World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:4, No:10, 2010

Alternative to M-Estimates in Multisensor Data Fusion

Nga-Viet Nguyen, Georgy Shevlyakov, and Vladimir Shin

Abstract—To solve the problem of multisensor data fusion under non-Gaussian channel noise. The advanced M-estimates are known to be robust solution while trading off some accuracy. In order to improve the estimation accuracy while still maintaining the equivalent robustness, a two-stage robust fusion algorithm is proposed using preliminary rejection of outliers then an optimal linear fusion. The numerical experiments show that the proposed algorithm is equivalent to the M-estimates in the case of uncorrelated local estimates and significantly outperforms the M-estimates when local estimates are correlated.

Keywords—Data fusion, estimation, robustness, M-estimates.

I. Introduction

DATA fusion [1] refers to a family of techniques that combine data from multiple local sources to achieve inferences being potentially more efficient and more accurate than what can be achieved by using a single source. To cope with outliers in non-Gaussian channel noise, advanced robust methods such as M-estimate based methods can be applied with some trade-off of accuracy. The challenge is how to improve this accuracy while still maintaining the robustness.

The nature of multisensor data fusion is that there are a number of sensors physically distributed around an environment that can extend in a large area. Each sensor has its own local processor that optimally estimates an object using sensor measurements prior to communication then only local estimates need to be transfered to and processed by the central processor [1], [2].

In this fusion architecture, many methods are available for the fusion of the local estimates. Most of them result in the linear fusion of sensor estimates (see [1]–[4] and references therein). Obviously, linear fusion well-defined under some restrictive conditions is useful to reduce the computation load at the fusion center. However, it is a well-known fact that linear (with respect to observations) methods drastically fail in the presence of outliers in the data and/or under heavy-tailed noise distributions and, in this case, they should be substituted by their robust analogs such as the advanced M-estimates [5]–[10].

On the other hand, it is unreasonable to fully refuse classical linear methods optimal in regular conditions and relatively easy in implementation. On the contrary, it is desirable to make these classical tools resistant to possible deviations from the restrictive initial assumptions about noise distributions. Hence, in this work, a combination of a classical optimal fusion method and a rejection rule of outliers is presented and it is showed that this alternative method to M-estimates can indeed replace M-estimates, especially in the case of correlated local estimates.

N.-V. Nguyen and V. Shin are with the School of Information and Mechatronics, Gwangju Institute of Science and Technology, Gwangju 500-712, Republic of Korea. Emails: ngaviet@gist.ac.kr, vishin@gist.ac.kr.

G. Shevlyakov is with the Department of Applied Mathematics, St. Petersburg State Polytechnical University, Saint Petersburg, Russia. Email: georgy.shevlyakov@gmail.com.

Manuscript received April 30, 2010; revised May 31, 2010.

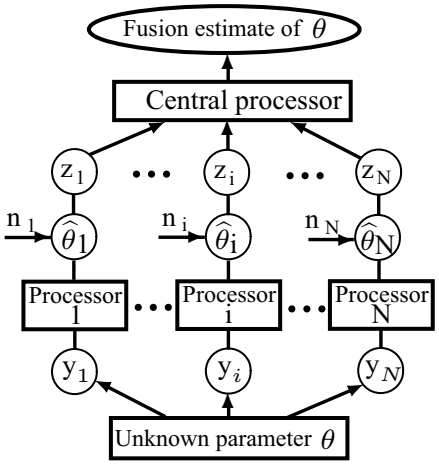

Fig. 1. Distributed multisensor data fusion system with noisy channels

II. Optimal Estimation Fusion

Our problem is illustrated in Fig. 1. The unknown constant parameter is locally observed by different sensors, each producing its own local estimate. All local estimates are then sent to the central processor where they are fused to obtain a more accurate estimate [1], [2].

A. Local Estimates

The popular fusion scheme in Fig. 1 uses a network of N distributed sensors to estimate an unknown constant vector parameter $\theta \in \mathcal{R}^p$. Each sensor produces a noisy vector measurement:

$$\mathbf{y}_i = h_i(\theta) + \mathbf{v}_i, \qquad i = 1, \ldots, N, \tag{1}$$

where $\mathbf{y}_i \in \mathcal{R}^{m_i}$, $h_i(\cdot)$ is a known measurement function relating the unknown parameter to the ith sensor measurement, and $\mathbf{v}_i$ is the measurement error vector, assumed to be Gaussian with zero mean and known covariance matrix. The essential thing is that the observation model given by (1) is a regression model, linear or non-linear, and there exist many methods, e.g.,

the maximum likelihood, the least squares, with linearization when necessary etc., solving the problem of estimation of an unknown parameter θ [11].

Now, the simplest version of model (1) is considered with a scalar parameter $\theta \in \mathcal{R}$ common for all the observations $\mathbf{y}_i$, $i = 1, \ldots, N$. In this case, the local estimates take the following form:

$$\widehat{\theta}_i = \theta + e_i, \qquad i = 1, \ldots, N, \tag{2}$$

where $e_i = \widehat{\theta}_i - \theta$ are the local estimate errors assumed to be Gaussian with zero means and known covariance matrix $C = [\sigma_{ij}]$.

B. Optimal Linear Fusion Formula

Let channel noises n_i, $i = 1, \ldots, N$ be independent random variables with zero means and finite variances $\sigma^2_{n_i}$. It is also assumed that channel noises are independent of local sensor estimate errors e_i, $i = 1, \ldots, N$. Hence, the channel outputs obtained at the central processor can be written as:

$$z_i = \widehat{\theta}_i + n_i = \theta + e_i + n_i = \theta + \zeta_i, \;\; \zeta_i = e_i + n_i, \;\; i = 1, \ldots, N, \tag{3}$$

the covariance matrix $\widetilde{C} = [\widetilde{\sigma}_{ij}]$ of the channel output errors $\{\zeta_i\}_1^N$ has the entries:

$$\widetilde{\sigma}_{ij} = \sigma_{ij}, \quad i \neq j, \quad \text{and} \quad \widetilde{\sigma}_{ii} = \sigma_{ii} + \sigma^2_{n_i} = \sigma^2_i + \sigma^2_{n_i}. \tag{4}$$

Thus, the channel outputs $\{z_i\}_1^N$ can be considered equivalent to the local estimates with known covariance matrix $\widetilde{C}$, and the well-known optimal linear fusion (LF) formula [12] can be adapted to this problem:

$$\widehat{\theta}_{LF} = \sum_{i=1}^{N} a_i \, z_i, \qquad \sum_{i=1}^{N} a_i = 1. \tag{5}$$

where $a_1, \ldots, a_N$ are the scalar weights minimizing the mean-square criterion $J = \mathrm{E}\left[(\theta - \widehat{\theta}_{LF})^2\right]$. The explicit formula for the optimal weights is given by:

$$\mathbf{a} = (\mathbf{1}_N^T \widetilde{C}^{-1} \mathbf{1}_N)^{-1} \widetilde{C}^{-1} \mathbf{1}_N, \tag{6}$$

where $\mathbf{a} = [a_1 \;\; \ldots \;\; a_N]^T$ and $\mathbf{1}_N = [1 \;\; \ldots \;\; 1]^T \in \mathcal{R}^N$. The fusion error variance J is equal to $J = \sum_{i,j=1}^{N} a_i \, a_j \, \widetilde{\sigma}_{ij}$.

III. Non-Gaussian Channel Noises and Robust Fusion

The specific of real-life fusion networks is such that, in general, there is no reliable information about channel noise distributions and their moments; the underlying noise distributions are not Gaussian and may vary in a wide range from light- to heavy-tailed forms; and even, if the assumptions of Gaussianity hold, there may be outliers and gross errors in channel noises. Apparently, under the conditions of uncertainty of noise distribution models, the usage of the linear fusion formula (5) is still not justified and appropriate. Thus, the aforementioned cases when the linear fusion can be used are more likely to be exceptions than a rule.

A. Non-Gaussian Noise Model

In this research, Tukey's gross-error model of contaminated Gaussian noise distribution [5] is used as follows:

$$f_{n_i}(x) = (1 - \varepsilon_i)\,\mathcal{N}(x; 0, \sigma_{n_i}) + \varepsilon_i \,\mathcal{N}(x; 0, \lambda_i \, \sigma_{n_i}), \qquad i = 1, \ldots, N, \tag{7}$$

where $\mathcal{N}(x; \mu, \sigma) = (2\,\pi)^{-1/2} \sigma^{-1} \exp\left\{-(x - \mu)^2/(2\,\sigma^2)\right\}$ is the Gaussian density, $0 \leq \varepsilon_i < 1/2$, $\lambda_i \geq 1$ are the parameters of contamination.

Tukey's model given in (7) describes the data samples as the mixture of the data from the two Gaussian distributions of "good" and "bad" data. The contamination parameter ε gives the probability of occurrence of outliers in the data. This model has been chosen for the following two reasons.

First, it is a standard model widely used in studies on robustness [5], [6]. Since the distributions of the main bulk of data and outliers can be described by the first and second term of (10), respectively, different types of outliers can be easily considered in the test by shifting the distribution location and scale.

Second, in this model, the contamination parameter ε can be considered as the probability of occurrence of outliers in the data sample with the number of outliers distributed according to the binomial law. An extended survey on the applicability of Tukey's model to description of the real-life data, in particular on the frequency of gross errors, is given in [6].

B. M-Estimates of Location

An M-estimate of location $\widehat{\theta}$ satisfies the following estimating equation:

$$\sum_{i=1}^{N} \psi(z_i - \widehat{\theta}) = 0, \tag{8}$$

where $\psi(u)$ is a score function [5]. Here the following score functions are considered:
Huber's linear bounded score function [5]:

$$\psi_H(u) = \begin{cases} u, & \text{for } 0 \leq |u| \leq c \quad (c > 0) \\ c\,\mathrm{sign}(u), & \text{for } c \leq |u| \end{cases} \tag{9}$$

and Hampel's redescending three-part score function [6]:

$$\psi_{25A}(u) = \begin{cases} u, & \text{for } 0 \leq |u| \leq a \\ a\,\mathrm{sign}(u), & \text{for } a \leq |u| \leq b \\ a\frac{r_H - |u|}{r_H - b}\,\mathrm{sign}(u), & \text{for } b \leq |u| \leq r_H \\ 0, & \text{for } r_H \leq |u| \end{cases} \tag{10}$$

with $0 < a \leq b < r_H < \infty$.

The simplest way of solving (8) is based on the so-called one-step M-estimates [6]:

$$\widehat{\theta} = \widehat{\theta}^{(0)} + \widehat{\sigma}^{(0)} \frac{\sum_{i=1}^{N} \psi(\frac{z_i - \widehat{\theta}^{(0)}}{\widehat{\sigma}^{(0)}})}{\sum_{i=1}^{N} \psi'(\frac{z_i - \widehat{\theta}^{(0)}}{\widehat{\sigma}^{(0)}})}, \tag{11}$$

where $\widehat{\theta}^{(0)}$ and $\widehat{\sigma}^{(0)}$ are the initial estimates of location and scale. These one-step estimates form the first step of the Newton-Raphson iterative procedure fortunately having the

same behaviour as their fully iterated versions (for details, see [6, p.6]). As the initial estimates of location and scale, the highly robust sample median $\widehat{\theta}^{(0)} = \mathrm{med}(z_i)$ and the median absolute deviation $\widehat{\sigma}^{(0)} = 1.483\,\mathrm{MAD}(z_i)$, where $\mathrm{MAD}(z_i) = \mathrm{med}(|z_i - \mathrm{med}(z_i)|)$, are used.

IV. Proposed Two-Stage Robust Estimation Fusion

In the distributed multisensor fusion system with noisy channels in Fig. 1, the information for fusion is given by the channel outputs $z_1, \ldots, z_N$, which are represented as the sum of an unknown scalar parameter θ and channel output errors $\{\zeta_i\}_1^N$: $z_i = \theta + \zeta_i,\ i = 1, \ldots, N$.

In this work, a two-stage robust fusion (TSRF) algorithm is proposed based on the preliminary rejection of the outliers from the channel output data with the subsequent application of conventional estimation procedures to the rest of observations. Thus, ths refers to the initial key sense of any robust procedure: the initial data are assumed to consist of two different kinds, "good" (regular, satisfying our assumptions) and "bad" (outliers, outputs distorted by gross errors), and the goal is to distinguish the "bad" from the "good" with the final inference basing mostly on the "good" data.

The first stage of this fusion algorithm, namely, the rejection rule, is formulated as follows: reject all the channel outputs z_i such that

$$|z_i - \mathrm{med}(z)| > K\,\mathrm{MAD}(z), \tag{12}$$

where $\mathrm{med}(z)$ is the sample median

$$\mathrm{med}(z) = \begin{cases} z_{(l+1)}, & N = 2\,l + 1, \\ \dfrac{z_{(l)} + z_{(l+1)}}{2}, & N = 2\,l; \end{cases} \tag{13}$$

$\mathrm{MAD}(z)$ is the median absolute deviation

$$\mathrm{MAD}(z) = \mathrm{med}\,(|z - \mathrm{med}(z)|)\,; \tag{14}$$

K is a free parameter to be adjusted to the data, henceforth called *the gate factor*. The sample median and the median absolute deviation are highly robust estimates of location and scale, respectively [6].

In the second stage, the fusion formula (5) is applied to the remaining channel outputs. Let N^* be their number, $\widetilde{C}$ becomes $\widetilde{C}^*$ (with the asterisk sign) after removing their columns and rows corresponding to discarded channel outputs. Then the fusion formula and its related equations can be rewritten as follows:

$$\widehat{\theta}_{TSRF} = \sum_{i=1}^{N^*} a_i\, z_i, \qquad \sum_{i=1}^{N^*} a_i = 1,$$

$$\mathbf{a} = (\mathbf{1}_{N^*}^T\ \widetilde{C^*}^{-1}\mathbf{1}_{N^*})^{-1}\widetilde{C^*}^{-1}\mathbf{1}_{N^*} \tag{15}$$

where $\mathbf{a} = [a_1\ \ldots\ a_{N^*}]^T$, $\widetilde{C^*} = [\widetilde{\sigma}_{ij}]_{i,j=1}^{N^*}$ with $\mathbf{1}_{N^*} = [1\ \ldots\ 1]^T$ in $\mathcal{R}^{N^*}$.

Although the remaining outputs being the central order statistics of the sample become strongly dependent between each other, the covariance matrix $\widetilde{C}^*$ is concisely used in (15) as a modification of $\widetilde{C}$ given by (4). We do this, firstly, for sake of preserving the low-complexity structure of the TSRF, and secondly, because of the principal unavailability of information about channel output distributions in real-life problems.

The pseudo-code of TSRF is written as follows:

Algorithm 1 TSRF algorithm

Require: $N; z_1, \ldots, z_N; C; \sigma_n; K$
Ensure: Estimation of θ
  $\widetilde{C}$ {from (4)}
  $\mathrm{MAD}(z) \Leftarrow \mathrm{med}(|z - \mathrm{med}(z)|)$
  $j \Leftarrow 0$
  for $i = 1$ to N **do**
    if $|z_i - \mathrm{med}(z)| \le K\,\mathrm{MAD}(z)$ **then**
      $j \Leftarrow j + 1$
      $I^*_{z_j} \Leftarrow i$ {Index of remaining element z_i}
    end if
  end for
  $\widetilde{C^*} \Leftarrow discard(\widetilde{C}, I_z^*)$
  $z^* \Leftarrow discard(z, I_z^*)$
  $\mathbf{a} \Leftarrow (\mathbf{1}_{N^*}^T\ \widetilde{C^*}^{-1}\mathbf{1}_{N^*})^{-1}\widetilde{C^*}^{-1}\mathbf{1}_{N^*}$ $\{\mathbf{a} = [a_1 \ldots a_{N^*}]\}$
  $\widehat{\theta}_{TSRF} \Leftarrow \sum_{j=1}^{N^*} a_j z_j^*$

V. Numerical Experiments

A. Experimental Setup

To compute the fusion mean square error (MSE), the Monte Carlo method is used with 5000 repetitions. The system of N local sensors is divided into the two approximately equal-in-size groups: one of size N_1 has higher accuracy, the other of size N_2 consists of local sensors with larger estimate error variances: $\sigma_1^2 = 0.1$ and $\sigma_2^2 = 1$, respectively.

In our Monte Carlo experiment, pseudo-random numbers are generated so that the correlation between local estimates is predetermined by the correlation matrix chosen as follows:

$$\mathrm{R}_{(N\times N)} = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \rho_{11} & \cdots & \rho_{1N_2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \rho_{N_21} & \cdots & \rho_{N_2N_2} \end{bmatrix}. \tag{16}$$

with $\rho_{ij} = 0.5(1 - |i - j|/N_2)$ for $i \neq j$ and $\rho_{ij} = 1$ for $i = j$. Thus, it is assumed that the sensors of higher accuracy are uncorrelated, the sensors of lower accuracy are correlated, and that there is no correlation between the sensors of these two groups.

To compare the performance of fusion estimates, the fusion efficiency is used:

$$\mathrm{eff} = \frac{\mathrm{MSE}_{ML}}{\mathrm{MSE}} \tag{17}$$

defined as the ratio of the mean square errors of the maximum likelihood (ML) and the fusion estimates, respectively.

ML estimate $\widehat{\theta}_{ML}$ is an optimal method, computable only when the probability density function (pdf) of each channel

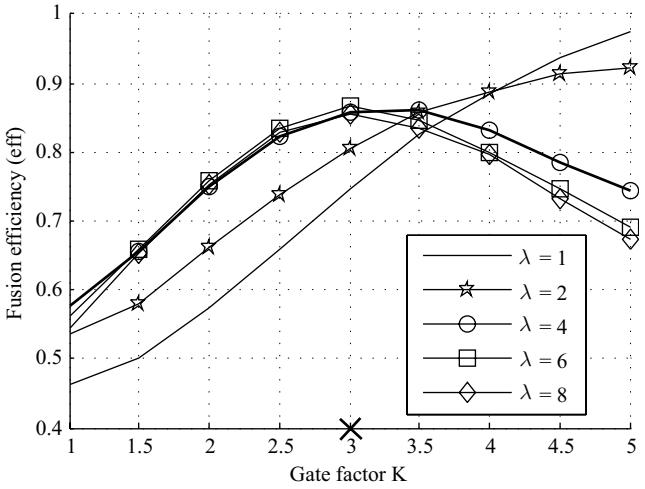

Fig. 2. Tuning the TSRF-algorithm. Total number of sensors: $N = 31$; numbers and variances of sensors of higher accuracy: $N_1 = 16$, $\sigma_1^2 = 0.1$, of lower accuracy: $N_2 = 15$, $\sigma_2^2 = 1$; channel noise variance: $\sigma_n^2 = 1$; contamination parameters: $\varepsilon = 0.2$, λ = 1–8.

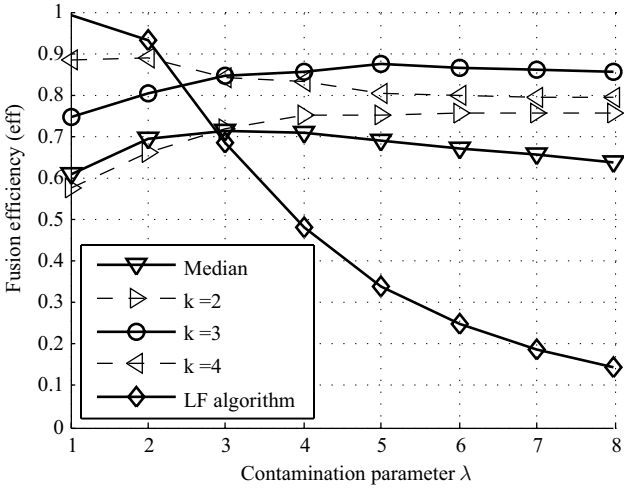

Fig. 3. Performance of TSRF algorithm under contaminated Gaussian noise. Total number of sensors: $N = 31$; numbers and variances of sensors of higher accuracy: $N_1 = 16$, $\sigma_1^2 = 0.1$, of lower accuracy: $N_2 = 15$, $\sigma_2^2 = 1$; channel noise variance: $\sigma_n^2 = 1$; contamination parameters: $\varepsilon = 0.2$, λ = 1–8.

output is known. Hence, a fusion estimate can be compared with ML estimate for evaluation. The computation of ML estimate is described as follows.

From (3) we have:

$$z_i = \theta + e_i + n_i. \tag{18}$$

Then, the assumed pdf of each channel output is:

$$f_{z_i}(\widehat{\theta}) = (1-\varepsilon_i)\mathcal{N}(\widehat{\theta}, \sqrt{\sigma_{e_i}^2 + \sigma_{n_i}^2}) + \varepsilon_i \mathcal{N}(\widehat{\theta}, \sqrt{\sigma_{e_i}^2 + \lambda_i^2 \sigma_{n_i}^2}), \tag{19}$$

where:

$$\mathcal{N}(\widehat{\theta}, \sigma_\theta) = \frac{1}{\sqrt{2\pi}\sigma_\theta} e^{\frac{-1}{2\sigma_\theta^2}(z_i - \widehat{\theta})^2}. \tag{20}$$

Finally, $\widehat{\theta}_{ML}$ is obtained when:

$$\mathcal{L}(\widehat{\theta}) = \prod_{i=1}^{N} (f_{z_i}(\widehat{\theta})) \rightarrow Max. \tag{21}$$

B. Tuning Gate Factor

The proposed two-stage algorithm is determined by the choice of the gate factor K. Fig. 2 exhibits the fusion efficiency with respect to K under different values of the contamination parameter λ. The experiment shows that, over a wide range of contamination, a reasonable choice of the gate factor is provided with $K = 3$.

The results displayed in Fig. 3 also confirm this choice. The curves corresponding to each value of the gate factor K gradually drop while it increases until $K = 3$ and rise while it passes this value and tends to infinity. The performance of the TSRF with the chosen $K = 3$ is slightly worse than the performance with larger values of the gate factor when $\lambda = 1$ (under Gaussian noise) or with small contamination. Paying for this trade-off, the TSRF achieves higher performance for a wide range of contamination parameters.

C. Comparison with M-Estimates

In this part, the proposed TSRF algorithm is compared with the well-known advanced robust M-estimates defined by Huber's and Hampel's score functions described in Section III-B.

These two estimates (shortly called the Huber and Hampel estimates) are tested under contaminated Gaussian channel noise with the same range of contamination parameters as in the previous subsection. The recommended parameters are: $c = 0.862$ for the Huber estimate [5] (the optimal choice for the ε-contaminated Gaussian distribution with $\varepsilon = 0.2$) and $a = 1.31,\ b = 2.039,\ r_H = 4$ for the Hampel estimate [6] (the best in the Princeton experiment [13]).

At first, the case of uncorrelated local estimates is considered. This condition assures the best performance for M-estimates. Fig. 4 shows that TSRF is comparable to the M-estimtes. The proposed algorithm is even more accurate for big values of λ. To be more specific, let us take $\lambda = 4$ to divide the total range of λ into small λ range and big λ range. Huber estimate is the best in small λ range and TFRF is the best in big λ range.

So, for uncorrelated local estimates, TSRF can just be equivalent to M-estimates; however, it really shows the advantage in the case of correlated local estimates. In Fig. 5, TSRF is now chosen as the reference for calculating relative efficiency, M-estimates are therefore compared to TFRF: they are more accurate if the relative efficiency is greater than 1, and vis versa. However, this experiment shows that this value is only around 0.8, meaning that TFRF outperforms the M-estimates in this case.

VI. Conclusion

The problem of multisensor data fusion was considered with non-Gaussian channel noise. The well-known advanced M-estimates were applied, resulting in a robust fusion method. In order to improve the estimation accuracy while maintaining

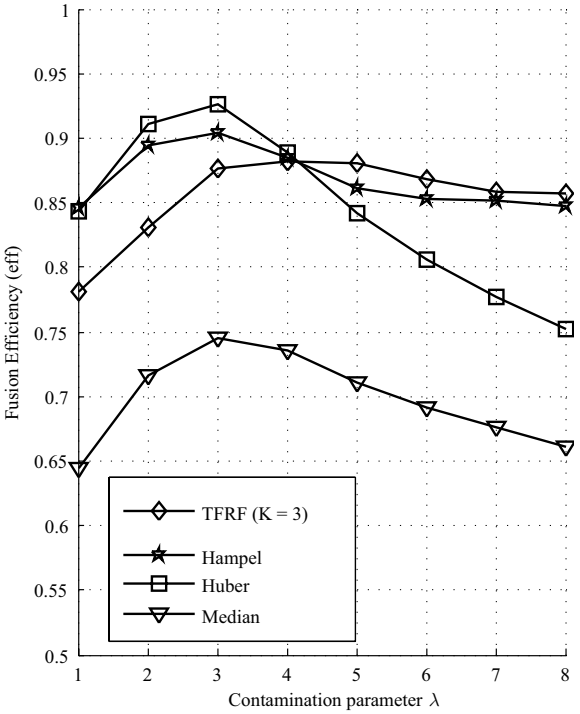

Fig. 4. Efficiency comparison of TSRF, M-estimates and Median for uncorrelated local estimates. Total number of sensors: $N = 31$; number of sensors and their variance for the higher accuracy group: $N_1 = 16$, $\sigma_1^2 = 0.1$, for the lower accuracy group: $N_2 = 15, \sigma_2^2 = 1$; channel noise variance $\sigma_n^2 = 1$; contamination parameters: ε = 0.2, λ = 1–8.

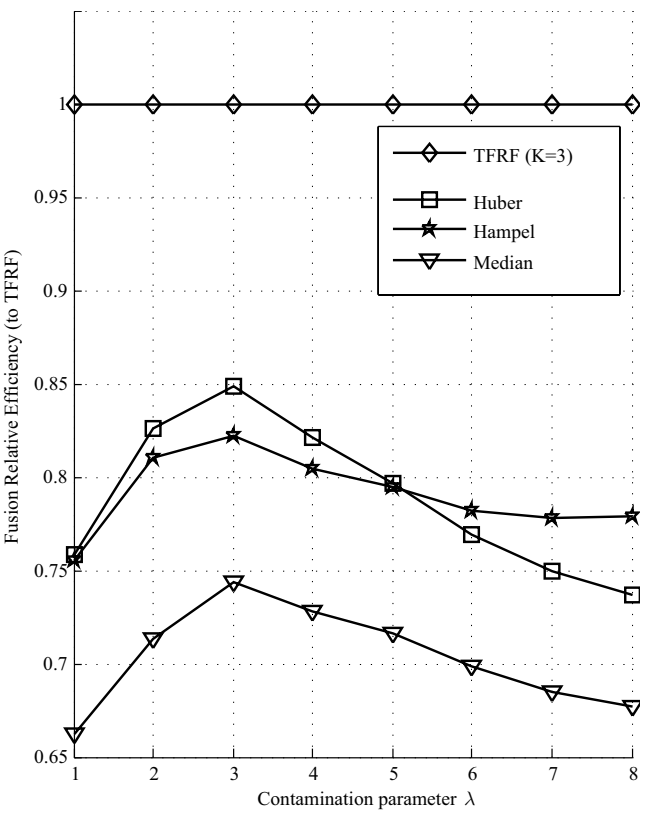

Fig. 5. M-estimates in comparison with TSRF for correlated local estimates. Total number of sensors: $N = 31$; number of sensors and their variance for the higher accuracy group: $N_1 = 16$, $\sigma_1^2 = 0.1$, for the lower accuracy group: $N_2 = 15, \sigma_2^2 = 1$; channel noise variance $\sigma_n^2 = 1$; contamination parameters: ε = 0.2, λ = 1–8.

an equivalent robustness, the TSRF algorithm was proposed using a preliminary rejection of outliers then an optimal linear fusion. The numerical experiments show that TSRF is equivalent to the M-estimates in the case of uncorrelated local estimates and significantly outperformed the M-estimates when local estimates are correlated.

ACKNOWLEDGMENT

This work was suppoted by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No.2010-0009925) and the Basic Research Project through a grant provided by GIST in 2010.

REFERENCES

[1] Yunmin Zhu, *Multisensor Decision and Estimation Fusion*, Kluwer, Boston, MA, 2002.
[2] Yunmin Zhu, Enbin Song, Jie Zhou, and Zhisheng You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *Signal Processing, IEEE Transactions on*, vol. 53, no. 5, pp. 1631–1639, May 2005.
[3] Yaakov Bar-Shalom and Xiao-Rong Li, *Multitarget-multisensor Tracking: Principles and Techniques*, YBS Publishing, Storrs, CT, 1995.
[4] Vladimir Shin, Younghee Lee, and Tae S. Choi, "Generalized Millman's formula and its application for estimation problems," *Signal Process.*, vol. 86, no. 2, pp. 257–266, 2006.
[5] Peter J. Huber, *Robust Statistics*, Wiley, New York, 1981.
[6] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, 1986.
[7] S.A. Kassam and H.V. Poor, "Robust techniques for signal processing: A survey," *Proceedings of the IEEE*, vol. 73, no. 3, pp. 433–481, March 1985.
[8] Georgy L. Shevlyakov and Nikita O. Vilchevski, *Robustness in Data Analysis: Criteria and Methods*, VSP, Utrecht, 2002.
[9] U. K. Bhargava and R. L. Kashyap, "Robust parametric approach for impulse response estimation," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 10, pp. 1592–1601, 1988.
[10] D. D. Lee, R. L. Kashyap, and R. N. Madan, "Robust decentralized direction-of-arrival estimation in contaminated noise," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 3, pp. 496–505, 1990.
[11] George A. F. Seber and Alan J. Lee, *Linear Regression Analysis*, Wiley, New York, 2003.
[12] Vladimir Shin, Georgy Shevlyakov, and Kiseon Kim, "A new fusion formula and its application to continuous-time linear systems with multisensor environment," *Comput. Stat. Data Anal.*, vol. 52, no. 2, pp. 840–854, 2007.
[13] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey, *Robust Estimates of Location*, Princeton University Press, Princeton, 1972.