CLUSTER FILTER

Jari Yli-Hietanen, Konsta Koppinen, Katriina Halonen
Signal Processing Laboratory, Tampere University of Technology,
P.O.Box 553, FIN-33101 Tampere, FINLAND
Tel: +358 3 365 3993; fax: +358 3 365 3857
e-mail: jari.yli-hietanen@tut.fi

ABSTRACT

In this paper, a new form of clustering method is presented where a priori knowledge of the reliability of different samples is used. This knowledge can be inserted into the cluster-finding based computation of the estimator output in the form of sample weights. This kind of method is needed in time-delay based angle of arrival estimation with nonuniform linear sensor arrays.

1 INTRODUCTION

Robust filtering is a largely investigated field [5, 4, 1]. Especially, robust methods are used in removing an impulsive type of noise, paying at the same time attention to detail preservation, especially edges. An edge can be modeled as a case where the majority of the window samples can be thought of as pseudo outliers [2]. Proposed methods for this situation are based on the idea of finding a cluster of samples with the largest sample density [3]. In this paper, we introduce a filter for finding an optimal sample cluster so that the knowledge of the a priori reliability of the samples can also be used. This knowledge can be presented in the form of sample weights. The filter presented in this paper finds a natural application in time-delay based angle of arrival estimation with nonuniform linear sensor arrays.

2 CLUSTER FILTER

Let us assume that we are acquiring data using different sensors and we have received samples

$$x_{ik},\ i = 1, 2, \ldots, N_k,\ k = 1, 2, \ldots, K \tag{1}$$

where N_k is the number of samples from sensor k and K is the number of sensors. Here it is assumed that we have some a priori information on the accuracy of each sensor. The goal is to find the value which represents the samples as well as possible. In this case that means that we are looking for a cluster of samples, i.e. a value y maximizing the number of samples x_{ik} which are at most Δ apart from y is searched. Each sample is weighted by a weight w_k which is assigned to each sensor based on the a priori information about the accuracy of the sensors. Then we maximize the sum of weights of the samples which are at most Δ apart from y instead of the number of samples. Based on that, the output of the cluster filter with x_{ik} as input is defined as follows:

$$\hat{x} = \arg max_y \sum_{k=1}^{K} \sum_{i=1}^{N_k} w_k \rho(x_{ik} - y), \tag{2}$$

where $\rho(x)$ is a windowing function representing the interval $[y - \Delta, y + \Delta]$ in the simplest case.

Procedure to calculate the output of the cluster filter

1. Sort the samples x_{ik}.

2. Go through the sorted samples. At each sample x_j, find all samples falling within the distance Δ from sample x_j. Compute the sum of the weights for this sample set S_j.

3. Find the set S_j with the maximum weight sum.

4. The output of the cluster filter is the weighted mean of this sample set.

3 ANALYSIS OF THE CLUSTER FILTER

To analyse the distributional robustness of the cluster filter, we use the influence function as a tool. The influence function measures the change in the output of a filter in case of slight deviation from the assumed distribution, for instance, in the presence of an outlier. The influence function for a filter is defined as

$$IF(x, T, F) = \lim_{t \to 0+} \frac{T((1-t)F + t\delta_x) - T(F)}{t}, \tag{3}$$

where T is the functional form of the filter in question, F is the assumed distribution of the input samples, and δ_x is distribution of a single sample at point x [4].
In order to write the cluster filter in the form of a functional, we rewrite Equation (2) as follows

$$\hat{x} = \arg max_y \sum_{k=1}^{K} w_k \sum_{i=1}^{N_k} \rho(x_{ik} - y), \tag{4}$$

so that

$$\begin{aligned} T(F) &= \arg max_y \sum_{k=1}^{K} w_k \int \rho(x-y) f_k(x) dx \\ &= \arg max_y \int \rho(x-y) \left(\sum_{k=1}^{K} w_k f_k(x) \right) dx \\ &= \arg max_y \int \rho(x-y) f(x) dx \end{aligned} \tag{5}$$

where

$$f(x) = \sum_{k=1}^{K} w_k f_k(x). \quad (6)$$

The notation $f_k(x)$ stands for the distribution for the samples from sensor k. The final row in Equation (5) is, in fact, the functional form of an M-estimator, whose influence function is [4]

$$IF(x, T, F) = \frac{\psi(x)}{\int \psi' dF}, \quad (7)$$

where $\psi(x)$ is the derivative of $\rho(x)$. So, we see that the cluster filter inherits its robustness properties from the well studied class of M-estimators.

4 APPLICATION

The need for this modification of clustering stems from a specific application. Assume a case where the angle of arrival of a signal is estimated using a linear array of sensors at a varying distance from each other (nonuniform linear array). The angle of arrival is computed using estimated time delays between signals received at different sensors [7]. In an ideal case, a single measure of time delay is sufficient to determine the angle of arrival, but in practice, several time delay estimates are needed to compensate disturbances. Time delay estimation produces outliers, for example, in case of sensor malfunctioning or in the presence of correlating noise at different sensors. Robust methods for time-delay based anegle-of-arrival estimation have been addressed in [8].The nonequal distances of the sensors gives different reliability status to the time delay estimates. This creates an environment where weighting of the time delay estimates is a natural choice. More specifically, the desired output gives us a reliable estimate of time delay normalized with respect to the sensor distances.

5 SIMULATION RESULTS

This section presents simulation results to demonstrate the performance of the proposed method. The simulations consisted of 10000 realisations of a situation with a small sensor array of five omnidirectional sensors, a signal source at a random position. Independent Gaussian noise was added to the signals received by the sensors to model measurement and background noise. The possible malfunctioning of the sensors was also modelled.

The array used in the simulations consisted of five sensors located at a line with distances $8l$, $18l$, $30l$ and $40l$ from the first sensor where l is the velocity of propagation times the sampling interval. The source signal used was a sum of 5 sinusoids with random frequencies between 0.003π and 0.07π with respect to the sampling frequency 2π and with a common amplitude. The source signal was set to propagate as a plane wave from a random direction. For each of the 10000 realisations, the direction of propagation was randomly selected such that θ was chosen from a uniform distribution on $[0,\pi]$ and ϕ from a uniform distribution on $[0,2\pi]$. The polar coordinate system used is shown in Figure 1.

The signals received by the sensors were the sum of the source signal delayed according to their respective direction of propagation and white independent Gaussian noise with variance $0.1v_s$ where v_s was the power of the test signal. Malfunctions in the sensors were simulated by setting the signal received by one randomly selected sensor to white independent Gaussian noise with variance $0.1v_s$.

All the 10 possible time delays between the sensor signals were estimated using polarity coincidence correlation [6]. The length of each realisation was 5000 samples. The time delay estimates were normalized by multiplying them with the product of the velocity of propagation and the sampling interval and dividing them with the corresponding distance between the sensors. For each realisation, representative values for normalized time delay estimates were formed using the mean of time delay estimates, the median of time delay estimates, and the cluster filter. The range for cluster filter was set to 0.3 and the distance between the corresponding sensor pair was used as a weight of the time delay estimate.

The error criterion used was the difference between the estimated and true normalized time delays. The cumulative distribution functions of the difference values estimated from the 10000 realisations for the different methods are shown in Figure 2. The dotted line represents the mean, the dashed line represents the median, and the solid line represents the cluster filter. As seen from the figure, the probability to have small error is the largest for the cluster filter.

6 CONCLUSION

In this paper, a new filter, cluster filter, for finding sample clusters was presented. The filter suits, for instance, time-delay based angle of arrival estimation with nonuniform linear sensor arrays, where outliers may well occur nonsymmetrically. The robustness study of the cluster filter revealed its connection to M-estimators.

The behavior of the cluster filter was simulated in the above mentioned application. For comparison, the mean and the

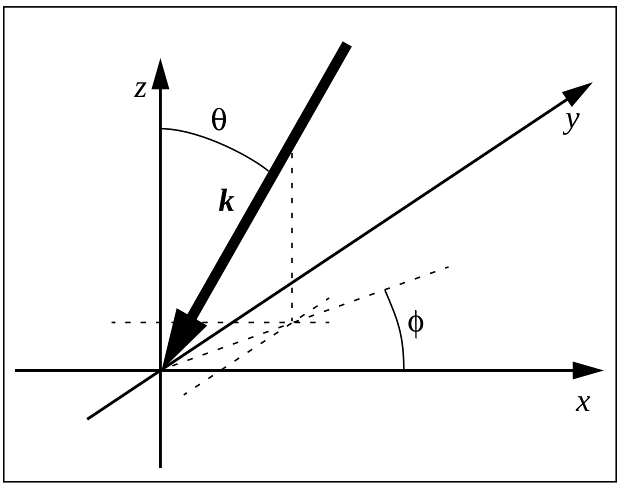

Figure 1. The polar coordinate system used for angle of arrival

median filters were applied, too. In this application, the cluster filter gave better results.

REFERENCES

[1] J. Astola, P. Kuosmanen, Fundamentals of Nonlinear Digital Filtering, New York, CRC Press LLC 1997.

[2] C. Stewart, “Bias in Robust Estimation Caused by Discontinuities and Multiple Structures,“ IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 19, no. 8, pp. 818-833, 1997.

[3] H. Longbotham, D. Eberly, “The WMMR Filters: A Class of Robust Edge Enhancers,“ *IEEE Transactions on Signal Processing*, vol 41, no. 4, pp. 1680-1685, 1993.

[4] F. Hampel, Robust Statistics, New York, Wiley, 1986.

[5] P. Huber, Robust Statistics, New York, Wiley, 1981.

[6] R. Cusani, "Performance of fast time delay estimators," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no 5, pp. 757-759, May 1989.

[7] J. Yli-Hietanen, K. Kalliojärvi, J. Astola, "Low-complexity angle of arrival estimation of wideband signals using small arrays," in *Proc. 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, Corfu, Greece, June 1996, pp. 109-112.

[8] J. Yli-Hietanen, K. Kalliojärvi, J. Astola, "Robust Time-Delay Based Angle of Arrival Estimation," in *Proc. 1996 IEEE Nordic Signal Processing Symposium*, Espoo, Finland, September 1996, pp. 219-222.

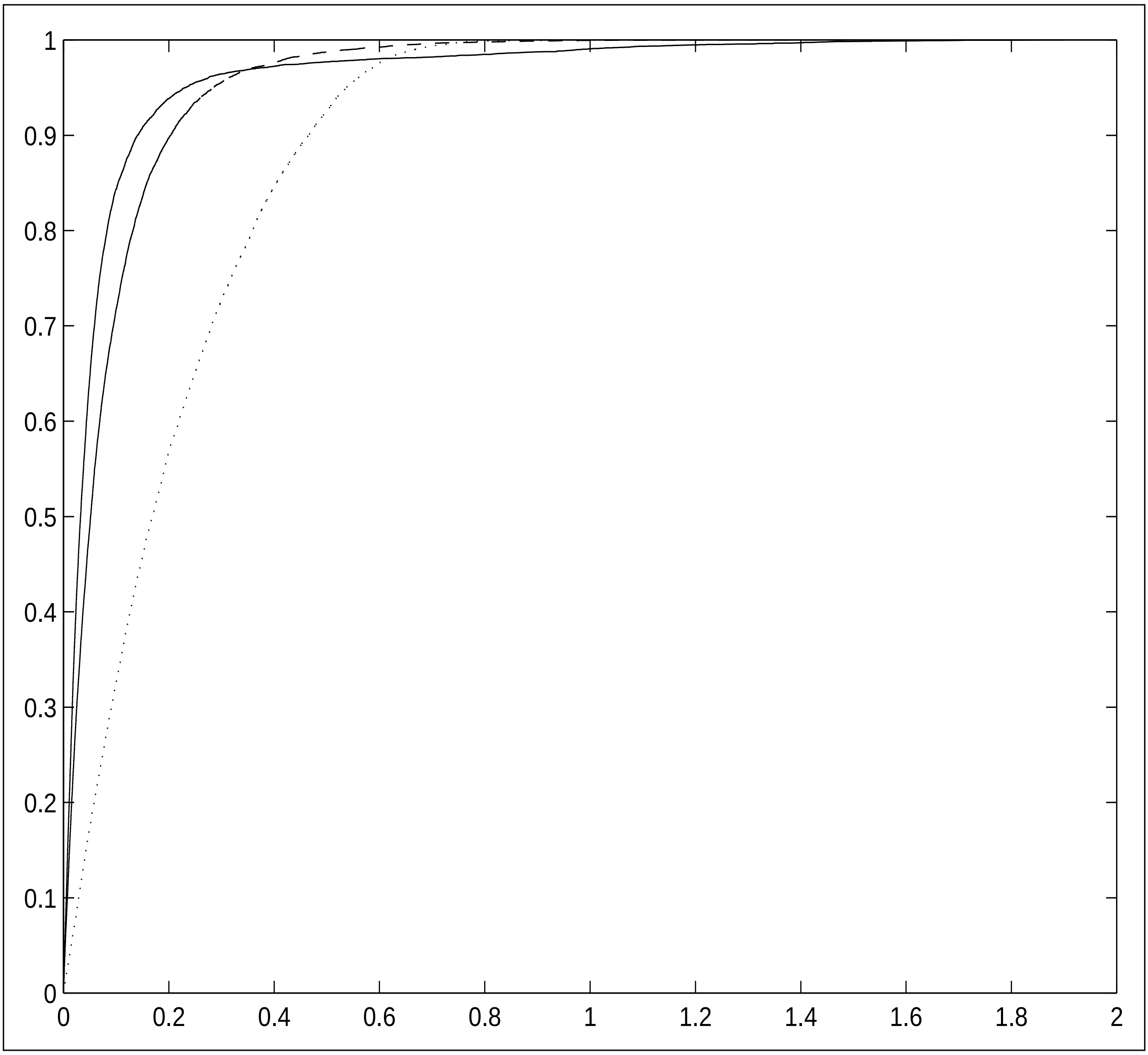

Figure 2. Estimated cumulative distributions of errors. Solid line: Cluster filter, dashed line: median, dotted line: mean