TRACKING MANEUVERING SOURCES IN ICA

S.A. Banani [1], M.A. Masnadi-Shirazi [2], A. Masnadi-Shirazi [3]

[1]Dept of Electrical Engineering, Shiraz University, Shiraz, Iran
[2]Dept of Electrical Engineering, Shiraz University, Shiraz, Iran
[3]Dept of Electrical Engineering, University of California, San Diego, USA

ABSTRACT

The problem addressed in this work is to separate the signals of moving sources with independent component analysis (ICA) and tracking the kinematics (position, velocity, acceleration) of each individual source in the working space using Particle filters. To identify the unpredictable movement of the speaker over time, the new proposed state switching scheme handles the uncertainty of the speaker‿s motion by incorporating multiple motion models in the tracking process instead of using the conventional IMM algorithm. The algorithm performance has been verified by illustrating some simulation results.

Index Terms— ICA; Particle filter; Source separation; tracking; positioning

1. INTRODUCTION

The problem of tracking sources in reverberating environments is relevant in several applications, including seismology, sonar and speech. Localizing and tracking multiple speakers talking in the same room can be used, for example, to automatically steer camera sensors in video-conferencing applications.

Several works with deferent strategies have been done in the field of tracking and source separation. Some works like [1] focuses on TRINICON algorithm in the noise free environment. But in real life, there is always some kind of noise present in the observations. Noise can correspond to the actual physical noise in measuring devices or accuracies of the model used.

In [2] an algorithm based on IMM-PDA filters has been proposed. Each of the speakers‿ state equation describing their movement and the observation equation has been assumed to be a linear function of the state. However, any of the equations may be a nonlinear function of the states. In such a case, using Kalman filters are not suggested. Furthermore, when there are severe nonlinearities in either of the state or observation equations, extended Kalman filters falls beneath its suboptimal performance. In this cases, using Particle filters as nonlinear state estimators are more suitable.

In this work we will present a novel, general framework that can deal with both cases, that is, dealing with the nonlinearities of the state and observation equations for tracking the sources and separating the voices of multiple, possibly moving, speakers in the noisy environment. In order to be able to cover the unpredictable movement of the speaker over time, the new proposed state inference scheme, handles the uncertainty of the speaker‿s motion by incorporating multiple motion models in the tracking process.

2. PROBLEM DESCTRIPTON

2.1. ICA Model

In the standard noisy ICA, the noise is assumed to be additive. This is a rather realistic assumption used in factor analysis and signal processing, and allows for a simple formulation of the noisy model. Thus, the noisy ICA model can be expressed as

$$\mathbf{o}_k = \mathbf{A}_k \mathbf{s}_k + \mathbf{w}_k \qquad (1)$$

where the vector $\mathbf{s}_k$ is the vector of independent sources and $\mathbf{o}_k$ is the observation vector in each iteration and $\mathbf{w}_k$ is the vector of additive noise that in general can have any non-Gaussian distribution.

Assume that we have L independent source components and M observations at each iteration. The indices k shows that in each iteration, the mixing matrix $\mathbf{A}$ is changing due to the movement of sources or possibly non-stationary environment of the work space.

As the elements of matrix A, i.e. a_{ij}, are some parameters that depend on the distances between the microphones and the sources, we may write any desirable nonlinear relation between a_{ij} and the distances. Thus we may write

$$a_{ij} = f_{ij}(r_{ij}), \quad i = 1,2,...,M \quad j = 1,2,...,L \qquad (2)$$

where r_{ij} is the distance between source j and microphone i.

Note that in general, r_{ij} depends on x, y and z which are the related distances in three dimensions of working space. For example if source i is located in the origin, i.e. (0, 0, 0),

We may write

$$r_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2 + z_{ij}^2} \quad , \; j = 1,2,...,L \tag{3}$$

Thus for example in the case of two independent sources with two microphones we may have

$$\begin{bmatrix} o_{1k} \\ o_{2k} \end{bmatrix} = \begin{bmatrix} f_1(r_{11}(x,y,z)) & f_2(r_{12}(x,y,z)) \\ f_3(r_{21}(x,y,z)) & f_4(r_{22}(x,y,z)) \end{bmatrix}_t \begin{bmatrix} s_{1k} \\ s_{2k} \end{bmatrix} + \begin{bmatrix} w_{1k} \\ w_{2k} \end{bmatrix} \tag{4}$$

2.2. Source Models

In order to be able to separate light-tailed sources a more flexible source model than the traditional 1/cosh density is needed [3]. It is difficult to use a switching model [4] in this context. We used generalized exponentials that provide a good deal of flexibility and do not suffer from the combinational complexities associated with mixture models. Each source density is modeled by

$$p(s^m | \theta_m) = b \exp - \left| \frac{s^m - v_m}{\omega_m} \right|^{r_m} \tag{5}$$

where the normalizing constant is

$$b = \frac{r_m}{2\omega_m \Gamma(1/r_m)}$$

Clearly this density is Laplacian when $r_m = 1$ (super-Gaussian) and uniform distribution as $r_m \to \infty$ (sub-Gaussian).

Generalized exponential source models in static ICA are able to separate mixtures of Laplacian, Gaussain and uniformly distributed sources while methods using static tanh nonlinearity are unable to separate such mixtures [3].

2.3. Source Dynamics

Here, we consider the speaker's movement as two different dynamic models describing constant velocity and constant acceleration motions.

For each source, we may write the constant velocity motion in each direction as

$$\begin{bmatrix} u_i \\ \dot{u}_i \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_i \\ \dot{u}_i \end{bmatrix}_k + \begin{bmatrix} v_{1cv} \\ v_{2cv} \end{bmatrix}_k \quad , \; u = x, y \; or \; z \quad , \; i = 1,2,...,L \tag{6}$$

where T is the sampling time and u_i and $\dot{u}_i$ are the position and the velocity of source i as the states respectively and v_{cv} 's are the additive model noise.

For the constant acceleration mode we have

$$\begin{bmatrix} u_i \\ \dot{u}_i \\ \ddot{u}_i \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & T & \frac{1}{2}T^2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_i \\ \dot{u}_i \\ \ddot{u}_i \end{bmatrix}_k + \begin{bmatrix} v_{1ca} \\ v_{2ca} \\ v_{3ca} \end{bmatrix}_k \quad , \; u = x, y \; or \; z \quad , \; i = 1,2,...,L \tag{7}$$

where $\ddot{u}_i$ is the state which we take as the acceleration of source i and v_{ca} 's are the additive model noise.

In order to relate the aforementioned motion equations and ICA model to a state-estimation problem, we may take the position, velocity and the acceleration of each source in every direction as a state variable. Due to the nonlinearities in the ICA model, the observation equations are nonlinear function of the states. Thus, we have to use a nonlinear state-estimator for the filtering problem. It is preferable to use nonlinear state-estimators such as Particle Filters rather than any of the Kalman Filter's family estimators like Extended Kalman Filters.

2.4. Particle Filtering

The problem is now to track A_k , as new observation $\mathbf{o}_k$ is recorded. If O_k denotes the set of observations $\{\mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_k\}$ then the goal of filtering methods is to estimate the probability density function of the states $p(X_k | O_k)$, where X_k denotes all the state variable vector at iteration k. Particle filters, which date back to the Sampling Importance Resampling (SIR) filter of Gorden *et al* [6] represent the state density $p(X_k | O_k)$ using the prediction and the update stages by a N_p swarm of "particles" each with a importance weights.

In the update stage, the normalized weights are obtained recursively by the following normalized weight updating equation [6]:

$$w_k^n = \frac{w_{k-1}^n p(\mathbf{o}_k | X_{k|k-1}^n)}{\sum_{n=1}^{N_p} p(\mathbf{o}_k | X_{k|k-1}^n)}, \quad n = 1,..., N_p \tag{10}$$

Note that the likelihood $p(\mathbf{o}_k | X_{k|k-1})$ is given by

$$p(\mathbf{o}_k | X_{kt|k-1}) = \int_{\mathbf{s}} p(\mathbf{o}_k | X_{kt|k-1}, \mathbf{s}) p(\mathbf{s}) d\mathbf{s} \tag{11}$$

Due to the independency of original sources, we get

$$p(\mathbf{o}_k | X_{kt|k-1}) = \int_{\mathbf{s}} p(\mathbf{o}_k | X_{k|k-1}, \mathbf{s}) \prod_{m-1}^{L} p_m(s_m) d\mathbf{s} \tag{12}$$

Laplace approximation [7] have been used to approximate the above integral for any fixed $X_{k|k-1}^n$ when the observation noise is small. This approximation has also been used in [8] for the non-stationary ICA problem.

2.5. Source Estimation

The maximum a posteriori (MAP) or mean estimate of $\mathbf{A}_k$ is used to estimate $\mathbf{s}_k$ [7]. Each $\mathbf{s}_k$ is found by maximizing $\log p(\mathbf{o}_k | A_k, \mathbf{s}_k)$.

2.6. Switching Between Different Dynamic Modes

A new technique has been used here for switching between different sources movement modes. When there are a huge number of dynamic modes for the states, it is hard to set the transitional matrix in the traditional techniques such as interacting multiple models (IMM). Furthermore, accurate assignment of the probabilities to all the dynamic modes in each iteration may not be feasible.

In the proposed technique we make use of one additional observation device (sensor or microphone) as well as one artificial source that generates a known signal at each instant of time and for simplicity is located constantly without movement at a known point. Adding the artificial source is to provide a reference measure for identifying the true dynamic mode of the system in each iteration. Thus in the ICA model, we augment the vector of real sources by the artificial source and update the mixing matrix. For example in the case of two real sources, we make use of three microphones and one artificial source. Thus the ICA model becomes

$$\begin{bmatrix} o_{1k} \\ o_{2k} \\ o_{3k} \end{bmatrix} = \begin{bmatrix} f_1(r_{11}(x,y,x))_k & f_2(r_{12}(x,y,x))_k & f_3(r_{13}(x,y,x)) \\ f_4(r_{21}(x,y,x))_k & f_5(r_{22}(x,y,x))_k & f_6(r_{23}(x,y,x)) \\ f_7(r_{31}(x,y,x))_k & f_8(r_{32}(x,y,x))_k & f_9(r_{33}(x,y,x)) \end{bmatrix} \begin{bmatrix} s_{1k} \\ s_{2k} \\ s_{ar}(k) \end{bmatrix} + \begin{bmatrix} w_{1k} \\ w_{2k} \\ w_{3k} \end{bmatrix} \tag{22}$$

As it can be seen, the functions f_3 , f_6 , f_9 are denoted without any iteration index k. in other words they are not changing over time because the artificial source is always located at one specific point.

At each iteration, suppose there are D different dynamic modes for every maneuvering source in each direction of the three dimensional space. For example if L real sources are maneuvering such that they switch between D dynamic modes in each direction of the three dimensional space, so in view of the system the total number of combinational maneuvering modes for the L sources would be equal to

$$DM = \underbrace{D_1^3}_{due\ to\ Source1} \times \underbrace{D_2^3}_{due\ to\ Source2} \times \ldots \times \underbrace{D_L^3}_{due\ to\ Source\ L} = \prod_{i-1}^{L} D_i^3 \tag{23}$$

To identify the dynamic mode at each iteration, first we generate $N_c << N_p$ swarm particles using the prediction stage for each mode. Considering the artificial source as one real source we evaluate the primary source vector estimation for each of its modes using (18), i.e. $\{\tilde{\mathbf{s}}_i^r\}_{r-1}^{N_c}$, $i = 1,\ldots,DM$. Now, we calculate the root mean square Error (RMSE) of the augmented element of vector $\tilde{\mathbf{s}}_i$ which is related to the artificial source, i.e. $\left[\tilde{s}_{L+1}\right]_i$, as the reference measure of accuracy in estimation of the artificial source for the ith mode:

$$RMSE_i(k) = \sqrt{\frac{1}{N_c}\sum_{r-1}^{N_c}\left(\left[\tilde{s}_{L+1}^j\right]_i - s_{ar}(k)\right)^2} \quad , \ i = 1,\ldots,DM \tag{23}$$

The mode which has the least RMSE, is selected as the identified dynamic mode in that iteration.

$$\mathrm{mod}\,e\ \ j = \min_i\{RMSE_i\} \tag{24}$$

After selecting the mode j as the true dynamic mode, the filtering steps begin by generating N_p swarm particles for the identified mode in the prediction stage.

3. SIMULATION RESULTS

Consider we want to track two moving sources in a 2-dimentional room with the working space of $100 \times 100\ m^2$. As it was mentioned before, we make use of one additional artificial source where for simplicity it is located constantly at a point say the center of the room, i.e. $P_{S_3} = P_{S_{ar}} = (79.29\ m, 79.29\ m)$. Three observing devices are also located in three edges on the working space as shown in Fig. (1).

A Laplacian source ($p(s) \propto e^{-|s|}$) and a source with uniform density and the artificial source that generates a signal with amplitude one "+1" at each sample are mixed with the mixing matrix whose components vary with location as follows:

$$\begin{bmatrix} o_{1k} \\ o_{2k} \\ o_{3k} \end{bmatrix} = \begin{bmatrix} 150\ \ \sqrt{x_1^2 + y_1^2} & 150\ \ \sqrt{x_2^2 + y_2^2} & 79.29 \\ 150\ \ \sqrt{(100\ \ x_1)^2 + (100\ \ y_1)^2} & 150\ \ \sqrt{(100\ \ x_2)^2 + (100\ \ y_2)^2} & 79.29 \\ 150\ \ \sqrt{(100\ \ x_1)^2 + (y_1)^2} & 150\ \ \sqrt{(100\ \ x_2)^2 + (y_2)^2} & 79.29 \end{bmatrix} \begin{bmatrix} s_{1k} \\ s_{2k} \\ s_{ar} \end{bmatrix} + \begin{bmatrix} w_{1k} \\ w_{2k} \\ w_{3k} \end{bmatrix} \tag{25}$$

Note that the above choice of the mixing matrix is optional, and we could choose any reasonable nonlinear functions of the distances between the sources and microphones for each of its components. It is worth mentioning that addressing the separation and tracking the independent sources using ICA in 3-dimnessional space is the same as in 2-D frame.

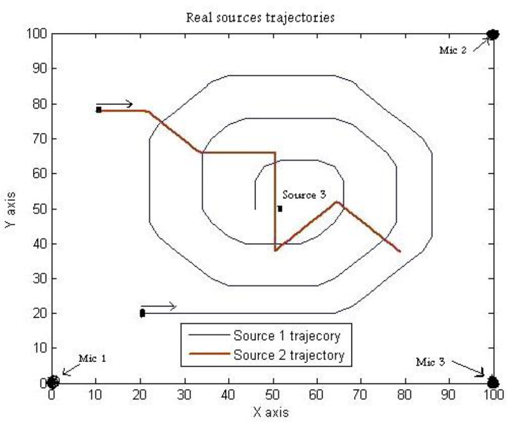

Figure 1. Real target trajectories

The two sources are tracked for 120 samples and the sampling time are considered to be $T = 0.05\ s$. The covariance of the process noise for both constant velocity dynamics and constant acceleration dynamics is $0.01I$ and the covariance of the observation noise is assumed to be 0.1I where I is the identity matrix. The real targets trajectories are shown in Fig (1).

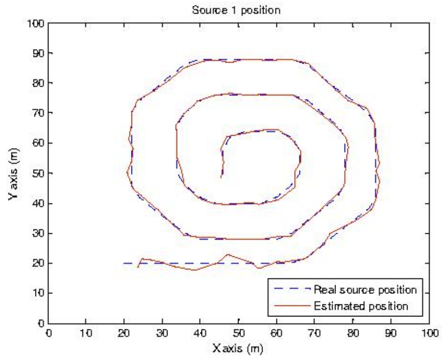

Figure 2. Source 1 real and estimated position

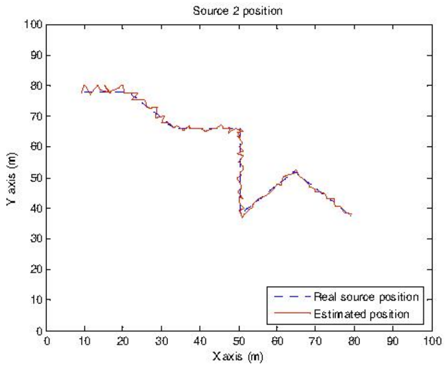

Figure 3. Source 2 real and estimated position

To provide an initial estimate of the mixing matrix, static ICA was run on the 20 samples received by the microphones. After 20 iterations, the program switches to dynamic ICA and the particle filtering algorithm starts its function. Consequently, the two sources are tracked for 100 iterations and the sources are estimated at the end of each iteration.

Figs (2) and (3) show the estimated positions of the two sources in x-y space.

50 Monte Carlo runs are carried out and the average is represented by the root mean square error (RMSE) criterion as a measure of the performance of the estimation of the kinematics (position, velocity, acceleration) in this simulation:

$$RMSE(k) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(r_k - \hat{r}_k^i)^2} \quad , \quad k = 1,2,...,100\,;\; m = 50$$

where $\hat{r}_k^i$ denotes the state estimate of the ith Monte Carlo run for the kth sample. Fig. (4) shows the resultant kinematics estimation by RMSE criterion.

To evaluate the performance of source estimation, we define the ratio of the real source signal power to the mean square error of source estimation in dB. In the average the source estimation performance would be 22.4 dB.

4. CONCLUSIONS

In this paper, we have presented a general scheme that can track and separate the signals of multiple moving sources in noisy environment. Using particle filter in its algorithm deals very well with the nonlinearities of the state and observation equations in tracking the sources. The main feature of this work is proposing a new technique for smooth switching between huge number of dynamic modes of the maneuvering system.

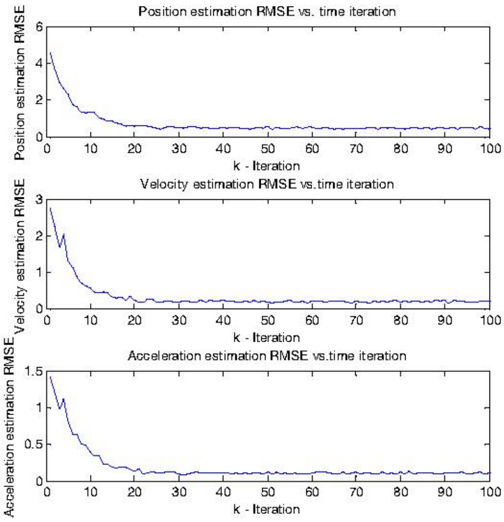

Figure 4. Kinematics estimation RMSE vs. time iteration

5. REFERENCES

[1] H. Buchner, R. Aichner, W. Kellermann, "TRINICON: a versatile framework for multichannel blind signal processing", Proc. IEEE international conference on Acoustic, Speech and Signal Processing. vol. 3, 17-21 May 2004

[2] I. Potamitis, N. Fakotakis, "Tracking and Voice Separationof Moving Speakers based on IMM-PDA Filters" The 7th International Conference on Information Fusion, June 28 to July 1, 2004

[3] R.M. Everson and S.J. Roberts, "ICA: A flexible non-linearity and decorrelating manifold approach.", *Neural Computation,* 11(8), 1999.

[4] T-W. Lee, M. Girolami, and T.J. Sejnowski, "Independent Component Analysis using an Extended Infomax Algorithm for Mixed-Sub-Gaussian and Super-Gaussian Sources." Neural Computation, 11: pp. 417-441, 1999.

[5] N. Gordon, D. Salmond, and A.F.M. Smith, "Novel approach to Nonlinear/Non-Gaussian Bayesian state estimation." IEEE Proceeding-*F,* 140: pp. 107 ت113, 1993.

[6] J.J.K. O' Ruanaidth and W.J. Fitzgerald, "Numerical Bayesian Methods Applied to Signal Processing", Springer, 1996.

[7] R. Everson and S. Roberts, "Independent Component Analysis, Principles and Practice", Cambridge University press 2001

[8] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian non-linear state space models", Journal of Computational and Graphical Statistics, vol. 5, no.1, pp. 1-25, 1996.