ISSN: 0976-3104
SPECIAL ISSUE: Computer Science

ARTICLE

MUSICAL INSTRUMENT SOUND CLASSIFICATION

R Rayar[1], M Anto Bennet[2*], A Nazreen Banu[3], A Sushanthi[4], M Rajasekar[5]

[2]Department of ECE , VEL TECH,Chennai-600062
[1,3,4,5]Department of CSE, Veltech Multitech RR.SR Engineering College, Chennai, INDIA

ABSTRACT

Music instrument classification is essential in music indexing systems. Today Digital Audio Applications is a part of everyday life. Audio in the form of CD's, DVD's and broadcast data, is available in the internet for public access. In this project an automatic music instrument classification system is developed using Discrete Wavelet Transform DWT features. Proximal Support Vector Machine (PSVM) are based on the principle of structural risk minimization. DWT features are extracted from different classes of musical instruments namely flute, guitar, violin and piano. PSVM is trained and tested by using DWT features and the system shows satisfactory results with an accuracy of 89.00%.Index Terms—Discrete Wave Transform, Musical Instrument Sound Classification, Proximal Support vector machine(SVM)

KEY WORDS
Schistosoma nasale, cross H.F cow, Anthiomaline

Received: 24 October 2016
Accepted: 20 December 2016
Published: 15 February 2017

***Corresponding Author**
Email:
bennetmab@gmail.com

INTRODUCTION

A musical instrument is an instrument created or adapted to is make musical sounds. In general, any object that produces sound can be a musical instrument. It is through purpose that the object becomes a musical instrument. There was history of musical instruments dates to the beginning of human culture. Early musical instruments were used for ritual, such as a trumpet to signal success on the hunt, or a drum in a religious ceremony. Cultures eventually developed composition and performance of melodies for entertainment. Musical instruments evolved in step with changing applications.

Music signals represent a large class of audio data where several sound sources are usually present at the same time. Depending on the genre, the instrument may consist of electric guitars, bass, drums, and vocals or saxophone, piano, strings and percussion, For example, there is a wide variety of instruments in Western music alone, representing different sound production mechanisms and timbre [1],[2]. Automatic recognition of the instruments in recorded music has several direct applications, including music retrieval based on the instrumentation and audio management in recording studios. Even more importantly, sound source recognition and modeling is an essential part of making sense of complex audio signals. When listening to polyphonic music, human listeners are able to perceptually organize the component sounds to their sources, largely based on timbre information. Similarly, source models are an integral part of music transcription and sound separation systems, where the source identity enables the use of source specific models and assumptions and allows the organization of sounds events to "streams" that can be attributed to certain instruments [2] [3].

In addition to practical applications, a system that can automatically classify recordings by genre has significant theoretical musicological interest as well. There is currently a relatively limited understanding of how humans construct musical genres, the mechanisms that they use to classify music and the characteristics that are used to perceive the differences between different genres. A system that could automatically classify music and reveal what musical dimensions it is using to do so would therefore be of great interest.

Low-level signal processing based features are of little use in this respect, something that further emphasizes the importance of studying the use of high-level features This kind of research also has applications beyond the scope of genre classification. The techniques developed for a genre classification system could be adapted for other types of classifications, such as by compositional style or historical period. Once a classification system is implemented, one needs to only modify the particular training recordings and taxonomy that are used in order to perform arbitrary types of classification[9].

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 [9],[10]. The identification of the instruments that compose a musical signal has received increasing attention in the last years. Such an interest is fed by the potential benefits that an accurate instrument classifier can bring to other digital audio applications. In particular, musical genre classification can be greatly improved if the instruments present in a given song are known, since this information can be used to narrow down the set of potential musical genres. Sound source separation algorithms can also explore such information, particularly if they deal with underdetermined signals. In this case, the knowledge about the instruments can be used to create instrument specific rules to improve the quality of the sound source separation. Early work in the area was mainly devoted to the identification of instruments in monophonic signals. This

THE IIOAB JOURNAL
Editor: Dr. K. Sakthisudhan

problem is in general, less challenging than the polyphonic case, since the instrument to be classified is isolated from the interference of any other sound source. Most of the proposals those deal with general instruments while a few others deal with specific cases like classification of woodwinds and discrimination between piano and guitar [4].

By automatic musical genre classification we mean the classification of music signals into a single unique class based computational analysis of music feature representations. Automatic music genre classification is a fundamental component of music information retrieval systems. The process of genre categorization is described in two steps namely: feature extraction and multiclass classification. In the feature extraction step, extract from the music signals information representing the music. The features extract should be comprehensive (representing the music very well), compact (requiring a small amount of storage), and effective (not requiring much computation for extraction). To meet the first requirement the design has to be made so that the both low-level and high-level information of the music is included [7].

FEATURE EXTRACTION

The Discrete Wavelet Transform

The Wavelet Transform (WT) is a technique for analyzing signals. It was developed as an alternative to the Short Time Fourier Transform (STFT) to overcome problems related to its frequency and time resolution properties. More specifically, unlike the STFT that provides uniform time resolution for all frequencies the DWT provides high time resolution and low frequency resolution for high frequencies and high frequency resolution and low time resolution for low frequencies. In that aspect it is similar to the human ear which exhibits similar time-frequency resolution characteristics.

The Discrete Wavelet Transform (DWT) is a special case of the WT that provides a compact representation of a signal in time and frequency that can be computed efficiently.

The DWT is defined by the following equation:

$$w(j,k) = \sum_{j}\sum_{k} x(k)2^{-\frac{j}{2}}\Psi(2^{-j}n-k)$$

Where t ψ is a time function with finite energy and fast decay called the mother wavelet. The DWT analysis can be performed using a fast, pyramidal algorithm related to multi rate filter banks. As a multi rate filter bank the DWT can be viewed as a constant Q filter bank with octave spacing between the centers of the filters. Each sub band contains half the samples of the neighboring higher frequency sub band. In the pyramidal algorithm the signal is analyzed at different frequency bands with different resolution by decomposing the signal into a coarse approximation and detail information. The coarse approximation is then further decomposed using the same wavelet decomposition step. This is achieved by successive high pass and low pass filtering of the time domain signal and is defined by the following equations.

$$y_{high}[k] = \sum_n x[n]g[2k-n]$$

$$y_{low}[k] = \sum_n x[n]h[2k-n]$$

where $y_{high}[k]$ is the high pass filter
$y_{low}[k]$ is the low pass filter

The output respectively after sub sampling. Because of the down sampling the number of resulting wavelet coefficients is exactly the same as the number of input points. A variety of different wavelet families have been proposed in the literature. In our implementation, the 4 coefficient wavelet family (DAUB4) proposed by Daubechies is used.

Wavelet representation for audio signals

An adaptive DWT and DWPT signal representation is considered in this work because of its highly flexible family of signal representations that may be matched to a given signal and it is well applicable to the task of audio data compression. In this case the audio signal will be divided into overlapping frames of length 2048 samples. [3] When designing the wavelet decomposition considered some restrictions to have compact support wavelets, to create orthogonal translates and dilates of the wavelet (the same number of coefficients than the scaling functions), and to ensure regularity (fast decay of coefficients controlled by

choosing wavelets with large number of vanishing moments). The DWT will act as anorthornormal linear transform. The wavelet transform coefficients are computed recursively using an efficient pyramid algorithm. In particular, the filters given by the decomposition are arranged in a tree structure, where the leaf nodes in this tree correspond to sub bands of the wavelet decomposition. This allows several choices for a basis. This filter bank interpretation of the DWT is useful to take advantage of the large number of vanishing moments. [3]

Wavelets with large number of vanishing moments are useful for this audio compression method, because if a wavelet with a large number of vanishing moments is used, a precise specification of the pass bands of each sub band in the wavelet decomposition is possible. Thus, it can be approximate the critical band division given by the auditory system with this structure and quantization noise power could be integrated over these bands.

Wavelet packet representation

Given a wavelet packet structure, a complete tree structured filter bank is considered. Once I find the "best basis" for this application, a fast implementation exists for determining the coefficients with respect to the basis. However, in the "best basis" approach, they do not subdivide every sub band until the last level. The decision of whether to subdivide is made based on a reasonable criterion according to the application (further decomposition implies less temporal resolution). The cost function, which determines the basis selection algorithm, will be a constrained minimization problem. The idea is to minimize the cost due to the bit rate given the filter bank structure, using as a variable the estimated computational complexity at a particular step of the algorithm, limited by the maximum computations permitted. At every stage, a decision is made whether to decompose the sub band further based on this cost function. Another factor that influences this decomposition is the tradeoff in resolution. If it is decomposed further down, it will sacrifice temporal resolution for frequency resolution.

The last level of decomposition has minimum temporal resolution and has the best frequency resolution. The decision on whether to decompose is carried out top-down instead of bottom-up. Following that way, it is possible to evaluate the signal at a better temporal resolution before the decision to decompose. It is proved in this paper that the proposed algorithm yields the "best basis" (minimum cost) for the given computational complexity and range of temporal resolution.

Feature Extraction & Classification

The extracted wavelet coefficients provide a compact representation that shows the energy distribution of the signal in time and frequency. In order to further reduce the dimensionality of the extracted feature vectors, statistics over the set of the wavelet coefficients are used. That way the statistical characteristics of the "texture" or the "Environmental sound" of the piece can be represented. The distribution of energy in time and frequency for music is different for every environment. The mean of the absolute value of the coefficients in each sub band. These features provide information about the frequency distribution of the audio signal. The standard deviation of the coefficients in each sub band. These features provide information about the amount of change of the frequency distribution. Ratios of the mean values between adjacent sub bands. These features also provide information about the frequency distribution. Points on the wrong side of and as training errors. However, in proximal SVM, all the points not located on the two planes are treated as training errors. In this case the value of training error ξi in [2] may be positive or negative. The second part of the objective function in [2] uses a squared loss function instead of to capture this new notion of error

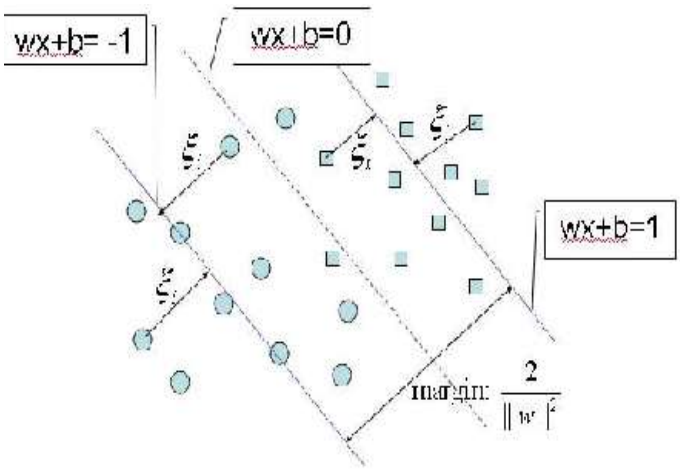

Fig. 1: Proximal SVM

TECHNIQUES

Proximal Support Vector Machine (PSVM)

The proximal SVM also uses a hyper plane as the separating surface between positive and negative training examples. But the parameter w and b are determined by solving the following problem.

$$Min\ \frac{1}{2}\left(\|w\|^2 + b^2\right) + c\sum_i \xi_i^2 \ \ldots\ldots\ldots\ldots\ldots 4$$

$$s.t.\ \forall_i y_i(w.x_i + b) + \xi_i = \ldots 5$$

The main difference between standard SVM [1] and proximal SVM [2] is the constraints. Standard SVM employs an inequality constraint where as proximal SVM employs an equality constraint. The intuition of Proximal SVM is shown in Figure 2. We can see that standard SVM only considers . We show the reason why the original proximal SVM is not suitable for classifying unbalanced data in this section. To the unbalanced data, without lose of generality, suppose the amount of positive data is much fewer than the negative data. In this case the total accumulative errors of negative data are much higher than that of positive data. Consequently, the bounding plane will shift towards the direction opposite to the negative data to produce a larger margin at the price of increasing the positive errors. Since the positive data are rare, this action will lower the value of objective function [2]. Then the separating plane will be biased to the positive data and result in a higher precision and a lower recall for the positive training data

The linear multicategory proximal support vector machine (MPSVM)

To motivate our MPSVM we begin with a brief description of the 2-category proximal support machine formulation (Fung & Mangasarian, 2001). We consider the problem, depicted in figure 1, of classifying m points in the n-dimensional real space , represented by the m × n matrix A, according to membership of each point Ai in the class A+ or A− as specified by a given m × m diagonal matrix D with plus ones or minus ones along its diagonal. For this problem, the proximal support vector machine (Fung & Mangasarian,2001) with a linear kernel is given by the following quadratic program with parameter v > 0 and linear equality constraint:

$$\frac{\nu}{2}\|y\|^2 \ \| \left\|\begin{bmatrix} w \\ \gamma \end{bmatrix}\right\|^2 \ldots\ldots\ldots 6$$

RESULTS

DATASET

The database for the experiments contains 400 samples which are taken from television broadcast database. The recordings are categorized into general classes according to common characteristics of the scenes (100 flute, 100 guitar, 100 violin, 100 piano) and events The categorization of the scenes was somewhat ambiguous, some of the recordings are associated with more than one higher-level class. The recordings are manually labelled and are separated into 1-second, 2-second and 3-second fragments. Every sound signal was stored with some properties that are also the initial conditions and criteria for the well-functioning of the algorithm. The sample database is split into training sets and test sets. In this work on randomly select 80% sounds of each class for the training set. The remaining 20% sounds form the test set. It is have taken different proportion of samples based on class dependency in each category as shown in table.

Category of Musical Instruments sound	No. Of Samples
Flute	96%
Guitar	95%
Violin	94%
Piano	92%

PREPROCEESING

The database is collected from the Television broadcast database. A window size of 16000 samples at 16KHz sampling rate with hop size of 1 second which is used as input for the feature extraction. The training data's are segmented into fixed length overlapped frame (in our experiment 20 ms frames with 10 ms overlapping is used).Since a 16KHz sampling rate is deployed, 20 ms frames consists of 320 values which are converted into 6 dimension for one frame. Here 400 clips used for training data, 40 clips for testing data and each clips must be mono channel.

Fig. 5.2.3: Result displayed

ACOUSTIC DATABASE DESCRIPTOR

CONCLUSION

In this work, "Musical instrument classification" using PSVM modeling techniques and DWT features are extracted to model the music instrument. Features for music instrument are extracted and those models were trained successfully. Music from four different instruments were modeled Using PSVM. in this work,400 database were chosen from television broad caste data, which is considered for training and testing 300 music samples are trained and testing for 100 samples data.

The characteristic of the sound signal collected from the television broad caste database were analyzed. PSVM shows an accuracy of 85% for Flute, 90% for guitar, 88% for violin, 91% for piano ,The results shows the overall performance of accuracy 88.5% using multi class PSVM.

CONFLICT OF INTEREST
There is no conflict of interest.

ACKNOWLEDGEMENTS
None

FINANCIAL DISCLOSURE
None.

REFERENCES

[1] C Joder,S Essid and G Richard, [2009] Temporal integration for audioclassification with application to musical instrument classification, IEEE Trans. Audio, Speech, Lang. Process. 17(01): 174–186.

[2] T Kitahara, M Goto, K Komatani, T Ogata, and H G Okuno, [2007] Instrument identification in polyphonic music: Feature weighting to minimize influence of sound overlaps, EURASIP J. Appl. Signal Process. 1–15.

[3] P Herrera-Boyer, A Klapuri and M Davy, [2006] Automatic classification of pitched musical instrument sounds, in Signal Processing Methods for Music Transcription, A.KlapuriandM.Davy,Eds. NewYork NY, USA: Springer. 163–200

[4] S Essid, G Richard, and B David, [2006] Instrument recognition in poly-phonic music based on automatic taxonomies, IEEE Trans. Audio,Speech, Lang. Process. 14 (01): 68–80.

[5] A A Livshin and X Rodet, [2004] Musical instrument identification in continuous recordings , in Proc. Int. Conf. Digital Audio Effects, Naples, Italy.
[6] T Kitahara, M Goto, K Komatani, T Ogata and H G Okuno, [2006] Musical instrument recognizer “instrogram” and its application to music retrieval based on instrumentation similarity, in Proc. IEEE Int. Symp. Multimedia. 265–274.
[7] A Eronen, [2001] Comparison of features for musical instrument recognition, in Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. 19–22.
[8] G Agostini, M Longari, and E Poolastri, [2003] Musical instrument timbres classification with spectral features, EURASIP J. Appl. Signal Process. 1: 5–14.
[9] A Eronen and A Klapuri, [2000] Musical instrument recognition usingcepstral coefficients and temporal features, in Proc. IEEE Int. Conf Acoust Speech, Signal Process. 753–756.
[10] S. Essid, G. Richard, and B. David, [2006] “Musical instrument recognition bypairwise classification strategies,” IEEE Trans. Audio, Speech, Lang.Process., vol. 14, no. 4, pp. 1401–1412.
[11] S. Essid, G. Richard, and B. David, [2006] “Musical instrument recognitionby pairwise classification strategies,” IEE Trans. Audio, Speech, Lang.Process., vol. 14, no. 4, pp. 1401–1412.
[12] KOSTEK B., CZYZEWSKI A., [2001] "Automatic Recognition of Musical Instrument Sound Further Developments", 110th Audio Eng. Soc. Conv., Amsterdam, 12-15.
[13] T. Kitahara, M. Goto, and H.G. Okuno,[2004] “Category-level identification of non-registered musical instrument sounds,” in International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Montreal, Canada.
[14] Z. Zhang and B. Schuller, [2012] “Semi-supervised learning helps in sound event classification," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 333-336, IEEE.
[15] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fis-sore, P. Laface, A. Mertins, C. Ris, R. Rose, C. Tyagi, and C. Wellekens, [2007]“Automatic speech recognition and speech variability: A review," Speech Communi-cation, vol. 49, no. 10-11, pp. 763,786.