[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice–Hall, 1998.
[9] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2006.
[10] L. S. Gradshteyn and L. M. Ryzhik, *Table of Integrals, Series, and Products*. New York: Academic, 2000.
[11] W. S. Jeon, D. G. Jeong, J. A. Han, G. Ko, and M. S. Song, "An efficient quiet period management scheme for cognitive radio systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 505–509, Feb. 2008.

Target Tracking Using Particle Filters With Support Vector Regression

Nihat Kabaoğlu, *Member, IEEE*

***Abstract*—This paper presents a numerical Bayesian approach for the direction-of-arrival (DOA) tracking of multiple targets using a linear and passive sensor array. In this paper, support vector regression (SVR) method is employed, together with particle filters (PFs), to obtain an effective proposed distribution utilizing observed phenomena to propose a new sample. Two PF algorithms are presented: One is based on SVR for a large sample set, and the other is based on sequential SVR for a small sample set. The simulation results present the superiority of the proposed method while considering a small sample set and show that it is also competitive when a large sample set is considered.**

***Index Terms*—Direction-of-arrival (DOA) tracking, particle filters (PFs), support vector regression (SVR).**

I. Introduction

Target problems have attracted great interest in many different applications, such as radar, sonar, acoustics, image processing, automatic large-area surveillance, and wireless sensor networks [1]–[20]. Since particle filter (PF)-based methods are adequate for online processing, they have recently been successfully applied to target-tracking problems [10]–[20]. In the PF methods, how to select the importance density is an important issue. As the importance density is, for simplicity, mostly selected as the prior distribution, it is therefore ineffective and leads to poor filtering performance. In such a case, since the importance density offers no information from the current observation from which to propose new samples, the state space is explored without the knowledge of the observation, which causes that filter to be sensitive to the outliers. Effective importance density-selection methods proposed by different researchers can be found in the literature, such as [14], [15], and [20]. However, they have some disadvantages, namely, the implementation of the method in [14] is usually quite tricky. In the case of large process noise, the method in [15] degrades the performance. The method in [20] requires matrix inversion, gradient, and Hessian computations.

In this paper, a PF method based on support vector regression (SVR) is proposed as a tracker within the sequential Monte Carlo framework. A similar method for a small sample set was proposed to track wideband moving sources in [6], where SVR was employed as a preestimator because of the structure of the signal model, which requires buffered data. Differently from the method in [6], in the method proposed in this paper, SVR is used to predict a mean value of the particles to reach an acceptable approximation for the optimal importance density. This method is proposed as two algorithms, for both a large sample set and a small sample set. Furthermore, the performance of both proposed algorithms and their comparison with methods such as auxiliary sampling importance resampling (ASIR), the Gaussian sum PF (GSPF), the reversible jump Markov chain Monte Carlo (RJMCMC), the method in [5], and the method in [20] are investigated in terms of the narrow-band target-tracking problem. Moreover, in this method, if an autoregressive model is constituted for time evolving for the number of sources, the number of sources and their direction of arrival (DOA) can jointly be estimated. Another PF method using SVR was proposed for the visual tracking problem in [5], where an algorithm is presented for tracking an object in a video sequence. In the algorithm in [5], SVR is utilized to reweigh the particles sampled by one of the classic resampling methods. Since this method directly focuses on the particles and their weights, SVR may destroy the convergence of the particles.

This paper is organized as follows: Section II gives the state-space model. In Section III, the joint posterior density function for constituting the methods and algorithms used in this paper is obtained. Section IV-A presents the proposed method for a large sample set. The proposed method for a small sample set is given in Section IV-B. In Section V, simulation results are presented. After an analysis of the simulation results, a discussion of the proposed method is presented in Section VI.

II. Data Model

In this paper, a data model is used under the narrow-band assumption. A scenario is considered in which M signal wavefronts from unknown time-varying directions $\boldsymbol{\theta}(t) = [\theta_1(t), \theta_2(t), \ldots, \theta_M(t)]^T$ received onto a uniform linear array of K sensors. The time-varying direction vector is modeled as

$$\boldsymbol{\theta}(t) = \boldsymbol{\theta}(t-1) + \mathbf{w}(t) \tag{1}$$

where $\mathbf{w}(t)$ is normally distributed with zero mean and the covariance matrix $\sigma_w^2\mathbf{I}$. Here, σ_w^2 represents the noise spectral parameter, and $\mathbf{I}$ is the identity matrix. The emitted signals are assumed that they evolve according to

$$\mathbf{s} \sim \mathcal{N}\left(\mathbf{0}, \sigma_n^2\mathbf{I}\right) \tag{2}$$

where $\mathbf{s}(t) = [s_1(t), \ldots, s_M(t)]^T \in \mathcal{C}^{M\times 1}$, and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

For a collection of observed output data of K sensors in the array, $\mathbf{x}(\mathbf{t}) = [\mathbf{x}_1(t), \ldots, \mathbf{x}_K(t)]^T \in \mathcal{C}^{K\times 1}$, the matrix formulation of outputs of the sensors at time t is obtained as follows:

$$\mathbf{x}(t) = \mathbf{A}\left(\boldsymbol{\theta}(t)\right)\mathbf{s}(t) + \mathbf{n}(t), \qquad t = 0, T_s, \ldots, (N-1)T_s. \tag{3}$$

The time-varying steering matrix

$$\mathbf{A}\left(\boldsymbol{\theta}(t)\right) = \left[\mathbf{a}\left(\theta_1(t)\right), \ldots, \mathbf{a}\left(\theta_M(t)\right)\right], \qquad 1 \le m \le M \tag{4}$$

consists of M steering vectors $\mathbf{a}(\theta_m(t)) \in \mathcal{C}^{K\times 1}$. $M < K$ is assumed to avoid ambiguity. In addition, the measurement noise process $\mathbf{n}(t) \in \mathcal{C}^{K\times 1}$ is independent identically complex and normally distributed with zero mean and covariance matrix $\sigma_n^2\mathbf{I}$, where σ_n^2 represents the noise spectral parameter, and $\mathbf{I}$ is the identity matrix.

Manuscript received July 31, 2007; revised February 29, 2008, August 20, 2008, and August 27, 2008. First published September 12, 2008; current version published May 11, 2009. The review of this paper was coordinated by Prof. T. J. Lim.

The author is with the Department of Electronics Engineering, Maltepe University, 34857 Istanbul, Turkey (e-mail: nihatk@maltepe.edu.tr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2008.2005723

III. Derivation of the Posterior Density

$\boldsymbol{\Psi}_{1:t} \triangleq (\boldsymbol{\theta}_{1:t}, \mathbf{s}_{1:t}, \sigma_w^2, \sigma_n^2, m_{1:t})$ is defined as a parameter vector, where the subscripts $(\cdot)_{1:t}$ denote the collection of parameters from time 1 to time t. The distribution of interest is $p(\boldsymbol{\Psi}_{1:t}|\mathbf{x}_{1:t})$, which can then be expanded using appropriately selected prior distributions of parameters, according to Bayes' theorem, as

$$p(\boldsymbol{\Psi}_{1:t}|\mathbf{x}_{1:t}) \propto p\left(\mathbf{x}_{1:t}|\boldsymbol{\theta}_{1:t}, \mathbf{s}_{1:t}, \sigma_w^2, \sigma_n^2, m\right) p\left(\boldsymbol{\theta}_{1:t}|\sigma_w^2\right) \times p\left(\mathbf{s}_{1:t}|\boldsymbol{\theta}_{1:t}, \sigma_n^2\right) p\left(\sigma_w^2\right) p\left(\sigma_n^2\right) p(m) \quad (5)$$

where $p(\mathbf{x}_{1:t}|\boldsymbol{\theta}_{1:t}, \mathbf{s}_{1:t}, \sigma_w^2, \sigma_n^2, m)$ is the likelihood function, and the other distributions compose the joint prior distribution for the parameter vector $\boldsymbol{\Psi}$.

Given the states, observations are assumed to be independent identically distributed, together with the state conditional update likelihood. The distributions for each of the terms in (5) can be assigned as

$$p\big(\mathbf{x}_{1:t}|\boldsymbol{\theta}_{1:t}, \mathbf{s}_{1:t}, \sigma_w^2, \sigma_n^2, m\big) = \prod_{j=1}^{t} \mathcal{N}\big(\mathbf{A}(\boldsymbol{\theta}_j)\mathbf{s}_j, \sigma_n^2\mathbf{I}\big) \quad (6)$$

$$p\big(\boldsymbol{\theta}_{1:t}|\sigma_w^2\big) = \prod_{j=1}^{t} \mathcal{N}\big(\boldsymbol{\theta}_{j-1}, \sigma_w^2\mathbf{I}\big) \quad (7)$$

$$p\big(\mathbf{s}_{1:t}|\boldsymbol{\theta}_{1:t}, \sigma_n^2\big) = \prod_{j=1}^{t} \mathcal{N}\Big(\mathbf{0}, \big[\mathbf{A}(\boldsymbol{\theta}_j)^H\mathbf{A}(\boldsymbol{\theta}_j)\big]^{-1}\Big). \quad (8)$$

The choice made in [19] vis-a-vis the prior distribution for the source amplitude vector $\mathbf{s}$ is acceptable. The prior distributions of the noise variances are both assumed to be the inverse Gamma distribution, which is the conjugate of the normal distribution, and can be written as follows:

$$p\left(\sigma_w^2\right) = \mathcal{IG}\left(\frac{\alpha_0}{2}, \frac{\gamma_0}{2}\right) \quad (9)$$

$$p\left(\sigma_n^2\right) = \mathcal{IG}\left(\frac{\alpha_1}{2}, \frac{\gamma_1}{2}\right). \quad (10)$$

The distribution for the model order m is assigned a discrete uniform distribution that is equal to $\Lambda = 1/m_{\max}$, $m = 0, 1, \ldots, m_{\max}$, where $m_{\max}$ is the maximum allowable model order, which is equal to $M - 1$.

Substituting all of these prior distributions into (5), the posterior probability density function can then be represented as follows:

$$\begin{aligned} p\left(\boldsymbol{\Psi}_{1:t}|\mathbf{x}_{1:t}\right) &\propto \prod_{j=1}^{t} \frac{1}{\pi^M|\mathbf{C}_{s_j}|} \exp\Big\{-(\mathbf{s}_j - \mathbf{m}_{s_j})^H\mathbf{C}_{s_j}^{-1}(\mathbf{s}_j - \mathbf{m}_{s_j})\Big\} \\ &\times \prod_{j=1}^{t} \frac{1}{\pi^K\left(\sigma_n^2\right)^K} \exp\left\{\frac{-1}{\sigma_n^2}\mathbf{x}_j^H\mathbf{C}_A^{\perp}(\boldsymbol{\theta}_j)\mathbf{x}_j\right\} \\ &\times \prod_{j=1}^{t} \frac{1}{\left(2\pi\sigma_w^2\right)^{M/2}} \exp\left\{\frac{-1}{2\sigma_w^2}(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j-1})^T(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j-1})\right\} \\ &\times \left(\sigma_w^2\right)^{\left(\frac{-\alpha_0}{2}-1\right)} \exp\left\{\frac{-\gamma_0}{\sigma_w^2}\right\} \left(\sigma_n^2\right)^{\left(\frac{-\alpha_1}{2}-1\right)} \\ &\times \exp\left\{\frac{-\gamma_1}{\sigma_n^2}\right\} \prod_{j=1}^{t} \boldsymbol{\Lambda} \end{aligned} \quad (11)$$

where

$$\mathbf{C}_{s_j} = \left[\mathbf{A}(\boldsymbol{\theta}_j)^H\mathbf{A}(\boldsymbol{\theta}_j)\right]^{-1} \quad (12)$$

$$\mathbf{m}_{s_j} = \mathbf{C}_{s_j}\mathbf{A}(\boldsymbol{\theta}_j)^H\mathbf{x}_j \quad (13)$$

$$\mathbf{C}_A^{\perp}(\boldsymbol{\theta}_j) = \mathbf{I} - \mathbf{A}(\boldsymbol{\theta}_j)\mathbf{C}_{s_j}\mathbf{A}(\boldsymbol{\theta}_j)^H. \quad (14)$$

From (11) and (13), it can be seen that $\mathbf{m}_{s_j}$ is a maximum *a posteriori* (MAP) estimate of amplitudes $\mathbf{s}_j$. Since the amplitudes can be estimated by using $\mathbf{m}_{s_j}$ after sampling the other parameters, there is no need to include amplitude estimation in the PF estimation algorithm. Therefore, (11) can analytically be integrated out over $\mathbf{s}_j$, which yields

$$\begin{aligned} p\left(\boldsymbol{\Psi}_{1:t}|\mathbf{x}_{1:t}\right) &\propto \prod_{j=1}^{t} \frac{1}{\left(\pi\sigma_n^2\right)^K} \exp\left\{\frac{-1}{\sigma_n^2}\mathbf{x}_j^H\mathbf{C}_A^{\perp}(\boldsymbol{\theta}_j)\mathbf{x}_j\right\} \\ &\times \prod_{j=1}^{t} \frac{1}{\left(2\pi\sigma_w^2\right)^{M/2}} \exp\left\{\frac{-1}{\sigma_w^2}(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j-1})^T(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j-1})\right\} \\ &\times \left(\sigma_w^2\right)^{\left(\frac{-\alpha_0}{2}-1\right)} \exp\left\{\frac{-\gamma_0}{\sigma_w^2}\right\} \left(\sigma_n^2\right)^{\left(\frac{-\alpha_1}{2}-1\right)} \\ &\times \exp\left\{\frac{-\gamma_1}{\sigma_n^2}\right\} \prod_{j=1}^{t} \boldsymbol{\Lambda} \end{aligned} \quad (15)$$

a relatively simpler posterior distribution in terms of the remaining parameters. As a result, the parameter vector $\boldsymbol{\Psi}_{1:t}$ can be written as follows:

$$\boldsymbol{\Psi}_{1:t} \triangleq \left(\boldsymbol{\theta}_{1:t}, \sigma_w^2, \sigma_n^2, m_{1:t}\right). \quad (16)$$

It is also possible to obtain the MAP estimates of σ_w^2 and σ_n^2 as in [11] using the fact that the mode of the inverted Gamma distribution is $\gamma/(\alpha + 1)$. Thus

$$\sigma_{w_{\mathrm{MAP}}}^2 = \frac{\frac{\gamma_0}{2} + \frac{1}{2}\sum_{j=1}^{t} \mathbf{x}_j^H\mathbf{C}_A^{\perp}(\boldsymbol{\theta}_j)\mathbf{x}_j}{\frac{\alpha_0}{2} + \frac{1}{2}Kt + 1} \quad (17)$$

$$\sigma_{n_{\mathrm{MAP}}}^2 = \frac{\frac{\gamma_1}{2} + \frac{1}{2}\sum_{j=1}^{t}(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j-1})^T(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j-1})}{\frac{\alpha_1}{2} + \frac{1}{2}\sum_{j=1}^{t} m(j) + 1}. \quad (18)$$

In the present case, (15) has still included the nuisance parameters σ_w^2 and σ_n^2. It is possible to integrate out these nuisance parameters, but this process requires a quite difficult computational procedure. Hence, they are kept in the expression of the posterior distribution. However, since the parameters σ_w^2 and σ_n^2 can be estimated through their MAP procedure, the parameter vector is now reduced to

$$\boldsymbol{\Psi}_{1:t} \triangleq (\boldsymbol{\theta}_{1:t}, m_{1:t}). \quad (19)$$

To get a good approximation of the posterior probability density function, the importance function to be proposed must be proportional to (15). In the next section, derivation of the importance density function is the main concern.

IV. Proposed Method

A. PFs With SVR for a Large Sample Set

While the PF method presents a facility to study on nonlinear/non-Gaussian models, it requires that particles constitute a large sample set and that the importance density must properly be chosen to get the desired approximation of the posterior density. Because the importance

density is mostly selected as the state transition prior distribution due to its simplicity, this selection is therefore ineffective and leads to poor filtering performance. In such a case, since the importance density has no information from the current observation in proposing new samples, the state space is explored without knowledge of the current observation, which causes the filter to be sensitive to outliers. The effective importance density selection methods proposed by different researchers have some disadvantages, such as computational difficulties, difficulties in the implementation, and a small process noise requirement. To tackle these problems, a new PF algorithm based on SVR is proposed in this paper. The proposed method does not require computational complexity and does not have implementation issues and a small process noise requirement.

To find a map that fits the training input set into the form of a given output set, let us consider the problem of approximating the set of data D_j

$$D_j = \left\{ \left(\left[\mathrm{Re}(\mathbf{x}_{t-T_b})^T \quad \mathrm{Im}(\mathbf{x}_{t-T_b})^T \right]^T, \tilde{\mathbf{\Psi}}^{(j)}_{t-T_{b+1}} \right), \ldots, \left(\left[\mathrm{Re}(\mathbf{x}_{t-1})^T \quad \mathrm{Im}(\mathbf{x}_{t-1})^T \right]^T, \tilde{\mathbf{\Psi}}^{(j)}_{t} \right) \right\} \quad (20)$$

where $j = 1, \ldots, M$; $t = 1, \ldots, N$; $\mathbf{x}_{t-l}$ and $\tilde{\mathbf{\Psi}}^{(j)}_{t-l+1}$ denote the input data vector and the corresponding output value for $l = 1, \ldots, T_b$, respectively; and T_b indicates the number of the elements in the training set. Thus, we can construct the regression function $f_r^{(j)}(.)$ over $\{(\mathbf{y}_{t-l}, \tilde{\mathbf{\Psi}}^{(j)}_{t-l+1}),\ l = 1, \ldots, T_b\}$ at time step t, i.e.,

$$f_r^{(j)} : \mathcal{R}^{2K} \rightarrow \mathcal{R} \quad (21)$$

where $\mathbf{y}_{t-l} = [\mathrm{Re}(\mathbf{x}_{t-l})^T \quad \mathrm{Im}(\mathbf{x}_{t-l})^T]^T$. Thus, SVR is employed to determine $f_r^{(j)}$ each time a new data sample is received, i.e.,

$$\begin{aligned} f_r^{(j)}(\mathbf{y}_{t+1}) &= \hat{\mathbf{\Psi}}^{(j)}_{t+1} \\ &= \sum_{v=0}^{T_b-1} (\beta_v - \beta_v^*)_{t+1} K(\mathbf{y}_{t-v}, \mathbf{y}_{t+1}) + b^{(j)}_{t+1} \end{aligned} \quad (22)$$

where $K(.,.)$ is a kernel function, β_v and β_v^* are the Lagrange multipliers, $f_r^{(j)}(.)$ is the corresponding output to $\mathbf{y}_{t+1}$ in the training phase, and $b^{(j)}_{t+1}$ is the bias term. Here, (22) implies how to obtain a prediction of the state value at time t by using the T_b number of previous buffered array output vectors. It can be expressed in compact form as

$$\hat{\mathbf{\Psi}}^{(j)}_{t+1} = \left[1 \quad \mathbf{K}(\mathbf{y}_{t-T_b+1:t}, \mathbf{y}_{t+1})^T \right] \mathbf{p}_{n+1} \quad (23)$$

where $\mathbf{K}(.) \in \mathcal{R}^{T_b}$, and $\mathbf{p} \in \mathcal{R}^{T_b+1}$ has its first entry equal to b_{t+1} and remaining entries equal to $(\beta_v - \beta_v^*)_{t+1}$. Once the prediction $\hat{\mathbf{\Psi}}_{t+1}$ is obtained, sample particles $\{\mathbf{\Psi}_t^{(i)},\ i = 1, \ldots, N\}$ are drawn from a Gaussian approximation to the optimal importance function $\mathcal{N}(\boldsymbol{\mu}_\pi, \mathbf{C}_\pi)$, which has the following mean and covariance:

$$\begin{aligned} \boldsymbol{\mu}_\pi &= \hat{\mathbf{\Psi}}_{t+1} \\ \mathbf{C}_\pi &= \frac{1}{T_b} \left\{ [\tilde{\mathbf{\Psi}}_{t-T_b} - \boldsymbol{\mu}_\pi, \ldots, \tilde{\mathbf{\Psi}}_t - \boldsymbol{\mu}_\pi] \right. \\ &\qquad \left. \times\ [\tilde{\mathbf{\Psi}}_{t-T_b} - \boldsymbol{\mu}_\pi, \ldots, \tilde{\mathbf{\Psi}}_t - \boldsymbol{\mu}_\pi]^T \right\} \end{aligned} \quad (24)$$

where $\tilde{\mathbf{\Psi}}_{t-v}$ stands for the estimates of $\mathbf{\Psi}_{t-v}$ for $v = 0, \ldots, T_b - 1$. Later, a classical resampling method can be applied to these particles, if necessary. All these processes are executed for each time step t. To obtain a new training data set for the current time step, the current measurement is included in the previous training data set, and the first measurement is then excluded from this set. Thus, the amount of data in the training set remains the same at each time step.

Consequently, a smoothed and well-approximated posterior density of the state for each time step t can be obtained by using this importance density function. Moreover, ineffective selection of the importance density and poor filtering performance are prevented through the use of the current measurement data within the SVR framework.

B. PFs With SVR for a Small Sample Set

In this part, a PF method based on sequential SVR [21] is proposed as a tracker within the sequential Monte Carlo framework to obtain sufficient performance of the PF method with a small sample set. In addition, to overcome the covariance issue, a separate kernel smoothing method is not needed because the kernel function and previous estimates of the parameter vector used in SVR provide the smoothing. This PF method is fundamentally the same as the PF method for a large sample set presented in the preceding part. The main difference of this method from the other appears in the formation of the regression function because it has a sequential form. The other difference is revealed while calculating the mean and the covariance of a Gaussian approximation to the optimal importance density function. In addition, to implement this method in the DOA tracking problem, the same T_b must be chosen as M. In this context, SVR is employed to determine a regression function each time a new data vector is received, i.e.,

$$f_r^{(j)}(\mathbf{y}_{t+1}) = \sum_{v=0}^{T_b-1} (\beta_v - \beta_v^*)_{t+1}\, \tilde{\mathbf{\Psi}}^{(j)}_{t-v} K(\mathbf{y}_{t-v}, \mathbf{y}_{t+1}) + b^{(j)}_{t+1}. \quad (25)$$

It can be seen that (25) includes the past T_b estimates of the parameter vector. Inclusion of these estimates in the regression function provides more accurate and smooth predictions for the parameter vector when a small sample set is available. It results in a good approximation of the optimal importance density function since particles are chosen from a set of particles that have closer values to the optimal ones, and thus, the proposed PF method does not usually require a resampling process.

At this point, (25) can be written in a more compact form as

$$\hat{\mathbf{\Psi}}^{(j)}_{t+1} = \left[1 \quad \left(\tilde{\mathbf{\Psi}}^{(j)}_{t-T_b+1:t} \otimes \mathbf{K}(\mathbf{y}_{t-T_b+1:t}, \mathbf{y}_{t+l}) \right)^T \right] \mathbf{q}_{t+1} \quad (26)$$

where $\tilde{\mathbf{\Psi}}^{(j)}_{t-T_b+1:t} \in \Re^{T_b}$, $\mathbf{K}(.) \in \Re^{T_b}$, $\otimes$ denotes component-wise multiplication, and $\mathbf{q} \in \Re^{T_b+1}$ has its first entry equal to $b^{(j)}_{t+1}$ and remaining entries equal to $(\beta_v - \beta_v^*)_{t+1}$.

Once the prediction of the parameter vector at time $t+1$, i.e., $\hat{\mathbf{\Psi}}_{t+1}$, is obtained, sample particles are drawn from a Gaussian approximation of the optimal importance function $\mathcal{N}(\boldsymbol{\mu}_\pi, \mathbf{C}_\pi)$, which has the following mean and covariance:

$$\boldsymbol{\mu}_\pi = \tilde{\mathbf{\Psi}}_{t-1} + \mathbf{C}_\pi(\hat{\mathbf{\Psi}}_{t+1} - \tilde{\mathbf{\Psi}}_{t-1}) \quad (27)$$

$$\mathbf{C}_\pi = \frac{1}{M}(\hat{\mathbf{\Psi}}_{t+1} - \tilde{\mathbf{\Psi}}_{t-1})(\hat{\mathbf{\Psi}}_{t+1} - \tilde{\mathbf{\Psi}}_{t-1})^T \quad (28)$$

where $\tilde{\mathbf{\Psi}}_{t-1}$ stands for the estimate of $\mathbf{\Psi}_{t-1}$. Later, a classical resampling method can be applied to these particles if necessary. All these processes are executed for each time step t. The method used to obtain a new training data set for the current time step is the same as that of the preceding method.

The proposed PF method for a small sample set has the same superiority as that for a large sample set in the selection of the importance density function. In addition, this method does not have the computational cost and sample degeneracy problem caused by

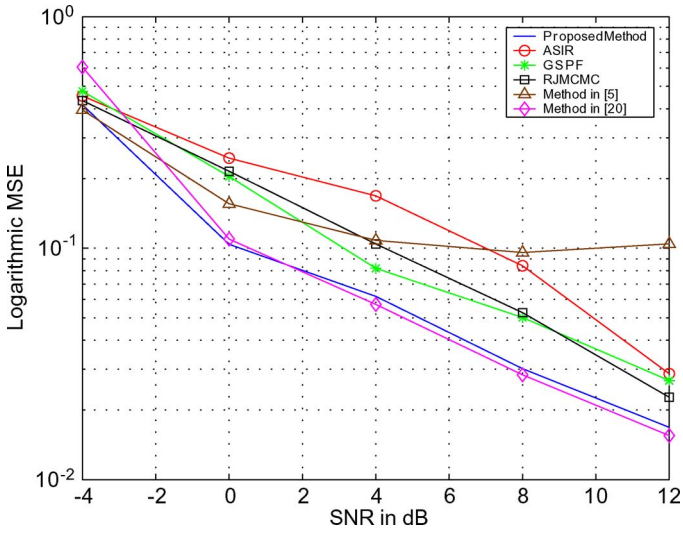

Fig. 1. MSEs of the estimates versus SNRs for a large sample set.

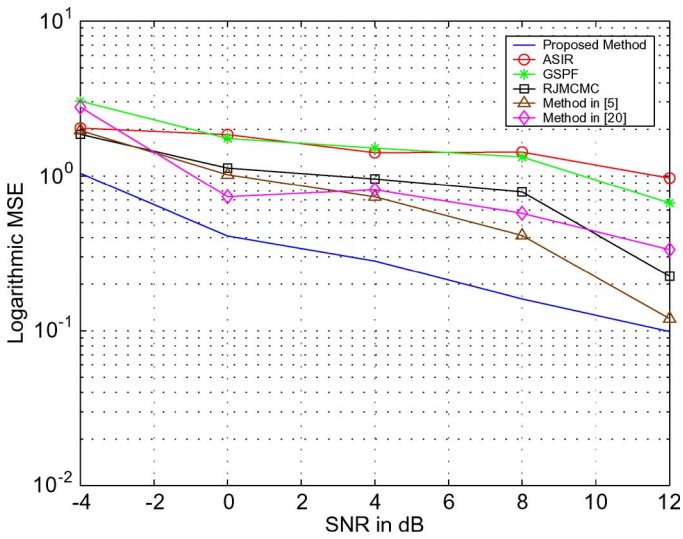

Fig. 2. MSEs of the estimates versus SNRs for a small sample set.

considering a large sample set. However, lower performance than that for a large sample set is unavoidable.

V. Simulations

In this section, some simulation results under large process noise are presented to verify the performance of the proposed algorithm while working with small and large sample sets. To test and to train SVR, a Gaussian radial basis kernel function is used.

In the following simulation, a uniform linear array composed of $K = 10$ omnidirectional antenna elements with half-wavelength spacing of the elements is used. $M = 2$ Gaussian processes for the moving sources that are zero mean with variance $\sigma_n^2 = 0.72^\circ$ are generated. Initial values for the time-varying DOAs of the moving sources are considered as $\boldsymbol{\theta} = [-10^\circ \quad 15^\circ]$. The time-varying DOAs evolve according to (1). Observations are generated using (3). The number of particles for a large sample set and for a small sample set, the number of Monte Carlo simulation, and the observation period are chosen as 500 and 100, 50, and 50, respectively. The data used to train support vector machines (SVMs) for small and large sample sets are 10 and 5, respectively. If noise variances σ_w^2 and σ_n^2 were unknown, they would sequentially be estimated by using their MAP procedure in (17) and (18), as well the signal amplitudes in (13).

To exhibit the performance of the estimator, the preceding scenario is used for both large and small sample sets. The performance of the proposed method is compared with the performance of the ASIR method, the GSPF method, the RJMCMC method, the method in [5], and the method in [20]. The prior density is selected as the importance density for both the RJMCMC method and the method in [5]. Since the measurement noise is additive white Gaussian noise, the number of Gaussian mixture models is selected as 1. The mean square errors (MSEs) of the estimates obtained using these methods versus different signal-to-noise ratios (SNRs) are shown in Figs. 1 and 2, respectively. In the case where a large sample set is considered, the proposed method performs better than the others, except for the method in [20]. When the computational requirements of the method in [20] such as matrix inversion, gradient vector, and Hessian matrix are considered, the performance of the method proposed here appears to have an advantage for a large sample set. In the case of a small sample set, the reduction in the MSE of the proposed method for increasing SNRs can be seen from the related figure.

While assigning 5 dB as the SNR and choosing the other related simulation parameters as in the previous simulation, the performance of the proposed PF method for a small sample set over a set of process noise variances is shown in Table I, which shows that the increase in noise variances does not have an overpowering effect on the performance of the proposed method. This result is expected because the prediction performance of SVR mainly depends on the variance of the kernel function.

TABLE I
Performance of the Proposed Algorithm for Different Process Noise at SNR = 5 dB

σ_n^2	MSE for source 1	MSE for source 2
0.1°	0.0524°	0.0513°
0.3°	0.1537°	0.1491°
0.5°	0.1509°	0.1468°
0.7°	0.1497°	0.1507°
0.9°	0.1512°	0.1521°
1.1°	0.1584°	0.1574°

VI. Conclusion

The main concern of this paper is to improve the performance of the PF method by using the current observation in proposing new samples. For this purpose, SVR is incorporated with a sequential Monte Carlo method. The SVR-based particle-producing scheme provides a more accurate approximation of the importance density function because an approximation of it obtained by the regression function is proportional to the product of the likelihood function with the prior density.

The efficiency and performance of the presented method are demonstrated in DOA tracking of the multiple moving sources problem in comparison with the ASIR method, the GSPF method, the RJMCMC method, the method in [5], and the method in [20]. The simulation results support the effectiveness of the method presented in this paper and demonstrate reliable tracking of the DOA of moving targets in a white-noise environment with a uniform linear array of sensors when a small sample set is used. Moreover, the performance of the proposed algorithm cannot be disregarded when a large sample set is available. Considering the importance density in [20], gradient and Hessian derivatives and matrix inversion are not desired due to their computational costs, because they are quadratic in terms of the number of particles. Thus, the performance of the proposed algorithm while considering a large sample set is capable of competing with it. Nevertheless, the performance of both algorithms proposed in this paper depends on choosing the parameters of SVR.

REFERENCES

[1] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration for tracking applications," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 68–77, Mar. 2002.

[2] X. Sheng, Y.-H. Hu, and P. Ramanathan, "Distributed particle filter with GMM approximation for multiple targets localization and tracking in wireless sensor network," in *Proc. Inf. Process. Sens. Netw.*, Los Angeles, CA, Apr. 2005, pp. 181–188.

[3] G. Zhu, D. Liang, Y. Liu, Q. Huang, and W. Gao, "Automatic multi-player detection and tracking in broadcast sports video using support vector machine and particle filter," in *Proc. IEEE Int. Conf. Multimed. Expo.*, Toronto, ON, Canada, Jun. 2006, pp. 1629–1632.

[4] G. Zhu, D. Liang, Y. Liu, Q. Huang, and W. Gao, "Action recognition in broadcast tennis video," in *Proc. Int. Conf. Pattern Recog.*, Hong Kong, Aug. 2006, vol. 1, pp. 251–254.

[5] G. Zhu, D. Liang, Y. Liu, Q. Huang, and W. Gao, "Improving particle filter with support vector regression for efficient visual tracking," in *Proc. IEEE Int. Conf. Image Process.*, Genova, Italy, Sep. 2005, vol. 2, pp. 422–425.

[6] N. Kabaoğlu and H. A. Cirpan, "Wideband target tracking by using SVR-based sequential Monte Carlo method," *Signal Process.*, vol. 88, no. 11, pp. 2804–2816, Nov. 2008.

[7] Q.-H. Meng, X. Yun, Y.-H. Wu, and J. Wang, "Particle filter based small mobile robot indoor localization using a rotary sonar," in *Proc. Robot. Appl.*, 2005.

[8] S. Fleck, F. Busch, P. Biber, and W. Straber, "3D surveillance—A distributed network of smart cameras for real-time tracking and its visualization in 3D," in *Proc. Comput. Vis. Pattern Recog. Workshop*, New York, Jun. 2006, pp. 118–126.

[9] G. Poyiadjis, A. Doucet, and S. S. Singh, "Particle methods for optimal filter derivative: Application to parameter estimation," in *Proc. IEEE ICASSP*, Philadelphia, PA, Mar. 2005, vol. 5, pp. 925–928.

[10] W. F. Leven and A. D. Lanterman, "Multiple target tracking with symmetric measurement equations using unscented Kalman and particle filters," in *Proc. 36th Southeastern Symp. Syst. Theory*, Atlanta, GA, Mar. 2004, pp. 195–199.

[11] J. R. Larocque, J. P. Reilly, and W. Ng, "Particle filters for tracking an unknown number of sources," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 2926–2937, Dec. 2002.

[12] C. Kumar, R. Rajagopal, and R. Kiran, "An optimal integrated tracking (ITS) for passive DOA tracking using unscented Kalman filter," in *Proc. IEEE Conf. Decision Control*, Las Vegas, NV, Dec. 2002, pp. 253–258.

[13] W. Ng, J. P. Reilly, T. Kirubarajan, and J.-R. Larocque, "Wideband array signal processing using MCMC methods," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 411–426, Feb. 2005.

[14] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proc. Amer. Control Conf.*, Seattle, WA, Jun. 1995, pp. 1628–1632.

[15] M. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *J. Amer. Stat. Assoc.*, vol. 94, no. 446, pp. 590–599, 1999.

[16] A. Giremus, A. Doucet, V. Calmettes, and J.-Y. Tourneret, "A Rao–Blackwellized particle filter for INS/GPS integration," in *Proc. IEEE ICASSP*, Montreal, QC, Canada, May 2004, vol. 3, pp. 964–967.

[17] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

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

[19] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2667–2676, Oct. 1999.

[20] M. Orton and W. Fitzgerald, "A Bayesian approach to tracking multiple targets using sensor array and particle filters," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 216–223, Feb. 2002.

[21] N. Freitas, M. Milo, and P. Clarkson, "Sequential support vector machines," in *Proc. IEEE NNSP*, Aug. 1999, pp. 31–40.

A Novel SFBC-OFDM Scheme for Doubly Selective Channels

Sili Lu, *Student Member, IEEE*,
Balachander Narasimhan, *Student Member, IEEE*,
and Naofal Al-Dhahir, *Fellow, IEEE*

***Abstract*—Effective intercarrier interference (ICI) mitigation for multiple-input–multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) requires accurate channel estimation, which is very challenging due to the large number and fast time-varying nature of the channel parameters to be estimated using scattered pilots. We present a novel space-frequency block coding OFDM (SFBC-OFDM) scheme for doubly selective channels and a reduced-complexity channel-estimation algorithm, which exploits the banded and sparse structure of the channel in the frequency and time domains, respectively. Furthermore, we design a finite-impulse-response minimum-mean-square-error (FIR-MMSE) ICI cancellation algorithm for mobile SFBC-OFDM and demonstrate its effectiveness for digital video broadcasting-handheld (DVB-H) systems.**

***Index Terms*—Alamouti scheme, Doppler effect, intercarrier interference (ICI), orthogonal frequency-division multiplexing (OFDM), space-frequency block coding (SFBC).**

I. INTRODUCTION

Orthogonal frequency-division multiplexing (OFDM) is very sensitive to channel time selectivity, which manifests itself as variations in the channel impulse response (CIR) within each OFDM symbol. This so-called Doppler effect results in intercarrier interference (ICI), whose energy is proportional to the Doppler frequency due to receiver mobility and is a significant performance-limiting impairment for OFDM systems.

In most previous studies, a quasi-static channel is assumed, where the CIR is assumed constant over each OFDM symbol, and a conventional one-tap-per-subcarrier frequency-domain equalizer (FEQ) is used, which becomes highly suboptimal under high Doppler effect. In [1], we proposed a finite-impulse-response minimum-mean-square-error (FIR-MMSE) FEQ with few (typically three) taps per subcarrier, which utilizes the approximately banded structure of the channel matrix in the frequency domain.

OFDM in a multiple-input–multiple-output (MIMO) system, which is known as MIMO-OFDM, is an attractive transmission scheme for high-rate reliable broadband wireless communication. A prime example is the Alamouti space-time block code (STBC) [2], which can be integrated with OFDM systems to achieve spatial and multipath diversity gains and reduce the error floor due to ICI. However, under high mobility, implementing the Alamouti STBC over adjacent OFDM symbols is not effective due to the significant channel time variations [3]. Instead, space-frequency block coding (SFBC) was considered in [4] and [5], where the Alamouti scheme is implemented over adjacent subcarriers within the same OFDM symbol and the channel frequency

Manuscript received April 10, 2008; revised August 1, 2008. First published September 26, 2008; current version published May 11, 2009. This work was supported in part by Semiconductor Research Corporation under Contract 2005-HJ-1328 and in part by a gift from Texas Instruments Incorporated. The review of this paper was coordinated by Dr. A. Ghrayeb.

The authors are with the University of Texas at Dallas, Richardson, TX 75080-3021 USA (e-mail: sxl059000@utdallas.edu; bxn062000@utdallas.edu; aldhahir@utdallas.edu).

Digital Object Identifier 10.1109/TVT.2008.2006271