TRACKING MULTIPLE ACOUSTIC SOURCES IN REVERBERANT ENVIRONMENTS USING REGULARIZED PARTICLE FILTER

F.Antonacci, M.Matteucci, D.Migliore, D.Riva, A.Sarti, M.Tagliasacchi, S.Tubaro

Dipartimento di Elettronica ed Informazione, Politecnico di Milano,
Via Ponzio, 34/5 - 20133 Milano (Italy)

ABSTRACT

This paper concerns the problem of tracking acoustic sources in reverberant environments by using a particle filter. The localization problem is transformed into the retrieval of the unobservable state of a dynamical model through noisy measures. Though effective, two problems are related to particle filter: the degeneracy phenomenon (all particles but one are not significative) and the loss of diversity (all particles collapse on the same point). By using Regularized particle filter (RPF) and Expectation Maximization (EM) we propose a solution to both problems. Experimental results validate the proposed solution: Regularized Particle Filter enables to obtain a RMS error lower than 0.2m with a reverberation time of 0.6s.

Index Terms— Particle filter – Regularized particle filter – Expectation Maximization – Time Differences of Arrival – Blind Source Separation

1. INTRODUCTION

The tracking acoustic sources is a problem that arises in a variety of applications like seismology, sonar and speech processing. This work is specifically concerned with the tracking of multiple acoustic sources in reverberant environments.

The localization is typically achieved in two steps: first, measures related to the source position are obtained by a suitable algorithm. In a second stage the source position is estimated by a state-space model, in which the source position is the unobservable state. In such a data model, the state is related to the noisy measures through nonlinear equations.

When more than one source is present, Time Differences of Arrival (TDOAs) are obtained through a separation algorithm as a preprocessing stage. In particular, when no information about source signals is available, Blind Source Separation (BSS) techniques are generally used. Recently, the TRINICON algorithm [1] has received a great deal of attention as an effective solution to the separation problem in mildly reverberant environments. The extrema of the de-mixing filters are related to the TDOAs of the sources [2]. In presence of reverberation it is difficult to achieve the correct separation, but the extrema of the de-mixing filters are reliable for the most part of their time. Thus TDOAs can be used by a filter to fulfill the localization task.

Due to the strong nonlinearity of the measurement-state relation, the Kalman filtering (KF) and the Extended Kalman Filtering (EKF) (see [3]) are both unsuitable for the purpose of retrieving source position from noisy TDOAs. Recently particle filter [4] has gained a great deal of attention as a solution for the problem of state estimation in nonlinear, multi-modal and non-Gaussian PDFs.

An introductive survey on particle filter for tracking acoustic sources was presented in the work of Ward et al. [5], where TDOAs are obtained with Generalized Cross Correlation (GCC) and Adaptive Eigenvalue Decomposition Algorithm (AEDA). Though effective, this approach suffers from being applicable only when one source is active.

Saiu et al. [6] have shown that localization of multiple acoustic sources with particle filter is possible through a suitable clustering algorithm: the algorithm presented in [5] is modified in the sense that particles are labeled by a k-means algorithm. Each particle is assigned to one of the sources. The state-space model is the same as in [5]: particles are shifted along pseudo-random trajectories. A problem related to the particle filter is the degeneracy phenomenon: all but one particle are not significative. In order to overcome the degeneracy problem, different strategies have been proposed in the literature. The first solution consists in re-sampling the set of particles every time the fraction of significative particles is below a threshold (see [7] for details) and by adopting a suitable importance density function during the re-sampling step ([8]).

The re-sampling step introduces the problem of loss of diversity among the particles, since particles are drawn from a discrete distribution and not a continuous one. A possible solution proposed in literature is the Regularized particle filter (RPF): RPF is identical to the traditional particle filter except for the re-sampling stage: the RPF re-samples from a continuous approximation of probability density function ([9]).

Croene et al. [10] have recently proposed an alternative solution to the degeneracy problem by using a different dynamical model: the main idea is that at each iteration of the PF, the particle that best explains the observations is assigned the role of “master”. All the other particles change their velocity in order to follow the master with some momentum. The degeneracy problem is overcome but the set of particles concentrate in the proximity of the “master”, and it exhibits a significative loss of diversity.

In this paper we propose an extension of the approach followed by Saiu et al. [6] in order to account for the specificity of the acoustic source tracking problem. In particular, we present a new method based on the combination of Regularized particle filter (for details see [4]) and Expectation Maximization which allows us to modify in real-time the number of particles according to the uncertainty of localization. In this way it is possible to track acoustic sources when measures are not reliable.

The rest of this paper is organized as follows: Section 2 shows the essence of the TRINICON algorithm and explains the motivations of using a tracking technique to obtain reliable estimations of the source positions. Section 3 focuses on the Regularized particle filter, from both a theoretical and an implementation points of view. Finally Section 4 shows the experimental results obtained with Regularized particle filter compared with those of traditional and Swarm particle filter.

2. TRINICON ALGORITHM

In the BSS literature, the data model is a convolutive mixture, as it represents the signal received by each of the P microphones as a sum of filtered replica of the sources:

$$x_p(n) = \sum_{q=1}^{Q} \sum_{k=0}^{M-1} h_{pq}(k) s_q(n-k), \quad (1)$$

where Q is the number of active acoustic sources and $h_{pq}(k)$, $k = 0, ..., M-1$ denote the coefficients of the finite impulse response (FIR) from the q-th source to the p-th microphone. In the following it will be assumed that the number of microphones equals the number of sources ($Q = P$). The goal of BSS is to find a de-mixing system where the output signals $y_q(n)$, $q = 1, \ldots, Q$ are described by:

$$y_q(n) = \sum_{p=1}^{P} \sum_{k=0}^{L-1} w_{pq}(k) x_p(n-k), \quad (2)$$

where $w_{pq}(k)$, $k = 0, ..., L-1$ is the de-mixing filter weighting the p-th sensor contribution to the q-th output signal.

The fundamental assumption of the TRINICON algorithm is that sources are non-Gaussian and statistically independent. With this assumptions in mind, adaptive estimation process of de-mixing filters converges to the correct solution when the overall probability density function of the outputs can be factorized out in the product of the marginal PDFs. Once de-mixing filters have been estimated, they can be used to retrieve the DOAs of the active sources according to following equations, in the case of two sources and two microphones:

$$\widehat{\tau}_1 = (\arg_n \max |w_{12}(n)| - \arg_n \max |w_{22}(n)|) f_s^{-1} \quad (3)$$

$$\widehat{\tau}_2 = (\arg_n \max |w_{11}(n)| - \arg_n \max |w_{21}(n)|) f_s^{-1} \quad (4)$$

where f_s is the sampling frequency. TDOAs determine a locus of potential positions consistent with the observations. Such a locus is an hyperbola but can be confused with a straight line when the distance from the microphones is much larger than the distance between the sensors. The triangulation of DOAs measured from different microphone pairs can be used to assess the source position. Due to the permutation problem between different pairs, at least three microphones pairs are needed in order to correctly localize sources in space. To achieve source localization, the position of global maxima/minima is used, thus the information contained in de-mixing filters is only partially exploited to determine the TDOAs. Figure 1) represents the TDOA error with respect to the true source location measured by TRINICON: we can observe that the sources are often confused, as there is an error of more than 40 degrees, which is the angular separation between the sources, therefore a simple localization algorithm (i.e. triangulation) cannot be used to estimate the position of the sources and a more sophisticated algorithm is needed. Moreover it is possible to notice that the measurements error is caused not only by the presence of reverberation but also by the discrete sampling rate, that induces a finite set of TDOAs. Kalman filter assumes that the relation between the state and the measures is linear and that the probability density function of the state exhibits only one peak; Extended Kalman filter removes the linear assumption but fails with multi-modal pdf's. For such reasons tracking by a traitional Kalman filter or Extended Kalman filter are both unsuitable, making the particle filter approach the most suitable solution.

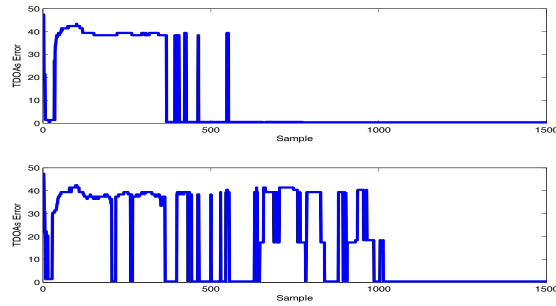


Fig. 1. Example of TDOAs error of a microphones pair

3. REGULARIZED PARTICLE FILTERING FOR SOURCE TRACKING

3.1. Background on particle filter

In this section we will denote the state at time t and the set of states from time 1 to time t with α_t and $\alpha_{1:t}$ respectively. Analogously z_t and $z_{1:t}$ are the observations at time t and the set of observations from time 1 to time t, respectively.

The goal of the particle filter is to estimate the posterior probability $p(\alpha_t|z_{1:t})$. Observing this problem from the point of view of the Bayesian estimation framework, we can break it up in two recursive steps [4]:

Step 1 $\quad p(\alpha_t|z_{1:t-1}) = \int p(\alpha_t|\alpha_{t-1}) p(\alpha_{t-1}|z_{1:t-1}) d\alpha_{t-1}$, (5)

Step 2 $\quad p(\alpha_t|z_{1:t}) = \frac{p(z_t|\alpha_t) p(\alpha_t|z_{1:t-1})}{p(z_t|z_{1:t-1})}$. (6)

Equation (5) is referred in literature as prediction step as it is based on the knowledge of the state evolution $p(\alpha_t|\alpha_{t-1})$ (e.g. a dynamic model). Equation (6) is known as measurement step as it represents the likelihood $p(z_t|\alpha_t)$ of the measure with respect to the state .

The basic idea behind particle filter is to recursively estimate $p(\alpha_t|z_{1:t})$. Starting with a weighted set of particles $\{w_{t-1}^n, \alpha_{t-1}^n\}_{n=1}^{N_s}$ approximatively distributed according to $p(\alpha_{t-1}|z_{1_{t-1}})$, new samples are generated from a suitably designed proposal distribution: $\alpha_t^n \sim q(\alpha_t|\alpha_{t-1}^n, z_t)$ for $n = 1, ..., N_s$.

The function $q(\alpha_t|\alpha_{t-1}^n, z_t)$ is known in literature as importance function and the choice of the correct importance function is crucial for the correctness of estimations. To maintain a consistent sample the new importance weights are set to

$$w_t^n \propto w_{t-1}^n \frac{p(z_t|\alpha_t^n) p(\alpha_t^n|\alpha_{t-1}^n)}{q(\alpha_t^n|\alpha_{t-1}^n, z_t)} . \quad (7)$$

In order to grant that the new set of importance weights approximate the posterior $p(\alpha_t|z_{1_t})$, a normalization step is performes:

$$\sum_{n=1}^{N_s} w_t^n = 1. \quad (8)$$

Many different solutions have been proposed in the past years to implement the particle filter, each of them being different from the others for the choice of the importance density function. In particular the Sequential Importance Sampling approach (SIS) ([11]) forms the basis for other methods. SIS tries to approximate the posterior density through a set of discrete samples, whose weights are computed according to the Importance Sampling theory. The main drawback related to Sampling Importance Sampling is the degeneracy problem: after a few steps of recursion (5) and (6), all but one particles have negligible weight. In order to characterize the degeneracy phenomenon, Liu and Chen have introduced the effective sample size N_{eff}, defined as

$$N_{eff} = \frac{N_s}{1 + \mathrm{Var}(w_t^{*n})} \tag{9}$$

An exact evaluation of N_{eff} is impossible, so an estimate $\hat{N}_{eff}$ is used.

Different approaches have been proposed to overcome the degeneracy problem. The first solution resides in appropriately choosing the importance density function: in fact the choice $q(\alpha_t|\alpha_{t-1}^n, z_t) = p(\alpha_t|\alpha_{t-1}^n, z_t)$ yields $\mathrm{Var}(w_t^{*n}) = 0$. The second solution adopted in literature consists in re-sampling the set of particles whenever $\hat{N}_{eff}$ is below a threshold N_T (see [8]).

The major drawback related to re-sampling is that the particles suffers from loss of diversity: the new set contain many related points. This problem, known also as sample impoverishment, is severe in the case of small noise in the observations. A solution consists in the regularization of the set of particles.

3.2. Regularized Particle Filter

The loss of diversity arises during the re-sampling step: this is due to the fact that new samples are drawn from a discrete distribution rather than a continuous one. A modified particle filter, known as Regularized Particle Filter (RPF) was proposed as a potential solution to the above problem ([9]). The RPF re-samples from a continuous approximation of the posterior density:

$$p(\alpha_t|z_{1:t}) \approx \sum_{n=1}^{N_s} w_t^n K_h(\alpha_t - \alpha_t^n) \ , \tag{10}$$

where

$$K_h(\alpha) = \frac{1}{h^{s_\alpha}} K\left(\frac{\alpha}{h}\right)$$

is the re-scaled Kernel density $K(\cdot)$, $h > 0$ is the Kernel bandwidth, s_α is the dimension of the state vector α. The Kernel density is a symmetric PDF with finite variance. The Kernel and the bandwidth are chosen to minimize the Mean Integrated Square Error (MISE) between the true posterior density and the regularized representation. In the special case of an equally weighted sample, the optimal choice of the Kernel is the Epanechnikov Kernel [9]:

$$K_{opt} = \begin{cases} \frac{s_\alpha + 2}{2c_{\alpha_x}}(1 - ||\alpha||^2) & \text{if } ||\alpha|| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where c_{α_x} is the volume of the unit hypersphere in $\mathbb{R}^{s_\alpha}$. When the underlying density is Gaussian, with unit covariance matrix the optimal choice for the bandwidth is

$$h_{opt} = [8c_{s_\alpha}^{-1}(s_\alpha + 4)(2\sqrt{\pi})^{s_\alpha}]\frac{1}{s_\alpha + 4} \ . \tag{12}$$

The results are optimal only in the special case of equally weighted particles and underlying Gaussian density. However the regularization can still be used in the general case to obtain a suboptimal filter. The RPF differs from generic particle only for the addition of the regularization steps when conducting the re-sampling.

3.3. Particle Filter for Source tracking

Let us call with α_t the state of the underlying dynamical model at time t. The state is composed by the position and velocity of the source:

$$\alpha_t = [X(t), Y(t), \dot{X}(t), \dot{Y}(t)]^T \ . \tag{13}$$

At time t a new set of TDOAs becomes available. The source motion is modeled as a Langevin process, which is specified, for the x coordinate by the following equations:

$$\dot{X}_t = a_X \dot{X}_{t-1} + b_X F_{X_t} \ , \tag{14}$$

$$X_t = X_{t-1} + \Delta T \dot{X}_t \ , \tag{15}$$

where $F_{X_t} \sim N(0, 1)$, ΔT is the discretized time step and

$$a_X = \exp(-\beta_X \Delta T) \ , \tag{16}$$

$$b_X = \overline{v}_X \sqrt{1 - a_X^2} \ , \tag{17}$$

with $\overline{v}_X$ the steady state root mean square velocity. Equations from (14) to (17) specify a dynamic model according to the first-order Markov assumption. In particular we can write that

$$\begin{aligned} p(\alpha_t|\alpha_{t-1}) &= p(X_t|X_{t-1}, \dot{X}_t)p(\dot{X}_t|\dot{X}_{t-1}) \\ &\quad p(Y_t|Y_{t-1}, \dot{Y}_t)p(\dot{Y}_t|\dot{Y}_{t-1}) \ . \end{aligned} \tag{18}$$

The system we have used is composed by four microphone pairs indexed by m. Each pair provides two TDOAs, denoted as $\hat{\tau}^{(m,1)}$ and $\hat{\tau}^{(m,2)}$. The likelihood function is thus

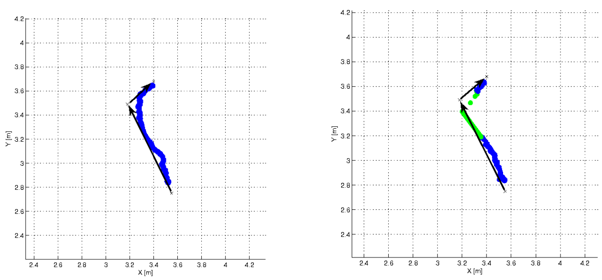
$$p(z_t|\alpha_t) = \prod_{m=1}^{M} \sum_{k=1,2} q_k N(\tau_{\alpha_t}; \hat{\tau}^{(m,k)}, \sigma^2) + q_0 \ , \tag{19}$$

where $N(x, \mu, \sigma^2)$ is the probability of extracting x from a Gaussian distribution having mean μ and variance σ^2, z_t is the measure set (the collection of TDOAs). The value of q_0 is the prior probability that none of the potential locations is due to the source location and $q_k = (1-q_0)/2$. It is important to notice that the TRINICON needs an initialization period to reach the numerical convergence. Consequently, before using TDOAs, it is necessary to perform a check of measures stability. In order to initialize the Regularized particle filter, we decided to use an initialization step based on the Expectation Maximization algorithm. During the startup of the system, each source in the environment is localized and distinguished. After this, a Regularized particle filter is assigned to each source. In order to achieve a reliable estimation of the source position in the initialization phase, we applied a particle filter similar to [6]. When particles are concentrated in a small portion of space a clustering algorithm is applied, calculating mean anda variance of each cluster. The clustering algorithm used in this work is the Expectation Maximization (EM). EM is used in statistics to find maximum likelihood estimates of parameters in probabilistic models, where the model depends on unobserved variables. EM is called in this way because it alternates between an expectation (E) step, which computes an expectation of the likelihood by including the latent variables as if they were observed, and a maximization (M) step, which computes the maximum likelihood of the parameters by estimating the expected likelihood found in the E step. The parameters found on the E step are then used to begin another E step and the process is repeated. In this work we use EM to estimate the parameters of the Gaussian Mixture Model related to the likelihood function $p(z_t|x_t)$ of the measures with respect to the state.

4. EXPERIMENTS AND RESULTS

To evaluate the algorithm explained in this paper we use a simulated dataset, using speech male segments sampled at $fs = 44.1KHz$ as original source signals. The impulse responses from each source to each microphone have been simulated using a fast beam tracing algorithm every $0.125s$ along the source path. In particular, the dataset was obtained considering two sources moving in a room with dimensions $5m \times 5m \times 2.7m$ and using four couples of microphones each positioned at the side of the room.

The trajectory followed by sources and an example of localization in mildly reverberating conditions (T60 = 250 ms) is depicted in Figure 2.



(a) (b)

Fig. 2. Tracking results obtained with synthetic data without (a) and with (b) sources pause

These experiments are conducted with different reverberation times (from $0.11s$ to $0.61s$) and the results are compared with those of Swarm particle filter [10], SIS particle filter and triangulation. In the special case of this work, the comparison index is the Root Mean Square localization error (RMS) computed as the average distance between the ground truth data (the real position) and the localization results. The Figure 3 shows how the algorithm proposed, based on RPF, outperforms the other techniques in environments with reverberations, keeping the localization error under $0.23m$.

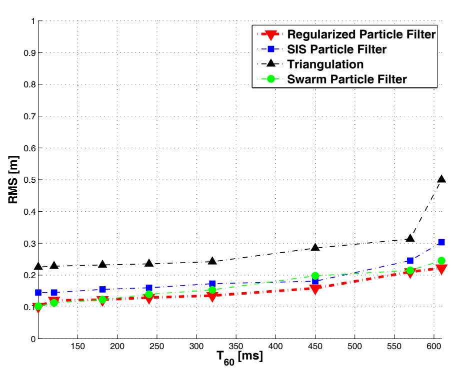


Fig. 3. RMS localization error of different algorithms as a function of reverberation time

In Figure 2b it is possible to observe the same experiment conducted in presence of pauses. The pause occurs in proximity of a direction change of the target in the top-right of the room. It is important to notice that when the source returned active the RPF recovered quickly the correct source pose.

5. CONCLUSIONS

In this paper we have presented a new algorithm for localizing and tracking acoustic sources in reverberant environment. The experimental results demonstrate that the solution proposed works well both with active and inactive sources.

In order to make the algorithm more general, the a priori knowledge about the number of sources in the environment can be removed, and it can be retrieved by analyzing the data and estimating the number of sources in the initialization step.

6. REFERENCES

[1] R. Aichner H. Buchner and W. Kellermann, "Trinicon: A versatile framework for multichannel blind signal processing," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, May 2004, vol. 3, pp. 898–92.

[2] J. Stenglein H. Teutsch W. Kellermann H. Buchner, R. Aichner, "Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Philadelphia, PA, Mar. 2005, vol. 3, pp. 97–100.

[3] John B. Moore Brian D.O. Andersson, *Optimal Filtering*, Englewood Cliffs, Prentice Hall, 1979.

[4] N.Gordon M.S.Arulampalam, S.Maskell and T.Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, pp. 174–188, Feb. 2002.

[5] R.C. Williamson D.B. Ward, E.A. Lehmann, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Trans. on Speech and Audio Processing,*, vol. 11, pp. 826–836, Nov. 2003.

[6] D.Riva D.Saiu A.Sarti M.Tagliasacchi S.Tubaro F.Antonacci, "Tracking multiple acoustic sources using particle filtering," in *In European Signal Processing Conference, Florence, Italy*, Sep. 2006.

[7] J.S. Liu and R.Chen, "Sequential monte carlo methods for dynamics systems," *Journal of the American Statistical Association*, vol. 93, pp. 1032–1044, 1998.

[8] P.Clifford J.Carpenter and P.Fearnhead, "Improved particle filter for nonlinear problems," in *IEEE Proceedings on Radar and Sonar Navigation*, 1999.

[9] C. Musso, N. Oudjane, and F. Le Gland, *Improving regularized particle filters*, chapter 12, pp. 247–271, Springer–Verlag, 2001.

[10] P. Croene R. Parisi and A. Uncini, "Particle swarm localization of acoustic sources in the presence of reverberation," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2006, pp. 4739 – 4742.

[11] S. J. Godsill A. Doucet and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.