of musical instrument.

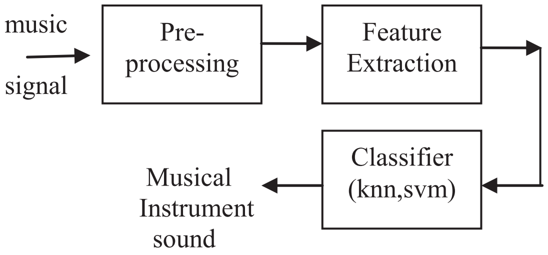
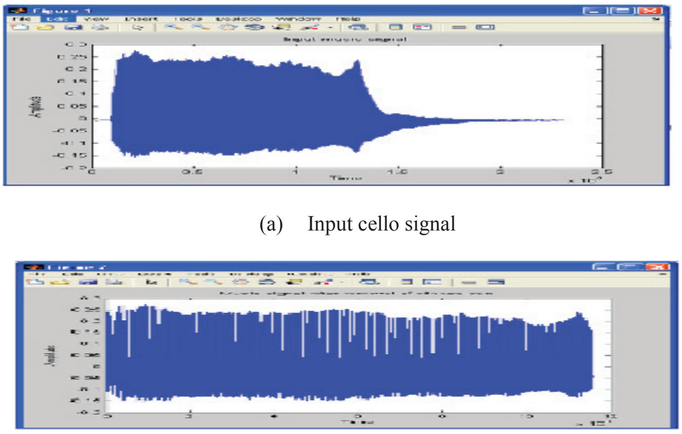


Fig.1 Block diagram of musical instrument recognition system

Music signals usually contain many areas of silence or noise. Therefore, in music analysis it is needed to first apply a silence removal method, in order to detect "clean" signal. The signal is first divided into frames of 23.2 milliseconds in length. Silence removal algorithm is carried out based on energy feature as in [7]. A simple threshold based algorithm is applied to extract music signal. A threshold is calculated based on energy of signal. We considered threshold as median of energy. Fig.2 shows graphical representation of silence removed signal.



(a) Input cello signal

(b) Silence removed Cello Signal

Fig. 2 Graphical representation of input signal and silence removed signal

III. FEATURE EXTRACTION

The field of music feature extraction is a wide research area, for improving feature extraction will most likely have the major impact on the performance of an instrument classification system .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, spectral features, cepstral features and wavelet based entropy. The 31 features from four categories are shown in Table 1.

TABLE 1. 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 as in [5], [4].

1) Energy:

The summation of amplitudes present in frame is simply defined as Energy (1).

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

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

2) Zero-Crossing Rate:

The count of instances when the signal crosses zero during the frame is taken as measure of noisiness in the signal. It is defined as in (2).

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

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

3) 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. It is defined as in (3).

$$Lat=log_{10}(stop\_attack - start\_attack) \qquad (3)$$

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 as in [5],[4].

1) Spectral Centroid:

The amplitude-weighted average, or centroid, of the frequency spectrum, is equivalent of a human perception of 'brightness'. It is arrived by multiplying the value of each frequency by its magnitude in the spectrum, then taking the sum of all these. Dividing values by the summation of all the magnitude fives the value in (4).

$$Spectral\ centroid = \left(\frac{(\sum mag[i]) \times freq[i]}{\sum mag[i]}\right) \quad (4)$$

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

2) Spectral flux:

This is a calculations of the value of local spectral change. This is defined as the squared difference between the normalized magnitude spectra of successive frames as in (5).

$$spectral\ flux = \sum(norm_f[i] - norm_f[i])^2 \quad (5)$$

3) Spectral spread:

The spectral spread is a measure of variance (or spread) of the spectrum around the mean value μ .It is given in (6).

$$Spectral\ spread = \sqrt{\frac{\sum_{k=0}^{N/2}(freq_k - SC)^2 mag^2}{\sum_{k=0}^{N/2} mag^2}} \quad (6)$$

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

4) 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. It is defined as in (7).

$$Spectral\ skewness = \frac{\sum(freq - SC)^3 \times mag}{\sum mag} \quad (7)$$

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

C. Cepstral feature

1) Mel frequency cepstral coefficients:

Mel Frequency Cepstral Coefficients (MFCCs) are cepstralcoefficients used for definingaudio in a way that mimics the physiological properties of the human auditory system [5],[9]. 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 enhanced logarithmically using the Mel scale. A mel is a psychoacoustic unit of frequency which relates to human perception, the mel scale can approximated from a Hz value as in (4)

$$Melfrequency = 2595 \times log_{10}\left(\frac{1+x}{700}\right) \quad (8)$$

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

The wavelet analysis gives spectro-temporal Information. The wavelet 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 many types of filters that can be used for this purpose. In this experiment we considered fifth level decomposition of Daubencies wavelet

IV. CLASSIFICATION

Classification is taking a decision about the class membership of an object. Such a decision can be correct or incorrect; the goal is to maximize the chance of making the right decision. The proposed system has been experimented with k-Nearest Neighbor

A. K-nearest neighbor classifier

In k-nearest neighbor classification, the training dataset is used to classify a testing dataset. The algorithm is described as follows:

1. For each case in the testing dataset to be classified, locate the k nearest neighbors of the training dataset. A Euclidean Distance measure is used to calculate how close each member of the training set is to the target row that is being examined.
2. Examine the k nearest neighbors. Assign this category to which most of them belong to, to the case being examined.
3. Repeat this procedure for the remaining cases in the target set.

The k-nearest neighbour method is intuitively a very attractive method. A disadvantage of this method is its large computing power requirement, since for classifying an object its distance to all the objects in the learning set has to be calculated.

There are two significant ways to fine-tune the performance of the k-NN classifier:

1. Modify the distance function,
2. Change k, the number of neighbours conferred in each classification.

B. Support vector machine

The SVM is a learning method which is supervised. We compute the most effective boundary that divides the data into each class. A nonlinear SVM is applied, because we aim to mark off the boundary complexly. The fundamental idea of the nonlinear SVM is to transform input vectors into a high-dimensional feature space using a kernel function, and then to separate in feature space linearly. The discrimination function for the nonlinear SVM is described as in (9).

$$f(x) = sgn(w^T K(x_k, x) + b) \qquad (9)$$

Where xk is the support vectors in the data x, the weighting vector w and the threshold b are parameters that decide the discrimination function. In the learning step, the support vectors xk and the optimal parameters in the discrimination function (the weighting vector w and the threshold b) is decided from the learning data using for the Lagrange's method of undetermined multipliers. In order to make the SVM classifier, the radial basis function(RBF), exponential radial basis function(ERBF) and the Gaussian(GRBF) kernel are used. A radial basis function is described as in (10).

$$K(x, x^{'}) = exp\left(-\frac{\left\|x - x^{'}\right\|}{2\sigma^2}\right) \qquad (10)$$

1) The multi-class classification method:

In general, it is known that the SVM is able to classify into 2 classes. We propose the multi-class classification method to combine some SVM classifiers. The multiclass classification method contains the one-against-rest method and the one-versus-one method as in [2],[3].

Fig.3 shows the flowchart of the one-against-rest method when classifying into 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 arbitrary 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.4 shows the flowchart of the one-versus-one method by examples for classifying into Class A, Class B, and Class C. When data is classified into k classes, the one-versus-one method is used kC2 SVM classifiers, where kC2 represents the number of the combination selected. Then, classification is done by the calculated values. The value is calculated as follows. If the output for the discrimination function of a SVM classifier is the positive value, 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.

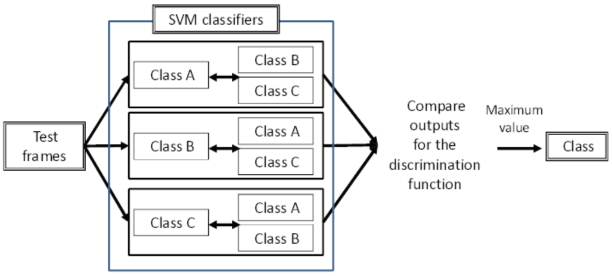


Fig. 3 Flowchart of the one-against-rest method

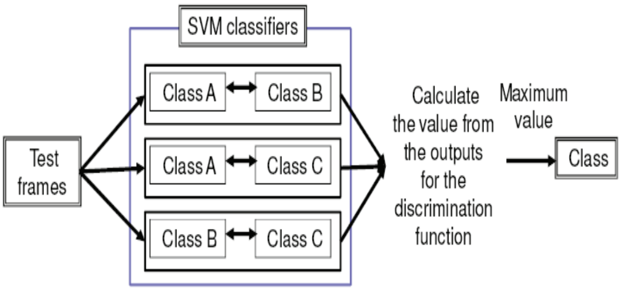


Fig. 4 Flowchart of the one vs. one method

V. Simulation Results

We show computer simulations to indicate the effectiveness of the proposed method. The samples used in our experiment consists of 50% of single instrument files from 10 instruments as a training samples and 50% of testing samples. Table 2 shows the category of instrument those are used in this experiment.

TABLE 2. The musical instrument collection

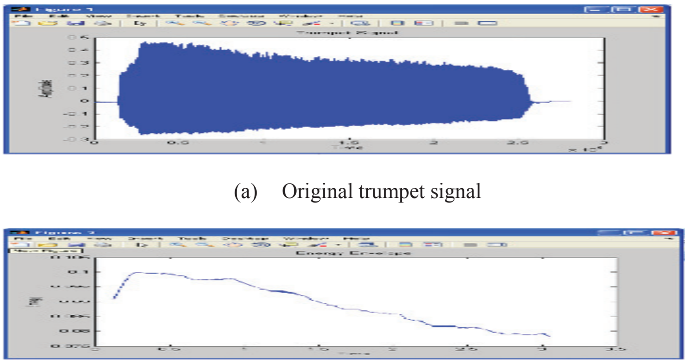
Instrument Family	Instrument Example
String	Violin, viol, viola, cello, bass, harp, guitar
Brass	Trombone, Trumpet
Keyboard	Piano

Table 3 shows the details of the experimental condition for the analysis in the computer simulation. Every audio file is divided into frames of 1024 samples. Each frame is hamming-windowed and temporal, spectral, perceptual and wavelet based features are extracted for each frame. The system is tested with KNN and SVM classifier using each feature scheme and finally all features are combined into a single vector which consists of 31 numeric values

TABLE 3. The Details of the experimental condition for analysis

Property	Value
Sampling frequency	44.1 KHz
Frame duration	23 ms
Samples per frame	1024 points
Window function	Hamming

The energy envelope is required to find out temporal features such as log attack time and temporal centroid. The graphical representation of energy envelope of trumpet signal is shown in Fig.5



(a) Original trumpet signal

(b) Energy envelope of trumpet

Fig. 5 Graphical representation of energy envelope

The spectral features are calculated using frequency spectrum. We calculated spectral centroid, spectral flux, spectral rolloff, spectral spread and spectral skewness using frequency spectrum. Fig.6 shows the spectral features of cello signal. Also perceptual features are shown in Fig.7.

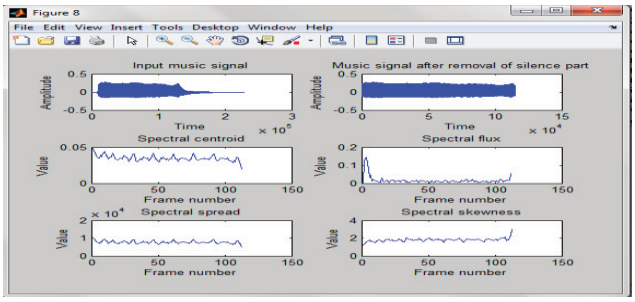


Fig.6 Spectral features of cello

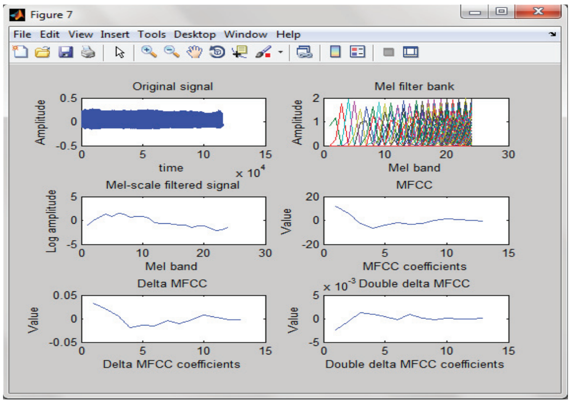


Fig.7 Perceptual features of cello

Table 3 shows the recognition rate using KNN classifier. Using all features the accuracy of system is 65%.

TABLE 3.Average Accuracy of KNN

Feature Scheme	Recognition Rate(%)
Temporal	73.84
Spectral	59.56
Perceptual	73.52
Wavelet	53.90
All features	60.43

Table 4 and 5 shows the accuracy of system using SVM-one against rest and SVM-one vs. one method for different types of kernel. The recognition rate achieved with all feature scheme is 73.73% in both classifier using radial basis function.

TABLE 4. Accuracy of SVM-one against rest

Feature Scheme	Avg. Accuracy(%) of SVM (One Against Rest)		
	RBF	ERBF	GRBF
Temporal	30	60	75
Spectral	78	77	78
Perceptual	64	78	80
Wavelet	70	79	81
All features	73.73	72.17	67.48

TABLE 5. Accuracy of SVM-one vs.one

Feature Scheme	Avg. Accuracy(%) of SVM (One Vs. One)		
	RBF	ERBF	GRBF
Temporal	50	65	70
Spectral	76	77	78
Perceptual	69	74	79
Wavelet	80	86	82
All features	73.73	72.71	67.48

Feature selection techniques are often applied to optimize the feature set used for classification. This way, redundant features are removed from the classification process and the dimensionality of the feature set is reduced to save computational time. We note that care has to be taken that not too many features are removed. The effect of multiple features compensating each other could be desirable, since it is not exactly clear how musical timbre is described best.

Feature combination schemes generated from the selection rankings are then further assessed using classifiers and cross validated. The KNN and Support Vector Machine, classification algorithm has been implemented.

To implement the robust system, we observed the results of feature selection method and decided to increase the weight of the feature which shows the maximum accuracy. The feature which shows the maximum accuracy has given the large weight.

The ranked feature set considers best 18 features which are used to test the system. The results of ranked feature set and weight factor method are compared with all features. The results of KNN, SVM-one against rest and SVM-one vs one are listed in Table 6, Table 7 and Table 8 respectively.

The results with ranked feature set and weight factor method are improved over the all feature set. The recognition rate with weight factor method using KNN is 73%. SVM in both cases achieve the same recognition rate.

TABLE 6. Performance of KNN using different feature set

Features	KNN (Accuracy %)
All 31	60.43
Best 18	70
Weight factor	73

TABLE 7. Performance of SVM-one against rest using different feature set

Feature Scheme	Accuracy(%) of SVM (One against rest)		
	RBF	ERBF	GRBF
All 31	73.74	72.18	67.49
Best 18	84.64	84.64	77.93
Weight factor	88.39	90.3	89.89

TABLE 8. Performance of SVM-one vs one using different feature set

Feature Scheme	Accuracy(%) of SVM (One vs one)		
	RBF	ERBF	GRBF
All 31	73.74	72.18	67.49
Best 18	84.64	84.64	77.93
Weight factor	88.39	90.3	89.89

VI. CONCLUSION

The simulation results show that k-nearest neighbor classifier achieve the 60.43% of recognition rate with all features. The recognition rate of SVM-one against rest and SVM-one vs one is same which is 73.73% using radial basis function with all features. As the features are reduced the recognition rate is increased. Using weight factor method KNN shows 73% accuracy while SVM shows 90.3% accuracy using exponential kernel function. In simulation, some instrument shows the same feature similarity. So future work must consider the combination of some features to build the robust system.

REFERENCES

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

[2] 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

[3] Harya Wicaksana, Septian Hartono, & Foo Say Wei, “Recognition of Musical Instruments” IEEE TRANSACTIONS 2006

[4] Qian Ding and Nian Zhang, “Classification of Recorded Musical Instruments Sounds Based on Neural Networks” Proc eedings of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing

[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] 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

[7] Theodoros Giannakopoulos , “ A method for silence removal and segmentation of speech signals ”, implemented in Matlab, 2010