Available online at www.sciencedirect.com

ScienceDirect

Applied Acoustics 69 (2008) 673–680

applied acoustics

www.elsevier.com/locate/apacoust

Robust tracking of moving sound source using scaled unscented particle filter

Zhiwei Liang *, Xudong Ma, Xianzhong Dai

Department of Automatic Control, Southeast University, Nanjing, Jiangsu Province, China

Received 1 December 2006; received in revised form 9 March 2007; accepted 2 April 2007
Available online 23 May 2007

Abstract

Tracking an active sound source involves the modeling of non-linear non-Gaussian systems. To address this problem, this paper proposed scaled unscented particle filter (SUPF) algorithm for tracking moving sound source. The particle filter part of the SUPF provides the general probabilistic framework to handle non-linear non-Gaussian systems, and the scaled unscented Kalman filter (SUKF) part of the SUPF generates better proposal distributions by taking into account the most recent observation. Meanwhile, models used in SUPF algorithm were also explored for the sound source motion, observation and the likelihood of the sound source location in the light of the Langevin process. Compared with the conventional PF approach, the simulated results demonstrated that the SUPF algorithm had superior tracking performance.

Keywords: Sound source localization; Time delay of arrival; Scaled unscented particle filter; Kalman filter

1. Introduction

Reliable sound source tracking has been an active research topic with the increasing applications in multimedia field, including automatically guiding the camera to pick up the faces of speakers for video-conferencing, the auditory system of a robot and providing steering information for speech enhancement. Unfortunately, even after years of research, robust and efficient tracking is still an open problem.

The traditional approaches to address the problem collect data from several microphones and use a frame of data obtained at the current time to estimate the current source location. These traditional methodologies can be divided into two categories: (1) time-delay estimation (TDE) methods such as the well-known generalized cross-correlation (GCC) function [1,2], which estimate location based on the time delay of arrival of signals at the receivers and (2) direct methods such as steered beamforming [3]. Each method transforms the received frame of data into a function that exhibits a peak in the location corresponding to the source. We will refer to this function as the localization function. The practical disadvantage of these traditional approaches is that reverberation causes spurious peaks to occur in the localization function. These spurious peaks may have greater amplitude than the peak due to the true source, so that simply choosing the maximum peak to estimate the source location may not give accurate results.

A promising technique that overcomes the drawback of traditional methods is to use a state-space approach based on particle filtering (PF), as recently described in [4,5]. The key to these new techniques is that peak due to the true source follows a dynamical model from frame to frame, whereas there is no temporal consistency to the spurious peaks. When the particles are properly placed, weighted, propagated, posteriors can be estimated over time using a sequential Monte Carlo method. However, most of them (e.g. CONDENSATION) use the state transition prior as the proposal distribution does not take into account the most recent observation, the particles drawn from transition prior

* Corresponding author. Tel.: +86 25 8379 2418.
E-mail addresses: lzhw_ly@hotmail.com, lzhw_ly@163.com (Z. Liang), xudma@seu.edu.cn (X. Ma), daixianzhong@163.com (X. Dai).

0003-682X/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2007.04.001

may have very low likelihood, and their contributions to the posterior estimation become negligible. This type of particle filters is prone to be distracted by background clutters [6,7]. To design better proposal distributions for particles filters, in general, there are two approaches: the direct approach and the indirect approach. The indirect approach attacks this problem indirectly by using an auxiliary tracker to generate the proposal distribution for the main tracker. The direct method, on the other hand, addresses this problem directly in its original space by taking into account the most recent observation. The indirect method is adopted in the CONDENSATION algorithm [8], where a color auxiliary tracker is used to generate the proposal distribution for the main contour tracker. While better than the conventional particle filters, this indirect approach has two major limitations. First, in audio-based speaker localization, there is simply no easy auxiliary tracker or sensing modality available. Second, and more importantly, the auxiliary tracker itself needs a good proposal distribution if it plans to use particle filters, or it falls back to ad hoc approaches.

Merwe and Doucet have recently developed the scaled unscented particle filter (SUPF) in the field of filtering theory [7]. Based on this new development, in this paper, we introduce a direct approach to generate better proposal distributions for moving sound source tracking. The SUPF is a parametric and non-parametric hybrid of SUKF and particle filters. The particle filter part of the SUPF provides the general probabilistic framework to handle non-linear non-Gaussian systems, and the SUKF part of the SUPF generates better proposal distributions by taking into account the most recent observation.

The remainder of this paper is organized as follows: Section 2 described the particle filters framework focused on the importance of the proposal distribution. In Section 3, we present the SUPF algorithm, which apply the scaled unscented Kalman filter to generate better proposal distributions that seamlessly integrate the current observation. According to Langevin process, the target's motion model for SUPF is built in Section 4, with a tracking simulation following in Section 5. Finally, the findings of the paper are summarized in Section 6.

2. Particle filter

In the pioneering work of CONDENSATION [8], extend factored-sampling is used to formulate the particle filter framework. Even though easy to follow, it obscures the role of proposal distributions. In this section, we present a new formulation of particle filtering theory that is centered round proposal distributions. This new formulation illustrates how to improve the particle filter's performance by designing better proposal distributions.

In Monte Carlo simulation, a set of weighted particles (samples), drawn from the posterior distribution, is used to map integrals to discrete sums. More precisely, the posterior can be approximated by the following empirical estimate

$$\hat{p}(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = \frac{1}{N}\sum_{i=1}^{N}\delta_{\mathbf{x}_{0:t}^{(i)}}(\mathrm{d}\mathbf{x}_{0:t}) \tag{1}$$

where the random samples $\{\mathbf{x}_{0:t}^{(i)};\ i = 1,\cdots,N\}$ are drawn from the posterior distribution $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ and $\delta(\mathrm{d}\cdot)$ denotes the Dirac delta function. Consequently, it follows from the strong law of large numbers that as the number or samples N increases, expectations can be mapped into sums. Unfortunately, it is often impossible to sample directly from the posterior density function. However, we can circumvent this difficulty by sampling from a known, easy-to-sample, proposal distribution $q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$.

Definition 1 [9,14]: A set of random samples $\{\mathbf{x}_{0:t}^{(i)}, w_t(\mathbf{x}_{0:t}^{(i)});\ i = 1,\cdots,N\}$ drawn from a distribution q is said to be properly weighted with respect to p if

$$\lim_{N\to\infty}\sum_{i=1}^{N} h\left(\mathbf{x}_{0:t}^{(i)}\right)w_t\left(\mathbf{x}_{0:t}^{(i)}\right) = E_p(h(\mathbf{x}_{0:t})) \tag{2}$$

for any integrable function h. In a practical sense we can think of p as being approximated by the discrete distribution supported on the $\mathbf{x}_{0:t}^{(i)}$ with probabilities proportional to the weights $w_t(\mathbf{x}_{0:t}^{(i)})$. Furthermore, as N tends to infinity, the posterior distribution p can be approximated by the properly weighted particles drawn from q[7,9]:

$$\hat{p}(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = \sum_{i=1}^{N} w_t\left(\mathbf{x}_{0:t}^{(i)}\right)\delta_{x_{0:t}^{(i)}}(\mathrm{d}\mathbf{x}_{0:t})$$

The weights are further given by

$$\tilde{w}_t\left(\mathbf{x}\left(\mathbf{x}_{0:t}^{(i)}\right)\right) = p\left(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}^{(i)}\right)p\left(\mathbf{x}_{0:t}^{(i)}\right)\Big/q\left(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{1:t}\right) \tag{3}$$

$$w_t\left(\mathbf{x}_{0:t}^{(i)}\right) = \tilde{w}_t\left(\mathbf{x}\left(\mathbf{x}_{0:t}^{(i)}\right)\right)\Big/\sum_{i=1}^{N}\tilde{w}_t\left(\mathbf{x}\left(\mathbf{x}_{0:t}^{(i)}\right)\right) \tag{4}$$

where the particles $\{\mathbf{x}_{0:t}^{(i)}, w_t(\mathbf{x}_{0:t}^{(i)}); i = 1,\cdots,N\}$ are drawn from the known distributions q, $\tilde{w}_t(\mathbf{x}(\mathbf{x}_{0:t}^{(i)}))$ and $w_t(\mathbf{x}_{0:t}^{(i)})$ are the unnormalized and normalized importance weights. In order to compute a sequential estimate of the posterior distribution at time t without modifying the previously simulated states $\mathbf{x}_{0:t}$, proposal distribution of the following form can be used:

$$q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = q(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1})q(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t}) \tag{5}$$

Here we are making the assumption that the current state is not dependent on future observations. Besides, under our assumptions that the states correspond to a Markov process and that the observations are conditionally independent given the states [7], i.e.:

$$p(\mathbf{x}_{0:t}) = p(\mathbf{x}_0)\prod_{j=1}^{t} p(\mathbf{x}_j|\mathbf{x}_{j-1}) \quad \text{and}$$

$$p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}) = \prod_{j=1}^{t} p(\mathbf{y}_j|\mathbf{x}_j) \tag{6}$$

By substituting Eqs. (5) and (6) into Eq. (3), a recursive estimate for the importance weights can be derived as follows: