Signal compensation of AC MMW radiometer based on DCT and RVM

Luan Yinghong, Li Yuehua, Luo Lei
(Institute of Near Sensing Technique with Millimeter Wave & Optical-wave, Nanjing University of Science & Technique, Nanjing 210094, China)

Abstract: Aimed at the shortcoming that the loss of low-frequency information of alternating current millimeter-wave radiometer signal, relevance vector machine (RVM) algorithm is used to compensate the lost component in discrete cosine transform (DCT) domain, and through inverse discrete cosine transform (IDCT) we can receive the compensated signal. RVM exploits Bayesian learning framework, which has dramatically fewer kernel functions than comparative support vector machine. So that accurate prediction models can be acquired. Experimental results also show that this method can obtain good compensation effect.

Key words: alternating current radiometer; discrete cosine transform; relevance vector machine

1 Introduction

The millimeter-wave (MMW) radiometer is a high-sensitive detector for millimeter wave passive detection, which is widely applied to military, remote sensing, civil areas and so on[1]. Alternating current (AC) radiometer use direct current (DC) isolation circuit to filter the system noise in signal. But simultaneously DC isolation circuit causes the distortion of AC signal, which contains target information. The loss of part of energy will cause more trouble in feature extraction and target identification. Ref. [2] researched low-frequency compensation of AC signal of the MMW total power radiometer through transcendental function. As the AC signal can be considered as Gaussian signal, with the information of undistorted parts, the whole spectrum can be recovered through solving the nonlinear equations by separately real and imaginary parts.

Therefore a compensation algorithm is designed for AC radiometer signal. After discrete cosine transform (DCT) of signal, relevance vector machine (RVM) is used to predict the filtered low-frequency ingredient in DCT domain and then utilize inverse discrete cosine transform (IDCT) to get the compensated radiometer signal. After DCT, RVM algorithm only carries on real part, but the effect is actually almost or fairly good. It simplifies the process of compensation and obtains better results.

2 Discrete cosine transform

Ahmed introduced a definition of DCT, that is: for a given signal $s(n)$ with length of M, its cosine transforming coefficient $f(k)$ is[3]:

$$f(k) = a(k)\sum_{n=0}^{M-1} s(n)\cos\left[\frac{(2n+1)k\pi}{2M}\right],\quad 0\leqslant k\leqslant M-1$$

Where, $a(k)=\begin{cases}(1/M)^{1/2}, & k=0\\ (2/M)^{1/2}, & k\neq 0\end{cases}$, $0\leqslant k\leqslant M-1$

From above definition, it can be seen that DCT of input signal is to decompose the sequence into linear adding of a group of basic cosine elements. If $s(n)$ is real, then its DCT is also real. Compared with discrete Fourier transform (DFT), DCT avoids the complex operation[4,5].

The inverse transform is[6]:

$$s(n) = \sum_{k=0}^{M-1} a(k)f(k)\cos\left[\frac{(2n+1)k\pi}{2M}\right],\quad 0\leqslant n\leqslant M-1$$

Fig. 1(a) shows the DC MMW radiometer signal and AC MMW radiometer signal whose frequency is no less than 7 Hz is in Fig. 1(b). Fig. 1(c) is the DCT coefficients (DCTC, first 50 sequences) of DC MMW radiometer and the same for AC signal is given in Fig. 1(d). There is difference in smaller sequence between Fig. 1(c) and Fig. 1(d), but they are nearly the same in higher sequence. So, the difference of DCTC is due to the loss of low-frequency component, and the loss information can be compensated through

Received 11 *October* 2007

high-frequency component.

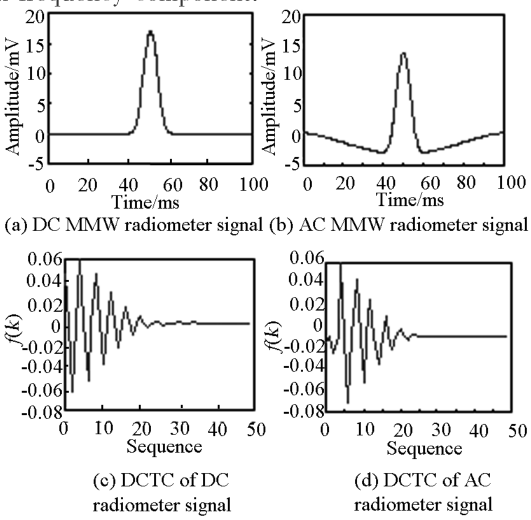


Fig. 1 DC and AC signals and their DCT coefficients

3 Relevance vector machine (RVM)

RVM is a nonlinear machine learning algorithm based on positive nucleus which is proposed by Michael E. Tipping[7, 8]. It exploits Bayesian learning framework. Compared with support vector machine (SVM), RVM has sparser solutions. Besides, RVM hasn't needed to adjust any parameters of model, and doesn't have any restriction on the selection of kernel function.

Tipping took Lagrange coefficient vector $\boldsymbol{w}$ as weight parameters to establish corresponding significant framework, which inferred datum corresponding non-zero w_i and called relevance vector (RV). RV is equal to the support vector in SVM. For a same dataset, the number of RV is much fewer than SV, so the test time used by RVM is much shorter than that of SVM.

3.1 Model specification

The RVM makes predictions based upon the function $y(x)$ defined over the input space. A flexible and popular set of candidates for $y(x)$ is that of the form:

$$y(x, w) = \sum_{i=1}^{N} w_i \Phi_i(x) = w^{\mathrm{T}} \Phi(x) \quad i = 1, \ldots, N \tag{1}$$

Where, $\Phi(x)$ is nonlinear kernel function and w is the model weights. Gaussian function of every datum is chosen as kernel function. Learned from minimize structure risk principle, that is, if the parameters are not defined but maximizing likelihood function will lead to severe over-fitting. A prior probability distribution $p(w_j | \alpha_j)$ is defined for each weight in RVM:

$$p(w_j | \alpha_j) = \left(\frac{\alpha_j}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\alpha_j w_j^2\right) \tag{2}$$

Where, α is the hyper parameter which is associated with the distribution of weight w[9].

Given a dataset of input-target pairs $\{x_i, t_i\}_{i=1}^{N}$, together with input vectors $\boldsymbol{X} = \{x_i\}$ and corresponding target values $T = \{t_i\}$, so that $t_i = y(x_i, w) + \varepsilon_i$. Suppose the targets t_i are independent and the noise ε_i is assumed to be mean-zero Gaussian with variance σ^2. The likelihood $p(\boldsymbol{t} | \boldsymbol{w}, \sigma^2)$ of the dataset can be written as:

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

Where, $\boldsymbol{t} = (t_1, \ldots, t_N)^{\mathrm{T}}$, $\boldsymbol{w} = (w_1, \ldots, w_N)^{\mathrm{T}}$, Φ is matrix whose lines are the response to every input x_i with $\Phi_i = [1, \Phi_1(x_i), \ldots, \Phi_N(x_i)]$.

3.2 Inference

Having defined the prior and likelihood, from Bayes rule, the posterior of weight w, $P(w | \boldsymbol{t}, \alpha, \sigma^2)$ is given the data:

$$P(w | \boldsymbol{t}, \alpha, \sigma^2) = p(\boldsymbol{t} | w, \sigma^2) \frac{p(w | \alpha)}{p(\boldsymbol{t} | \alpha, \sigma^2)} \tag{4}$$

Where, the denominator is not related to w, so the posterior distribution over the weight is submit ted to Gaussian distribution, that is:

$$P(w | \boldsymbol{t}, \alpha, \sigma^2) = N(\mu, \Sigma) \tag{5}$$

with mean and covariance:

$$\mu = \sigma^{-2} \Sigma \Phi^{\mathrm{T}} \boldsymbol{t}, \quad \Sigma = (\sigma^2 \Phi^{\mathrm{T}} \Phi + \boldsymbol{A})^{-1} \tag{6}$$

Where, $\boldsymbol{A}$ is diagonal matrix $(a_0, \ldots, a_N)$, when $a_i \to \infty$, $\mu_i = 0$.

3.3 Optimizing the hyper parameters

According to Bayesian framework, marginal likelihood distribution of hyper parameters is written as:

$$P(\boldsymbol{t} | \alpha, \sigma^2) = \oint p(\boldsymbol{t} | w, \sigma^2) \cdot p(w | \alpha) \, \mathrm{d}w = N(0, \boldsymbol{C}) \tag{7}$$

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

Values of α_{MP} and σ^2_{MP} can be estimated through maximizing the likelihood distribution of hyper parameters, but Eq. (7) cannot be obtained by analytic form, so iterative re-estimation is considered. For α, differentiation of Eq. (7), equating to zero and rearranging, following the approach of MacKay[10], gives:

$$\alpha_i^{\mathrm{new}} = \gamma_i / \mu_i^2 \tag{8}$$

Where, $\gamma_i \equiv 1 - \alpha_i \Sigma_{ii}$, μ_i is the i^{th} posterior mean weight from Eq. (6) and Σ_{ii} is the i^{th} diagonal element of the posterior weight covariance from Eq. (6) computed with the current α and σ^2 values. The same with noise variance σ^2, differentiation leads to the re-estimate:

$$(\sigma^2)^{new} = |\boldsymbol{t} - \Phi\mu|^2 / (N - \sum_{i=1}^{N} \gamma_i),$$
$$i = 1, \ldots, N \quad (9)$$

The learning algorithm thus proceeds by repeated application of Eq. (8) and Eq. (9), concurrent with updating Eq. (6), until some suitable convergence criteria have been satisfied. In practice, during re-estimation, many of the α_i are generally tended to infinity. From Eq. (5), this implies that $P(w|\boldsymbol{t},\alpha,\sigma^2)$ becomes highly peaked at zero and those corresponding w_i are posteriori zero. The corresponding basis function can thus be "pruned", and sparsity is realized.

3.4 Making predictions

At convergence of the hyper parameter estimation procedure, predictions are made based on the posterior distribution over the weights, conditioned on the maximizing values α_{MP} and σ_{MP}^2, the predictive distribution can then be computed. For a new datum x^*:

$$P(\boldsymbol{t}^* | \boldsymbol{t}, \alpha_{MP}, \sigma_{MP}^2) = \oint p(\boldsymbol{t}^* | w, \sigma_{MP}^2) \cdot p(w | \boldsymbol{t}, \alpha_{MP}, \sigma_{MP}^2) \mathrm{d}w \quad (10)$$

Since both terms in the integrand are Gaussian, this is readily computed, giving:

$$P(\boldsymbol{t}^* | \boldsymbol{t}, \alpha_{MP}, \sigma_{MP}^2) = N(\mu^*, \sigma^{2*}) \quad (11)$$

with predictive mean and variance:

$$\mu^* = \mu^{\mathrm{T}} \Phi(x^*)$$
$$\sigma^{2*} = \sigma_{MP}^2 + \Phi^{\mathrm{T}}(x^*) \Sigma \Phi(x^*) \quad (12)$$

So the predictive mean is intuitive $y(x^*, \mu)$. A probabilistic Bayesian learning framework is utilized in RVM to solve the selection of model parameters, which has better applicability. RVM used into prediction can possess more perfect performance.

4 Compensate algorithm

Fig. 2 is the flow chart of compensate algorithm. Discrete cosine transformation is taken on processing signal and can receive a DCT sequence $\{f_1, \ldots, f_N\}$. DCT is closely related with DFT, so that the signification of DCT sequence is similar to DFT and can be seen as a frequency spectrum.

According the low frequency of DCT sequence is almost zero, it is made all-zeros and use the high-frequency left to train RVM. How to distinguish low or high frequency? Looking for the serial number n correspond to the maximum absolute value D_{max} in signal DCT coefficients sequence. The coefficient whose serial number bigger than n is defined as high-frequency sequence. High-frequency $\{f_{n+1}, \ldots, f_N\}$ is used to training RVM network. For more information, input/output is written as matrix form:

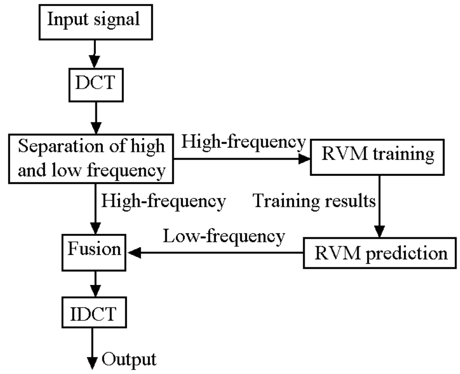


Fig. 2 Flow chart of compensation algorithm

$$\boldsymbol{X} = \begin{bmatrix} f_{n+1} & \cdots & f_l \\ f_{n+2} & \cdots & f_{l+1} \\ \vdots & & \vdots \\ f_{N+n-l} & \cdots & f_{N-1} \end{bmatrix}, \boldsymbol{Y} = \begin{bmatrix} f_{l+1} \\ f_{l+2} \\ \vdots \\ f_N \end{bmatrix},$$
$$n + 1 < l < N$$

If non-linear basis function is chosen as Gaussian function ($\sigma^2 = 1$), RVM network can be expressed as:

$$y_i = (x, w^*) = \sum_{j=1}^{N-n} w_j^* \exp(-|x_j - x_k|) + w_0^*,$$

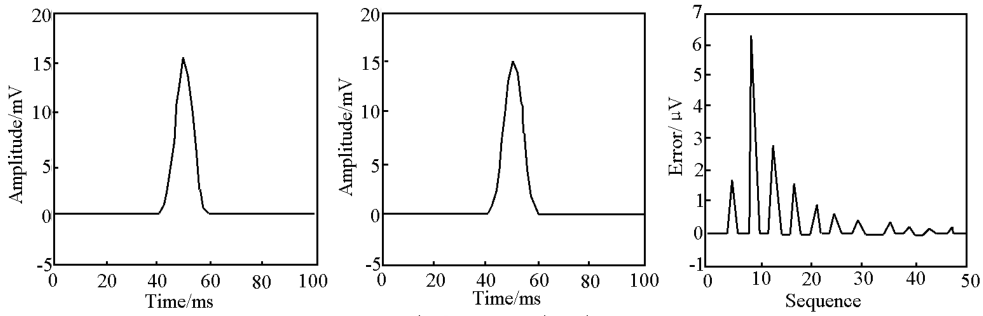
where $k = l + 1, \ldots, N$. After that, the trained RVM network is used to predict low-frequency sequence $\{f_1^*, \cdots, f_n^*\}$.

At last, high-frequency of signal sequence and low-frequency of signal sequence integrate into one sequence $\{f_1^*, \ldots, f_n^*, f_{n+1}, \ldots, f_N\}$. IDCT is taken on the integrated sequence and the compensated results can be obtained.

5 Experiment result

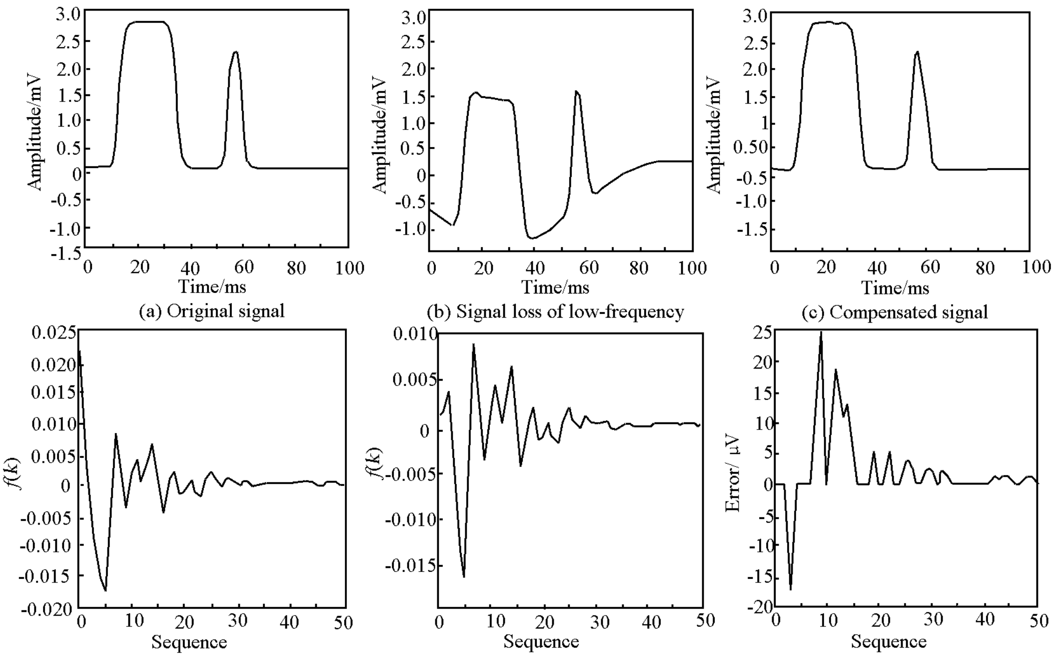
Fig. 3(a) shows the original DC radiometer signal and Fig. 3(b) is the compensated signal. The error of compensated and original signal in DCT domain displays in Fig. 3(c). The pictures show that the error between them is very small. The computed MSE is 4.52×10^{-6}, which is 3 ~ 4 measure levels smaller than the original signal. Compared with Ref. [2], the process of transform, prediction and re-transform is much simple, and the results can also be satisfied to practical use.

Fig. 4 is the experiment of platform signal. Fig. 4(a) is the original platform signal and Fig. 4(b) is filtered signal. The compensated signal is shown in Fig. 4(c). Fig. 4(d) and Fig. 4(e) is the DCT coefficients of the two signals and the error is in Fig. 4(f). From above, we can see the compensated signal is similar to original signal totally. Because of the error of high-frequency ingredient in DCT, the signal is not compensated very well somewhere.



(a) DC radiometer signal (b) Compensated results (c) Error of compensated and original signal in DCT domain

Fig. 3 Compensation results



(a) Original signal (b) Signal loss of low-frequency (c) Compensated signal

(d) DCT coefficients of original signal (e) DCTC of signal loss of low-frequency (f) Error in DCT domain

Fig. 4 Compensation results of platform signal

6 Conclusions

DC isolation network of AC radiometer removes low-frequency ingredient of MMW radiometer signal, but causes the distortion of signal at the same time. In order to overcome this disadvantage of AC radiometer, we propose a compensated method which exploits RVM to predict low-frequency coefficients of signal in DCT domain. After the fusion of the high and low frequency coefficients, IDCT is taken to obtain the AC MMW radiometer signal. RVM exploits Bayesian learning framework. It has excellent generalization properties and good prediction results. Therefore, signal compensation system designed in this paper can compensate the low-frequency ingredient of millimeter-wave radiometer signal effectively.

References

[1] Li Xingguo. Millimeter-wave Near Sensing Technique and Its Application[M]. Beijing: National Defence Industry Press, 1991. (in Chinese)

[2] Qian Songsong, Li Xingguo. Single recovery of AC total power MMW radiometer[J]. Journal of Nanjing University of Science and Technology, 2007, 31(3):350-354. (in Chinese)

[3] Ahmed N, Natarajan T, Rao K R. Discrete cosine transform[J]. Computers, 1974, 23(1): 90-93.

[4] Mansour M F. On the odd-DFT and its applications to DCT/IDCT computation[J]. Signal Processing, 2006, 54(7):2819-2822.

[5] Komatsu K, Sezaki K. Reversible discrete cosine transform[C]. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1998, 3: 1769-1772.

[6] Feng Chunhuan, Tu Jianping, Guo Jian. Infrared targets recognition algorithm based on discrete cosine transformation[J]. Journal

of System Simulation, 2005, 17(6): 1363-1365.

[7] Tipping M E. The relevance vector machine[A]. Advances in Neural Information Processing Systems[C]. Cambridge: MIT Press, 2000. 652-658.

[8] Tipping M E. Sparse Bayesian learning and the relevance vector machine[J]. Journal of Machine Learning Research, 2001, 1(3):211-244.

[9] Chen S, Gunn S R, Harris C J. The relevance vector machine technique for channel equalization application[J]. IEEE Trans on Neural Networks, 2002, 12(6):1529-1532.

[10] MacKay. Bayesian interpolation[J]. Neural Computation, 1992, 4(3):415-447.

Author

Luan Yinghong, female, was born in 1983. She is now working towards the Ph. D. degree in electronic engineer & optoelectronic technology in Nanjing University of Science & Technique. Her current research interests include signal detection and target identification.

Li Yuehua, male, was born in 1959, Ph. D. tutor and professor of the Institute of Near Sensing Technique with Millimeter Wave & Optical-wave in NUST. His research interests include millimeter wave near sensing detection and target recognition.

Foundation item: National Defence Foundation under Grant No. 9140A05070107BQ0204

(cont. from p. 71)

[7] Mayseless M, Erlich Y. Interaction of shaped charge jets with reactive or passive cassettes[A]. 14th ISB[C]. Canada, 1993:439-447.

[8] Held M. Stopping power of explosive reactive armours against different shaped charge diameters or at different angles[J]. Propellants, Explosives, Pyrotechnics, 2001, 26:97-104.

Author

Li Rujiang, male, born in 1978, graduated from USTC and now is a tutor in North University of China, Taiyuan, China. Mr. Li has published over 10 papers. His current research is modern explosion mechanics. He can be reached by E-mail:liru7841@mail.ustc.edu.cn