OUTLIER COMPENSATION IN SENSOR NETWORK SELF-LOCALIZATION VIA THE EM ALGORITHM

Joshua N. Ash, Randolph L. Moses

Ohio State University
Department of Electrical and Computer Engineering
Columbus, Ohio 43210

ABSTRACT

Self-localization is an important component of distributed sensor systems. The presence of a few highly erroneous measurements, or outliers, results in erroneous sensor location estimates. In this paper we employ the EM algorithm to iteratively detect outlier measurements and provide robust position estimates of the sensors. The derivation of the algorithm is given, and Monte-Carlo simulations are presented to compare this estimator to others. The performance of the EM-based algorithm is also shown to be close to the Cramér-Rao lower bound for position estimation when perfect knowledge of the outlier process is known.

1. INTRODUCTION

Self-localization of individual sensor positions is an essential prerequisite for the utility of most sensor networks. Localization estimates are usually obtained by processing of time-of-arrival (TOA), angle-of-arrival (AOA), or received signal strength (RSS) measurements between nodes.

In practice, measurements used for self-localization often contain outliers with large errors – frequently due to low SNR conditions [1] or multipath components of the calibration signal. Most self-localization algorithms do not account for outliers and therefore produce highly erroneous localization solutions when outliers are present. In this paper we develop and evaluate a robust self-localization algorithm that effectively mitigates the effects of outlier measurements.

Previous work by Ward *et al.* [2] removed outliers by eliminating measurements with the largest studentized residuals until a minimum threshold variance was reached. Savarese *et al.* [3] suggest a refinement phase to initial localization estimates which establishes a confidence weight that effectively removes inconsistent measurements.

Prepared through collaborative participation in the Advanced Sensors Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-02-0008.

If a more complete statistical characterization than a point estimate is desired, nonparametric belief propagation (NBP) may be used to provide posterior probability density estimates of the sensor locations [4]. Outliers are inherently handled by including a high variance component capable of representing an erroneous measurement.

In this paper we pose the self-localization task as a statistical estimation problem that includes an outlier measurement model. We employ the EM algorithm [5] to jointly determine outlying measurements and estimate sensor positions. Our presentation takes a centralized approach; however the EM algorithm may be implemented in a distributed fashion as described in [6]. Additionally, although we only consider TOA measurements, the results are easily extended to AOA and RSS observations.

2. PROBLEM FORMULATION AND EM ALGORITHM

Consider N sensors with unknown locations randomly distributed in a planar region. In turn, each sensor emits a signal which is detected by a subset of the other sensors. Sensors that detect this signal process the received waveform to obtain an estimate of the time of arrival of that signal. Let the arrival time at sensor i of an emission from sensor j be denoted t_{ij}. Under normal (non-outlier) conditions $t_{ij} \sim f_1(t_{ij};\boldsymbol{\theta})$, which is assumed to happen with probability μ. With probability $1-\mu$ we have $t_{ij} \sim f_2(t_{ij};\boldsymbol{\theta})$, where the outlier distribution $f_2(t_{ij};\boldsymbol{\theta})$ typically has a much larger variance than the typical distribution $f_1(t_{ij};\boldsymbol{\theta})$. Here, $\boldsymbol{\theta} = [\{x_i, y_i, \tau_i\}_{i=1}^N]$ is the unknown parameter vector to be estimated, where x_i, y_i, τ_i are the x-coordinate, y-coordinate, and emission time of node i respectively. The unknown emission times are nuisance parameters in this problem.

For convenience, we define the indicator random variable I_{ij} to indicate whether measurement t_{ij} is selected from the typical or outlier pdf:

$$\begin{aligned} f(t_{ij};\boldsymbol{\theta}|I_{ij}=1) &= f_1(t_{ij};\boldsymbol{\theta}) \\ f(t_{ij};\boldsymbol{\theta}|I_{ij}=0) &= f_2(t_{ij};\boldsymbol{\theta}), \end{aligned} \tag{1}$$

where $p(I_{ij} = 1) = \mu$ and $p(I_{ij} = 0) = (1 - \mu)$.

For a given measurement, we define the *complete data* in the EM context to be $\{t_{ij}, I_{ij}\}$, which has distribution

$$f(t_{ij}, I_{ij}; \boldsymbol{\theta}) = f(t_{ij}; \boldsymbol{\theta}|I_{ij})p(I_{ij}). \tag{2}$$

The *incomplete data* is the observation $\{t_{ij}\}$. The TOA measurements are collected into the matrix $\boldsymbol{T} = \{t_{ij}\}$, and the indicator variables are similarly collected into $\boldsymbol{I} = \{I_{ij}\}$.

Let the set $\mathcal{S}$ be composed of node pairs for which TOA measurements are observed; that is, $(i,j) \in \mathcal{S}$ if j's emission is heard by i. Then, the full pdf of the complete data is

$$f(\boldsymbol{T}, \boldsymbol{I}; \boldsymbol{\theta}) = \prod_{(i,j)\in\mathcal{S}} f(t_{ij}; \boldsymbol{\theta}|I_{ij})p(I_{ij}). \tag{3}$$

It is desirable to treat $\boldsymbol{I}$ as an additional parameter and maximize (3) jointly over $\boldsymbol{I}$ and $\boldsymbol{\theta}$. However, the large size and discrete nature of $\boldsymbol{I}$ makes this impractical even for small networks. Therefore, we choose to maximize the marginal $f(\boldsymbol{T}; \boldsymbol{\theta})$ which is accomplished with the EM algorithm by treating the elements of $\boldsymbol{I}$ as missing measurements. The EM algorithm proceeds as follows.

With $\hat{\boldsymbol{\theta}}^{(n)}$ denoting the current estimate of $\boldsymbol{\theta}$, the expectation step of EM is to compute

$$\begin{aligned} U(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(n)}) &= E\big[\ln f(\boldsymbol{T}, \boldsymbol{I}; \boldsymbol{\theta})|\boldsymbol{T}; \hat{\boldsymbol{\theta}}^{(n)}\big] \\ &= E\big[\textstyle\sum_{(i,j)\in\mathcal{S}} \ln f(t_{ij}; \boldsymbol{\theta}|I_{ij}) \\ &\qquad + \ln p(I_{ij})|\boldsymbol{T}; \hat{\boldsymbol{\theta}}^{(n)}\big]. \end{aligned} \tag{4}$$

Taking the expectation we obtain

$$\begin{aligned} &U(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(n)}) = \\ &\sum_{(i,j)\in\mathcal{S}} \Big[\ln f(t_{ij}; \boldsymbol{\theta}|I_{ij} = 1)p(I_{ij} = 1|\boldsymbol{T}; \hat{\boldsymbol{\theta}}^{(n)}) \\ &\quad + \ln f(t_{ij}; \boldsymbol{\theta}|I_{ij} = 0)p(I_{ij} = 0|\boldsymbol{T}; \hat{\boldsymbol{\theta}}^{(n)}) \\ &\quad + \ln p(I_{ij} = 1)p(I_{ij} = 1|\boldsymbol{T}; \hat{\boldsymbol{\theta}}^{(n)}) \\ &\quad + \ln p(I_{ij} = 0)p(I_{ij} = 0|\boldsymbol{T}; \hat{\boldsymbol{\theta}}^{(n)})\Big]. \end{aligned} \tag{5}$$

Let $\bar{I}_{ij}^{(n)}$ be the expected value of I_{ij} at the n^{th} step, $\bar{I}_{ij}^{(n)} = p(I_{ij} = 1|t_{ij}; \hat{\boldsymbol{\theta}}^{(n)})$. From Bayes rule we have

$$\bar{I}_{ij}^{(n)} = \frac{\mu f_1(t_{ij}; \hat{\boldsymbol{\theta}}^{(n)})}{\mu f_1(t_{ij}; \hat{\boldsymbol{\theta}}^{(n)}) + (1 - \mu) f_2(t_{ij}; \hat{\boldsymbol{\theta}}^{(n)})}. \tag{6}$$

Because $U(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(n)})$ is to be maximized with respect to $\boldsymbol{\theta}$, we can eliminate the last two terms in (5) which are independent of $\boldsymbol{\theta}$. The resulting function, after substituting (1) and (6) is

$$\begin{aligned} \tilde{U}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(n)}) = \textstyle\sum_{(i,j)\in\mathcal{S}} \big[&\ln f_1(t_{ij}; \boldsymbol{\theta})\bar{I}_{ij}^{(n)} \\ &+ \ln f_2(t_{ij}; \boldsymbol{\theta})(1 - \bar{I}_{ij}^{(n)})\big] \end{aligned} \tag{7}$$

The resulting EM algorithm is, for $n = 0, 1, 2, \ldots$

E-step: Compute

$$\bar{I}_{ij}^{(n)} \ \forall\, (i,j) \in \mathcal{S} \quad \text{from (6)} \tag{8}$$

M-step: Find $\hat{\boldsymbol{\theta}}^{(n+1)}$ as

$$\arg\max_{\boldsymbol{\theta}} \tilde{U}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(n)}). \tag{9}$$

3. EXAMPLE AND DISCUSSION

As an example, we consider $f_1(t_{ij}; \boldsymbol{\theta})$ and $f_2(t_{ij}; \boldsymbol{\theta})$ to be Gaussian distributions with means equal to the true arrival time

$$\eta_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}/c + \tau_j, \tag{10}$$

and standard deviations $\sigma_1 = 1\,\text{ms}$ and $\sigma_2 = 30\,\text{ms}$ respectively. We take the propagating signals to be acoustic with propagation velocity $c = 340\,\text{m/s}$. The standard ranging error corresponding to the above timing standard errors is $0.34\,\text{m}$ for typical measurements and $10.2\,\text{m}$ for outlier measurements.

Figure 1 illustrates the sample network considered, with solid links denoting quality measurements and dashed links denoting outlier measurements. The four corner nodes are designated as *anchor nodes* having known locations. The reception range is limited to $90\,\text{m}$, so that not all sensors make measurements of each others' emitted signal. In this example, the number of measurements made is $|\mathcal{S}| = 162$, 12 of which are outliers[1]. The *a priori* assumption on the probability of an outlier used in the EM algorithm is $1 - \hat{\mu} = 0.05$; note that this is mismatched from the actual value of $1 - \mu = 0.0741$.

Figure 2 illustrates the position estimates from the EM algorithm corresponding to 200 random realizations of the arrival time matrix, $\boldsymbol{T}$. In order to quantify the performance of the estimates we consider the scene RMS localization error defined as

$$e_{rms} = \sqrt{1/N \sum_{i=1}^{N} E[d_i^2]}, \tag{11}$$

where d_i is the distance between the true position of node i and its estimate. Using the 200 position estimates to empirically approximate each $E[d_i^2]$, we obtain a scene RMS error of $0.29\,\text{m}$.

For comparison, we consider the error of a "Blind" maximum likelihood estimator which uses the same TOA measurements but assumes that none of them are outliers

$$\hat{\boldsymbol{\theta}}_{blind} = \arg\max_{\boldsymbol{\theta}} \prod_{(i,j)\in\mathcal{S}} f_1(t_{ij}; \boldsymbol{\theta}). \tag{12}$$

[1]The links are bidirectional and for ease of presentation a dashed line represents an outlier measurement in both directions.

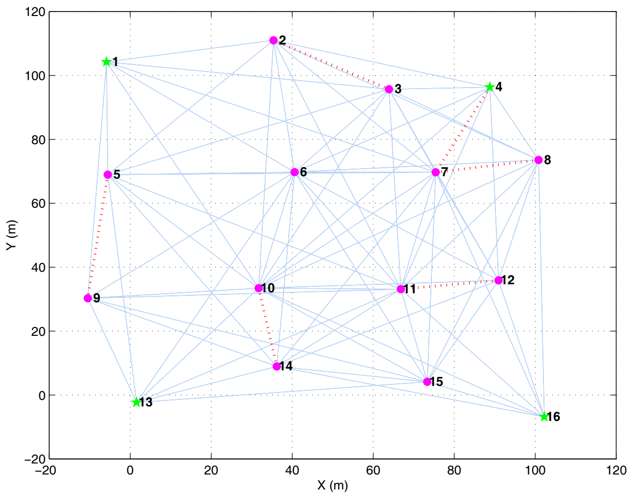

Fig. 1: Sample network in a crude grid configuration to be localized. $\bullet$'s denote sensors with unknown positions, $\bigstar$'s denote sensors with known positions. Solid links between sensors indicate quality measurements distributed via $f_1(\cdot)$, and dashed links indicate outlier measurements distributed via $f_2(\cdot)$.

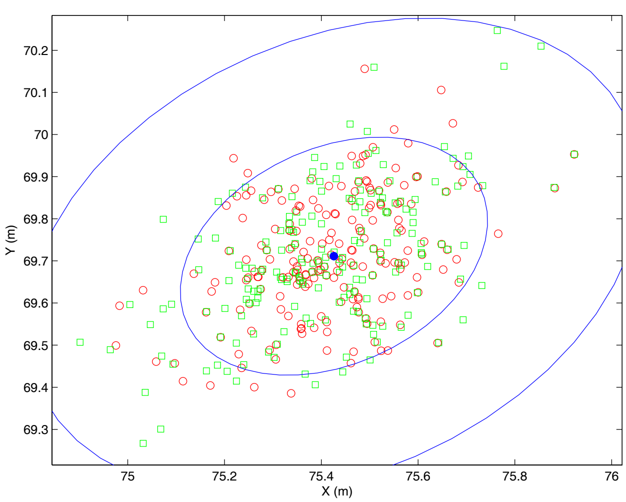

Fig. 2: Sample localization results for node #7 of Figure 1 for the EM algorithm ($\square$) and Genie ML ($\circ$) estimators corresponding to $\sigma_1 = 1$ ms and $\sigma_2 = 30$ ms. The 2-sigma and 4-sigma ellipses derived from the CRLB are also plotted (-). Not shown are three EM estimates, resulting from poor initialization, which lie significantly outside of the plot range.

We also consider a "Genie" estimator which performs a maximum likelihood estimate based on knowledge of the true value of $\boldsymbol{I}$. Finally, we compare the performance of these estimators to the Cramér-Rao Lower Bound (CRLB) for the estimation of $\boldsymbol{\theta}$ from the set of TOA observations and knowledge of the true $\boldsymbol{I}$ (see [7] for the derivation of the CRLB for this problem). The RMS error for each of the three estimators, and that derived from the CRLB, is presented in Table 1. For the EM algorithm, the reported results correspond to 5 EM iterations, as only negligible improvement was observed for greater numbers. Each estimator was initialized with the same values resulting from a variant of multidimensional scaling [8] modified for the TDOA case.

Method	RMS error (m)
CRLB	0.18
Genie ML	0.18
EM	0.29
Blind ML	1.39

Table 1: Estimator performance comparison for the network in Figure 1 with $\sigma_1 = 1$ ms and $\sigma_2 = 30$ ms.

Since the Genie estimator utilizes knowledge of the true $\boldsymbol{I}$ and is a maximum likelihood estimator, it is expected to achieve the CRLB for sufficiently low noise levels (σ_1 and σ_2). We observe this behavior from Table 1 where both have an RMS error of 0.18 m. In Figure 3 we plot the performance of the estimators as a function of the standard deviation of the outlier measurements, σ_2. Here the efficiency of the Genie estimator is further confirmed by being indistinguishable from the lower bound over the range of σ_2.

The Genie estimator is of course not realizable in practice; however from Table 1 and Figure 3 we see that the EM algorithm achieves comparable performance over a range of σ_2. For $\sigma_2 = 30$ ms, the lower bound on RMS error is 0.18 m and the EM estimator achieves an RMS error of 0.29 m, for a relative efficiency of 1.6. The Blind estimator, which naively assumes that none of the measurements are outliers, has an RMS error of 1.39 m. So for this example, the EM algorithm achieves approximately a 79% reduction in RMS error over the Blind estimator and is reasonably close to the CRLB.

A common problem in EM implementations is avoiding local maxima of the likelihood function. In this case, we observe our solutions to be somewhat sensitive to the initial coordinates used, especially for large σ_2. If an erroneous initial coordinate overly favors an outlier measurement, the outlier will be classified as a good measurement while most of the remaining measurements get incorrectly characterized as being outliers. As such, the maximization phase will move the coordinate estimate very little and subsequent iterations will have negligible effect. We found the output of the Blind ML estimator to provide adequate performance as an initializer to EM and used that for the results presented in this paper. However, with this initializer the algorithm did not always converge to the global maximum. For example, in the $\sigma_2 = 30\,\text{ms}$ scenario, 8 of the 200 cases converged to a local maximum. If we ignore these, the error reduces from $0.29\,\text{m}$ to $0.19\,\text{m}$ - nearly obtaining the lower bound. As such, a better initialization routine would improve upon the results presented here.

Finally we observe that the EM estimator is insensitive to the parameter $\hat{\mu}$ as illustrated by the results in Table 2. The actual fraction of quality links in our sample network was 0.9259, however mismatched values had negligible effects on performance. We also note that $\hat{\mu}$ can be adaptively estimated during the E-step of the algorithm as $\hat{\mu}^{(n)} = \sum_{(i,j)\in\mathcal{S}} I_{ij}^{(n)}/|\mathcal{S}|$. The performance of this approach is comparable to the others and is also given in Table 2. $\hat{\mu}$ enters the estimation problem in Equation (6). Because the typical and outlier distributions have a large difference in variance, there is frequently a significant difference in the magnitudes of $f_1(t_{ij};\hat{\boldsymbol{\theta}}^{(n)})$ and $f_2(t_{ij};\hat{\boldsymbol{\theta}}^{(n)})$. This dominates (6) to $\bar{I}_{ij}^{(n)} \approx 0$ or $\bar{I}_{ij}^{(n)} \approx 1$, regardless of μ.

$\hat{\mu}$ used in EM alg.	RMS error (m)
0.70	0.33
0.9259	0.29
0.95	0.29
adaptive	0.29

Table 2: EM estimator performance comparison for different values of the parameter $\hat{\mu}$. The actual value of μ was 0.9259 in this example.

4. CONCLUSION

We have applied the EM algorithm to the sensor network self-localization problem in the presence of outlier measurements. The unobserved measurements of the complete data set in the EM algorithm were taken as indicator variables $\{I_{ij}\}$ denoting whether a particular measurement was an outlier or not. Through successive refinements on the posterior distribution of $\{I_{ij}\}$, the EM algorithm proved successful in compensating for outlier measurements. In the example considered, the localization error was reduced 79% over a naive maximum likelihood estimator which assumed high fidelity measurements throughout. Both the Blind and EM algorithms require the maximization in (9), however the additional cost of the EM algorithm is that this must be done for each iteration.

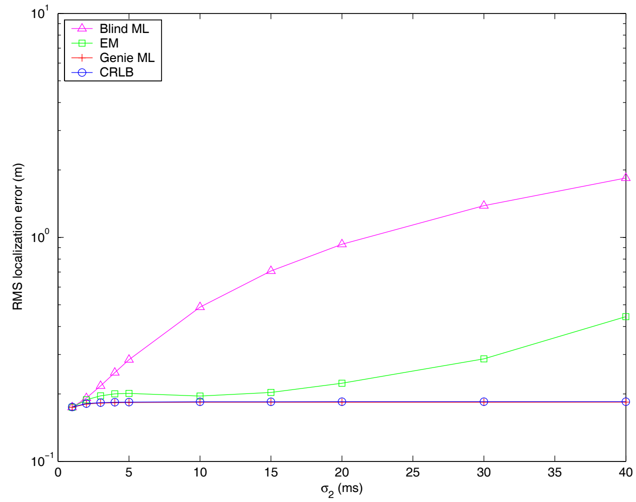

Fig. 3: Estimator performance as a function of outlier standard deviation, σ_2.

5. REFERENCES

[1] A. J. Weiss and E. Weinstein, "Fundamental limitations in passive time delay estimation, part I: Narrow-band systems," *IEEE Transactions on Acoustics, Speech, and Signal-Processing*, vol. 31, no. 2, pp. 472–485, April 1983.

[2] A. Ward, A. Jones, and A. Hopper, "A new location technique for the active office," *IEEE Personal Communications*, vol. 4, no. 5, pp. 42–47, Oct. 1997.

[3] C. Savarese, K. Langendoen, and J. Rabaey, "Robust positioning algorithms for distributed ad-hoc wireless sensor networks," *USENIX Technical Annual Conference*, Monterey, CA, June 2002.

[4] A. T. Ihler, J. W. Fisher III, R. L. Moses, and A. S. Willsky, "Nonparametric belief propagation for sensor self-calibration," *Proc. ICASSP '04*, vol. 3, pp. 861–864, May 2004.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the em algorithm," *J. Roy. Stat. Soc.*, vol. 39, pp. 1–17, 1977.

[6] R. Nowak, "Distributed EM algorithms for density estimation and clustering in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2245–2253, 2003.

[7] R. L. Moses, D. Krishnamurthy, and R. Patterson, "A self-localization method for wireless sensor networks," *Eurasip Journal on Applied Signal Processing, Special Issue on Sensor Networks*, vol. 2003, pp. 348–358, March 2003.

[8] W. S. Torgerson, "Multidimensional scaling: I. theory and method," *Psychometrika*, vol. 17, pp. 401–419, 1952.