Musical Instrument Classification Based on Nonlinear Recurrence Analysis and Supervised Learning

Rui RUI, Changchun BAO

Speech and Audio Signal Processing Lab, School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China

rr_2006@bjut.edu.cn, baochch@bjut.edu.cn

Abstract. *In this paper, the phase space reconstruction of time series produced by different instruments is discussed based on the nonlinear dynamic theory. The dense ratio, a novel quantitative recurrence parameter, is proposed to describe the difference of wind instruments, stringed instruments and keyboard instruments in the phase space by analyzing the recursive property of every instrument. Furthermore, a novel supervised learning algorithm for automatic classification of individual musical instrument signals is addressed deriving from the idea of supervised non-negative matrix factorization (NMF) algorithm. In our approach, the orthogonal basis matrix could be obtained without updating the matrix iteratively, which NMF is unable to do. The experimental results indicate that the accuracy of the proposed method is improved by* 3% *comparing with the conventional features in the individual instrument classification.*

Keywords

Phase space reconstruction, recurrence analysis, dense ratio, supervised learning, musical instrument classification.

1. Introduction

Nowadays, the classification of musical instrument has become an interesting research field. It could be treated as the first step in music information retrieval (MIR) systems. The technique of machine learning for signal processing enhanced the classification study on musical instrument [1], [2].

As we know, music signals are usually produced by the instruments. Over the last decade, there has been a great deal of work on musical instrument classification. A. Eronen adopted Mel-frequency cepstral coefficients (MFCC), spectral and temporal features for instrument classification, and the identification of 35% for 29 instruments classes was achieved [3]. B. Kostek trained wavelet features and MPEG-7 descriptors with multilayer neural networks and made the mean recognition rate up to 70% for 12 instruments [4]. Recently, E. Benetos et al. utilized a supervised non-negative matrix factorization (NMF) algorithm yielding a correct classification rate of 95.2% for 6 instrument classes [5]. But, the basis matrices extracted by NMF are not orthogonal. Consequently, an improved method was presented to perform Gram-Schmite (GS) orthogonalization on the basis matrix by utilizing QR decomposition [6]. Experimental results demonstrated that the improved method outperformed the supervised NMF algorithm.

Although good performance has been achieved in musical instrument classification, there is a problem which has not been solved, i.e. the wrong classification between the different instrumental families often occurs. For instance, piano is classified as oboe, clarinet and oboe are classified as cello, guitar is classified as trumpet and piano, etc. [3], [7]. These phenomena do not accord with the auditory system of human ears. The reason is that the conventional features cannot capture the unique properties of instruments.

With the deeper studies of nonlinear dynamics theory, the nonlinear signal analysis has got an extensive application in the field of audio signal processing. It has been proved that the time series of audio signals, including musical instrument signals, obviously have the typical nonlinear characteristics [8], [9]. Therefore, this paper presents a new idea, i.e. we can apply the concept of nonlinear dynamics into the musical instrument classification. More concretely, musical instruments usually have some unique properties that can be revealed by the phase space reconstruction of the time series produced by the instruments. Furthermore, a novel recurrence parameter, dense ratio, is proposed to describe the difference of wind instruments, stringed instruments and keyboard instruments in the phase space by analyzing the recurrence characteristics of all instruments.

In addition, a novel supervised learning algorithm which could obtain the orthogonal basis matrix without updating the matrix iteratively is addressed. For musical instrument classification, each class is trained individually to gain the orthogonal basis matrix and save it. Afterwards, the test data are projected onto each trained basis matrix. Feature selection for varying dimensions is also considered. Moreover, nearest neighbors (NN), Gaussian mixture model (GMM) and radial basis function (RBF) networks have been employed for classification and their performance is evaluated. The results indicate that the classification

accuracy of the proposed method is comparable to the performance of supervised NMF algorithm using in [6] for the same experiments.

The paper is organized as follows. The nonlinear dynamic theory that will be used for musical instrument classification is introduced in Section 2. The recurrence characteristics analysis of musical instrument signals is given in Section 3. The proposed method of supervised learning algorithm is described in Section 4. Feature extraction and selection are briefly presented in Section 5. The experimental results for musical instrument classification are shown in Section 6. Finally, the conclusions are drawn in Section 7.

2. Nonlinear Dynamic Theory

Nonlinear dynamics theory has recently been adopted as a new nonlinear approach for audio signal processing. The nonlinear characteristics analysis in audio signals can be employed by the phase space reconstruction. Then, the basis concept of phase space reconstruction will be given.

2.1 Phase Space Reconstruction

The fundamental concept of nonlinear dynamics is phase space reconstruction. Each state of the dynamics is represented unambiguously by one point in a multi-dimension space.

According to the phase space reconstruction method proposed by F. Takens [10], the one-dimension nonlinear time series of audio signal $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_K)^{\mathrm{T}}$, where T denotes transposition, can be reconstructed to a state matrix $\mathbf{Y}$ which is given by:

$$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N) = \begin{pmatrix} x(1) & x(2) & \cdots & x(N) \\ x(1+\tau) & x(2+\tau) & \cdots & x(N+\tau) \\ \vdots & \vdots & & \vdots \\ x(1+(m-1)\tau) & x(2+(m-1)\tau) & \cdots & x(N+(m-1)\tau) \end{pmatrix} \quad (1)$$

where τ is the time delay and m is the embedding dimension. $N = K - (m - 1)\tau$ denotes the number of phase points in the phase space. And $\mathbf{y}_i$ is one of points in m-dimension phase space which represents the i^{th} state of the system. Notice that for $m = 1$, equation (1) reduces $\mathbf{Y}$ to the signal $\mathbf{x}$.

2.2 Parameter of Phase Space Reconstruction

It is seen that two parameters need to be predefined firstly for the phase space reconstruction. They are the time delay τ and the embedding dimension m. In this paper, τ and m are calculated by autocorrelation method and false nearest neighbor method [11], respectively. For each kind of instruments, τ and m are all different because the distributions of them differ from each other in the structure of phase space. However, it is unrealistic to compute τ and m for large data in the study of pattern recognition. Therefore, the statistical method is adopted by calculating the probability of the values of τ and m for each frame. Then, the optimal values of τ and m are taken by finding the position of the maximum probability. Finally, we set them manually as $m = 6$, $\tau = 9$.

2.3 Phase Space Reconstruction of Musical Instrument Signals

According to the nonlinear dynamic theory, the instrumental signals are reconstructed in the phase space and projected onto a three-dimension space in order to analyze the characteristics of signals visually. The temporal waveform and three-dimension phase space trajectory of three common instruments are depicted in Fig. 1, where Fig. 1(a), Fig. 1(b) and Fig. 1(c) are the temporal waveforms of clarinet, cello and piano, and Fig. 1(d), Fig. 1(e) and Fig. 1(f) are the corresponding phase space trajectories. From the figures of temporal waveforms, we can see that clarinet, cello and piano have the characteristics of quasi-periodic signals obviously, whereas it is difficult to distinguish them from each other by the temporal waveforms. However, it can be seen from the three-dimension phase space that the trajectories of clarinet have strong property of periodic recursion. It is very easy to find the number of samples T which is the corresponding period of the temporal signal, for example by the autocorrelation method [12]. These T points compose a complete trajectory. In the phase space, the total number of trajectories can be expressed as n which is calculated by $n = \lfloor N / T \rfloor$, and the trajectories of clarinet are close to each other. Furthermore, the trajectories of cello in the three-dimension phase space have some property of periodic recursion, but each trajectory is incompact. In the phase space of piano, the trajectories are randomly distributed, while the all trajectories are in a relative steady area.

The different working principles of wind instrument (include woodwind and brass), string instrument and keyboard instrument lead to different distribution in the phase space. It is to be noticed that the changes of pitch, rhythm and style by players will impact the distribution in the phase space, but the characteristics of phase space trajectories of wind instruments, string instruments and keyboard instruments are still similar with Fig. 1(d), Fig. 1(e) and Fig. 1(f), respectively. The aforementioned discoveries give us an inspiration that it is possible to classify the different instrument families by the nonlinear dynamics theory.

3. Recurrence Characteristics Analysis of Musical Instrument Signals

3.1 Recurrence Plot

Eckmann [13] et al. have introduced a tool which enables us to investigate the m-dimension phase space trajectory through a two-dimension representation of its recur-

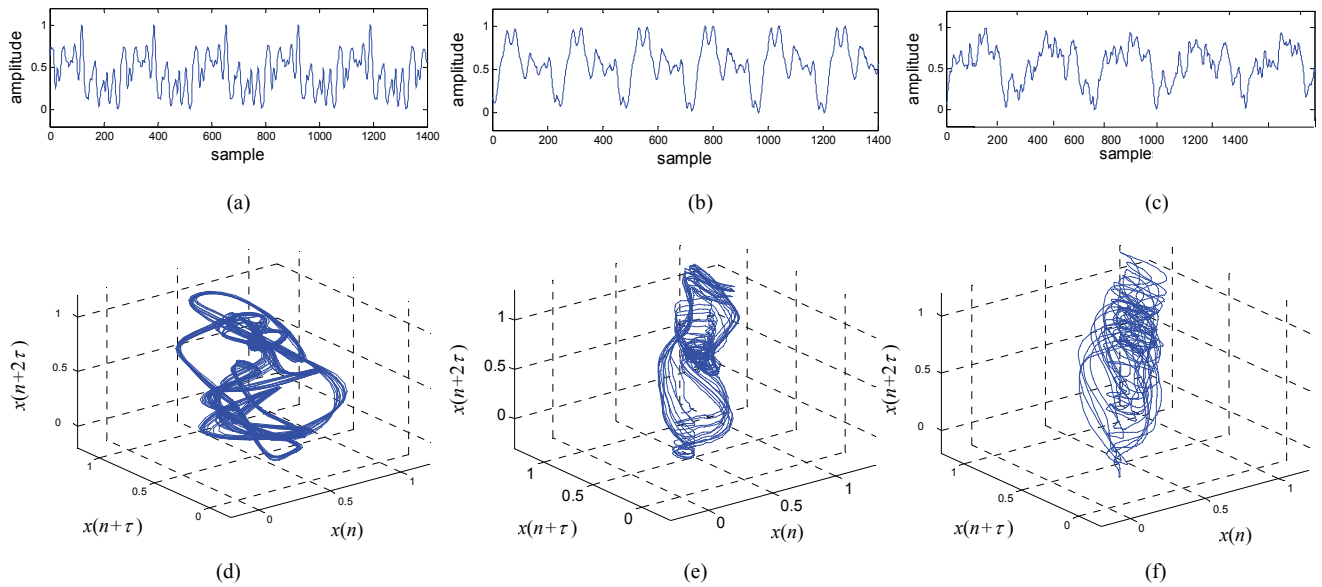

Fig. 1. The temporal waveforms and three-dimension phase space trajectories of clarinet, cello and piano.

rences. This representation is called recurrence plot (RP) and it can be mathematically expressed as

$$r_{i,j} = \Theta\left(\varepsilon - \left\|\mathrm{y}_i - \mathrm{y}_j\right\|\right), \quad i, j = 1, \cdots N \tag{2}$$

where N is the number of the considered states y_i, ε is a predefined threshold, $\|\bullet\|$is the norm (e.g. the Euclidean norm) and $\Theta(\bullet)$ is the Heaviside function, which is defined as

$$\Theta(z) = \begin{cases} 1 & z \geq 0 \\ 0 & z < 0 \end{cases} \tag{3}$$

RP intuitively represents the m-dimension phase space trajectory of the dynamical system through a two-dimension figure, which reveals the variation regularity of the internal structure for the system.

The recurrence of a state at time i with regard to time j is represented as a two-dimension square matrix with black and white points. If the vectors $\mathbf{y}_i$ and $\mathbf{y}_j$ are falling into a region whose centre is $\mathbf{y}_i$ and radius is ε, $r_{i,j}$ is represented as a black point in terms of coordinate (i, j) in the RP. On the contrary, if the vectors $\mathbf{y}_i$ and $\mathbf{y}_j$ are not falling into this region, $r_{i,j}$ is represented as a white point in terms

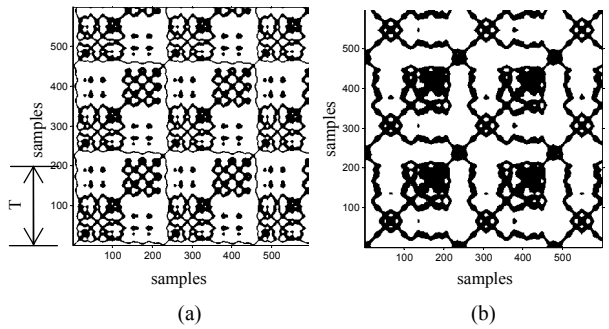

Fig. 2. Examples of the recurrence plot obtained from clarinet (a) and cello (b). Parameters $m = 6$, $\tau = 9$, and $\varepsilon = 1.5\sigma$ are used.

clarinet and cello are given in Fig. 2, where σ is the variance of the instrument signals. Both of them have the obvious property of periodic recursion, since the distances between the lines which parallel the main diagonal line with a 45 angle are almost equal. The lag of each line is the number of samples T which corresponds to the period of temporal signal.

3.2 A Novel Recurrence Quantification Parameter

Zbilut and Webber developed the recurrence quantification analysis (RQA) [14] to quantify the afore-mentioned structures in the RP. However, these RQA parameters reveal the different properties of the random signals and the periodic signals effectively [15]. But the signals analyzed in this paper are quasi-periodic signals (shown in Fig. 2). And the traditional recurrence quantification parameters are useless to classify the different instrument families. For this problem, a novel recurrence quantification parameter is proposed in this section.

3.2.1 Dense Ratio

In order to describe the difference among the instrument families, a novel recurrence quantification parameter is developed by analyzing the recursive property of every instrument family. This parameter is named as dense ratio (DR), since it represents the density of every trajectory distributed in the phase space. DR is defined by

$$DR = \frac{2}{n(n-1)T} \sum_{i=1}^{T} \sum_{\substack{j,k=0 \\ j<k}}^{n-1} r_{i+jT,\ i+kT} \tag{4}$$

where T is the number of samples corresponding to the period of the temporal signals and $n = \lfloor N / T \rfloor$. If the distance between the vectors $\mathbf{y}_{i+jT}$ and $\mathbf{y}_{i+kT}$ is close, the value of $r_{i+jT,\ i+kT}$ will be 1, and the position $(i+jT, i+kT)$ in the RP

is represented as a black point. All the points of $(i + jT, i + kT)$, $i = 1,\ldots,T$, j, $k = 1,\ldots,n - 1$, and $j \neq k$, compose the lines which parallels the main diagonal line with a 45 angle. Thus, DR implies the percentage that the coordinates $(i + jT, i + kT)$ are represented by black points in the RP. From the definition of RP, we know that $r_{i,j}$ is symmetrical with respect to the main diagonal line. So the area of $i < j$ in the RP is considered for reducing half computational quantity.

3.2.2 Discussion of the Threshold ε

The capability of DR for distinguishing three instrument families depends on the value of the threshold ε. If ε is too large, $r_{i+jT,\, i+kT}$ of all musical instrument signals is represented as a black point in the RP. And if ε is too small, $r_{i+jT,\, i+kT}$ of all musical instrument signals is represented as a white point in the RP. Therefore, the choice of ε is so important that we predefined it from a range of $\varepsilon \in \{0.1\sigma, 0.2\sigma, 0.3\sigma, 0.4\sigma, 0.5\sigma, 0.6\sigma\}$, where σ is the variance of the instrument signals. Finally, $\varepsilon = 0.3\sigma$ is chosen by the experimental results.

3.2.3 Classification of Instrument Families Using Dense Ratio

RP of note a4 played by clarinet, cello and piano are depicted in Fig. 3. The recurrence parameters are chosen as $m = 6$, $\tau = 9$, and $\varepsilon = 0.3\sigma$. In Fig. 3(a), the black points, that is $r_{i+jT,i+kT}$, compose the lines which parallel the main diagonal line. It is seen that the trajectories in the phase space of wind instruments are dense and the distance of each trajectory is less than the threshold ε. Therefore, the positions paralleling the main diagonal line are all the black points. In Fig. 3(b), the trajectories in the phase space of stringed instruments are incompact. So, there are numerous isolated recurrence points distributed in the position of paralleling the main diagonal line. This indicates that the distance of each trajectory is fluctuant around the threshold ε. In Fig. 3(c), the trajectories in the phase space of piano are randomly distributed. There are few black points in the positions of paralleling the main diagonal line. This can be explained that the distance of each trajectory is more than the threshold ε.

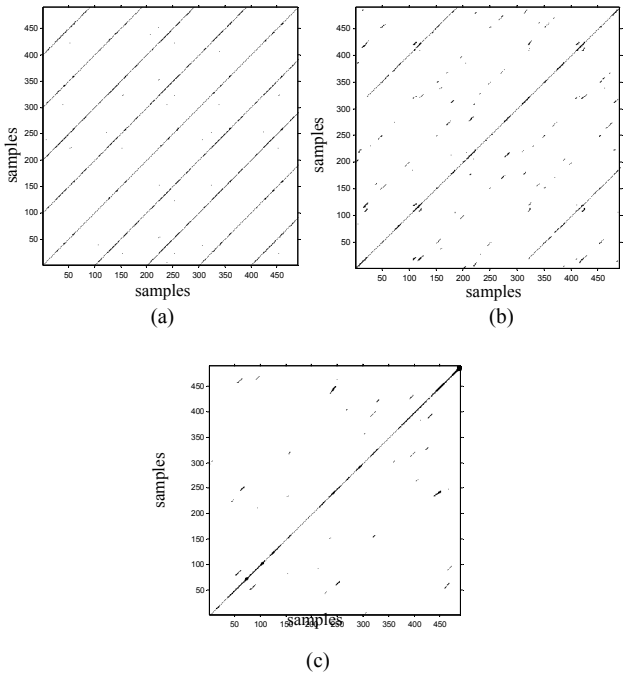

Fig. 3. RP of a4 played in clarinet, cello and piano. Parameters $m = 6$, $\tau = 9$, and $\varepsilon = 0.3\sigma$ are used.

The larger the DR is, the denser the trajectories in the phase space are. It is certified that the DR of wind instruments has the maximum value which is from 0.8 to 1, the DR of piano has the minimum value which is below 0.3, and the DR of stringed instruments has a value which is from 0.3 to 0.7. Thus, three instrument families are distinguished by DR effectively.

4. Supervised Learning Algorithm for Musical Instrument Classification

In this section, a novel supervised learning algorithm will be addressed inspiring by the idea of supervised NMF algorithm. Then, the optimization of the feature subsets will be also discussed. First of all, we will review the supervised NMF algorithm.

4.1 Review of Supervised NMF Algorithm

For a given non-negative $n \times m$ matrix $\mathbf{V}$ (can be regarded as the musical instrument features consisting of n vectors of dimension m), NMF finds the non-negative $n \times r$ matrix $\mathbf{W}$ (basis matrix) and non-negative $r \times m$ matrix $\mathbf{H}$ (encoding matrix) in order to approximate the matrix $\mathbf{V}$ as [16]:

$$\mathbf{V} \approx \mathbf{WH}. \tag{5}$$

Usually, r is chosen so that $(n + m)\, r < nm$. To find an approximate factorization in (5), Kullback-Leibler divergence between $\mathbf{V}$ and $\mathbf{WH}$ is used frequently, and the optimization problem can be solved by the iterative multiplicative rules. But, the basis vectors defined by the columns of matrix $\mathbf{W}$ are not orthogonal. Thus, QR decomposition was utilized on $\mathbf{W}$ in [7], that is $\mathbf{W} = \mathbf{QR}$, where $\mathbf{Q}_{n\times r}$ is an orthogonal matrix and $\mathbf{R}_{r\times r}$ is an upper triangular matrix. At this time,

$$\mathbf{V} = \mathbf{QH}'. \tag{6}$$

$\mathbf{V}$ can be written as a linear combination between an orthogonal basis and a new encoding matrix, where $\mathbf{Q}$ contains the orthogonal basis and $\mathbf{H}' = \mathbf{RH}$ becomes the new encoding matrix. This method, however, costs a mass of computation for updating $\mathbf{W}$ and $\mathbf{H}$ iteratively and QR decomposition. Thus, a novel supervised learning algorithm is proposed in the next part.

4.2 A Novel Supervised Learning Algorithm

We assume that

$$\mathbf{X}=(\mathbf{x}_1,\mathbf{x}_2,\cdots\mathbf{x}_m)=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}, n>>m \quad (7)$$

stands for the musical instrument features. An auxiliary matrix **F** is constructed as follows:

$$\mathbf{F}=\mathbf{X}^{\mathrm{T}}\mathbf{X}. \quad (8)$$

The form of normalization of **X** is:

$$\mathbf{X}^{*}=(\mathbf{x}_1^{*},\cdots,\mathbf{x}_m^{*})=\left(\frac{\mathbf{x}_1-E(\mathbf{x}_1)}{\sqrt{D(\mathbf{x}_1)}},\cdots,\frac{\mathbf{x}_m-E(\mathbf{x}_m)}{\sqrt{D(\mathbf{x}_m)}}\right) \quad (9)$$

where $E(\mathbf{x}_i)$ and $D(\mathbf{x}_i)$ represent the mean and variation of i^{th} feature, respectively. Consequently, (8) becomes a correlation coefficient matrix (CCM) which is given by:

$$\mathbf{R}=(\mathbf{x}_i^{*})^{\mathrm{T}}(\mathbf{x}_j^{*})=(\rho_{ij}),\ i,j=1,2,\cdots,m, \quad (10)$$

where

$$\rho_{ij}=\left(\frac{x_i-E(x_i)}{\sqrt{D(x_i)}}\right)^{\mathrm{T}}\left(\frac{x_j-E(x_j)}{\sqrt{D(x_j)}}\right) \quad (11)$$

are the cross-correlation coefficients between $\mathbf{x}_i$ and $\mathbf{x}_j$.

It should be noted that matrix **R** satisfies: $\mathbf{R}^{\mathrm{T}}=\mathbf{R}$. As we know, a real symmetrical matrix is consequentially diagonalized. So, there exists a $m \times m$ orthogonal matrix **U** such that:

$$\mathbf{U}^{\mathrm{T}}\mathbf{R}\mathbf{U}=\mathbf{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_{\mathrm{m}}) \quad (12)$$

where **Λ** is a diagonal matrix, its elements $\lambda_1, \lambda_2, \ldots, \lambda_m$ are the eigenvalues of **R**, and $\alpha_1, \alpha_2, \ldots, \alpha_m$ are the corresponding eigenvectors.

We suppose that:

$$\mathbf{H}=\mathbf{X}\mathbf{U} \quad (13)$$

where $\mathbf{U}=(\alpha_1, \alpha_2, \ldots, \alpha_m)$ is an orthogonal matrix and $\mathbf{U}^{-1}$, inverse of **U**, is also an orthogonal matrix. According to (13), the original data matrix X can be decomposed by:

$$\mathbf{X}=\mathbf{H}\mathbf{U}^{-1}. \quad (14)$$

Since matrix **U** has the property of $\mathbf{U}^{-1}=\mathbf{U}^{\mathrm{T}}$, (14) can be written by the transposition as:

$$\mathbf{X}^{\mathrm{T}}=\mathbf{U}\mathbf{H}^{\mathrm{T}}. \quad (15)$$

This representation is very similar to (6), where **U** contains the basis vectors and the column vectors of $\mathbf{H}^{\mathrm{T}}$ contains the weights which approximate the corresponding column of matrix $\mathbf{X}^{\mathrm{T}}$ as a linear combination of the columns of **U**.

In the problem of classification, **X** is regarded as the features extracted from the original data. The creation of the supervised learning method is performed for each data class individually as:

$$\mathbf{H}_i=\mathbf{X}_i\mathbf{U}_i,\quad i=1,2,\cdots N, \quad (16)$$

where N is the number of different classes, $\mathbf{X}_i$ is the features of class i and $\mathbf{U}_i$ is the orthogonal basis matrix for each class. It is clear that this method does not need any iteration for training.

During the test procedure, each test sequence is represented by the feature vector $\mathbf{x}_{test}$. Afterwards, $\mathbf{x}_{test}$ is projected onto basis matrix $\mathbf{U}_i$ of each class:

$$\mathbf{h}_{test}^{(i)}=\mathbf{x}_{test}\mathbf{U}_i. \quad (17)$$

For each class, the vector $\mathbf{h}_{test}$ is compared to each column of $\mathbf{H}_i$ by the cosine similarity measure (CSM). The vector which maximizes the CSM of $\mathbf{H}_i$ is computed as a measure of similarity for the class:

$$CSM_i=\max_{j=1,2,\cdots,r}\frac{\mathbf{h}_{test}^{(i)\,T}\mathbf{h}_j^{(i)}}{\left\|\mathbf{h}_{test}^{(i)}\right\|\left\|\mathbf{h}_j^{(i)}\right\|} \quad (18)$$

where $\mathbf{h}_j^{(j)}$ represents the j^{th} column of matrix $\mathbf{H}_i$, and the class label of the sequence is determined by the maximum CSM_i:

$$k=arg\max_{i=1,2,\cdots,N} CSM_i. \quad (19)$$

The flow chart of the proposed supervised learning algorithm is described in Fig. 4, where the training phase is linked by solid line and the testing phase is linked by dashed line.

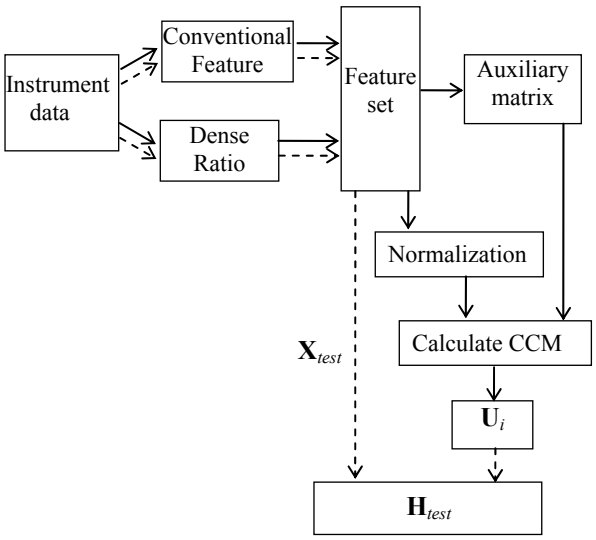

Fig. 4. Training and testing procedure of the proposed supervised learning algorithm.

5. Feature Extraction and Selection

The musical instrument signals from the instruments are windowed into frames of 30 ms which are half overlapped with each other. Four groups of features, listed in Tab. 1, are used in our feature analysis. They are the eight perceptual features, eighty four cepstral features, seven timbre features and one recurrence quantification parameter [8]. The mean and standard deviation of these features are also employed. This results in 200 features in total. Here, the instrument discriminative information is quantified using Fisher's F-ratio in each MFCC frequency region (More information can be found in [9]). We improve the resolution in those frequency regions with high F-ratio values and the ninety eight sub-band filters are used to form 42 MFCC.

Features	Abbreviation	Dimension	Types
Zero-crossing rate	ZCR	1	Perceptual features
Sub-band energy	SE	4	
Root mean square	RMS	1	
Bandwidth	B	1	
Spectral flux	SF	1	
Mel-frequency cepstral coefficients	MFCC	42	Cepstral features
The first time derivatives of MFCC	Δ MFCC	42	
Harmonic Spectral Centroid	HSC	1	Timbre features
Harmonic Spectral Deviation	HSD	1	
Harmonic Spectral Spread	HSS	1	
Harmonic Spectral Variation	HSV	1	
Spectral Centroid	SC	1	
Log Attack Time	LAT	1	
Temporal Centroid	TC	1	
Dense Ratio	DR	1	Recurrence Quantification Parameter

Tab. 1. Feature descriptions.

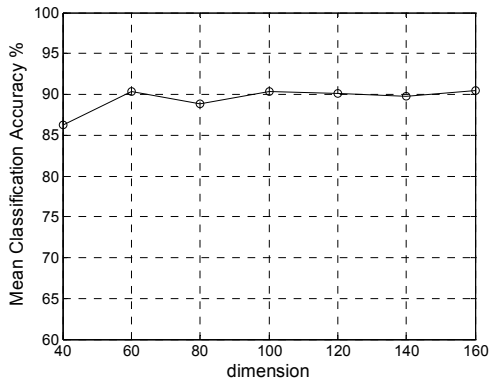

Fig. 5. Performance of several feature subsets used in musical instrument classification.

In order to reduce the dimension of features, a discriminative feature subset should be chosen. The rule of selection is to maximize [6]:

$$J = tr\left(\mathbf{S}_w^{-1}\mathbf{S}_b\right) \tag{20}$$

where $tr(\cdot)$ stands for the trace of a matrix, $\mathbf{S}_w$ is the within-class scatter matrix and $\mathbf{S}_b$ is the between-class scatter matrix.

The details of feature subset selection can be found in [17]. Using the subsets of sizes 40, 60, 80, 100, 120, 140, and 160, respectively, to estimate the classification accuracy, depicted in Fig. 5, we can see that the set of 60 should be considered as the most suitable feature subset for musical instrument classification.

6. Experiments

6.1 Database

Audio files from CD collections in terms of mono format are sampled at 44.1 kHz. Overall three hours audio data, 127 recordings, contain 11 different instrument classes: cello, violin, guitar, clarinet, oboe, saxophone, flute, trumpet, horn, trombone, and piano. The specific length of each instrument is shown in Tab. 2. The recordings contain sound segments instead of isolated instrument tones. Each test sequence has the duration of about 4 seconds. The classification accuracy is estimated using the leave-one-out method [11] which can achieve a least bias evaluation. When using the leave-one-out method, the learning algorithm is trained multiple times, employing all but one of the training data.

Instrument	Data(min)	Instrument	Data(min)
cello	25	flute	20
violin	23	trumpet	15
guitar	20	horn	8
clarinet	10	trombone	15
oboe	15	piano	22
saxophone	8	total	181

Tab. 2. Audio sources of instruments.

6.2 Performance Evaluation

The supervised NMF algorithm and the proposed method are utilized by aforementioned audio data. About 370 iterations for the supervised NMF algorithm are needed for convergence in the training, while the proposed method does not need any iteration.

To evaluate the performance of different algorithm, three classifiers: NN, GMM, and RBF are considered. We can better assess the accuracy improvements by several classifiers and not just by a particular one. The mean value of the classification accuracy for the two algorithms and three classifiers is shown in Fig. 6. The results indicate that two methods show a comparable performance to GMM and RBF. The NMF algorithm makes a few improvements of 0.2% than the proposed method whose mean classification accuracy is 90.4%. An obvious advantage of the proposed method is that the orthogonal basis matrix can be gained without any iteration, although there is not too much improvement in the classification accuracy. Hence, the proposed supervised learning algorithm seems to be

an effective method for the musical instrument classification. Next, the individual musical instrument classification will be discussed.

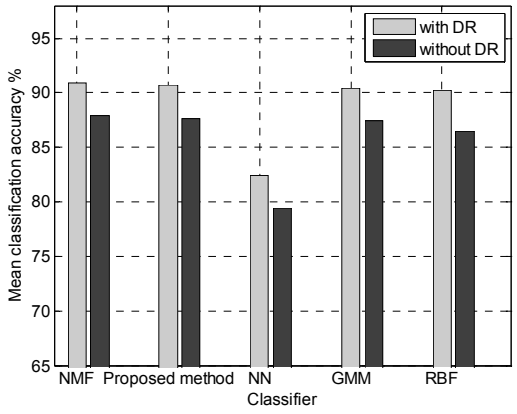

Fig. 6. Mean classification accuracy for the proposed method, supervised NMF algorithm and three classifiers.

6.3 Individual Musical Instrument Classification

Eleven instruments are directly distinguished from each other. As far as the proposed method is concerned, the results of individual instrument classification are given in Tab. 3. Cello, violin, guitar, clarinet, oboe, saxophone, flute, trumpet, horn, trombone, and piano are represented by letters of *a*-*k*, respectively. The main diagonal line gives the classification accuracy of each instrument. The data, inside the bracket, represent the results using conventional features alone. Apparently, the cases of wrong classification among different instrument families are common. For example, 6.5% of trombone sounds are classified as cello. The data, outside the bracket, represent the results combining DR. It is seen that the percentage of wrong classification among different instrument families decreases obviously while employing the proposed method. The approach based on conventional features seems to have some problems in identifying different instruments, contrarily to DR

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*
a	**97.7** (95.1)	1.4 (1.7)	0.6 (0.8)	0	0	0	0	0	0	0.3 (2.4)	0
b	0	**95.3** (92.4)	4.1 (4.4)	0	0	0.4 (1.8)	0	0	0	0	0.2 (1.4)
c	4.5 (4.5)	0	**93.6** (90.2)	0	0	0	0	0	0.9 (2.5)	0	1.0 (2.8)
d	0.9 (5.4)	0	0	**80.6** (75.1)	0	9.4 (10.3)	4.8 (4.9)	0	0	4.3 (4.3)	0
e	0.8 (4.0)	0	0	3.8 (3.8)	**87.5** (81.7)	0	2.7 (2.7)	0	0	5.2 (5.2)	0 (2.6)
f	0	0.6 (6.9)	0	4.6 (4.6)	0	**88.6** (82.2)	4.0 (4.1)	0	2.2 (2.2)	0	0
g	0	0	0	5.6 (6.0)	1.9 (2.7)	4.3 (4.3)	**85.2** (84.0)	0	3.0 (3.0)	0	0
h	0	0	0	0	0	2.2 (2.2)	0	**91.2** (91.2)	2.6 (2.6)	3.9 (3.9)	0
i	0	0	0	3.8 (3.8)	0	0	0.9 (0.9)	4.3 (4.3)	**84.7** (84.7)	6.3 (6.3)	0
j	0.4 (1.5)	0	0	0	0	0	2.3 (2.3)	3.2 (3.2)	3.3 (3.3)	**89.8** (87.6)	0.9 (2.0)
k	0	0	0	0	0 (2.4)	0	0	0	0	0	**100** (97.6)

Tab. 3. Confusion matrix for 11 instruments (in percentage).

descriptor which is very efficient means. Especially this may be observed in the cases of wrong instruments classification, such as piano, where its wrong classification rate is less than 1%. The overall average classification accuracy is 90.4%. The experimental results indicate that the accuracy of the proposed method is improved by 3% in the individual instrument classification.

7. Conclusions

In this paper, the unique properties of different instrument families are revealed by the phase space reconstruction based on the nonlinear dynamic theory. The dense ratio, a novel quantitative recurrence parameter, is proposed to describe the difference between wind instruments, stringed instruments and keyboard instruments in the phase space by analyzing the recursive property of each instrument. In addition, a novel supervised learning algorithm for classifying musical instrument classification using a traditional features and dense ratio is proposed. The performance of the proposed method is comparable to the supervised NMF algorithm. The most important advantage of the proposed method is that the orthogonal basis matrix can be obtained without any iteration. The experimental results indicate that the accuracy of the proposed method is higher than the conventional features in the individual instrument classification.

Future work will focus on how to decrease the misclassification from other families and improve the accuracy of musical instrument from the same family.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 61072089), Beijing Natural Science Foundation Program and Scientific Research Key Program of Beijing Municipal Commission of Education (No. KZ201110005005), and the Funding Project for Academic Human Resources Development in Institutions of Higher Learning under the Jurisdiction of Beijing Municipality.

References

[1] EVERY, M. R. Discriminating between pitched sources in music audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 2008, vol. 16, no. 2, p. 267 - 277.

[2] ZLATINTSI, A., MARAGOS, P. AM-FM modulation features for music instrument signal analysis and recognition. In *Proc. 20th European Signal Processing Conference*. 2012, p. 2035 - 2039.

[3] ERONEN, A. Comparison of features for musical instrument recognition. In *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.* 2001, p. 19 - 22.

[4] KOSTEK, B. Musical instrument classification and duet analysis employing music information retrieval techniques. *Proc. of IEEE*, 2004, vol. 92, no. 4, p. 712 - 729.

[5] BENETOS, E., KOTTI, M., KOTROPOULOS, C. Musical instrument classification using non-negative matrix factorization algorithms and subset feature selection. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP*. 2006, p. V221 - V224.

[6] BENETOS, E., KOTTI, M., KOTROPOULOS, C. Large scale musical instrument identification. In *Proc. of 4th Sound and Music Computing Conf.* 2007, p. 283 - 286.

[7] BARBEDO, J. G. A., TZANETAKI, G. Musical instrument classification using individual partials. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, vol. 19, no. 1, p. 111 - 122.

[8] SHA, Y.-T., BAO, CH., JIA, M.-S., LIU, X. High frequency reconstruction of audio signal based on chaotic prediction theory. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. 2010, p. 381 - 384.

[9] SERRÀ, J., DE LOS SANTOS, C. A., ANDRZEJAK, R. G. Nonlinear audio recurrence analysis with application to genre classification. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. 2011, p. 169 - 172.

[10] TAKENS, F. Detecting strange attractors in turbulence. *Lecture Notes in Math.*, 1981, vol. 898, p. 366 - 381.

[11] KANTZ, H., SCHREIBER, T. *Nonlinear Time Series Analysis*. 2nd edition. Cambridge (UK): Cambridge University Press, 2004.

[12] KITAHARA, T., GOTO, M., KOMATANI, K., OGATA, T., OKUNO, H. G. Musical instrument recognizer instrogram and its application to music retrieval based on instrumentation similarity. In *Proc. IEEE International Symposium on Multimedia*. 2006, p. 265 - 272.

[13] ECKMANN, J.-P., OLIFFSON KAMPHORST, S., RUELLE, D. Recurrence plots of dynamical systems. *Europhys. Lett*., 1987, vol. 4, p. 973 - 977.

[14] ZBILUT, J. P., WEBBER JR., C. L. Embeddings and delays as derived from quantification of recurrence plots. *Phys. Lett. A*, 1992, vol. 171, p. 199 - 203.

[15] ZHANG, L., BAO, CH., LIU, X., ZHANG, X., BAO, F., BU, B. Audio classification algorithm based on nonlinear characteristics analysis. In *Proc. Asia Pacific Signal and Information Processing Association Annual Summit and Conference*. 2011, p. 214 - 217.

[16] LEE, D. D., SEUNG, H. S. Algorithm for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, 2001, vol. 13, p. 556 - 562.

[17] VAN DER HEIJDEN, F., DUIN, R. P. W., DE RIDDER, D., TAX, D. M. J. *Classification, Parameter Estimation and State Estimation: An Engineering Approach Using MATLAB*. London (UK): Wiley, 2004.

About Authors ...

Rui RUI received the B.S. degree in the Dept. of Electronic Information Engineering, Beijing Union University in 2005. She is currently pursuing the Ph. D degree in the Dept. of Circuit and System, Beijing University of Technology. Her research mainly focuses on musical instrument classification and music information retrieval.

Chang-chun BAO received the B.S., degree in Telecommunication Engineering from Chang Chun Inst. of Posts and Telecommunications, M.S., and Ph. D degrees in Communication and Electronic System from JiLin University of Technology in 1987, 1992 and 1995, respectively. From August, 1992 to June, 1993, he was a visiting scholar in the Dept. of Electrical Engineering, Tsinghua University. From December, 1995 to November, 1997, he was a Postdoctoral Research Fellow and an Associate Professor in the School of Communication Engineering, Xidian University. He joined Beijing University of Technology as an Associate Professor in November, 1997, and was promoted to a Professor from July, 1999 in the School of Electronic Information and Control Engineering. From July to September, 1998, he was a senior researcher in Digital System Technology Lab, Radio Products Research Group, Land Mobile Products Sector Motorola, Florida, USA. From March to August, 2004, he was a visiting professor at the University of Wollongong. His research interests are in the areas of speech & audio signal processing, speech coding, speech enhancement, speech transcoding, audio coding, audio enhancement, bandwidth extending for speech and audio signals and 3D audio signal processing. He has published the book "Principles of Digital Speech Coding" (Xian, China: Xidian University Press, 2007). He is an author or co-author of over 170 papers in journals and conferences and holds 7 patents. He is currently an Associate Editor for the Journal on Communications, an Editor for Signal Processing and an Editor for Journal of Data Acquisition & Processing, all in Chinese. Prof. Bao is a Board and Senior Member of Chinese Inst. of Electronics (CIE), a Board member of the Acoustical Society of China (ASC), a Board Member of Signal Processing Academy of CIE, a member of International Speech Communication Association (ISCA), Vice-Chairman of APSIPA SLA TC and Vice-Chairman of National Conference on Man-Machine Speech Communication-Standing Committee in China.