14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

DOA ESTIMATION BASED ON CROSS-CORRELATION BY TWO-STEP PARTICLE FILTERING

Mitsunori Mizumachi, Norikazu Ikoma, and Katsuyuki Niyada

Kyushu Institute of Technology
1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka 804-8550, Japan
E-mail: mizumach@comp.kyutech.ac.jp

ABSTRACT

This paper proposes a two-step particle filtering method for achieving robust DOA estimation under noisy environments. The proposed method aims at fusing the advantages of traditional cross-correlation (CC) and generalized (whitened) cross-correlation (WCC) methods, which are a CC method taking power information into account and a WCC method having a sharp peak in a cross-correlation function. We regard rectified CC and WCC functions as likelihood. The two-step filtering is carried out on bi-dimensional, spectro-spatial state space. Experimental results under noisy and slightly reverberant environments show that the proposed method is superior to CC and WCC methods both in accuracy and stability.

1. INTRODUCTION

Direction-Of-Arrival (DOA) plays an important role in the field of acoustic signal processing, for example, beam-steering for speech enhancement [1]. Cross-correlation (CC) between two received signals has been widely used for estimating the DOA due to its simplicity and efficiency. A CC function roughly estimates a DOA with a blunt peak, thus the estimate is not accurate. There are many approaches to improve the accuracy. Among those approaches, generalized cross-correlation (GCC) method [2] is one of the most popular techniques. For GCC in this paper, whitened filter is introduced for calculating a whitened cross-correlation (WCC) function with wide-band signals such as speech signals. Wide-band signals usually enable to yield sharp mainlobes in correlation functions, so we can have accurate estimates. WCC method assumes that the band-width of the target signal is wider than those of interferences. Under wide-band interference environments such as multiple speaker environments, however, it is hard for the WCC method to correctly find the DOA of the target signal. This is because of the whitening operation carried out in each frequency bin leading to a global DOA dominated by the interference with the widest band-width even in high Signal-to-Noise Ratio (SNR) conditions.

We deal with a cross-correlation-based DOA finder by using a particle filtering technique. Particle filter is a state-of-the-art stochastic Bayesian filter, because it can estimate posterior distributions with relax assumptions on stochastic behaviors of observations [3]. Target tracking is one of the suitable problems to be solved by particle filters, since trajectories of moving targets can be modeled by Markov process even under non-Gaussian, non-stationary, practical noise environments. Recently, a number of particle filtering approaches have been proposed in various fields, and they have been expected to achieve robust DOA estimation in adverse environments. As a pioneering work, Vermaak and Blake show that non-linear filtering is effective for speaker tracking under reverberant environments [4]. In this work, they model the dynamics of DOAs in the form of random model, and the DOA estimation itself is done by a rather simple method. Ward and Williamson proposed the particle filter beamforming technique for sound source localization [5]. Generally, it is necessary for beamformer-based approaches to prepare a lot of carefully-positioned microphones. Asoh *et al.* aimed at fusing audio and video information by particle filtering for multiple speaker tracking [6]. In the method, MUSIC algorithm [7] is employed to obtain DOA estimates from audio inputs, however the parametric MUSIC algorithm requires to estimate the number of sources in advance. Further investigation is needed to achieve robust DOA estimation in adverse conditions.

In this paper, we propose a two-step particle filtering method to fuse the advantages of both CC and WCC methods. Posterior distribution over bi-dimensional, spectro-spatial state space, which is specified by frequency and time lag, gives DOA by finding the maximum peak in time lag direction. This is effectively achieved by particles as follows. We regard rectified CC and WCC functions as likelihoods. The likelihood is used to determine weights for particles in a two-step way: CC likelihood leads particles to a rough DOA region, and the second filtering with WCC likelihood aims at concentrating the particles near the true DOA on spectro-spatial space. We regard kernel density in terms of time lag as a global correlation function, which gives a DOA estimate. The proposed method also yields a smooth DOA trajectory, since time evolution of DOA in each short-term frame is modeled by Markov process.

Rest of this paper is organized as follows. In Section 2, we formulate a signal model for DOA estimation, and review CC, GCC, and WCC methods. In Section 3, a new approach is proposed to fuse respective advantages of both CC and WCC methods by two-step particle filtering on bi-dimensional spectro-spatial space. In Section 4, we confirm the performance of the proposed method under an adverse environment, compared with the conventional CC and WCC methods. Finally, conclusion is given in Section 5.

2. DOA ESTIMATION BY CROSS-CORRELATION-BASED APPROACH

2.1 Signal model

Let us assume that a target signal $s(t)$ is received by a pair of spatially-separated microphones. The signals, $x_i(t)$ and $x_j(t)$, which are received by the microphones M_i and M_j, can

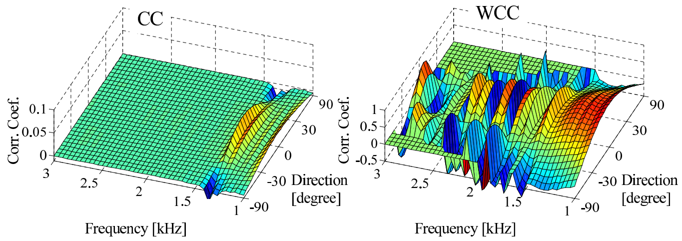

Figure 1: Examples of CC and WCC functions are plotted in left and right panels, respectively. Even in 15 dB SNR conditions, WCC functions may fail in finding the true DOA of the target signal in the presence of wide-band interference.

be modeled as follows.

$$\left\{ \begin{array}{lcl} x_i(t) & = & h_i(t) * s(t) + n_i(t), \\ x_j(t) & = & h_j(t) * s(t-\tau) + n_j(t), \end{array} \right. \tag{1}$$

where $h_i(t)$ represents a room impulse response between the sound source and the microphone M_i, $n_i(t)$ is additive noise as a mixture of measuring noise and acoustic signals delivered by non-target sound sources, and τ is the time lag of the signal arriving at the microphones M_i and M_j. For far-field sound sources, assuming that $h_i(t) \approx h_j(t)$, difference in phase is much important for DOA estimation than that in amplitude.

2.2 Cross-correlation-based DOA estimation

Concerning DOA estimation, cross-correlation-based approach is widely used due to its simplicity. DOA is given straightforwardly from the time lag τ with the peak in a CC function. CC function has a blunt peak. On the other hand, WCC function has the sharp peak in a correlation function [2]. The GCC function $R_{x_ix_j}(\tau)$ is defined as follows.

$$R_{x_ix_j}(\tau) = \int_{-\infty}^{\infty} \Psi(f) r(\tau, f) df, \tag{2}$$

$$r(\tau, f) = X_i(f) X_j^*(f) e^{j2\pi f \tau}, \tag{3}$$

where $X_i(f)$ and $X_j(f)$ are the short-term Fourier transforms (STFTs) with Hanning windows of received signals $x_i(t)$ and $x_j(t)$, respectively, and $*$ denotes the complex conjugate. Provided the generalizing operator $\Psi(f)$ takes a constant value over the whole frequency, $R_{x_ix_j}(\tau)$ becomes a conventional CC function. It is certain that a matched filter would be optimum as the operator $\Psi(f)$, if target signal could be exactly estimated.

The WCC method employs the whitening operator as

$$\Psi(f) = \frac{1}{|X_i(f)||X_j(f)|}. \tag{4}$$

In each frequency bin, WCC method accurately finds the local DOA of the signal having the highest energy in the bin. Figure 1 shows examples of CC and WCC functions in each bin, when a target speech signal comes from -10 degrees in a wide-band noise condition (15 dB SNR). In the first formant frequency around 1.3 kHz, the CC function gives the true DOA roughly, and is less affected by noise perspectively. On the other hand, the WCC function gives the true DOA more accurately at the formant frequency, but it may have improper DOA estimates in other frequencies in presence of wide-band noise even in high SNR condition.

3. ROBUST DOA ESTIMATION BY TWO-STEP PARTICLE FILTERING

3.1 Motivation

Generally, WCC method has the advantage of accuracy in DOA estimation compared with a conventional CC method [2]. When the band-width of an interference is wider than that of the target signal, however, the WCC function is not superior to the CC function in finding the true DOA. This is because a WCC function gives a DOA in each frequency irrelevant to energy distribution, and a global DOA is determined by vote of the DOA estimates in the whole frequency. Therefore, the signal with the widest band-width dominates the final decision of DOA estimation.

This paper proposes to fuse the advantages of CC and WCC methods by two-step particle filtering. Assuming SNR is positive, CC function gives rough DOA estimates as the first step. We expect that the rough DOA estimates contribute greatly to finding the true DOA by a WCC function in the second step. Particle filter is effective to yield smooth DOA trajectory considering the dynamics of DOA in each short-term frame [4].

3.2 Problem statement

Particle filter is flexible on modeling the system compared with other Bayesian filters such like traditional Kalman filters and extended Wiener filters, because it does not require any linearity or Gaussianity on the model. Temporal trajectory of DOA is modeled by a state space model, and it is estimated through the state estimation procedure using particle filters.

In the k-th frame, sampled signals are denoted as $x_{i,k} \equiv x_i(t_k : t_{k+1})$ and $x_{j,k} \equiv x_j(t_k : t_{k+1})$ with the observed continuous signals in time between t_k and t_{k+1}. Observation is simply noted as $\mathbf{x}_k = (x_{i,k},\, x_{j,k})$. State estimation is recursively done with the state $\mathbf{z} \equiv (\tau,\, f)$ on the spectro-spatial state space, that is specified by frequency f and time lag τ. Posterior distribution $p(\mathbf{z}_{1:k} \,|\, \mathbf{x}_{1:k})$ is estimated with sampled signals up to the k-th frame, $\mathbf{x}_{1:k}$, by particle filtering in the section **3.3**.

System model:

Time evolution of the state $\mathbf{z}_k$, which consists of τ_k and f_k, is assumed to be smoothly changed, and is modeled as

$$\left\{ \begin{array}{lcl} \tau_k & = & \tau_{k-1} + v_k^{(\tau)}, \\ f_k & = & f_{k-1} + v_k^{(f)}, \end{array} \right. \tag{5}$$

where, $v_k^{(\tau)} \sim N(0,\, \sigma_\tau^2)$ and $v_k^{(f)} \sim N(0,\, \sigma_f^2)$, are system noises.

Observation model:

Both CC and WCC functions are calculated from the observations $\mathbf{x}_k$ through their STFTs.

$$R_{\mathbf{x}_k}^{(\Theta)}(\tau) = \sum_{f \in \{F_1, F_2, \ldots, F_N\}} \Psi_k^{(\Theta)}(f) r_k(\tau, f), \tag{6}$$

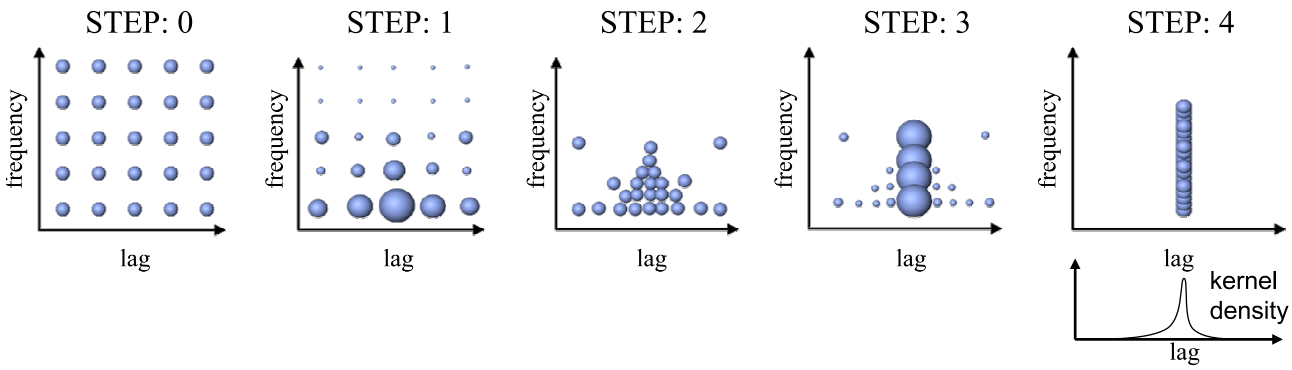

Figure 2: Explanatory illustration of two-step particle filtering on spectro-spatial state space. In **STEP 0**, uniform distribution is applied. In **STEP 1** and **STEP 3**, particles are filtered out with updated weights by CC and WCC likelihoods, respectively. In **STEP 4**, kernel density is obtained in terms of time lag, and it is regarded as global correlation function.

where $R^{(\Theta)}_{\mathbf{x}_k}(\tau)$ corresponds to either CC or WCC function depending on $\Theta \in \{CC, WCC\}$. We represent sub-band CC or sub-band WCC function as $R^{(\Theta)}_{\mathbf{x}_k}(\tau, b)$ in the b-th frequency bin by taking the summation over the bin. We regard rectified sub-band CC and WCC functions as likelihood such that

$$p^{(\Theta)}(\mathbf{x}_k \,|\, \mathbf{z}_k) \propto R^{+(\Theta)}_{x_k}(\tau, f), \tag{7}$$

where R^+ means a half-wave rectified correlation function.

State estimation:

State estimation is formally carried out, and the posterior distribution $p(\mathbf{z}_{1:k} \,|\, \mathbf{x}_{1:k})$ is recursively obtained as

$$p(\mathbf{z}_{1:k} \,|\, \mathbf{x}_{1:k}) \propto p(\mathbf{z}_{1:k-1} \,|\, \mathbf{x}_{1:k-1}) p(\mathbf{x}_k \,|\, \mathbf{z}_k) p(\mathbf{z}_k \,|\, \mathbf{z}_{k-1}). \tag{8}$$

In the two-step filtering scheme, we use

$$p(\mathbf{x}_k \,|\, \mathbf{z}_k) \approx \begin{cases} p^{(CC)}(\mathbf{x}_k \,|\, \mathbf{z}_k), & \text{for 1st step filtering,} \\ p^{(WCC)}(\mathbf{x}_k \,|\, \mathbf{z}_k), & \text{for 2nd step filtering.} \end{cases} \tag{9}$$

3.3 State estimation by particle filtering

Sequential state estimation is carried out by updating weighted particles according to Eq. (8). In the first filtering step, we employ a bootstrap filter which uses the system model as proposal (see [3] for more details). We expect that resulting particles are well arranged in this step. In the second filtering step, only weights are updated according to WCC likelihood. Thus, particles $\{z_k\}$ with weights $\{w_k\}$ are updated sequentially according to CC and WCC likelihood by turns

$$\{(z_{k-1}, w_{k-1})\} \xrightarrow{CC} \{(\tilde{z}_k, \tilde{w}_k)\} \xrightarrow{WCC} \{(z_k, w_k)\}. \tag{10}$$

The two-step particle filtering is performed as below. Figure 2 gives a conceptual illustration.

STEP 0: (initial distribution)

Uniform distribution is adopted, and particles are drawn according to the distribution.

STEP 1: (1st step filtering by CC)

Particles at time k are drawn from the system model with particles of the previous ime step, and the weights are updated.

$$\tilde{z}^{(l)}_k \sim p(\mathbf{z}_k \,|\, z^{(l)}_{k-1}). \tag{11}$$

$$\tilde{w}^{(l)}_k \propto w^{(l)}_{k-1} \cdot p^{(CC)}(\mathbf{x}_k \,|\, \tilde{z}^{(l)}_k). \tag{12}$$

The filtered particles $\{\tilde{z}^{(l)}_k\}^M_{l=1}$ distribute mostly according to signal energy. This corresponds to a rough DOA estimate.

STEP 2: (resampling)

The particles $\{\tilde{z}^{(l)}_k\}^M_{l=1}$ are resampled in proportion to the weight $\{\tilde{w}^{(l)}_k\}^M_{l=1}$. Then we have resampled particles $\{\dot{z}^{(l)}_k\}^M_{l=1}$ with uniform weights. In the next step, this set of particles is used as the proposal particle distribution.

STEP 3: (2nd step filtering by WCC)

WCC likelihood updates the weights for the particles $\{\dot{z}^{(l)}_k\}^M_{l=1}$ as

$$w^{(l)}_k \propto p^{(WCC)}(\mathbf{x}_k \,|\, \dot{z}^{(l)}_k). \tag{13}$$

The filtered particles $\{\dot{z}^{(l)}_k\}^M_{l=1}$ and weight $\{w^{(l)}_k\}^M_{l=1}$ are delivered to the next frame.

STEP 4: (finding DOA)

The particles are resampled, and kernel density as a global correlation function is obtained in terms of time lag. DOA is estimated from the time lag with the peak in the global correlation function.

STEP 5: go to STEP 1

No process is carried out during silence (low energy) period. Consequently, when signal onset is detected, we need to re-initialize particle set in an appropriate way. For the re-initialization, we use a DOA histogram formed by the whole DOA estimate in past. We adopt a distribution consisting of the DOA histogram in time lag axis and an uniform distribution in frequency axis.

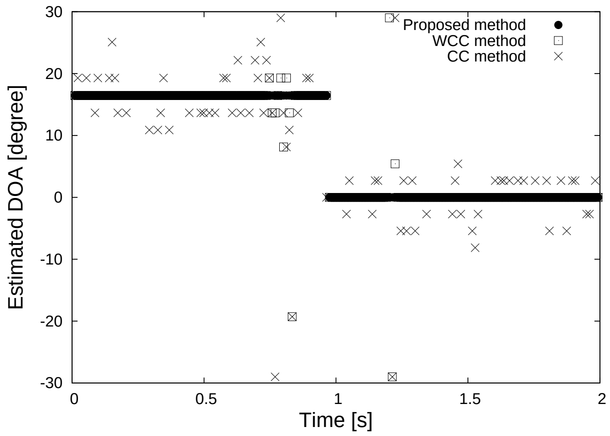

Figure 3: DOAs estimated in each short-term frame by the proposed ('•'), WCC ('□'), and CC ('×') methods in 10 dB SNR condition. The true DOA was changed from 16 degrees to 0 degree at 1.0 [s].

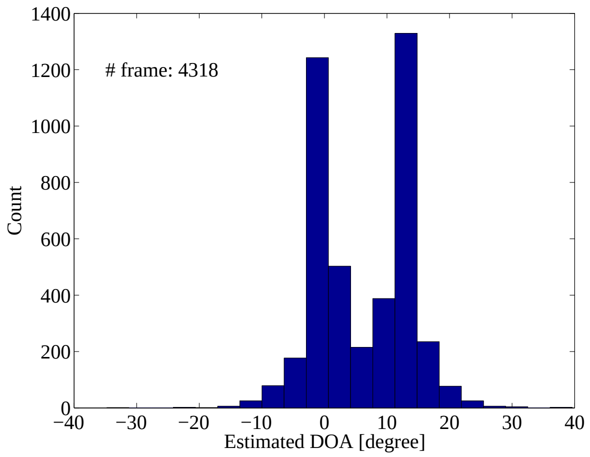

Figure 4: Histogram of DOA estimates obtained by the proposed method over 4,318 frames (total data size is 46 s in time). The true DOA was set at 0 degree or 16 degrees, while either of two sources generated speech signals.

4. PERFORMANCE EVALUATION

4.1 Experimental condition

Performance of the proposed method was evaluated by using data acquired by a pair of microphones in a real room (7.0 m × 4.1 m × 2.6 m, reverberation time was 0.23 s approximately). In the room, two microphones (audio-technica AT805F) were placed with the spacing of 0.15 m, and two loud speakers (BOSE 101) were placed at 0 degree (the front) and 16 degrees. Either loud speaker played target signals alternately. Connected digit speech utterances, which were partially selected from the TI-digit speech database [8], were prepared as target signals. The recorded signals were originally distorted by background noises and room reverberation. To obtain noisy speech signal at 10 dB SNR, channel-independent white Gaussian noises were added to the recorded speech signals in a computer. DOA estimation with two-step particle filtering was performed with 3,498 particles every 10.6 ms. DOA estimation was also carried out with CC and WCC methods as references.

4.2 Experimental result

Figure 3 shows a snapshot of DOA estimates by the proposed (marked by '•'), WCC (marked by '□'), and CC (marked by '×') methods, respectively. Histogram of the whole DOAs estimate over 4,318 frames is shown in Fig. 4. In the histogram, the skirts of two peaks are formed by estimation error, which was occurred near the time that an active source switches over. Afterward estimation proceeded through recursive procedure, the proposed method seldom failed in finding true DOAs as shown in Fig. 3. The experimental results indicate that the proposed method outperforms conventional CC and WCC methods over local accuracy and stability (global accuracy).

5. CONCLUSION

This paper proposes a robust DOA estimation technique by two-step particle filtering under adverse environments. The proposed method aims at fusing the advantages of CC and WCC methods on a spectro-spatial domain, which is specified by frequency and time lag. We regard both of rectified sub-band CC and WCC functions as likelihoods to update the weights for particles on the spectro-spatial space. The proposed method has significant advantage over conventional CC and WCC methods in terms of both accuracy and stability. Further investigation is to deal with concurrent multiple sound source with reasonable computational cost.

Acknowledgment
This work was partially supported by Grant-in-Aid for Scientific Research (No. 17760338) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

[1] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Springer-Verlag, Berlin, 2001.

[2] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process., vol. 24, pp. 320–327, 1976.

[3] A. Doucet *et al.*, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, 2001.

[4] J. Vermaark and A. Blake, 'Nonlinear Filtering for speaker tracking in noisy and reverberant environments," Proc. ICASSP '01, vol. 5, pp. 3021–3024, 2001.

[5] D. Ward *et al.*, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," IEEE Trans. Speech Audio Process., vol. 11, pp. 826–836, 2003.

[6] H. Asoh *et al.*, "An application of a particle filter to bayesian multiple sound source tracking with audio and video information," Proc. 7th Intl. Conf. on Information Fusion, pp. 805–812, 2004.

[7] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propagation, vol. AP-34, pp. 276–280, 1986.

[8] R. G. Leonard, "A database for speaker independent digit recognition," Proc. ICASSP '84, vol. 9, pp. 328–331, 1984.