A New Method Based on Local Integral Bispectra and SVM for Radio Transmitter Individual Identification

LU Xuan-min, YANG Ju
School of Electronic and Information
Northwestern Polytechnical University
Xi'an, China
e-mail: luxuanmin@nwpu.edu.cn

ZHOU Ya-jian
Information Security Center
Beijing University of Posts and Telecommunications
Beijing, China
e-mail: zhouyajian@126.com

***Abstract*—To resolve the difficult problem of identifying radio transmitters with the same model, a new method using support vector machine with mixtures of kernels is present for classification of individual transmitters. In this method, the selected local integral bispectra and parameters significant for classification of the received signal form the new identification feature vector. To optimize the classifier, different parameters of kernel function are discussed. The performance of classifier which based on mixtures of kernels is compared with which based on conventional kernel functions. The result of experiments on FM individual transmitters shows that this method is able to achieve better classification rate than conventional kernels even in low SNR.**

Keywords- Local Integral Bispectra; Support Vector Machine; Kernel Function; Transmitter Individual Identification

I. INTRODUCTION

Identification of individual communication transmitters with the same model and manufacturing lot using fingerprint feature of communication signals is a new issue on communication confrontation. Currently, the radio performance indicators become higher, and the radios' internal filter works better. With the spread spectrum communication systems widely used in military communications, the signals environment is more and more complex and transmitter individual identification become more difficult.

The features of "turn-on" transient signals are usually used for classifying the transmitters [1][2] because they have unique characteristics. However, classification based on transient features has strict requirements for integrity of transient signals. Embedded spread spectrum sets[3] is suitable for co-operation transmitter identification, but they pay little attention to identification of transmitters which have same model and manufacturing lot.

In this paper, the selected local integral bispectra and parameters significant for classification of the received signal form the identification feature vector. To realize the transmitter identification, some samples should be got, always these samples are not enough, while the support vector machine has many unique advantages for small samples, nonlinear and high-dimensional pattern recognition. Compared with other learning machine, SVM has good generalization and universal ability, it can be used for pattern recognition, regression estimation and etc. In this paper, a new method using support vector machine with mixtures of kernels for transmitter classification is present for identifying transmitters with the same model and manufacturing lot. The result of experiments shows that the method is able to classify the same model transmitter with an accuracy rate of about 90% even in lower SNR.

II. FEATURE VECTOR EXTRACTIONS USING LOCAL SURROUNDING-LINE INTEGRAL BISPECTRA

A. Bispectra

Bispectra is a powerful tool to deal with random signal which is non-Gaussian and non-linear. Let $x(n)$ be a real stationary signal. The 3-th order moment of $x(n)$ is defined as:

$$c_{3x}(\tau_1,\tau_2)=E\{x^*(n)x(n+\tau_1)x(n+\tau_2)\} \tag{1}$$

where bispectra is defined by:

$$B(\omega_1,\omega_2)=\sum_{\tau_1=-\infty}^{\infty}\sum_{\tau_2=-\infty}^{\infty}c_{3x}(\tau_1,\tau_2)e^{-j(\omega_1\tau_1+\omega_2\tau_2)} \tag{2}$$

*denotes the complex conjugate.

From the signal domain to bispectra domain, on the one hand, the transform make bispectra translation invariant, scale variant and part of the phase information keep; on the other hand, the direct utilization of bispectra requires a 2-dimensional matching template, which brings about very heavy computation, and the bispectra has great information redundancy, so it is necessary to consider to lower the number of dimensions. The integral bispectra is defined to transform the 2-dimensional bispectra into a 1-dimensional template. The approach can not only reflect the need of features but also reduce the computational complexity.

B. Integral bispectra

In this paper, the local surrounding-line integral bispectra is selected. Supposed that S transmitters need to be classified, and the number of observation records for each transmitter is $N_1, \mathrm{N}_2, \ldots, N_S$. The record $x_k^{(1)}(n)$ is defined by:

$$Y_k^i(l)=\sum_{R_l}B_k^i(\omega_1,\omega_2) \tag{3}$$

978-0-7695-4080-1/10 $26.00 © 2010 IEEE
DOI 10.1109/ICIE.2010.305

Where superscript i $(i = 1,2,..., S)$ denotes the i th class of signal, subscript k $(k = 1,2,..., N_i)$ denotes the k th observation record of the k th class of signal, and R_l is the integral path on the bifrequency plane $(l = 1,2,..., L)$, which is a square with the origin as the center where each point is a bispectra value, shown in Figure 1.

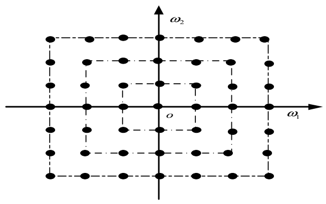

Figure1. Integral paths of SLIB

Surrounding-line integral bispectra will not miss or reuse any bispectra values, and this helps to extract meaningful information for target identification. It has some characteristics of translation invariant, scale variant and part of the phase information is kept. Another advantage is the better noise immunity than local bispectra. Considering that the cross-terms of bispectra will affect identification accuracy, SLIB with the most discriminant power should be selected as the main feature parameters for target identification. Considering that SLIB computation is reduced significantly, but still needs to calculate the whole bispectra plane to achieve a great amount of computation. In order to select the powerful SLIB as the feature set, Fisher's class-separability is used as the discriminant measure to judge a SLIB value. It avoids calculating the whole bispectra in the classification process, and will reduce the computation greatly. Besides, it needs not to consider polar coordinates spectrum interpolation and integral. As mentioned above, the selected SLIB for classification of the received signal form the basic identification feature vector.

C. *The basic feature vector extraction using local surrounding-line integral bispectra*

Integral bispectra can be obtained by integrating the bispectra, it avoids calculating the whole bispectra plane and the computation is reduced. In order to select the powerful SLIB as the feature set, Fisher's class-separability is used as the discriminant measure to judge a SLIB value.

A function for quantifying discriminant power is further defined by

$$Y(l) = \sum_{i \neq j} Y_{ij}(l) \tag{4}$$

where $Y_{ij}(l)$ is the Fisher measure function[4] :

$$Y_{ij}(l) = (m_i - m_j)^2 / (S_i + S_j) \tag{5}$$

where m_i , m_j are the mean value of i, j class object respectively. S_i , S_j are the variance of i, j class objects respectively.

The larger of $Y_{ij}(l)$, the more class-separability and class-cluster within i, j .

The set of SLIB values that has the largest $Y(l)$ forms the local surrounding-line integral bispectra and are used as the main feature vector, it can effectively eliminate the fewer cross-term interference and improve the efficiency of identification.

D. *Identification feature vector fusion*

For transmitters with the same model and same batch, individual difference with steady transmitted signal will grow out of uncertainty during the device manufacturing process. Therefore, these attribute parameters combined with individual transmitters should be interfused with SLIB to form the classification feature vector. It is necessary to improve the recognition rate. As for the FM radio stations used in the experiments, the kurtosis, the slope, carrier frequency, the carrier frequency and its rate of change, and the modulation factor are interfused with SLIB to form the identification feature vector, and it can achieve better classification rate.

Classification feature extraction methods and steps are as follows:(1) By equation (1) calculating the signal characteristics of bispectra.(2) By equation (3) calculating all the possible class combination (i,j) of Fisher's class-separability $Y_{ij}(l)$.(3) By equation (4) calculating quantifying discriminant power $Y(l)$.(4) The number of local surrounding line integral bispectra is N, selected N largest quantifying discriminant power $Y(l)$ corresponding to surrounding line integral bispectra to constitute local surrounding line integral bispectra feature.(5) Completed feature vector fusion. Selected extracted feature vector and parameters significant for classification of the received signal form the new identification feature vector.

III. MULTI-CLASSIFICATION METHOD BASED ON MIXTURES OF KERNEL

A. *Support Vector Machine*

Support Vector Machine (SVM) is a popular technique for pattern recognition and is based on statistical learning theory. Its core idea is to pre-selected through a non-linear mapping to a linear sub-space that can not be mapped to a separable space high-dimensional linear, in this space using structural risk minimization principle can simply looks for a separating hyperplane with largest margin. It can maximize the generalization ability of learning machine, while it would solve the optimization problem into a convex quadratic programming problem. Quadratic programming solution obtained is the only global optimal solution, so that the problem of local extremum does not exist in the general neural network. In addition, it also cleverly solved the problem dimension to the complexity of the algorithm

Let the training data input-output pairs be $\{x_i, y_i\}, i = 1,2,\ldots,l, y_i \in \{-1,1\}, x_i \in R^d$, Suppose that the predetermined feature extraction function is $\varphi(x) \in R^N$。So, the classification function of two classes is defined as：

$$y = \mathrm{sgn}(u(x)) = \mathrm{sgn}(\langle \omega, \varphi(x) + b \rangle) \tag{6}$$

where $\varphi(x) \in R^N$ is the weight vector, and $b \in R$ is the bias. ω and b determine the hyperplane of the feature space. That makes the maximum margin by minimizing $\|\omega\|^2$, which is subject to

$$y_i(\langle \omega, \varphi(x_i) \rangle + b) - 1 \geq 0 \tag{7}$$

If we use Lagrangian function to solve the problem, then:

$$L_p = \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{l} \alpha_i y_i (x_i.\omega + b) + \sum_{i=1}^{l} \alpha_i \tag{8}$$

Minimizing L_p means that minimizing for the variables ω and b, and simultaneously maximizing for the multipliers α_i. L_p is subject to the constraints:

$$\omega = \sum \alpha_i y_i x_i \tag{9}$$

$$\sum_i \alpha_i y_i = 0 \tag{10}$$

For the non-separable case, we should relax the constraints of （8） with the positive slack variables $\xi_i, i = 1, \ldots, l$ .Then the Lagrangian becomes:

$$L_p = \frac{1}{2}\|\omega\|^2 - C\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\alpha_i\{y_i(x_i.\omega + b) - 1 + \xi_i\} - \sum_{i=1}^{l}\mu_i\xi_i \tag{11}$$

where the μ_i are the Lagrange multipliers introduced to enforce positively of the ξ_i ,and C ($0 \leq \alpha_i \leq C$) is a cost parameter, it corresponding to assigning a higher penalty to errors.

According to the relevant functional theory, as long as the kernel function satisfies the Mercer condition, it corresponds to a change in the inner product space. Kernel function formula is as follows:

$$K(x_i, y_i) = \sum \varphi(x_i)\varphi(y_i) \tag{12}$$

where $K(x_i, y_i)$ is kernel function. The introduction of kernel function greatly improves the learning machine's non-linear processing power, while maintaining in high-dimensional space within a line of a learning machine, thereby enabling learners can easily be controlled. By introducing the kernel function, the low-dimensional sub-problems can not be divided into high-dimensional problems, while avoiding a dimension catastrophe of the calculation problems.

Most of the kernel functions of the present study are mainly four types [5]: (1) Linear kernel function: $K(x, x_i) = x.x_i$; (2) Polynomial kernel function: $K(x, x_i) = [(x.x_i) + 1]^q$; (3) Gaussian radial basis kernel function: $K(x, x_i) = \exp(-\|x - x_i\|^2 / 2\sigma^2)$; (4) Sigmoid function： $K(x, x_i) = \tanh(v(x.x_i) + c)$.

Type（2~4）, q, σ, c and other parameters are real constant. In practice, according to the specific situation of choice to construct a suitable kernel function and the corresponding parameters.

B. Parameters selection in the model

(1) $K(x, x_i)$. According to Mercer's theorem [6], as long as a real symmetric function satisfies the relevant conditions can be as a nuclear function. Because the signal was characteristic parameters non-linear related, so a variety of functions can be chose as kernel function. It was found that mixtures of kernels (mixtures of Gaussian radial basis kernel function) is best, the expression is:

$$K(x, x_i) = A\exp(-\|x - x_i\|^2 / 2\sigma^2)$$

The best results reflect the kernel function and the inherent laws of the sample data itself.

(2) The control upper bound of C. The kernel parameter σ and the error penalty parameter C, different combinations of pairs of classification effect shown in Figure 2.

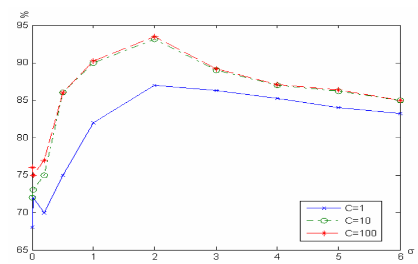

Figure 2. Recognition rates with different parameters σ ,C combinations

When the penalty factor C is small (C = 1),the classification results are poor, to increase the value of classifier C, the ability of generalization enhanced markedly .The smaller C value is given, the less of support vector number are used and computing time are spent. C values expressed experience of a small punishment for a small error, classification complexity is low, but experienced risk value larger, and the corresponding generalization ability is weak. But there is always has a C to make generalization ability become strong. When C exceeds a certain value, the classifier's complexity has reached the allowed maximum of data space, the experience risk and the promote ability are virtually no change in this condition. As the mixtures of Gaussian RBF kernel is based on the Gaussian RBF kernel through training, obtain the model parameters in Table I.

As the Mixtures of Gaussian RBF kernels is based on Gaussian RBF kernel, so we can obtain the model parameters through training, shown in Table I.

TABLE I. MODEL PARAMETERS

Parameters	selected
kernel $K(x,x_i)$	Mixtures of Gaussian RBF kernel
Control upper bound C	C=100
σ	2
A	20

C. Multi-classification of SVM based on Mixtures of Kernels

We have obtained the training feature vectors which constitutes the template library from the second section. The parameters of the mixtures of kernels model is determined in the first two parts of Section 3. Then the n(in this paper, n is 5.) class FM radio which are the same model and same batch can be right separated. For multi-classification, there are usually three strategies: OVA (One-VS.-All), AVA (All-VS.-All) and OVO(One-VS.-One) [7]. In this paper we use the OVA strategies. Let the Training set T be:

$$T=\{(x_1,y_1),\ldots,(x_l,y_l)\},x_i=R^n,y_i\in\{1,2,\ldots,M\},i=1,\ldots,l$$

In the training, when the first i is the sample as a class, the remaining samples of M-1 category as a negative category, with the SVM decision function derived as follows:

$$f(x)=\arg\max_i f_i(x) \tag{13}$$

Where $f_i(x)$ is the i-th classifier. In this paper,

$$f_i(x)=\operatorname{sgn}[\sum_{j=1}^{l} y_j\alpha_i K(x,x_j)+b_i] \tag{14}$$

where $K(x,x_i)$ is mixtures of kernels:

$$K(x,x_i)=A\exp(-\|x-x_i\|^2/2\sigma^2) \tag{15}$$

In the classification of the test samples, test samples were substituted into the decision-making function, according to the output value, the greatest $f_i(x)$ value of the corresponding subscript of the test samples shall be a class number.

IV. EXPERIMENTAL RESULTS

Experiments were carried out in the Matlab 7.0. The data is from the same model, same batch of 5 FM radio samples, by $T_i, i=1,2,..,5$ expressing them. The SNR of the sample is 15dB, and the signal sampling frequency is 84MHZ, sampling points in 4096, and data were normalized before the training. Each FM radio provides a wide range of $N_i=100$ observational data, the top 20 group are as the SVM the training set, and the rest for the test set. In the classification experiments, the first is to obtain fusion of the feature vector, and then feature vector classification using SVM based on mixtures of kernel.

The results of experiments on FM individual transmitters shows that using SVM based on mixtures of kernels can achieve better classification rate than which using conventional kernel functions.

TABLE II. RECOGNITION RATE USING DIFFERENT KERNEL FUNCTION

FM samples	Recognition Rate		
	Linear kernel	***Gaussian RBF kernel***	***Mixture of Gaussian RBF kernel***
T1	79.4%	84.2%	89.8%
T2	78.8%	86.1%	93.5%
T3	75.2%	85.3%	88.6%
T4	76.4%	82.8%	87.4%
T5	71.6%	87.4%	90.2%
Average rate	76.3%	85.2%	89.9%

From Table II, the average recognition rate is lower than 80% when using linear kernel, when using Gaussian RBF kernel, the percentage can be raised to above 85%. However, when mixtures of Gaussian RBF kernels are utilized, the average recognition rate is about 90%. The experiments results reflect the great difference between individual transmitters with the same model and same batch in low SNR, it has shown strong classification ability.

V. CONCLUSION

This paper has introduced a classification algorithm for identification of transmitters which have same model and same batch. The method using mixtures of kernels for support vector machine is proposed to solve transmitter classification problem. The selected local integral bispectraum and parameters significant for classification of the received signal form the new identification feature vector, it is not only able to classify different individuals of transmitters in low SNR, but also does not depend on individual characteristics and stability, so it has a strong record of samples of the time, with strong stability. The result of experiments on FM individual transmitters shows that using mixtures of kernels for SVM can achieve better classification rate than which using conventional kernel functions. It proves that the method is an effective solution of transmitter individual identification.

REFERENCES

[1] J. Toonstra, W. Kinsner. A Radio Transmitter Fingerprinting System ODO-1. Canadian Conference on Electrical and Computer Engineering[J], 1996, (1): 60-63.

[2] D. Shaw, W. Kinsner. Multifractal modelling of radio transmitter transients for classification[J], IEEE Proceedings . WESCANEX 97: Communications, Power and Computing, IEEE, 22-23 May 1997. 306-312.

[3] Xianbin Wang, Yiyan Wu, B. Caron. Transmitter identification using embedded spread spectrum sets[J], Proc. International Conference on Communication Technology, ICCT 2003, (2): 1517-1523.

[4] Bi-cheng Li, Mei-zhen Shao, Jie Huang. Pattern Recognition Theory and Application [M]. Xi'an: Xidian University Press, 2008.

[5] Li-chuan Guo, Shi-yu Sun, Xiu-sheng Duan. Study on support vector machine and kernel function [J]. Science, Technology and Engineering, 1671-1819(2008)

[6] Fu Liu. Some properties for Support Vector Machines Mercer kernels [J]. Changsha University, 2005. 41-46.

[7] Vapnik V N. Statistical Learning Theory[M]. New-York, John Wiley, 1998.