Wavelet Support Vector Machine

Li Zhang, Weida Zhou, and Licheng Jiao, *Senior Member, IEEE*

***Abstract*—An admissible support vector (SV) kernel (the wavelet kernel), by which we can construct a wavelet support vector machine (SVM), is presented. The wavelet kernel is a kind of multidimensional wavelet function that can approximate arbitrary nonlinear functions. The existence of wavelet kernels is proven by results of theoretic analysis. Computer simulations show the feasibility and validity of wavelet support vector machines (WSVMs) in regression and pattern recognition.**

***Index Terms*—Support vector kernel, support vector machine, wavelet kernel, wavelet support vector machine.**

I. Introduction

THE SUPPORT vector machine (SVM) is a new universal learning machine proposed by Vapnik *et al.* [6], [8], which is applied to both regression [1], [2] and pattern recognition [2], [5]. An SVM uses a device called kernel mapping to map the data in input space to a high-dimensional feature space in which the problem becomes linearly separable [10]. The decision function of an SVM is related not only to the number of SVs and their weights but also to the *a priori* chosen kernel that is called the support vector kernel [1], [9], [10]. There are many kinds of kernels can be used, such as the Gaussian and polynomial kernels.

Since the wavelet technique shows promise for both nonstationary signal approximation and classification [3], [4], it is valuable for us to study the problem of whether a better performance could be obtained if we combine the wavelet technique with SVMs. An admissible SV kernel, which is a wavelet kernel constructed in this paper, implements the combination of the wavelet technique with SVMs. In theory, wavelet decomposition emerges as a powerful tool for approximation [11]–[16]; that is to say the wavelet function is a set of bases that can approximate arbitrary functions. Here, the wavelet kernel has the same expression as a multidimensional wavelet function; therefore, the goal of the WSVMs is to find the optimal approximation or classification in the space spanned by multidimensional wavelets or wavelet kernels. Experiments show the feasibility and validity of WSVMs in approximation and classification.

II. Support Vector Machines (SVMs)

SVMs use SV kernel to map the data in input space to a high-dimensional feature space in which we can process a problem in linear form.

A. SVM for Regression [1], [2]

Let $\mathbf{x} \in R^N$ and $y \in R$, where R^N represents input space. By some nonlinear mapping Φ, $\mathbf{x}$ is mapped into a feature space in which a linear estimate function is defined

$$y = f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \Phi(\mathbf{x}) + b. \tag{1}$$

We seek to estimate (1) based on independent uniformly distributed data $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)$ by finding a function f with a small risk. Vapnik *et al.* suggested using the following regularized risk functional to obtain a small risk [6], [8]:

$$\frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{l}\sum_{i=1}^{l}|y_i - f(\mathbf{x}_i, \mathbf{w})|_\varepsilon \tag{2}$$

where $C > 0$ is a constant, and $\varepsilon > 0$ is a small positive number. The second term can be defined as

$$|y-f(\mathbf{x}, \mathbf{w})|_\varepsilon = \begin{cases} 0, & \text{if } |y - f(\mathbf{x}, \mathbf{w})| < \varepsilon \\ |y - f(\mathbf{x}, \mathbf{w})| - \varepsilon, & \text{otherwise.} \end{cases} \tag{3}$$

By using Lagrange multiplier techniques, the minimization of (2) leads to the following dual optimization problem. Maximize

$$W\left(\alpha^{(*)}\right) = -\varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) + \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)y_i - \tfrac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(\mathbf{x}_i, \mathbf{x}_j). \tag{4}$$

Subject to

$$\sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0$$
$$\alpha_i^{(*)} \in [0, C]. \tag{5}$$

The resulting regression estimates are linear. Then, the regression takes the form

$$f(\mathbf{x}) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)K(\mathbf{x}, \mathbf{x}_i) + b. \tag{6}$$

A kernel $K(\mathbf{x}_i, \mathbf{x}_j)$ is called an SV kernel if it satisfies a certain conditions, which will discussed in detail in Section II-C.

B. SVM for Pattern Recognition [2], [5]

It is similar to SVM for regression. The training procedure of SVM for pattern recognition is to solve a constrained quadratic optimization problem as well. The only difference between them is the expression of the optimization problem. Given an i.i.d. training example set $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)\}$, where $\mathbf{x} \in R^N$, $y \in \{-1, 1\}$. Kernel mapping can map the training examples in input space into a feature space in

Manuscript received February 3, 2001; revised November 5, 2001. This work was supported by the National Natural Science Foundation of China under Grants 60073053, 60133010, and 69831040. This paper was recommended by Associate Editor P. Bhattacharya.

The authors are with the National Key Laboratory for Radar Signal Processing, Xidian University, Xi'an, 710071, China.

Digital Object Identifier 10.1109/TSMCB.2003.811113

which the mapped training examples are linearly separable. For pattern recognition problem, SVM becomes the following dual optimization problem:

$$\text{Maximize} \quad W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i y_i \alpha_j y_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \tag{7}$$

$$\text{subject to} \quad \sum_{i=1}^{l} \alpha_i y_i = 0$$

$$\alpha_i \in [0, C], \qquad i = 1, \ldots, l.$$

The decision function becomes

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \right). \tag{8}$$

C. Conditions for Support Vector Kernel

The formation of an SV kernel is a kernel of dot-product type in some feature space $K(\mathbf{x}, \mathbf{x}') = K(\langle \mathbf{x} \cdot \mathbf{x}' \rangle)$. The Mercer theorem (see [7]) gives the conditions that a dot product kernel must satisfy.

Theorem 1: Suppose $K \in L_\infty(R^N \times R^N)$ (R^N denotes the input space) such that the integral operator T_K: $L_2(R^N) \rightarrow L_2(R^N)$

$$T_K f(\cdot) := \int_{R^N} K(\cdot, \mathbf{x}) f(\mathbf{x})\, d\mu(\mathbf{x}) \tag{9}$$

is positive. Let $\phi_j \in L_2(R^N)$ be the eigenfunction of T_K associated with the eigenvalue $\lambda_j \neq 0$ and normalized such that $\|\phi_j\|_{L_2} = 1$. Let $\overline{\phi_j}$ denote its complex conjugate. Then, we have the following.

1) $(\lambda_j(T_K))_j \in l_1$.
2) $\phi_j \in L_\infty(R^N)$ and $\sup_j \|\phi_j\|_{L_\infty} < \infty$.
3) $K(\mathbf{x}, \mathbf{x}') = \sum_j \lambda_j \overline{\phi_j(\mathbf{x})} \phi_j(\mathbf{x}')$ holds for almost all $(\mathbf{x}, \mathbf{x}')$, where the series converges absolutely and uniformly for almost all $(\mathbf{x}, \mathbf{x}')$.

In (9), $\mu(\mathbf{x})$ denotes a measure defined on some measurable set. This theorem means that if (Mercer's condition, [6], [9])

$$\iint_{L_2 \otimes L_2} K(\mathbf{x}, \mathbf{x}') g(\mathbf{x}) g(\mathbf{x}')\, d\mathbf{x}\, d\mathbf{x}' \geq 0$$
$$\forall\, g(\mathbf{x}) \in L_2(R^N),\ K(\mathbf{x}, \mathbf{x}') \in L_2(R^N \times R^N) \tag{10}$$

holds we can write $K(\mathbf{x}, \mathbf{x}')$ as a dot product $K(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}') \rangle$ in some feature space.

Translation invariant kernels, i.e., $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x} - \mathbf{x}')$ derived in [9] are admissive SV kernels if they satisfy Mercer's condition. However, it is difficult to decompose the translation invariant kernels into the product of two functions and then to prove them as SV kernels. Now, we state a necessary and sufficient condition for translation invariant kernels [1], [9].

Theorem 2: A translation invariant kernel $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x} - \mathbf{x}')$ is an admissible SV kernels if and only if the Fourier transform

$$F[K](\omega) = (2\pi)^{-N/2} \int_{R^N} \exp(-j(\omega \cdot \mathbf{x})) K(\mathbf{x})\, d\mathbf{x} \tag{11}$$

is non-negative.

The theorems stated above can be useful for both checking whether a kernel is an admissible SV kernel and actually constructing new kernels.

III. Wavelet Support Vector Machines

In this section, we will propose WSVMs and construct wavelet kernels, which are admissible SV kernels. It is the wavelet kernel that combines the wavelet technique with SVMs.

A. Wavelet Analysis

The idea behind the wavelet analysis is to express or approximate a signal or function by a family of functions generated by dilations and translations of a function $h(x)$ called the mother wavelet:

$$h_{a,c}(x) = |a|^{-1/2} h\left(\frac{x - c}{a} \right) \tag{12}$$

where $x,\ a,\ c \in R$, a is a dilation factor, and c is a translation factor (In wavelet analysis, the translation factor is denoted by b, but here, b is used for expressing the threshold in SVMs.) Therefore, the wavelet transform of a function $f(x) \in L_2(R)$ can be written as

$$W_{a,c}(f) = \langle f(x),\ h_{a,c}(x) \rangle. \tag{13}$$

In the right-hand side of of (13), $\langle \cdot, \cdot \rangle$ denotes the dot product in $L_2(R)$. Equation (13) means the decomposition of a function $f(x)$ on a wavelet basis $h_{a,c}(x)$. For a mother wavelet $h(x)$, it is necessary to satisfy the condition [3], [12]

$$W_h = \int_0^\infty \frac{|H(\omega)|^2}{|\omega|}\, d\omega < \infty \tag{14}$$

where $H(\omega)$ is the Fourier transform of $h(x)$. We can reconstruct $f(x)$ as follows:

$$f(x) = \frac{1}{W_h} \int_{-\infty}^{\infty} \int_0^\infty W_{a,c}(f) h_{a,c}(x)\, da/a^2\, dc. \tag{15}$$

If we take the finite terms to approximate (15) [3], then

$$\hat{f}(x) = \sum_{i=1}^{l} W_i h_{a_i, c_i}(x). \tag{16}$$

Here, $f(x)$ is approximated by $\hat{f}(x)$.

For a common multidimensional wavelet function, we can write it as the product of one-dimensional (1-D) wavelet functions [3]:

$$h(\mathbf{x}) = \prod_{i=1}^{N} h(x_i) \tag{17}$$

where $\{\mathbf{x} = (x_1, \ldots, x_N) \in R^N\}$. Here, every 1-D mother wavelet $h(x)$ must satisfy (14).

For wavelet analysis and theory, see [17]–[19].

B. Wavelet Kernels and WSVMS

Theorem 3: Let $h(x)$ be a mother wavelet, and let a and c denote the dilation and translation, respectively. $x,\ a,\ c \in R$. If $\mathbf{x},\ \mathbf{x}' \in R^N$, then dot-product wavelet kernels are

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{N} h\left(\frac{x_i - c_i}{a} \right) h\left(\frac{x_i' - c_i'}{a} \right) \tag{18}$$

TABLE I
RESULTS OF APPROXIMATIONS

Kernel	Parameter	Iteration times	Approximation error
Wavelet kernel	$a = 1$	2000	0.0555
Gaussian kernel	$\beta = 1$	2000	0.0666

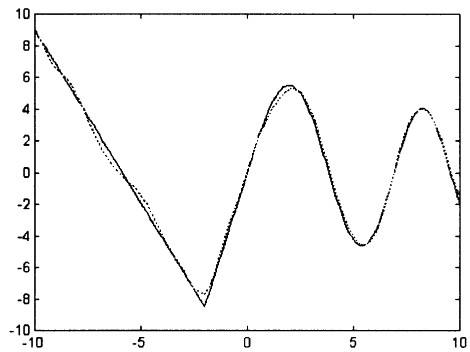


Fig. 1. Original function (solid line) and resulting approximation by Gaussian kernel (dotted line).

and translation-invariant wavelet kernels that satisfy the translation invariant kernel theorem are

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{N} h\left(\frac{x_i - x_i'}{a}\right). \tag{19}$$

The proof of Theorem 3 is given in Appendix A. Without loss of generality, in the following, we construct a translation-invariant wavelet kernel by a wavelet function adopted in [4].

$$h(x) = \cos(1.75x)\exp\left(-\frac{x^2}{2}\right). \tag{20}$$

Theorem 4: Given the mother wavelet (20) and the dilation a, $a, x \in R$. If $\mathbf{x}, \mathbf{x}' \in R^N$, the wavelet kernel of this mother wavelet is

$$\begin{aligned} K(\mathbf{x}, \mathbf{x}') &= \prod_{i=1}^{N} h\left(\frac{x_i - x_i'}{a}\right) \\ &= \prod_{i}^{N}\left(\cos\left(1.75 \times \frac{(x_i - x_i')}{a}\right)\exp\left(-\frac{\|x_i - x_i'\|^2}{2a^2}\right)\right) \end{aligned} \tag{21}$$

which is an admissible SV kernel.

The proof of Theorem 4 is shown in Appendix B. From the expression of wavelet kernels, we can take them as a kind of multidimensional wavelet function. The goal of our WSVM is to find the optimal wavelet coefficients in the space spanned by the multidimensional wavelet basis. Thereby, we can obtain the optimal estimate function or decision function.

Now, we give the estimate function of WSVMs for the approximation

$$f(\mathbf{x}) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\prod_{j=1}^{N} h\left(\frac{x^j - x_i^j}{a_i}\right) + b \tag{22}$$

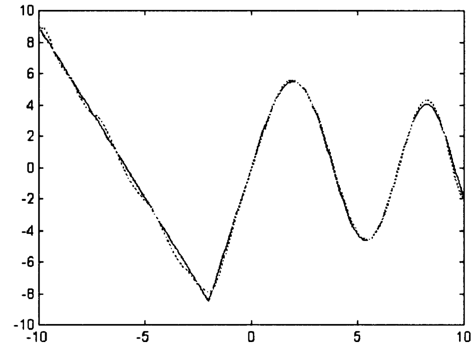


Fig. 2. Original function (solid line) and resulting approximation by wavelet kernel (dotted line).

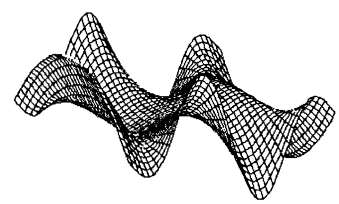


Fig. 3. Original function.

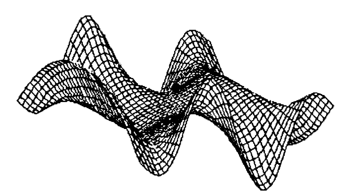


Fig. 4. Resulting approximation by Gaussian kernel.

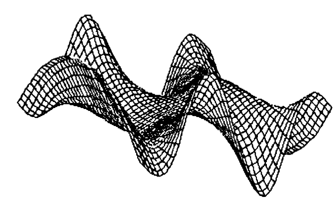


Fig. 5. Resulting approximation by wavelet kernel.

and the decision function for classification is

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{l}\alpha_i y_i \prod_{j=1}^{N} h\left(\frac{x^j - x_i^j}{a_i}\right) + b\right) \tag{23}$$

where the x_i^j denotes the jth component of the ith training example.

IV. Simulation Experiment

Now, we validate the performance of wavelet kernel by three simulation experiments, the approximation of a single-variable function and two-variable function, and the recognition of the 1-D images of radar target.

For comparison, we showed the results obtained by wavelet kernel and Gaussian kernel, respectively. The Gaussian kernel is one of the first SV kernels investigated for most of learning problems. Its expression is $K(\mathbf{x}, \mathbf{x}') = \exp(-\beta\|\mathbf{x} - \mathbf{x}'\|^2)$, where $\beta > 0$ is a parameter chosen by user. Since SVMs cannot optimize the parameters of kernels, it is difficult to determine

TABLE II
APPROXIMATION RESULTS OF TWO-VARIABLE FUNCTION

Kernel	Parameter	Iteration times	Approximation error
Gaussian kernel	$\beta = 0.03$	10000	0.0512
Wavelet kernel	$a = 4$	10000	0.0457

l parameters a_i, $i = 1, \ldots, l$. For the sake of simplicity, let $a_i = a$ such that the number of parameters becomes 1. The parameters a for wavelet kernel and β for the Gaussian kernel are selected by using cross validation that is in wide use [20], [21].

A. Approximation of a Single-Variable Function

In this experiment, we approximate the following single-variable function [3]

$$f(x) = \begin{cases} -2.186x - 12.864, & -10 \leq x < -2 \\ 4.246x, & -2 \leq x < 0 \\ 10e^{-0.05x-0.5}, & \\ \quad \cdot \sin((0.03x + 0.7)x), & 0 \leq x \leq 10. \end{cases} \quad (24)$$

We have a uniformly sampled examples of 148 points, 74 of which are taken as training examples and others testing examples. We adopted the approximation error defined in [3] as

$$\delta = \sqrt{\frac{\sum_{i=1}^{l}(y_i - f_i)^2}{\sum_{i=1}^{l}(y_i - \overline{y})^2}}, \qquad \text{where } \overline{y} = \frac{1}{l}\sum_{i=1}^{l} y_i \quad (25)$$

where y denotes the desired output for $\mathbf{x}$ and f the approximation output. Table I lists the approximation errors using the two kernels. The approximation results are plotted in Figs. 1 and 2, respectively. The solid lines represent the function f and the dashed lines show the approximations.

B. Approximation of Two-Variable Function

This experiment is to approximate a two-variable function [3]

$$f(\mathbf{x}) = (x_1^2 - x_2^2)\sin(0.5x_1) \quad (26)$$

over the domain $[-10, 10] \times [-10, 10]$. We take 81 points as the training examples, and 1600 points as the testing examples. Fig. 3 shows the original function f, and Figs. 4 and 5 show the approximation results obtained by Gaussian and wavelet kernel, respectively. Table II gives the approximation errors.

C. Recognition of Radar Target

This task is to recognize the 1-D images of three-class planes B-52, J-6, and J-7. Our data is acquired in a microwave anechonic chamber with imaging angle from 0 to 160°. Here, the dimension of the input space of the 1-D image recognition problem is 64. The 1-D images of B-52, J-6, and J-7 under 0°

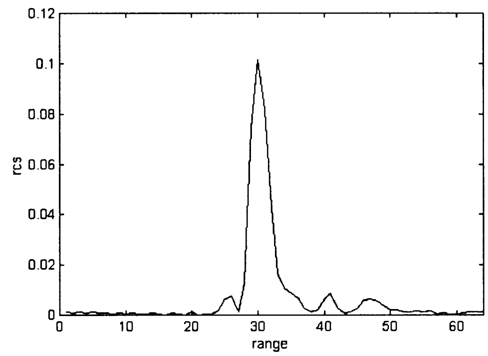


Fig. 6. One-dimensional image of B-52 plane model under 0°.

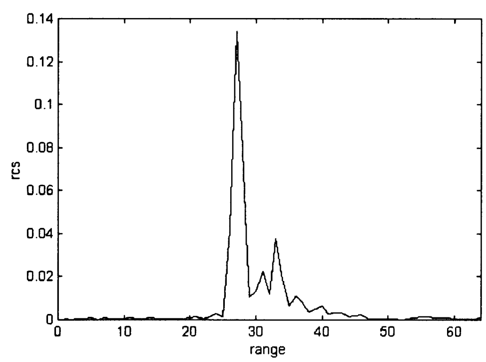


Fig. 7. One-dimensional image of J-6 plane model under 0°.

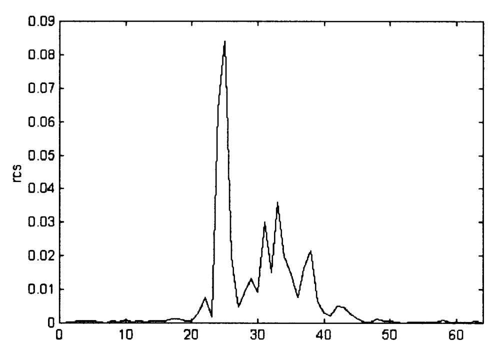


Fig. 8. One-dimensional image of J-7 plane model under 0°.

are shown in Figs. 6–8, respectively. We divided these examples into two groups shown in Table III. The imaging angle of the first group is from 0 to 100° and the second from 80 to 160°. The recognition rates obtained by Gaussian and wavelet kernel are shown in Table IV, which imply that wavelet kernel gives a comparable recognition performance with Gaussian kernel.

TABLE III
NUMBER OF TRAINING AND TESTING EXAMPLES

Group	No. of training examples			No. of testing examples		
	B-52	J-6	J-7	B-52	J-6	J-7
1	51	41	56	200	160	224
2	30	31	47	118	120	186

TABLE IV
RESULTS OF RADAR TARGET RECOGNITION

Kernel	Parameter	Group	Recognition rates for testing			Average optimization time over 30 runs (s)	Average recognition rates		
			B-52 (%)	J-6 (%)	J-7 (%)		B-52 (%)	J-6 (%)	J-7 (%)
Gaussian kernel	$\beta = 0.04$	1	98.00	96.25	96.88	40.2	98.16	95.63	96.02
		2	98.31	95.00	97.31	16.6			
Wavelet kernel	$a = 2.5$	1	100	96.25	98.21	33.8	100	95.63	98.26
		2	100	95.00	98.31	12.0			

We have compared the approximation and recognition results obtained by Gaussian and wavelet kernel, respectively. In the three experiments, our wavelet kernel has better results than the Gaussian kernel.

V. CONCLUSION AND DISCUSSION

In this paper, wavelet kernels by which we can combine the wavelet technique with SVMs to construct WSVMs are presented. The existence of wavelet kernels is proven by results of theoretic analysis. Our wavelet kernel is a kind of multidimensional wavelet function that can approximate arbitrary functions. It is not surprising that wavelet kernel gives better approximation than Gaussian kernel, which is shown by Computer simulations. From (22) and (23), the decision function and regression estimation function can be expressed as the linear combination of wavelet kernel as well as the Gaussian kernel. Notice that the wavelet kernel is orthonormal (or orthonormal approximately), whereas the Gaussian kernel is not. In other words, the Gaussian kernel is correlative or even redundancy, which is the possible reason why the training speed of the wavelet kernel SVM is slightly faster than the Gaussian kernel SVM.

APPENDIX A
PROOF OF THEOREM 3

Proof: We prove first that dot-product wavelet kernels (18) are admissible SV kernels. For $\forall\, g(\mathbf{x}) \in L_2(R^N)$, we have

$$\begin{aligned}&\iint_{L_2\otimes L_2} K(\mathbf{x},\,\mathbf{x}')g(\mathbf{x})g(\mathbf{x}')\,d\mathbf{x}\,d\mathbf{x}'\\&=\int_{L_2}\prod_{i=1}^{N} h\left(\frac{x_i-c_i}{a}\right)g(\mathbf{x})\,d\mathbf{x}\int_{L_2}\prod_{i=1}^{N} h\left(\frac{x_i'-c_i'}{a}\right)g(\mathbf{x}')\,d\mathbf{x}'\\&=\left(\int_{L_2}\prod_{i=1}^{N} h\left(\frac{x_i-c_i}{a}\right)g(\mathbf{x})\,d\mathbf{x}\right)^2 \geq 0.\end{aligned}$$

Hence, dot-product kernels (18) satisfy Mercer's condition. Therefore, this part of Theorem 3 is proved.

Now, we prove that translation-invariant wavelet kernels (19) are admissible kernels. Kernels (19) satisfy Theorem 2 [or condition (11)], which is a necessary and sufficient condition for translation invariant kernels; therefore, they are admissible ones.

This completes the proof of Theorem 3.

APPENDIX B
PROOF OF THEOREM 4

Proof: According to Theorem 2, it is sufficient to prove the inequality

$$F[K](\omega) = (2\pi)^{-N/2} \int_{R^N} \exp(-j(\omega \cdot \mathbf{x}))K(\mathbf{x})\,d\mathbf{x} \quad (27)$$

for all $\mathbf{x}$, where $K(\mathbf{x}) = \prod_i^N h(x_i/a) = \prod_{i=1}^N \cos (1.75x_i/a)e^{-(\|x_i\|^2/2a^2)}$. First, we calculate the integral term

$$\begin{aligned}
&\int_{R^N} \exp(-j\omega\mathbf{x})K(\mathbf{x})\,d\mathbf{x} \\
&\quad= \int_{R^N} \exp(-j\omega\mathbf{x}) \left(\prod_{i=1}^N \cos\Big(1.75\,\frac{x_i}{a}\Big) \exp\Big(-\frac{\|x_i\|^2}{2a^2}\Big) \right) d\mathbf{x} \\
&\quad= \prod_{i=1}^N \int_{-\infty}^{\infty} \exp(-j\omega_i x_i) \\
&\qquad \cdot \left(\frac{\exp(j1.75x_i/a) + \exp(-j1.75x_i/a)}{2} \right) \\
&\qquad \cdot \exp\left(-\frac{\|x_i\|^2}{2a^2}\right) dx_i \\
&\quad= \prod_{i=1}^N \frac{1}{2} \int_{-\infty}^{\infty} \left(\exp\left(-\frac{\|x_i\|^2}{2a^2} + \left(\frac{1.75j}{a} - j\omega_i a\right) x_i \right) \right. \\
&\qquad \left. + \exp\left(-\frac{\|x_i\|^2}{2a^2} - \left(\frac{1.75j}{a} + j\omega_i a\right) x_i \right) \right) dx_i \\
&= \prod_{i=1}^N \frac{|a|\sqrt{2\pi}}{2} \left(\exp\left(-\frac{(1.75-\omega_i a)^2}{2} \right) \right. \\
&\qquad \left. + \exp\left(-\frac{(1.75+\omega_i a)^2}{2} \right) \right). \qquad (28)
\end{aligned}$$

Substituting (28) into (27), we can obtain the Fourier transform

$$\begin{aligned}
F[K](\omega) = \prod_{i=1}^N \left(\frac{|a|}{2}\right) &\left(\exp\left(-\frac{(1.75-\omega_i a)^2}{2} \right) \right. \\
&\left. + \exp\left(-\frac{(1.75+\omega_i a)^2}{2} \right) \right). \qquad (29)
\end{aligned}$$

If $a \neq 0$, then we have

$$F[K](\omega) \geq 0 \quad (30)$$

and the proof is completed.

REFERENCES

[1] A. Smola and B. Schölkopf. (1998) A tutorial on support vector regression. Royal Holloway Coll., Univ. London, UK. [Online]NeuroCOLT Tech. Rep. NC-TR-98-030. Available: http://www.kernel-machines.org/.

[2] B. Schölkopf, A. J. Smola, R. Willianson, and P. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000.

[3] Q. H. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. Neural Networks*, vol. 3, pp. 889–898, Nov. 1992.

[4] H. H. Szu, B. Telfer, and S. Kadambe, "Neural network adaptive wavelets for signal representational and classification," *Opt. Eng.*, vol. 31, pp. 1907–1916, 1992.

[5] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Disc.*, vol. 2, no. 2, pp. 1–47, 1998.

[6] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[7] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equation," *Philos. Trans. R. Soc. London*, vol. A-209, pp. 415–446, 1909.

[8] C. Cortes and V. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[9] A. Smola, B. Schölkopf, and K.-R. Müller, "The connection between regularization operators and support vector kernels," *Neural Network*, vol. 11, pp. 637–649, 1998d.

[10] C. J. C. Burges, "Geometry and invariance in kernel based methods," in *Advance in Kernel Methods—Support Vector Learning*. Cambridge, MA: MIT Press, 1999, pp. 89–116.

[11] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 91, pp. 909–996, 1988.

[12] ——, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, pp. 961–1005, Sept. 1990.

[13] S. G. Mallat, "A theory for nuliresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.

[14] ——, "Multiresolution approximation and wavelets orthonormal bases of $L^2(R)$," *Trans. Amer. Math. Soc.*, vol. 315, no. 1, pp. 69–88, 1989.

[15] Y. Meyer, "Wavelet and operator," in *Proceedings of the Special Year in Modern Analysis*. London, U.K.: Cambridge Univ. Press, 1989.

[16] I. Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," *J. Math. Phys.*, vol. 27, pp. 1271–1283, 1986.

[17] X. D. Zhang and Z. Bao, *Non-Stable Signal Analysis and Processing*. Beijing, China: Nat. Defense Industry, 1998.

[18] G. Z. Liu and S. L. Di, *Wavelet Analysis and Application*. Xi'an, China: Xidian Univ. Press, 1992.

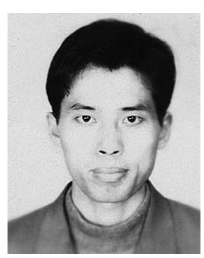
[19] S. G. Krantz, Ed., *Wavelet: Mathematics and Application*. Boca Raton, FL: CRC, 1994.

[20] T. Joachims, "Estimating the generalization performance of a SVM efficiently," in *Proc. 17th Int. Conf. Machine Learning*. San Francisco, CA: Morgan Kaufman, 2000. [Online]. Available: http://www.kernel-machines.org/.

[21] M. Kearns and D. Ron, "Algorithmic stability and sanity-check bounds for leave-one-out cross validation," in *Proc. Tenth Conf. Comput. Learning Theory*. New York: ACM, 1997, pp. 152–162.

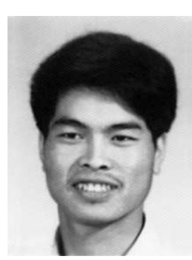
Li Zhang received the B.S. degree in electronic engineering and the Ph.D. degree from Xidian University, Xi'an, China, 1997 and 2002, respectively. She is currently pursuing the postdoctorate at the Institute of Automation of Shanghai, Jiao Tong University, Shanghai, China.

Her research interests have been in the areas of pattern recognition, machine learning, and data mining.



Weida Zhou received the B.S. degree in electronic engineering from Xidian University, Xi'an, China, in 1996. Since 1998, he has been working toward the M.S. and Ph.D. degrees at Xidian University.

His research interests include machine learning, learning theory, and data mining.



Licheng Jiao (SM'89) received the B.S. degree from Shanghai Jiaotong University, Shanghai, China, in 1982 and the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

He is currently Professor and Dean of the electronic engineering school at Xidian University. His research interests include neural networks, data mining, nonlinear intelligence signal processing, and communication.