# Relevance vector machine technique for the inverse scattering problem*

Wang Fang-Fang(王芳芳) and Zhang Ye-Rong(张业荣)†

*School of Electronic Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing* 210003, *China*

(Received 26 August 2011; revised manuscript received 25 October 2011)

A novel method based on the relevance vector machine (RVM) for the inverse scattering problem is presented in this paper. The nonlinearity and the ill-posedness inherent in this problem are simultaneously considered. The nonlinearity is embodied in the relation between the scattered field and the target property, which can be obtained through the RVM training process. Besides, rather than utilizing regularization, the ill-posed nature of the inversion is naturally accounted for because the RVM can produce a probabilistic output. Simulation results reveal that the proposed RVM-based approach can provide comparative performances in terms of accuracy, convergence, robustness, generalization, and improved performance in terms of sparse property in comparison with the support vector machine (SVM) based approach.

**Keywords:** inverse scattering problem, through-wall problem, relevance vector machine, finite-difference time-domain

**PACS:** 02.30.Zz, 29.40.Gx, 41.20.Jb, 81.70.Ex **DOI:** 10.1088/1674-1056/21/5/050204

## 1. Introduction

The detecting and imaging of objects in layered background using electromagnetic waves is an important research area with applications to geophysical exploration, environmental characterization, and subsurface sensing. In recent years, as one of the inverse scattering problems, the through-wall problem is a relatively new area that addresses the objective to detect, locate and characterize an occupant behind a wall.[1,2]

The solution of the inverse scattering problem is usually very challenging due to its inherent nonlinearity and ill-posedness. Nowadays, a wide variety of techniques have been developed to solve the inverse scattering problem. In order to circumvent the nonlinearity, some linearized techniques, like the Born or Rytov approximation are used.[3,4] However, this kind of technique is suitable only for the weak scattering objects. More recently, some iterative techniques such as Born iterative method (BIM) and the distorted Born iterative method (DBIM) are used as the inversion method to account for the inherent nonlinearity of the inverse scattering problem.[5−7] Nevertheless, for both BIM and DBIM, a numerical forward solver is required during the iterations, which is computationally expensive. In addition, some techniques such as the contrast source inversion (CSI) method are used to recast the original problem into an optimization one.[8−10] Unfortunately, nonlinear optimization methods for inverse scattering problems are often unsatisfactory. Because this kind of method requires a priori information about the connectivity or geometry of the scattering obstacle, which is usually unavailable in practical situations.

For all of the electromagnetic field inversion based algorithms mentioned above, the major difficulty is that it is relatively hard to obtain a closed-form solution from a Sommerfeld integral expression using the layered-medium Green's function in the spatial domain. This restricts the usage of the technique based on Green's function to inverse scattering problem. Therefore, an alternative technique based on a pattern recognition approach such as the support vector machine (SVM) was proposed.[11−13] Recently, a regression approach based on SVM has been widely used to deal with the inverse scattering problems such as the reconstruction of the target's dielectric permittivity in free space, the detection of the buried object, and the localization of the scatterer through a wall.[14−16] Besides, a classification approach based on

*Project supported by the National Natural Science Foundation of China (Grant No. 61071022) and the Graduate Student Research and Innovation Program of Jiangsu Province, China (Grant No. CXZZ11-0381).

†Corresponding author. E-mail: zhangyr@njupt.edu.cn

© 2012 Chinese Physical Society and IOP Publishing Ltd

SVM is used for subsurface sensing and human activity classification.[17,18]

Despite the good performance yielded by the SVM scheme, it does exhibit some disadvantages: 1) it is necessary to estimate the error/margin trade-off parameter, and in regression, the insensitivity parameter; 2) the kernel function must satisfy Mercer's condition; 3) sparsity is not always achieved, and a high number of support vectors are eventually obtained. Recently, Tipping[19] introduced a relevance vector machine (RVM) method, which is based on a Bayesian framework and has an identical functional form to the SVM. It can avoid over-fitting and provide good generalization performance. Compared with the SVM, the RVM can obtain a sparser model and use a wider kernel function. In addition, the RVM doesn't need to estimate several parameters which must be adjusted in SVM.

In this paper, an RVM-based approach to the through-wall problem is proposed. The problem is reformulated as a regression one. Thus, nonlinearity is considered and included in the relation between the scattered field and the target property, which is achieved through RVM training process. As mentioned previously, the inversion for characteristic parameters of an object based on measured scattering data is ill-posed. To address this problem, significant attention has been paid to developing the regularization techniques. Specifically, a limitation of the previous regularization procedures is the need to place a weight on the importance of the regularization relative to the importance of fitting the data to the underlying model. The RVM can produce probabilistic outputs and hence capture the uncertainty in the prediction. Consequently, rather than utilizing regularization, the ill-posed nature of the inversion is naturally accounted for, since a full-posterior density function is obtained rather than a single estimate. In this framework, the RVM-based approach deals with the nonlinearity and ill-posedness inherent in the inverse scattering problem simultaneously.

The rest of the present paper is organized as follows. In Section 2 the electromagnetic problem is described, the feature extraction is performed, and the RVM approach is briefly outlined. The simulation results obtained with RVM are given and compared with those achieved by SVM in Section 3. Finally, some conclusions and final remarks are provided.

# 2. Methodology

## 2.1. Problem description

The actual geometry of the through-wall problem is depicted in Fig. 1. The walls have relative dielectric permittivity $\varepsilon_{\rm r}$ and conductivity $\sigma$. The medium excluded wall structure is assumed to be free space whose dielectric permittivity is denoted as $\varepsilon_0$. The permeability of the wall is the same as that in free space $\mu_0$. A cylindrical scatterer residing in the investigation domain $D$ is illuminated by a transmitting antenna (TX) located at a known position near the front wall. The scattered field $\boldsymbol{E}_{\rm scat}$ is collected by a receiving antenna (RX) which moves along the walls in order to collect corresponding field at each sampling position $(x_{\rm r}, y_{\rm r})$, $r = 1, \ldots, R$ and consequently synthesizes a measurement line $L$. The spacing between two consecutive sampling positions is $d$.

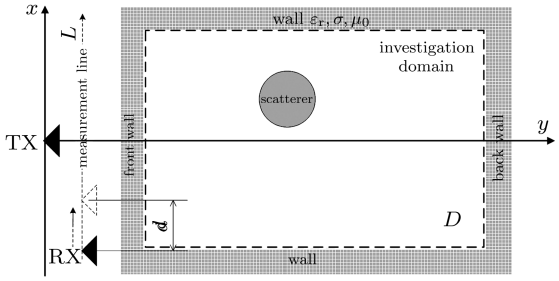



**Fig. 1.** Schematic of through-wall geometry.

Through-wall detection can be viewed as a regression issue, where it aimed at reproducing the relationship between a set of opportune features extracted from the scattered field $\boldsymbol{E}_{\rm scat}$ and the characteristic information about the target. The letter $t$ represents one of the parameters of the target such as diameter, $x$-coordinate or $y$-coordinate of the target centre, relative dielectric permittivity and conductivity. In this paper, we focus on the localization of the target through a wall (i.e., through-wall detection), thus $t$ represents the $x$-coordinate or $y$-coordinate of the target centre.

## 2.2. Feature extraction

In order to perform the regression task, some features should be extracted from the scattered field. In this paper, we only use the amplitude feature to represent the scattered field. At each sampling position, an amplitude $E^m$ is extracted from the scattered field.

Then, a vector $\boldsymbol{x} = \{E^m(x_r, y_r), r = 1, \ldots, R\}$ is formed. A solution based on the RVM will be obtained by assuming the knowledge of a set of known input-output pairs $\{(\boldsymbol{x}, t)^n; n = 1, \ldots, N\}$. These known examples are called a training dataset, and $N$ is the number of the training dataset.

## 2.3. Relevance vector machine

The structure of the RVM is described by the summation of the products of weights and kernel functions. It makes a prediction based on a function of the form

$$y(\boldsymbol{x}; \boldsymbol{w}) = \sum_{n=1}^{N} w_n K(\boldsymbol{x}, \boldsymbol{x}_n) + w_0 = \boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{x}) \cdot \boldsymbol{w}, \quad (1)$$

where $\boldsymbol{w} = [w_0, \ldots, w_N]^{\mathrm{T}}$, $\boldsymbol{\phi}(\boldsymbol{x}) = [1, K(\boldsymbol{x}, \boldsymbol{x}_1), K(\boldsymbol{x}, \boldsymbol{x}_2), \ldots, K(\boldsymbol{x}, \boldsymbol{x}_N)]^{\mathrm{T}}$, $K(\boldsymbol{x}, \boldsymbol{x}_n)$ is a kernel function which means a set of basis functions projecting the input data into a high-dimensional feature space.

By giving a set of input vectors $\{\boldsymbol{x}_n\}_{n=1}^N$ along with corresponding output scalars $\{t_n\}_{n=1}^N$, and assuming that the outputs are independent of each other and contaminated with mean-zero Gaussian noise $\varepsilon_n$ with variance $\sigma^2$, we have

$$t_n = y(\boldsymbol{x}_n; \boldsymbol{w}) + \varepsilon_n, \quad (2)$$

the likelihood of the dataset can be written as

$$p(\boldsymbol{t}|\boldsymbol{w}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \|\boldsymbol{t} - \boldsymbol{\Phi}\boldsymbol{w}\|^2\right\}, \quad (3)$$

where $\boldsymbol{t} = [t_1, \ldots, t_N]^{\mathrm{T}}$ is the target vector, $\boldsymbol{\Phi}$ is the $N \times (N + 1)$ design matrix with $\boldsymbol{\Phi} = [\boldsymbol{\phi}(\boldsymbol{x}_1), \cdots, \boldsymbol{\phi}(\boldsymbol{x}_N)]^{\mathrm{T}}$. Maximum-likelihood estimation of $\boldsymbol{w}$ and $\sigma^2$ from Eq. (3) will generally lead to severe over-fitting. To avoid this, a zero-mean Gaussian prior distribution over $\boldsymbol{w}$ with variance $\boldsymbol{\alpha}^{-1}$ is added as

$$p(\boldsymbol{w}|\boldsymbol{\alpha}) = \prod_{i=0}^{N} \mathrm{N}(w_i|0, \alpha_i^{-1}), \quad (4)$$

where hyper-parameters $\boldsymbol{\alpha} = [\alpha_0, \cdots, \alpha_N]^{\mathrm{T}}$. The introduction of an individual hyper-parameter for every weight is the key feature of the model, and is ultimately responsible for its sparsity properties.

The posterior distribution over the weight is then obtained from Bayes rule

$$\begin{aligned} & p(\boldsymbol{w}|\boldsymbol{t}, \boldsymbol{\alpha}, \sigma^2) \\ &= \frac{p(\boldsymbol{t}|\boldsymbol{w}, \sigma^2) p(\boldsymbol{w}|\boldsymbol{\alpha})}{p(\boldsymbol{t}|\boldsymbol{\alpha}, \sigma^2)} \\ &= (2\pi)^{-(N+1)/2} |\boldsymbol{\Sigma}|^{-1/2} \\ &\quad \times \exp\left(-\frac{1}{2}(\boldsymbol{w} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{w} - \boldsymbol{\mu})\right), \end{aligned} \quad (5)$$

where the posterior mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ are

$$\boldsymbol{\mu} = \sigma^{-2} \boldsymbol{\Sigma} \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{t}, \quad (6)$$

$$\boldsymbol{\Sigma} = (\sigma^{-2} \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi} + \boldsymbol{A})^{-1}, \quad (7)$$

respectively, with $\boldsymbol{A} = \mathrm{diag}(\alpha_0, \alpha_1, \cdots, \alpha_N)$.

The likelihood distribution over the training targets can be marginalized with respect to the weights to obtain the marginal likelihood, which is also Gaussian distributed as follows:

$$\begin{aligned} p(\boldsymbol{t}|\boldsymbol{\alpha}, \sigma^2) &= \int p(\boldsymbol{t}|\boldsymbol{w}, \sigma^2) p(\boldsymbol{w}|\boldsymbol{\alpha}) \mathrm{d}\boldsymbol{w} \\ &= (2\pi)^{-N/2} |\boldsymbol{C}|^{-1/2} \\ &\quad \times \exp\left\{-\frac{1}{2} \boldsymbol{t}^{\mathrm{T}} \boldsymbol{C}^{-1} \boldsymbol{t}\right\} \end{aligned} \quad (8)$$

with covariance $\boldsymbol{C} = \sigma^2 \boldsymbol{I} + \boldsymbol{\Phi} \boldsymbol{A}^{-1} \boldsymbol{\Phi}^{\mathrm{T}}$.

Values of $\boldsymbol{\alpha}$ and $\sigma^2$ that maximize the marginal likelihood cannot be obtained in closed form, and an iterative re-estimation method is required. The iterative procedure is as follows:

$$\alpha_i^{\mathrm{new}} = \frac{\gamma_i}{\mu_i^2}, \quad (9)$$

$$(\sigma^2)^{\mathrm{new}} = \frac{\|\boldsymbol{t} - \boldsymbol{\Sigma}\boldsymbol{\mu}\|^2}{N - \sum_i \gamma_i}, \quad (10)$$

with $\mu_i$ being the $i$th posterior means weight (6) and the quantities $\gamma_i = 1 - \alpha_i \boldsymbol{\Sigma}_{ii}$ being the $i$th diagonal element $\boldsymbol{\Sigma}_{ii}$ of the posterior weight covariance (7).

In the iterative maximization of marginal likelihood, many of hyper-parameter $\alpha_i$ tend to infinity, yielding a posterior distribution (5) of the corresponding weight $w_i$ that tends to be a delta function centred around zero. The corresponding weight is thus deleted from the model, along with its associated basis function. The remaining terms with nonzero associated weight are called the relevance vectors.

Given an input $\boldsymbol{x}_*$, the probability distribution of the corresponding output is given by the predictive distribution

$$\begin{aligned} & p(y_*|\boldsymbol{x}_*, \hat{\boldsymbol{\alpha}}, \hat{\sigma}^2) \\ &= \int p(y_*|\boldsymbol{x}_*, \boldsymbol{w}, \hat{\sigma}^2) p(\boldsymbol{w}|y, \hat{\boldsymbol{\alpha}}, \hat{\sigma}^2) \mathrm{d}\boldsymbol{w}, \end{aligned} \quad (11)$$

which has a Gaussian form, $p(y_*|\boldsymbol{x}_*, \boldsymbol{w}, \hat{\sigma}^2) = N(y_*, \sigma_*^2)$, where the mean and the variance of the prediction are

$$y_* = \boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{x}_*)\boldsymbol{\mu}, \tag{12}$$

$$\sigma_*^2 = \hat{\sigma}^2 + \boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{x}_*)\boldsymbol{\Sigma}\boldsymbol{\phi}(\boldsymbol{x}_*), \tag{13}$$

respectively. Here, $\hat{\boldsymbol{\alpha}}$ and $\hat{\sigma}^2$ are the estimations of hyper-parameters $\boldsymbol{\alpha}$ and noise $\sigma^2$respectively.

# 3. Simulation and results

## 3.1. Dataset description

To validate this proposed method based on RVM for through-wall detection experimentally, we conducted numerical experiments using finite-difference time-domain (FDTD) to collect the backscattered field at each observation location of receiving antenna. In our simulation, a metallic circular cylinder residing in a room is taken into account as shown in Fig. 1. The thickness of the walls is 0.2 m. The investigation domain is $D = [-1.08, 1.08] \times [0.25, 3.64]$ (unit: $\mathrm{m}^2$). The conductivity $\sigma$ and the relative permittivity $\varepsilon_{\mathrm{r}}$ of the walls are set to be standard values of 0.01 S/m and 8, respectively. The geometry of the room is discretized with the FDTD square cells of 1 cm in length. The time resolution is $19.25 \times 10^{-12}$ s. The TX is fixed at the position of (0, 0) which is 0.05 m away from the front wall outside the room. To model the electromagnetic (EM) illumination of the modeled room and the objects in it with an ultra wideband (UWB) short pulse, the transmitter dipole antenna is fed by a 4.5 ns Gaussian pulse modulated by a 0.5 GHz sine wave. The RX is 0.04 m away from the front wall, which moves and then synthesizes a measurement aperture $L = 2.4$ m. At each observation point, the scattered field from the target is obtained by subtracting, from the total electric field, the electric field computed in the absence of the scatterer. Then, a feature vector $\boldsymbol{x}$ is achieved after feature extraction procedure.

The RVM must be trained by exploiting some input–output patterns appropriately chosen so as to give an exhaustive representation of the through-wall detection problem being investigated. To do this the training dataset is achieved by repeated simulations with varying position of the target centre, as follows:

$$\begin{aligned} &x_n = -1.13 + n\Delta x, n = 0, 1, \ldots, \Delta x = 0.1 \text{ m}, \\ &y_n = 0.2 + n\Delta y, n = 0, 1, \ldots, \Delta y = 0.1 \text{ m}. \end{aligned} \tag{14}$$

In the same way, for testing samples, the positions of the target centre are made to vary in such a way

$$\begin{aligned} &x_n = -1.1 + n\Delta x, \quad n = 0, 1, \ldots, \quad \Delta x = 0.1 \text{ m}, \\ &y_n = 0.25 + n\Delta y, \quad n = 0, 1, \ldots, \quad \Delta y = 0.1 \text{ m}. \end{aligned} \tag{15}$$

Thus, the training and the testing datasets are obtained.

In the RVM regression, the kernel function is selected as the Gaussian kernel, which is given by

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\gamma \left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|^2), \tag{16}$$

where $\gamma$ is the variance of the kernel function which is determined using cross-validation in the training phase.

## 3.2. Results

In order to evaluate the effectiveness of RVM-based method for through-wall detection, after the training process of RVM, we randomly select 100 samples from the testing dataset to conduct the testing procedure. Since the RVM can produce probabilistic output, the predicted results with 95% confidence area are displayed in Figs. 2 and 3 for $x$-coordinate and $y$-coordinate of the target centre, respectively.

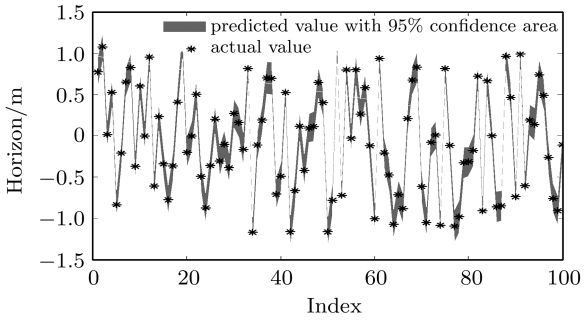



**Fig. 2.** Predicted results with 95% confidence area for $x$-coordinate of the target centre.

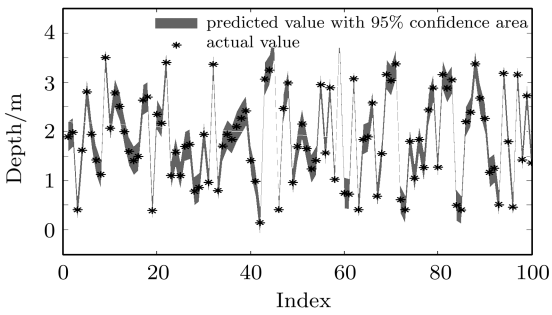



**Fig. 3.** Predicted results with 95% confidence area for $y$-coordinate of the target centre.

As shown in Figs. 2 and 3, the actual values of the target centre coordinates do not exceed the confidence area (95%). Therefore, the RVM-based approach can successfully estimate $x$-coordinate and $y$-coordinate of a target through walls.

In order to assess the accuracy of the RVM-based algorithm quantitatively, we define the relative mean square error (RMSE) for the position, i.e., $x$-coordinate and $y$-coordinate as

$$R_x = \sqrt{\frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \left( \frac{x_i - \tilde{x}_i}{l_x} \right)^2} \tag{17}$$

$$R_y = \sqrt{\frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \left( \frac{y_i - \tilde{y}_i}{l_y} \right)^2} \tag{18}$$

where $N_{\text{test}} = 100$ is the number of the samplings for testing procedure. $(x_i, y_i)$ and $(\tilde{x}_i, \tilde{y}_i)$ are the $i$th actual and predicted positions of the target centre respectively. $l_x$ and $l_y$ denote the total horizontal value and depth of room respectively. Firstly, the RMSE against the number of training samplings $N$ obtained using the RVM are illustrated in Fig. 4 , and the results obtained using the SVM are also plotted there for comparison. It shows that more than 300 samplings for training procedure are enough to converge for the RVM and the SVM predictors.

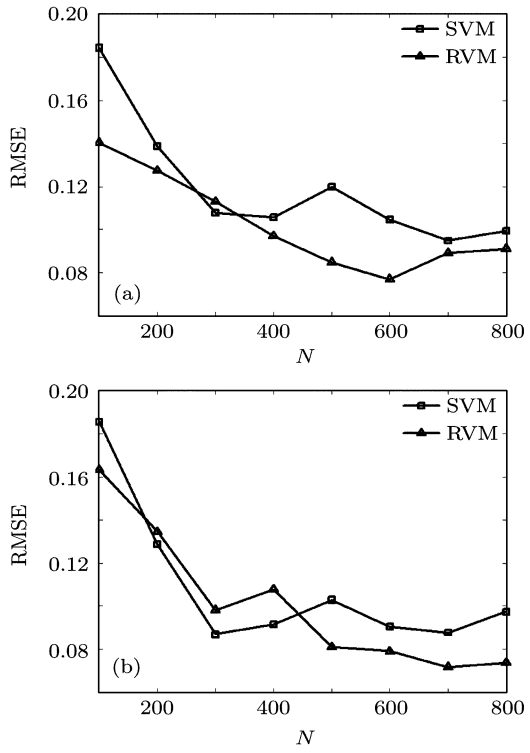



**Fig. 4.** Curves for RMSE versus the number of samples $N$ in training dataset for (a) $x$-coordinate and (b) $y$-coordinate of the target centre.

Then, we analyse the effect of the number of observation positions $R$ on predicted accuracy achieved by the RVM and compare the results with those achieved by the SVM. The results are depicted in Fig. 5. When $R$ is less than 40, the results obtained using RVM are comparable to those obtained using SVM. However, when $R$ exceeds 40, the RVM-based method is superior to the SVM-based method for through-wall target detection.

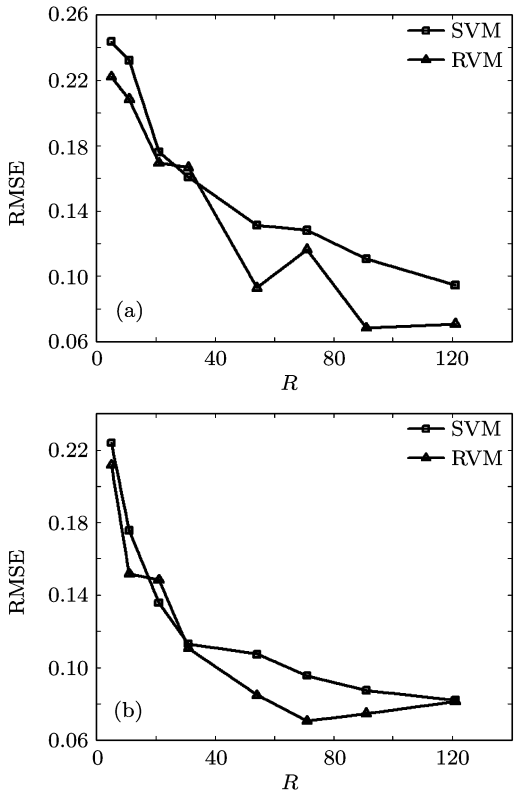



**Fig. 5.** Curves for RMSE versus the number of observation positions $R$ for (a) $x$-coordinate and (b) $y$-coordinate of the target centre.

The above-mentioned is for a noiseless circumstance. In reality, it is important that the reconstruction algorithm should be robust under noise contamination. Consequently, the scattered field is corrupted by an additive Gaussian noise with zero mean and fixed variance according to the desired signal-to-noise ratio (SNR). The RMSE as a function of SNR is exhibited in Fig. 6. It reveals that the RVM-based algorithm presents a competitive RMSE performance with the SVM-based algorithm at any level of SNR.

To confirm the sparser property of the RVM than that of the SVM, the relevance vectors (RVs) and support vectors (SVs) for the RVM-based and the SVM-based methods at given SNRs are listed in Table 1. It

shows that the number of RVs is smaller than that of SVs. Therefore, the computational complexity of the RVM predictor is below that of the SVM predictor.

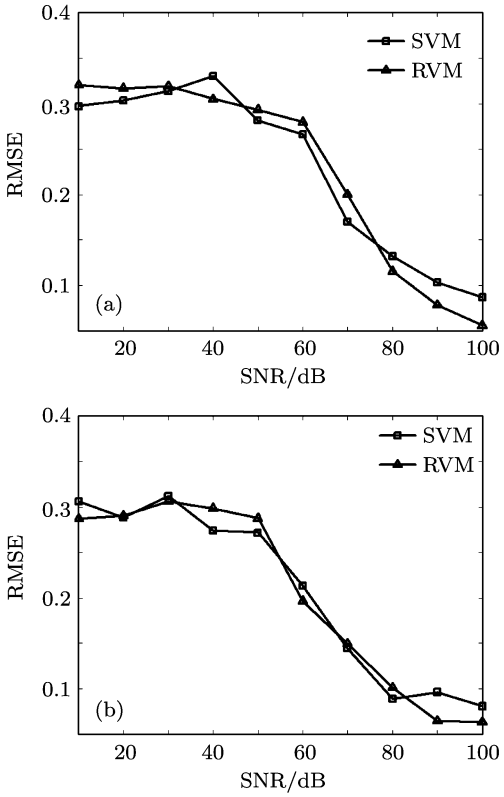



**Fig. 6.** Curves for RMSE versus SNR for (a) $x$-coordinate and (b) $y$-coordinate of the target.

**Table 1.** Comparison between the number of relevance vectors (RVs) and that of support vectors (SVs).

| SNR / dB | Horizon | | Depth | |
|---|---|---|---|---|
| | RVs | SVs | RVs | SVs |
| 10 | 43 | 700 | 83 | 758 |
| 20 | 50 | 700 | 41 | 761 |
| 30 | 45 | 700 | 52 | 759 |
| 40 | 49 | 738 | 51 | 748 |
| 50 | 48 | 704 | 63 | 743 |
| 60 | 58 | 706 | 115 | 718 |
| 70 | 183 | 628 | 183 | 669 |
| 80 | 292 | 503 | 193 | 581 |
| 90 | 284 | 405 | 284 | 484 |
| 100 | 272 | 372 | 272 | 452 |

# 4. Conclusion

In order to detect and locate the target behind a wall, a novel method based on RVM is proposed. Our results confirm that the RVM predictor can closely match the SVM predictor performances in terms of accuracy, convergence, robustness, and generalization. However, the RVM-based approach has a much sparser kernel representation than the SVM-based approach. Therefore, the computational time taken by the RVM algorithm is less than that of the SVM algorithm. In such a framework, the proposed method in this paper is more suitable for real-time detection, which is important in through-wall scenarios. The RVM method involves a nonlinear iterative optimization process. Nevertheless, we use the RVM method to train the dataset "off-line" to approximate the data-unknowns relation.

# References

[1] Ferris D D and Currie N C 1999 *Proc. SPIE* **3577** 62

[2] Baranoski E J 2008 *J. Frank. Inst.* **345** 556

[3] Soldovieri F and Solimene R 2007 *IEEE Geosc. Rem. Sens. Lett.* **4** 513

[4] Soldovieri F, Solimene R and Prisco G 2008 *IEEE Trans. Geosci. Remot. Sens.* **46** 1192

[5] Chew W C and Wang Y M 1990 *IEEE Trans. Med. Imag.* **9** 218

[6] Cui T J, Chew W C, Aydiner A A and Chen S Y 2001 *IEEE Trans. Geosci. Remot. Sens.* **39** 339

[7] Li F H, Liu Q H and Song L P 2004 *IEEE Geosc. Rem. Sens. Lett.* **1** 107

[8] Song L P, Yu C and Liu Q H 2005 *IEEE Trans. Geosci. Remot. Sens.* **43** 2793

[9] Song L P, Liu Q H, Li F H and Zhang Z Q 2005 *IEEE Trans. Antennas Propagat.* **53** 1556

[10] Zhang Z Q and Liu Q H 2004 *IEEE Trans. Biomed. Eng.* **51** 544

[11] Smola A J and Scholkopf B 2004 *Stat. Comput.* **14** 199

[12] Sun Z H and Jiang F 2010 *Chin. Phys.* B **19** 110502

[13] Meng Q F, Chen Y H and Peng Y H 2009 *Chin. Phys.* B **18** 2194

[14] Zhang Q H, Xiao B X and Zhu G Q 2007 *Microwave Opt. Technol. Lett.* **49** 372

[15] Bermani E, Boni A, Caorsi S and Massa A 2003 *IEEE Trans. Geosci. Remot. Sens.* **41** 927

[16] Wang F F and Zhang Y R 2011 *J. Electromagn. Waves Appl.* **25** 75

[17] Massa A, Boni A and Donelli M 2005 *IEEE Trans. Geosci. Remot. Sens.* **43** 2084

[18] Kim Y and Ling H 2009 *IEEE Trans. Geosci. Remot. Sens.* **47** 1328

[19] Tipping M E 2001 *J. Mach. Learn. Res.* **1** 211