A New Nonlinear Filtering Technique for Source Localization

Frederik Beutler, Uwe D. Hanebeck
Intelligent Sensor–Actuator–Systems
Institute of Computer Design and Fault Tolerance
Universität Karlsruhe
Kaiserstrasse 12, 76128 Karlsruhe, Germany
Beutler@ieee.org, Uwe.Hanebeck@ieee.org

Abstract

A new model-based approach for estimating the parameters of an arbitrary transformation between two discrete-time sequences will be introduced. One sequence is interpreted as part of a nonlinear measurement equation, the other sequence is typically measured sequentially. Based on every measured value, the probability density function of the parameters is updated using a Bayesian approach. For the evolution of the system over time, a system equation is included. The new approach provides a high update rate for the desired parameters up to the sampling rate with high accuracy. It will be demonstrated for source localization of a speaker, where the parameters describe the position of the source.

Keywords

Source Localization, Nonlinear Filter Technique, Bayesian Approach

INTRODUCTION

Estimating the position of a target object is a typical problem arising in acoustic localization systems. A sound source is placed in the environment and emits signals that are picked up by a sensor node equipped with microphones. If the signal is additionally transmitted through a different medium to the sensor node, e.g. wireless radio, the position of the source can be calculated. The conventional method for this problem consists of two steps. In the first step, the Times of Arrival (ToA) are estimated for example by means of generalized cross correlation (GCC) [6]. These estimates are then used in the second step, which converts Times of Arrival to object positions based on the known microphone arrangement. This is in general a nonlinear optimization procedure [5].

When the signal emitted by the sound source is not known a priori and no timing reference is given, the position of the source has to be calculated on the basis of Time Differences of Arrival (TDoA). In [8] a Sequential Monte Carlo (SMC) method for source tracking based on Auxiliary Particle Filter (APF) is presented for that purpose. The method makes use of a model for speaker motion. Other frameworks for source tracking based on particle filtering algorithms can be found in [7, 9, 10].

A method for source direction estimation based on hemisphere sampling is presented in [2], where a nonlinear transformation is used to convert the correlation vectors of the GCC for all TDoA measurements to a common coordinate system.

In [3] a fast Bayesian acoustic localization method is presented, a unifying framework for localization can be found in [1]. These approaches are similar to the nonlinear filtering technique presented here, but they use a window for source estimation from received samples. Furthermore, a motion model is not included.

This paper introduces a new model-based approach for estimating the state, e.g. the position of a source, directly from time sequences. Hence, a model for the evolution of the desired state over time can be used. The novelty is the new interpretation of a known reference sequence as part of a measurement equation of a nonlinear system. This new approach can be used as a framework for a multitude of procedures. An important application is the use of this new approach in sensor-actuator-networks, because the received sequence does not have to be stored on the sensor node, since every received measured value will be immediately processed.

In comparison to the conventional method there are several advantages. In the conventional method the estimate of the time delay from the time sequences is the result from processing overlapping or consecutive time windows. This disadvantage disappears in this new approach, since every measured value will be used immediately. Thus the computational complexity is lower and the processing delay for state estimation is shorter compared to the conventional method, because for every measured value the approach provides a result for the desired state. In addition, the structure and the density of measurement noise and process noise can be exploited. The search complexity compared to the conventional beamforming algorithm is reduced. In addition, prior knowledge can be exploited.

The structure of the paper is as follows. The problem of estimating the parameters of a transformation between two time sequences is formulated in the next section. For efficient estimation based on sequential measurements, a nonlinear filtering technique is presented. This approach is then specialized for the case of source localization. The performance of the proposed new approach is evaluated in comparison to

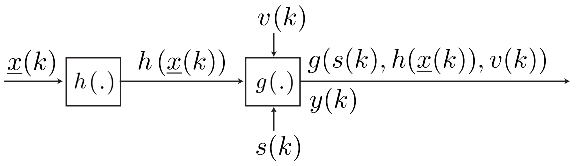


Figure 1. Diagram of the measurement equation.

the conventional method. Conclusions and some hints on future investigations are given in the last section.

PROBLEM FORMULATION

We consider a nonlinear discrete-time system with a system equation, which describes the evolution of the system state over time as

$$\underline{x}(k+1) = a(\underline{x}(k), \underline{w}(k)) \ ,$$

where $a(.)$ is the nonlinear system equation, $\underline{x}(k)$ is the system state at time t_k and $\underline{w}(k)$ is the process noise. The main problem of state estimation based on sequences is the reconstruction of the state $\underline{x}(k)$ from different measured values $y(k)$. In Fig. 1 a block diagram for the discrete-time measurement equation is shown. The parameters of an arbitrary transformation $h(.)$ represent the system state $\underline{x}(k)$, which is mapped to an intermediate state. The nonlinear function $g(.)$ maps the sequence $s(.)$ depending on this intermediate state on the sequence $y(.)$. The sequence $y(.)$ is typically measured sequentially and can be described as

$$y(k) = g(s(k), h(\underline{x}(k)), v(k)) \ , \qquad (1)$$

where $v(k)$ is the measurement noise.

NEW MODEL-BASED APPROACH FOR DIRECT NONLINEAR STATE ESTIMATION

The nonlinear state estimation technique for estimating the desired transformation parameters will be presented in this section.

Measurement Update, Filter Step

In the measurement update, or the filter step, the estimated density $f^e(\underline{x}(k))$ will be calculated according to Bayes' law

$$f^e(\underline{x}(k)) = c(k) \cdot f^J(\underline{x}(k)) \cdot f^p(\underline{x}(k)) \ ,$$

where $f^J(\underline{x}(k))$ is the joint density, which depends on the likelihoods corresponding to the measurement equation. $f^p(\underline{x}(k))$ is the predicted density resulting from the previous estimation step and $c(k)$ is a normalization constant.

Prediction Step

In the prediction step the new predicted density is given by

$$f^p(\underline{x}(k+1)) = \int_{\mathbb{R}} f^T(\underline{x}(k+1)) \cdot f^e(\underline{x}(k)) d\underline{x}(k) \ ,$$

where $f^e(\underline{x}(k))$ is the estimated density from the filter step and $f^T(\underline{x}(k+1))$ is the transition density, which depends on the system equation.

Calculation of Point Estimate

The point estimate of the state can be calculated according to

$$E(\underline{x}(k)) = \int_{\mathbb{R}} \underline{x}(k) f^e(\underline{x}(k)) \mathrm{d}\underline{x}(k) \ ,$$

where $E(\underline{x}(k))$ is the expected value of the estimated density.

SPECIAL CASE: SOUND SOURCE LOCALIZATION

We consider the far field of an acoustic point source located at $\underline{x} = [x \ y \ z]^{\mathrm{T}}$, e.g. a speaker. The source emits a known signal $s(k)$, which is detected by several microphones. The microphones are placed at known points $\underline{M}_i = [x_i \ y_i \ z_i]^{\mathrm{T}}$, $i = 1 \ldots N$ in the environment. For microphone i the received signal $y_i(k)$ according to (1) is given as

$$y_i(k) = g(s(k), h_i(\underline{x}(k), \underline{M}_i), v_i(k)) \ .$$

The speaker signal is delayed and attenuated depending on the distance from the microphones. Assuming additive measurement noise gives

$$y_i(k) = h_i(k, \underline{x}(k), \underline{M}_i) \star s(k) + v_i(k) \ ,$$

where $\star$ stands for convolution. We assume that there is no reflection. The time delay and attenuation of the signal can be described by means of a dirac-function

$$h_i(k) = D_i(\underline{x}(k)) \cdot \delta(k - \tau_i(\underline{x}(k))) \ ,$$

where the attenuation factor and the time delay are given as

$$D_i(\underline{x}(k)) = \frac{1}{\parallel \underline{M}_i - \underline{x}(k) \parallel}$$

and

$$\tau_i(\underline{x}(k)) = \frac{\parallel \underline{M}_i - \underline{x}(k) \parallel}{c} \ ,$$

where c is the acoustic velocity. The received signal at microphone i depending on the position of the microphone $\underline{M}_i$ and of the source $\underline{x}$ is given by

$$y_i(k) = D_i(\underline{x}(k)) \cdot s(k - \tau_i(\underline{x}(k))) + v_i(k)$$

or

$$y_i(k) = \frac{1}{\parallel \underline{M}_i - \underline{x}(k) \parallel} \cdot s\left(k - \left[\frac{\parallel \underline{M}_i - \underline{x}(k) \parallel}{c}\right]\right) + v_i(k) \ . \qquad (2)$$

Measurement Equation

For the derivation of the joint density, according to the measurement update we use (2) and can thus define the likelihood f^L as

$$f^L(\underline{x}(k)) = f^v(y_i(k) - D_i(\underline{x}(k)) \cdot s(k - \tau_i(\underline{x}(k)))) \ ,$$

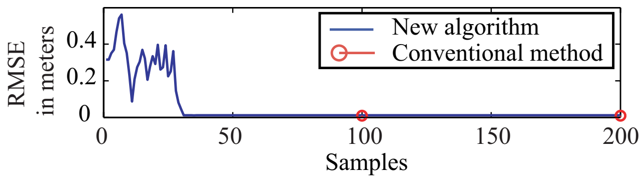


Figure 2. Comparison of the new algorithm and the conventional method: RMSE versus number of received samples.

where $s(.)$ is piecewise constant.
We assume the density f^v to be Gaussian, resulting in

$$f^L(\underline{x}(k)) = \exp(-\frac{1}{2}\frac{[y_i(k) - D_i(\underline{x}(k))s(k - \tau_i(\underline{x}(k)))]^2}{\sigma^2}) .$$

The likelihood function maps the scalar measurement $r_i(k)$ to the 3–dimensional speaker position.
When combining the likelihoods of N microphones to a joint density $f^J(\underline{x}(k))$, we obtain

$$f^J(\underline{x}(k)) = \prod_{i=1}^{N} \exp(-\frac{1}{2}\frac{[y_i(k) - D_i(\underline{x}(k)) \cdot s(k - \tau_i(\underline{x}(k)))]^2}{\sigma^2}) .$$

In this case we assume, that the measurement noise processes at different microphones are independent.

System Equation

We now derive the system equation, which describes the motion of the speaker over time. In the case of additive noise $\underline{w}(k)$ the system equation is given by

$$\underline{x}(k+1) = a(\underline{x}(k)) + \underline{w}(k) .$$

In this paper we consider a simple linear model according to

$$\underline{x}(k+1) = \underline{x}(k) + \underline{w}(k) ,$$

where the uncertainty of the source position is increasing over time.

IMPLEMENTATION

In order to evaluate the performance of the proposed new approach, it was implemented using a grid-based procedure. Although this kind of implementation is inefficient for real-time-tracking, it is sufficient for demonstrating the new algorithm with simulated and real data.

RESULTS

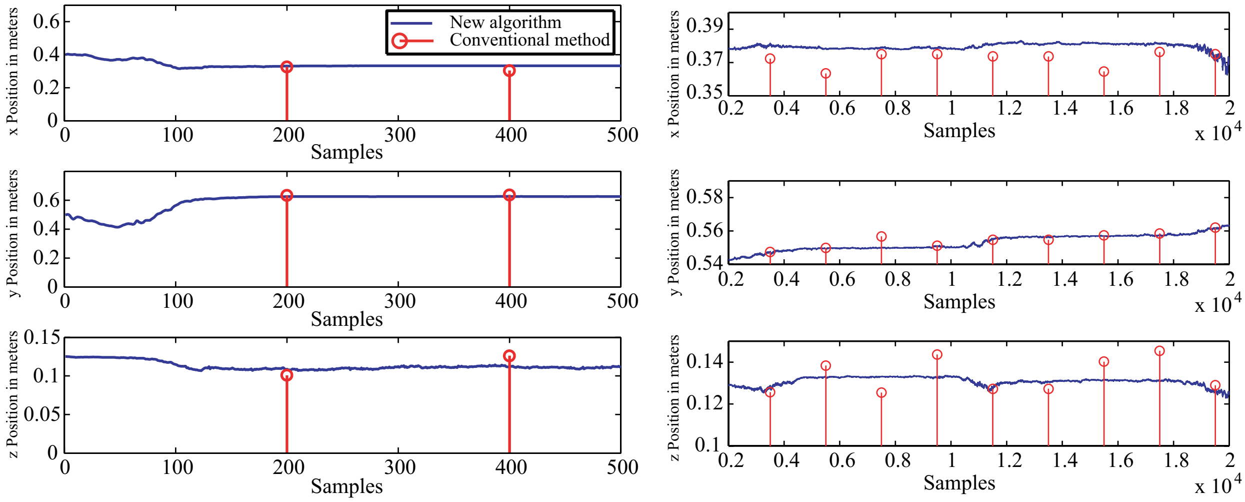
The performance of the new approach is evaluated in simulation as well as in experiments with real data. Furthermore, the new algorithm will be compared with the conventional method. In the simulation the locations of the microphones and of the speaker are given as $\underline{M}_1 = [2.5 \quad 2.8]^{\mathrm{T}}$, $\underline{M}_2 = [3 \quad 2]^{\mathrm{T}}$, $\underline{M}_3 = [2 \quad 1.8]^{\mathrm{T}}$, and $\underline{x} = [0.2 \quad 0.5]^{\mathrm{T}}$, respectively. The speaker signal is delayed depending on the distance between microphone and speaker. The signals are generated with zero-mean white Gaussian noise with the variance 1. In addition, the signals are disturbed by adding zero-mean white Gaussian noise of appropriate variance of 2. The estimated probability density function for the new algorithm is calculated on a two-dimensional grid comprising 50×50 grid points and ranging from 0 m to 1 m. For every received sample the position is calculated. In the conventional method, the Time of Arrival is estimated by means of cross correlation, where the block length for the received sequence is set to 100. In the second step, the estimated Times of Arrival are used to estimate the position of the speaker by using the algorithm presented in [5] as a starting guess for a gradient descent algorithm. In Fig. 2 the Root-Mean-Square-Errors between the estimation of the speaker position and its true value for the new algorithm and the conventional method are plotted for each received sample. The new algorithm converges after 31 samples to the true value, because for every received sample the estimated density is updated, whereas the conventional method provides a result after 100 samples, which depends on the block length.

Finally, we compare the new algorithm to the conventional method in a real experimental setups. The microphone positions are given as $\underline{M}_1 = [0.16 \quad -0.145 \quad 0.115]^{\mathrm{T}}$, $\underline{M}_2 = [0.485 \quad -0.007 \quad 0]^{\mathrm{T}}$, $\underline{M}_3 = [0.315 \quad -0.155 \quad 0.255]^{\mathrm{T}}$, and $\underline{M}_4 = [0 \quad 0 \quad 0]^{\mathrm{T}}$. The sampling frequency is 48 kHz and the acoustic velocity is $340\frac{\mathrm{m}}{\mathrm{s}}$. In the first setup the speaker position is fixed. In Fig. 3a, the speaker position is plotted as a function of the number of received samples. Both methods converge to the same value. The proposed method, however, provides accurate estimates for every new sample starting from sample 150, while the conventional method provides estimates whenever 200 samples have been received and processed as the block length is set to 200.

In the second setup the speaker moves on the y-axis forward to the microphones. The results are plotted in Fig. 3b, where the position of the estimated source position is depicted as a function of the received samples. The new approach provides good tracking results, whereas the conventional method produces outliers resulting from the assumption that the source position remains constant while a block is being received and processed.

CONCLUSIONS AND FUTURE WORK

A new model-based approach for estimating the parameters of an arbitrary transformation between two discrete-time sequences has been presented, where the corresponding model characterizes the evolution of the desired transformation parameters over time. The new approach provides probability densities describing the parameter estimates that are updated



(a) Comparison of the new algorithm and the conventional method for a fixed speaker.

(b) Comparison of the new algorithm and the conventional method for a dynamic speaker.

Figure 3. Experimental results.

based on every received measurement sample. It is demonstrated for the special case of source localization, where the source location parameterizes the transformation between the emitted and the received sound sequence. This parameter set, the source location, is directly estimated from the sound samples and updated with every new sample received. The results are compared with the conventional two-step procedure for source localization, which relies on time-delay estimation. A motion model for a moving object is also included in the new approach.

The presented implementation provides very good results compared to the conventional procedure, but is not efficient as the probability density functions of the parameters are represented on a grid. Future work is concerned with developing an efficient approximation scheme for the probability density functions based on the Progressive Bayes framework [4].

REFERENCES

[1] Birchfield, S. T. A unifying framework for acoustic localization. 12th European Signal Processing Conference (2004).

[2] Birchfield, S. T., and Gillmor, D. K. Acoustic source direction by hemisphere sampling. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 5 (2001), 3053–3056.

[3] Birchfield, S. T., and Gillmor, D. K. Fast bayesian acoustic localization. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2 (2002), 1793–1796.

[4] Hanebeck, U. D., Briechle, K., and Rau, A. Progressive bayes: A new framework for nonlinear state estimation. Proceedings of SPIE, AeroSense Symposium 5099 (2003), 256–267.

[5] Hanebeck, U. D., and Schmidt, G. Closed-form elliptic location with an arbitrary array topology. Proceedings of the 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'96) (1996), 3070–3073.

[6] Knapp, C., and Carter, G. C. The generalized correlation method for estimation of time delay. IEEE Transactions on Acoustics, Speech, and Signal Processing 4, 4 (1976), 320–327.

[7] Lehmann, E. A., Ward, D. B., and Williamson, R. C. Experimental comparison of particle filtering algorithms for acoustic source localization in a reverberant room. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 5 (2003), 177–180.

[8] Vermaak, J., and Blake, A. Nonlinear filtering for speaker tracking in noisy and reverberant environments. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 5 (2001), 3021–3024.

[9] Ward, D. B., Lehmann, E. A., and Williamson, R. C. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. IEEE Transactions on Speech and Audio Processing 11, 6 (2003), 826–836.

[10] Ward, D. B., and Williamson, R. C. Particle filter beamforming for acoustic source localization in a reverberant environment. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2 (2002), 1777–1780.