Signal Processing 96 (2014) 382–394

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Particle filtering for TDOA based acoustic source tracking: Nonconcurrent Multiple Talkers

Xionghu Zhong [a,*], James R. Hopgood [b]

[a] Centre for Multimedia and Network Technology, The School of Computer Engineering, College of Engineering, Nanyang Technological University, Singapore 639798, Singapore
[b] The Institute for Digital Communications, Joint Research Institute for Signal and Image Processing, School of Engineering, The University of Edinburgh, King's Buildings, Edinburgh EH9 3JL, UK

ARTICLE INFO

Article history:
Received 13 February 2013
Received in revised form
25 May 2013
Accepted 3 September 2013
Available online 13 September 2013

Keywords:
Acoustic source tracking
Room reverberation
Time-delay of arrival
Particle filtering
Extended Kalman filter

ABSTRACT

Room reverberation introduces multipath components into an audio signal and causes problems for acoustic source localization and tracking. Existing tracking methods based on the extended Kalman filter (EKF) and sequential importance resampling based particle filter (SIR-PF) usually assume that a single source is constantly active in the tracking scene. Assuming that multiple talkers may appear alternatively during a conversation, this paper develops an extended Kalman particle filtering (EKPF) approach for nonconcurrent multiple acoustic tracking (NMAT). Essentially, an EKF is introduced to obtain an optimum importance sampling, by which the particles are drawn according to the current time-delay of arrival (TDOA) measurements as well as the previous position estimates. Hence, the proposed approach can quickly adapt to the sharp position change when the source switches and the tracking lag in SIR-PF can be avoided. Moreover, the amplitude of the TDOA measurement is investigated to formulate a measurement hypothesis prior. Such a prior is fused into the tracking algorithm to enhance the tracking accuracy. Both simulations and real audio lab experiments are organized to study the tracking performance. The results demonstrate that the proposed EKPF approaches outperforms the SIR-PF and EKF in a broad range of tracking scenarios.

1. Introduction

Acoustic source (talker) localization and tracking in a room environment plays an important role in many speech and audio applications such as diarization, hearing aids, hands-free distant speech recognition and communication, and teleconferencing systems. Once the talker is localized and tracked, the position information can be fed into a higher processing stage for: high-quality speech acquisition; enhancement of a specific speech signal in the presence of other competing talkers; or keeping a camera focused on the talker in a video-conferencing scenario [1–6]. Usually, a distributed system equipped with a number of microphone pairs/arrays is employed to localize or track the source [7–11]. However, it is a challenge to provide an accurate position estimation since the received audio signal can be significantly distorted and its statistical properties drastically changed due to room reverberation. The difficulties also arise from the uncertainty in the source motion and the non-stationary characteristics of the speech signal.

1.1. Background overview of existing techniques

Existing acoustic source localization (ASL) approaches can be divided into two main categories depending on the

* Corresponding author. Tel.: +65 67904745.
E-mail addresses: xhzhong@ntu.edu.sg (X. Zhong), James.Hopgood@ed.ac.uk (J.R. Hopgood).

0165-1684/$ - see front matter
http://dx.doi.org/10.1016/j.sigpro.2013.09.002

measurement type: location measurement based approaches [12–20] and time-delay of arrival (TDOA) measurement based approaches [7–10,21,22]. The former ones are usually referred to as direct approaches since the location measurement, which is typically extracted using beamforming methods [12,13], directly links to the source position. The latter ones can be regarded as indirect methods since the measurement contains the position information in a nonlinear time-delay of arrival (TDOA) function. TDOA measurement can be extracted, for example, by employing the generalized cross-correlation (GCC) function [23] or an adaptive eigenvalue decomposition (AED) algorithm [24]. Since each TDOA yields half a hyperholoid of two sheets (see Eq. (3)) which, in the far field, can be approximated by an angular segment [7], multiple TDOA measurements from distributed microphone pairs/arrays are usually employed to triangulate a target position. The direct approaches have the advantage that the relationship between the measurement and state is linear [10]. However, extracting the position measurement requires a multi-dimensional search over the state space and is usually computationally expensive. In contrast, the TDOA measurements are simple and easily available in many applications, and are extensively studied and used for either localization or tracking [8–10,22,25–30]. Other measurement types such as range difference measurements [31,32], interaural level difference [33,34] and joint TDOA and vision [35–38] have also been employed for room acoustic source position estimation. Also, acoustic source tracking using a single acoustic vector sensor has been studied in [39,40].

If there is a time delay between the received signals, and the background noise has a Gaussian distribution, then the TDOAs can be accurately extracted from the GCC function, as the largest peak of the GCC function corresponds to the TDOA measurement. The measurements taken from all microphone pairs are then used to triangulate the position based on a maximum likelihood (ML) criterion [5]. Since the TDOA measurement function is nonlinear, this triangulation can be approximated either by using a linear intersection algorithm [7] or by using an extended Kalman filter (EKF) [9,22]. However, the performance of these algorithms can be seriously degraded due to the presence of reverberation and different kinds of noise in real-life. The sequential importance resampling based particle filter (SIR-PF) [10,21,39] was introduced into the room acoustic source localization and tracking (ASLT) problem to reduce TDOA errors caused by multipath reverberant components. The likelihood is formulated by using a bi-model: a Gaussian distribution for real TDOA measurements and a uniform distribution for false alarms. Generally, the SIR-PF is able to exploit both the temporal information from the source dynamic model and the spatial information from the TDOA measurements. It is therefore more accurate than the linear intersection based localization, which only takes the spatial information into account. Moreover, due to the incorporation of the bi-modal likelihood, the SIR-PF is less affected by the false TDOA measurements and is more robust than the EKF on its own in noisy and reverberant environments [9,22]. In [14], more advanced particle filter (PF) algorithms that incorporate a voice activity detector (VAD) have been developed for room ASLT. The VAD is employed to reduce the effect of heavy false alarms due to the weak source signals in silence gaps.

1.2. Proposed approach and contributions

In this paper, the nonconcurrent multiple acoustic tracking (NMAT) problem is addressed in which multiple talkers have distinct spatial locations and only speak one-at-a-time, as might occur in many parts of a polite conversation, or other scenarios such as a scripted scene in a production. While this is more specific than the problem of jointly detecting and tracking multiple concurrent talkers, as addressed in [20,40–42], it is both important where computational constraints are important, but also for investigating the tracking transition time among different individual talkers where multiple concurrent acoustic tracking (MCAT) approaches cannot offer any advantages. Therefore, in the NMAT case, considering the estimation problem as tracking the position of the current talker (the *target*), the source position may change drastically as different talkers speak due to their distinct spatial locations. This particular case requires the algorithm to capture the sharp change in position and lock onto the position of the new talker rapidly. Unfortunately, the SIR-PF suffers from tracking lags and losses when following a sharp change of target position, such as in the NMAT scenario; this is because the particles are only drawn according to the source dynamic model which does not explicitly model rapid changes in target position.

The central idea of the approach proposed in this paper is an extension of the work proposed in [43], and is that by employing an EKF in the particle filter, the optimal importance function is approximated and the particles sampled in a more relevant area compared with using the prior density function as in the SIR-PF. Since optimal importance sampling is achieved, the proposed approach can lock on to the rapid target position changes, thus avoiding the tracking lag in the SIR-PF which occurs in the NMAT scenario. Since multiple TDOA measurements with weighting information are collected across each microphone pair, the EKF cannot be applied directly and therefore two novel methods are developed in this paper to incorporate the EKF to the multiple-measurement case. The first approach uses only the TDOA from the highest peak in the phase transform GCC (PHAT-GCC) function as the measurement at each microphone pair. This is reasonable since the TDOAs from the highest peaks are, in most cases, more reliable than those from the remaining peaks. The second novel method takes all the TDOA measurements into account and incorporates amplitude information of the TDOA peaks in the tracking algorithm to provide a prior probability of the measurement hypothesis. Finally, a parameter is introduced in the innovation updating process to reduce the effect of false alarms. The advantage of the proposed approach in NMAT are assessed via simulated room environment experiments as well as real audio lab experiments.

The core contribution of this paper is that the nonconcurrent multiple acoustic tracking problem in a noisy and reverberant environment is addressed and accordingly, an extended Kalman particle filtering (EKPF) approach is