Comparing accuracy of Musical Instrument identification for different features with k-NN and SVM Classifiers

Amrit Pal Kaur **Prof. D.M. Chandwadkar** **Prof. M. S. Sutaone**

***Abstract*— Music signal is different from speech signal, therefore for recognizing musical instrument there is a need of designing a different system. Here we have designed an instrument identification system.In this paper we have discussed the role of various features for classifying musical instruments with k-NN and SVM classifiers.We have tried to identify musical instrumentsusing monophonic signals. Fivemusical instruments: Flute, Piano, Trumpet, Guitar and Xylophone are considered. From each signal of the database we have extracted different features and these features are used for training the classifiers. The trained classifiers are then tested to identify musical instruments using unseen signals. It is observed that the identification accuracy depends on the features used as well as the classifier. This analsis helps to select the desired classifier with all features and a desired feature with a classifier to design the musical instrument identifier system.**

***KeyWords* —Signal framing, Feature extraction, Feature selection, Classifier**

I. INTRODUCTION

The central topic in our paper is the analysis of musical instrument features for their recognition. Musical instrument identification has many applications in Music Information Retrieval (MIR) systems such as the automatic tagging and cataloguing of music for search and retrieval, automatic musical transcription and as part of larger systems (e.g. genre classification). This paper describes how to implement a musical instrument identification system using Matlab. Musical instrument identification is defined as the process of analyzing an acoustic musical signal so as to identify the instrument playing it. It obtains the name of the musical instrument from the sound sample under testing.

We can identify musical instruments by using the monophonic or polyphonic recordings. In this paper we have used the monophonic signals (isolated notes played by various orchestral instruments). The McGill University Master samples collection, a fabulous set of DVDs of instruments playing every note in their range, recorded in studio conditions are used as the database.For the training and testing of various classifiers, this database in .wav format is used. The classification is done using either k-NN classifier or SVM.

The complete process of identification is divided into three steps:

1. Signal framing
2. Feature Extraction and Feature selection.
3. Classification

II. SIGNAL FRAMING

The process involves dividing the signal into frames using windowing. The aim of windowing is to segment the signal into statistically stationary blocks because it is very difficult to analyze the whole signal at once. Hamming window is used to weight the frames. The Discrete Fourier transform (DFT) is calculated for each of these frames.

Time domain features are directly extracted from the segmented signal for all frames and frequency domain features are extracted from the calculated DFT of individual frames.

III. FEATURE EXTRACTION

The purpose of feature extraction is to obtain the relevant information from the input data to execute a certain task using this desired set of features. There are basically two types of features [1]:

1. Time domain features
2. Frequency domain features

Time domain features: temporal centroid and zero crossing rate, and frequency domain/ spectral features: spectral flux, spectral centroid, spectral irregularity and spectral flatness measure are used in this paper.

a) Temporal Centroid: The temporal centroidhas been found important as signal descriptors for highly transient and percussive sounds. It is defined as the energy weighted mean of the time signal given the in equation.

$$\mathrm{TC} = \frac{\sum_{k=0}^{L-1} kx[k]}{\sum_{k=0}^{L-1} x[k]} \tag{1}$$

$x[k]$is the input signal, $k$time index and L the length of the signal.

b) *Zero crossing rate*: Zero-crossing rate is a measure of the number of times in a given time interval that the amplitude of the audio signal passes through a value of zero [2].

ZCR

$$=\frac{1}{2}*\sum\left|\operatorname{sgn}(x[n])-\operatorname{sgn}(x[n-1])\right| \quad (2)$$

sgn(x)= -1 for x =0
1 for x ≥0

Zero-crossing rate is important because they abstract valuable information about the audio signal and they are simple to compute.

c) *Spectral Flux*: Spectral flux reflects the rate of change of the power spectrum. It is the measure of how quickly the power spectrum changes from frame to frame. It is calculated by comparing the power spectrum of one frame with the power spectrum of the previous frame. We may calculate the Euclidean distance between the normalized spectra. Spectral Flux can be used to determine Timbre of an Audio signal [3].

$$\text{Spectral Flux}=\sum_{k-2}^{k}\left|M(Fk)-M(Fk-1)\right| \quad (3)$$

d) *Spectral Centroid*: This parameter also characterizes the spectrum of signal. It indicates the location of the 'centre of gravity' of magnitude spectrum. Perceptually, it gives the impression of brightness of sound. It can be evaluated as the weighted mean of spectral frequencies. We find the FFT of the signal segment and then find the average energy distribution in steady state portion of the signal [3].It measures the spectral shape. Higher Centroid values correspond to brighter texture with more high frequencies [4].

$$\text{Spectralcentroid}=\frac{\sum_{k=0}^{N-1}f[k]x[k]}{\sum_{k=0}^{N-1}x[k]} \quad (4)$$

e) *Spectral irregularity*:Spectral irregularityalso referred to as spectral smoothnessbasically shows the irregularity of asignal usually computedwith the STFT where the average of the current, next and previous amplitudevalues are compared with the current amplitude value. It is calculated as:

$$\text{SI}=\frac{\sum_{k=0}^{N-1}(x[k]-x[k-1])^{\wedge}2}{\sum_{k=0}^{N-1}x[k]^{\wedge}2} \quad (5)$$

f) Spectral flatness measure: The spectral flatness measureis defined as the ratio between the geometricmean(Gm), and the arithmetic mean(Am). It givesinsight on the noise content of a signal and has been used in speech research toextractvoiced and unvoiced speech signals. As SFMapproaches 0 the signal becomes more sinusoidal and as SFM approaches 1 thesignal becomes more flat and de-correlated.

$$\text{SFM}=10log\frac{Gm}{Am} \quad (6)$$

IV. Feature Selection

Feature selection is one of the important steps to find the ideal feature set for the identification system. As 'N' numbers of features are extracted and the value of N varies from signal to signal, the features are normalized by their minimum and maximum value, mean, standard deviation and variance. This forms the feature vector of same size for each input signal which is used for training and testing the classifiers.

V. Classifier

For the classification purpose we have used k-Nearest Neighbor (k-NN) classifier and Support Vector Machine (SVM), where k-NN is an instance based classifier and SVM are Support Vector Machines based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships.

In k-NN classifier, classification of unknown instances can be done by relating the unknown to the known according to some distance/similarity function. It is simple matter of locating the nearest neighbor in instance space and labeling the unknown instance with same class label as that of the known neighbor [5]. Here we used value of k as 1 and used Euclidean distance for nearest rule. The classifier takes less time in training but more in predicting.

Support vector machine (SVM) is a non-linear classifier which is often reported as producing superior classification results compared to other methods. The idea behind the method is to non-linearly map the input data to some high dimensional space, where the data can be linearly separated, thus providing great classification (or regression) performance [5]. Here for training the SVM we have used linear kernel function.

VI. Procedure

We have implemented the musical instrument identification system in MATLAB environment. The database which is used for this is in .wav format, which is the standard format. The procedure used in the process of identification of Musical instruments involve following steps

1. The whole system is divided into two stages, training and testing stage. Both the stages are having same frame work.
2. The input signal is read using wavread instruction.
3. Then this input signal is divided into number of frames with a frame size of 23 ms and 50% hop size. Hamming window is used. The music signals used from the McGill University master samples are sampled at 44.1 KHz. Hence in 23 ms we get 1024 (2^{10}) samples (window size)

and in 11.5 ms we get 512 samples (overlap). The FFT length is taken as 1024. Hence computation of FFT becomes accurate. Time domain features are extracted for each frame directly whereas for spectral features first FFT of each frame is computed.

4. Then from these features we form the feature vector.

This feature vector is used for training the classifiers and for testing purpose also we followed the same steps.

VII. Experimental Results

The system was used for identification of five instruments: flute, piano, trumpet, guitar and xylophone. For flute we used 27 samples, for piano 32 samples, for trumpet 34 samples, for guitar 49 samples and for xylophone 39 samples. Therefore, total 181 data samples were used from five instruments.

A. *Classification Accuracy with individual features*: First we used each feature individually for classification with k-NN classifier and SVM.

TABLE 1: Classification accuracy with individual features

Feature	Classification Accuracy for Classifier (%)	
	k-NN	SVM
Temporal Centroid	78	45
ZCR	79	50
Spectral Centroid	84	54
Spectral Flux	79	57
Spectral Irregularity	85	54
Spectral Flatness Measure	87	52

From the above analysis it is observed that the k-NN classifier gives higher accuracy for all individual features as compared to SVM classifier and the feature spectral flatness measure gives higher accuracy compared to other features.

B. *Classification Accuracy with all features*: We combined all the features and computed classification accuracy using this feature vector. The table below shows the confusion matrix obtained when all features were used for training and testing the classifier.

TABLE 2: Confusion matrix forSVM as classifier using all features

a	b	c	d	e	→ classified as
27	0	0	0	0	a = FLUTE
1	27	0	1	3	b = PIANO
1	0	33	0	0	c = TRUMPET
0	1	0	48	0	d = GUITAR
0	0	0	0	39	e = XYLOPHONE

TABLE 3: Classification accuracy with all features

Classifiers	k-NN	SVM
All features used together	82%	96.13%

Out of 181 samples of different instruments only 147 samples were classified correctly by k-NN classifier and 174 were classified correctly by SVM classifier. Therefore, the system gives the accuracy of 82% for k-NN and 96.13% for SVM classifier. From this analysis we conclude that SVM is a better classifier as k-NN when more number of features are used.

Conclusion

From the above experimental results we have analyzed that SVM gives much higher accuracy than k-NN classifier for maximum number of features when used for classification, whereas for the individual features the k-NN classifier gives better accuracy than SVM.

We also observed that the k-NN classifier with 'k' as 1 gives higher accuracy than 'k' with higher value. As well as the spectral flatness measure feature gives higher accuracy compared to other features with k-NN classifier.

Thus the analysis computed that, individual features give more accuracy for k-NN classifier and with all features SVM is giving much higher accuracy. Also when all spectral features are used the classifiers give more accuracy.

References

[1] J. Sirisha Devi, Y. Srinivas& N. Murali Krishna, "A Study: Analysis of Music features for Musical Instrument Recognition and Music Similarity Search," in IJCSI, ISSN (PRINT): 2231-5292, Vol.II, Issue-1,2.

[2] Chandwadkar, D.M.; Sutaone, M.S., "Role of features and classifiers on accuracy of identification of musical instruments," *Computational Intelligence and Signal Processing (CISP), 2012 2nd National Conference on*, pp.66-70, 2-3 March 2012.

[3] V.S.Shelar, D.G.Bhalke "Musical instrument transcription using neural network", TJSRET Vol.I, Issue2.

[4] G. Peeters, "A large set of audio features for sound description (similiarity and classification) in the CUIDADO project," in Proc. 115th AES Convention, 2004.

[5] G. Agostini, M. Longari, and E. Pollastri, "Musical instrument timbres classification with spectral features," in IEEE Workshop on Multimedia Signal Processing, 2001, p. 97

Author's Profile

Amrit Pal Kaur
She is pursuing M.E under the guidance of Prof. Dinesh M. Chandwadkar.
Email:badesha.amrit@gmail.com

Dinesh M.Chandwadkar
He is Professor in E & TC Department at K. K. Wagh Institute of Engineering Education & Research, Nashik, India. He is pursuing Ph. D. under the guidance of Prof. Sutaone. Email: dmc.eltx@kkwieer.org

Dr. Mukul S. Sutaone
He is Professor in Electronics and Telecommunication department, and Dean, Alumni and International Affairs, at College of Engineering, Pune (COEP), India. He has Ph.D. in the field of Texture Analysis. His areas of research interests are Multimedia, Multidimensional Signal Processing and Multi-carrier Communication.
Email: mssutaone.extc@coep.ac.in