Measurement prioritization for optimal Bayesian fusion

Jason Matthew Aughenbaugh
jason@arlut.utexas.edu

Brian R. LaCour
blacour@arlut.utexas.edu

Applied Research Laboratories
The University of Texas at Austin
Austin, TX, U.S.A.

Abstract – ***This paper examines the ordering of measurement updates for a general Bayesian inference problem and its impact on the estimation of the posterior distribution. The approach used compares the expected improvement to the posterior from various types of potential measurements, taking into account the current estimated prior but not the actual measurements, to determine the optimal measurement to perform and/or incorporate. The expected improvement is quantified using both an entropy and a covariance-based measure, each of which is further approximated for computational expedience. Compared to a random ordering of measurements, the posterior is observed to converge more quickly, resulting in a significant improvement in performance.***

Keywords: Bayesian target tracking, data fusion, sensor management, active sonar, passive sonar, information theory, mutual information, entropy

1 Introduction

In fusing data from a network of sensors, it is possible to have many different types of sensors that provide different modalities and qualities of data. The measurement received from a particular sensor at a particular time depends on the sensor's location, the sensor's basic capabilities, the current target state, and the environment.

Combining information from such diverse sources involves both low-level and high-level challenges. The low-level challenges relate to the physics of the sensor systems and mathematics of signal processing. The high-level challenges involve questions such as how to select measurements, combine information, and plan additional testing in order to most efficiently estimate the states of potential targets. In this paper, the fusion of contact-level data is considered in a single target tracking application in which centralized data fusion is assumed.

The term "contact-level data" is used in this paper to refer to point measurements—e.g., a passive sonar bearing measurement of 45 degrees from receiver to target, or an active sonar time delay measurement of 9 seconds from transmitter to target to receiver. It is assumed that the signal processing has been completed at the sensor level. This assumption allows for the consideration of high-level fusion in a network of various sensors types and locations.

As is explained in more detail in Section 3, it is assumed that, for various reasons, not all of the potential measurement data can be used in the updating process. Limitations such as power and bandwidth may prevent the collection and transmission of measurements. Additionally, due to numerical implementation issues and the potential for erroneous measurements that fall outside the assumed sensor model, it may be advantageous to prioritize the incorporation of measurements in terms of those that are likely to provide the most insight into state estimation.

The main subject of this paper is a study of how the prioritization and sequencing of measurement updates improves the Bayesian updating and state estimation procedure. The approach is to predict and quantify the expected improvement to the posterior due to each candidate measurement and order the updates accordingly. An outline of the paper is as follows. Section 2 contains a brief overview of Bayesian tracking. The motivations for prioritizing measurements updates are presented in Section 3. The use of entropy and covariance as metrics for prioritization is discussed in Section 4. A particular bearings-only sensor scenario is explored in Section 5. Finally, the localization performance for randomly positioned bearing and time-delay sensors for the different metrics, including each one's computational complexity, is consider in Section 6.

2 Bayesian tracking overview

We consider the problem of measurement prioritization in the context of a general Bayesian inference problem.

2.1 Mathematics of Bayesian inference

The basic concept of Bayesian inference is to create a posterior (i.e., after observation) estimate of the state of the target using a likelihood model for sensor measurements and a prior distribution on the target state. More specifically, let $p(x\,|\,z)$ represent the probability that the state of the target is x, given the particular received measurement z. This is called the posterior distribution in that it is the distribution *after* the receipt of z. The posterior distribution can be calculated using Bayes' Theorem:

is described by probability density function $p(x)$ having a support region S. The entropy of X is then defined by

$$H(X) = -\int_S p(x) \log p(x) dx. \quad (3)$$

4.2 Entropy and mutual information in sensor selection

The first suggestion of using entropy in a problem related to sensor management and state estimation appears to have been [4]. Subsequently, Manyika and Durrant-Whyte suggested considering expected information gain in sensor management and data fusion problems [5]. Several information metrics are also discussed in [6], including entropy differences. If $H(X)$ is the prior entropy and $H(X \mid Z = z)$ is the posterior entropy after incorporating a particular measurement, z, then the increase in information is given by $H(X) - H(X \mid Z = z)$.

Several authors have considered selecting measurements according to the maximal mutual information, including [7, 8]. Let $I(X;Z)$ denote the mutual information between the state X and the measurement Z. Then $I(X;Z)$ is given by

$$I(X;Z) = \int p(x,z) \log \left[\frac{p(x,z)}{p(x) p_Z(z)} \right] dxdz. \quad (4)$$

Note that $-I(X;Z)$ is just the relative entropy between the joint density of X and Z and the product of their respective marginal densities and, thus, is a measure of statistical dependence. It can be shown (e.g. [9 p. 43]) that

$$I(X;Z) = H(X) - \int H(X \mid Z = z) p_Z(z) dz, \quad (5)$$

which may be interpreted as the *expected* increase in information due to a measurement of this type.

Because the calculation of the mutual information is computationally demanding, other authors have sought approximations, as described in the next subsection.

4.3 Approximations to mutual information

An alternative expression for the mutual information is

$$I(X;Z) = H(Z) - \int H(Z \mid X = x) p(x) dx, \quad (6)$$

where

$$H(Z \mid X = x) = -\int p(z \mid x) \log p(z \mid x) dz. \quad (7)$$

Wang and co-authors [10] have developed an approximation of this quantity, defined as follows. First, they assume that $p(z \mid x)$ varies slowly with x and, thus, can be approximated by considering only the values near the peak (or mean) of the prior distribution, denoted $\hat{x}$. Letting $p(x) = \delta(x - \hat{x})$, then, one obtains the following.

$$\int H(Z \mid X = x) p(x) dx \approx H(Z \mid X = \hat{x}) \quad (8)$$

Next, to compute $H(Z)$ they approximate $p(z)$ by assuming that the sensor errors are insignificant compared to the uncertainty in the prior distribution. Thus, if $z = h(x)$ (ignoring sensor errors) is the mapping from state x to measurement z, then, in the grid representation,

$$p_Z(z) \approx \sum_{n: z = h(x_n)} p(z \mid x_n) p(x_n) \Delta x_n. \quad (9)$$

4.4 Covariance as a measure of information

Covariance is a natural quantity to consider in assessing the quality of the current prior distribution and the potential impact that a given type of measurement is expected to have on the resulting posterior. Suppose the state x is viewed as a column vector. Then the expected covariance of the posterior following a measurement of Z is

$$\overline{C} = \int \int (x - \mu)(x - \mu)^T p(x \mid z) dx p_Z(z) dz, \quad (10)$$

where μ is the expectation relative to the posterior. Note that the inner integral is the covariance resulting from a particular realization of Z; the outer integral gives the expected value of this quantity. The quality of a given type of measurement may be summarized by, say, the determinant of $\overline{C}$, denoted $\det(\overline{C})$,which is related to the volume of the corresponding ellipsoid of uncertainty. Thus, types of measurements for which this quantity is small are to be given higher prioritization.

4.5 Linear Gaussian approximation

Numerical computation of the expected covariance can be slow, as the numerical integration requires integrating over all possible measurement realizations. A faster approach would be to approximate the posterior by a multivariate Gaussian.

In this approach, we approximate the likelihood function as a Gaussian probability density function with mean $Hx + b$ and covariance R. The matrix H and vector b are determined by a first-order expansion of $h(x)$ about the current target state estimate, where $h(\cdot)$ is the function that relates the state to the ideal measurements, i.e. $z = h(x)$ assuming no sensor errors. The covariance matrix R is determined by the sensor error model.

If the prior is approximated by a Gaussian with covariance C_0 (determined by numerical integration with respect to the prior) and the likelihood is approximated as described above, then basic Kalman filter theory [11] tells us that the resulting posterior is also Gaussian, with a covariance given by

$$C = \left(C_0^{-1} + H^T R^{-1} H \right)^{-1}. \quad (11)$$

Note that, since C is independent of the particular realization of the measurement, we may replace $\overline{C}$ with C in this approximation. Thus, types of measurements for which $\det(C)$ is small are to be preferred.

5 Bearings-only sensor example

For ease of exploration, we begin with an example of six bearings-only sensors, deployed as shown in Figure 1. The sensors are assumed to be identical, and therefore each has the same sensing error model: Gaussian errors with mean zero and standard deviation of three degrees. The actual errors in the simulation of the measurements are Gaussian with mean zero and standard deviation of two degrees. The target (shown with a black triangle) is assumed to be stationary. The scenario is discussed in three forms. First, an intuitive approach to prioritizing sensor selection is taken and the resulting posteriors examined. Second, the mutual information metric is considered. Third, the expected covariance matrix is considered.

5.1 Intuitive sensor selection

The purpose of this section is to demonstrate the behavior of the posterior as different measurements are selected. The measurements are chosen based on intuitive arguments, which can then be compared to the results for the two basic metrics.

Intuitively, one would expect sensors 1 and 2 to provide (on average) more information than any other sensors, since they are the closest to the true target position. However, it would not be wise to select only sensors 1 and 2, as the horizontal localization would be poor.

Initially, the prior is a uniform distribution across the shown grid, and zero elsewhere. It can be argued that sensor 3, being located near the center of the grid, is a reasonable first choice. After this sensor is queried and a particular (random) measurement from sensor 3 is incorporated, the posterior shown in Figure 1 results.

The localization provided by this posterior is much better in the x-direction than in the y-direction. This can be seen in both the grid shading and in the illustrated covariance ellipse, which was calculated from the grid representation of the posterior. The non-normality of the posterior is also evident. Intuitively, one desires a new measurement that will reduce the uncertainty in the y-direction.

Sensor 1 or 2 can be expected to provide a lot of information in the next update; they are both relatively close to the estimated target position, and they would each "cut across" the posterior, essentially giving a bearing reading close to perpendicular to the previous reading from sensor 3. There is a slight preference for sensor 2, since the mean of the prior happens to be closer to it. Sensor 4 might also be expected to provide more information than sensor 3, although the tradeoff between its better angle and farther distance is difficult to anticipate.

Choosing sensor 2 and incorporating a (random) measurement from it results in the posterior shown in Figure 2. With this posterior, it is difficult to develop an intuitive ranking between sensors 1-3, although they all appear better than sensors 4-6.

In Figure 3, the non-zero regions of the posteriors that result from "perfect" observed measurements from all sensors are shown. In other words, we assume for this example that the observed bearings are the true bearings in order to remove chance from the comparison. The Bayesian inference procedure still considers the possibility of error, as captured by the measurement models.

While an exact ordering of the overall quality of the posteriors is difficult, the nature of the results is revealing. Sensors 1 and 2 yield very similar results: specifically, posteriors that are elongated horizontally due to a repeated measurement along the same axis. Sensor 3 yields the most symmetric posterior, as it "cuts across" the previous posterior. Sensor 4 yields a posterior that is a bit in between those for sensor 3 and either sensor 1 or 2, as may

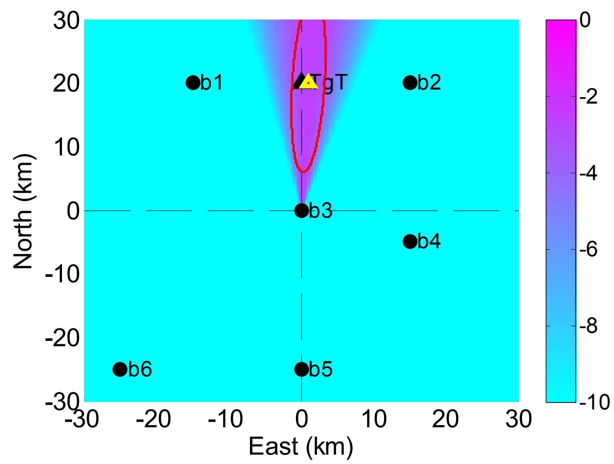

Figure 1: Sensor layout and posterior after one update. The color scale indicates the base-10 logarithm of the probability that the target is in a particular grid cell. The target is shown with a black triangle and the mean of the posterior is indicated with a yellow triangle. The red curve indicates the one-σ error ellipse.

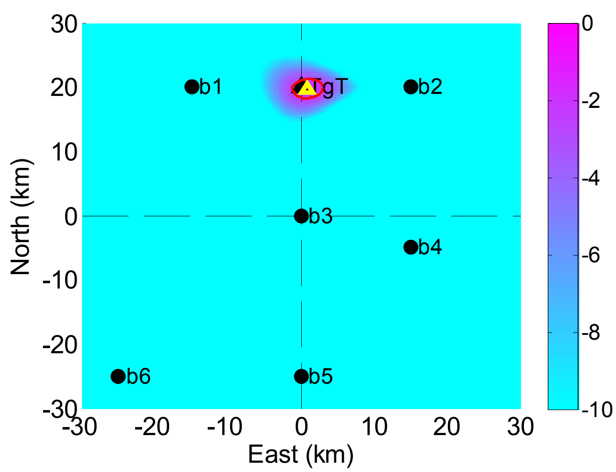

Figure 2: Posterior after two updates

the first ten measurements, although it will be sampled for the 13th and 15th measurements. This is not very surprising. Intuitively, the differences between sensors 1 and 2 should always be very small, since they have the same geometric relation to the target. The only difference between them is based on the shape of the posterior, which can become skewed by repeated sampling from the same sensor.

It appears that the expected covariance metric does not capture the same subtleties of the tradeoffs between sampling sensor 1 or sensor 2. Whether this is good, bad, or insignificant is difficult to say when considering just one random realization of a run. In the next section, the performance of the methods, in terms of state estimate error and computational cost, are explored for a mixture of randomly positioned bearing sensors and time-delay sensors.

6 General performance results

In this section, the actual localization performance of the different methods is compared across 275 different runs. In each run, there are 3 bearings-only sensors with identical sensing models (as described in Section 5), and three time-delay-only sensors with identical sensing models; specifically, the actual errors are Gaussian with mean zero and standard deviation of one second, and the error models assume a Gaussian error with mean zero and standard deviation of two seconds.

The positions of the sensors and the stationary target across runs were uniformly random in the interval [-25 km, +25 km] for each of the two dimensions, while the grid extended from -30 km to +30 km in each dimension. A grid size of $N = 150 \text{ x } 150 = 22{,}500$ cells was used to represent the state space. The measurement space for both time and bearing were divided into $M = 360$ uniform grid cells.

As a baseline, a random selection method was also

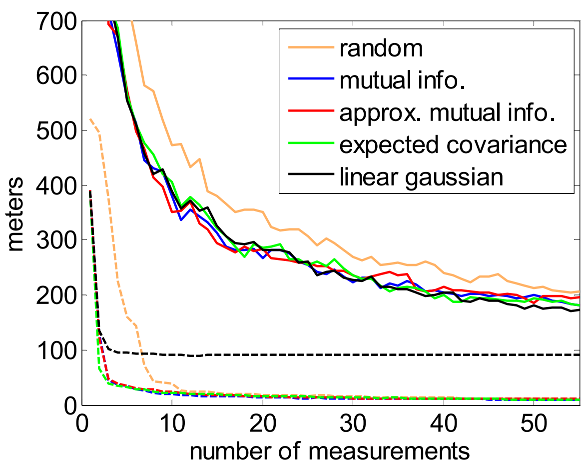

Figure 6: Comparing median (solid lines) and standard deviation (dashed lines) of estimate error across metrics

considered. This method simply randomly selects a sensor for each update, with each sensor having equal probability of selection.

6.1 Localization performance

The medians of the localization error over the 275 runs are shown in Figure 6 for the random method and all four metrics. Also shown (with dashed lines) are the standard deviations of the mean estimates. The median results indicate that generally, all four methods perform similarly and all exceed the performance of the random method. For the first five updates, the performance of the random method is much worse than that of each metric, so it is not shown in this region in order to highlight the differences on the smaller scale.

As the number of measurements increases, the performance of the random method approaches that of the other methods. This can be expected for two reasons. First, as the number of measurements increases, even a random selection will choose the valuable sensors a large number of times. Second, as the target estimate improves, the individual metrics may begin to break down due to numerical issues. For example, as the covariance of the posterior decreases, the grid-based model cannot capture it as accurately, since the number of grid cells with non-negligible probability decreases.

The standard deviations shown indicate that the linear Gaussian approximation has a much larger variation in performance than the other metrics. The median performance indicates that it generally performs very well, but the standard deviation suggests that it can deviate substantially. This is due to the possibility of degenerate behavior in very particular scenarios, such as a highly non-Gaussian prior distribution.

For example, of the 275 runs considered, for only 7 was the error using the linear Gaussian approximation method greater than one kilometer at the 30th update. Of these, all were less than 3 kilometers except for one particularly bad scenario that led to an error of around 20 kilometers.

This particular "bad" result is shown in Figure 7. Specifically, this is the posterior that resulted from selecting sensor B4 (bearing only) followed by selecting sensor T1 (time delay only). The source (S) is located very close to T1. The mean estimate is shown with the yellow triangle, and the actual target position with the black triangle. The covariance ellipse for the posterior is also shown. Note that the covariance ellipse is long in the x-direction due to the bimodality of the posterior.

The linearization approximation assumes a prior in the form of an elliptical cloud centered on the mean estimate (the red ellipse and yellow triangle in Figure 7). A good measurement would cut across this covariance ellipse, and thus the linear Gaussian approximation method anticipates a large reduction in the covariance by querying sensor T1.

only an estimate of the covariance matrix. Both the flexibility and low computational cost of the linear Gaussian approximation method make it attractive, and, as noted before, its performance is comparable to the other methods in most cases. Only the possibility of degenerate scenarios makes the method unattractive, further motivating the development of efficient methods for detecting degenerate scenarios.

7 Summary

In this paper, the impact of the ordering of measurement updates for a Bayesian inference problem in terms of the estimation of the posterior distribution was examined. To this end, we considered the actual localization performance of four methods for prioritizing the selection of sensors for data fusion. In general, all four methods performed similarly in terms of localization error and rates of convergence, and the all methods significantly outperformed a purely random selection method over the first few measurements.

The approximate expected covariance matrix method, which was based on a linearization of the measurements around the mean estimate and an update based on Kalman filtering, had the lowest computational cost and is the simplest to implement for a variety of filter types. In general, its performance was as good as the other methods. However, it was also subject to rare, yet significant errors in degenerate scenarios involving highly non-Gaussian prior distributions. These results will begin to help analysts assess the tradeoffs between the localization performance of the methods and their computational complexity.

Acknowledgements

This work was supported by the Office of Naval Research, Contract No. N00014-06-G-0218-01.

References

[1] L. D. Stone, C. A. Barlow, and T. L. Corwin, *Bayesian Multiple Target Tracking*, Artech House, Boston, 1999.

[2] C. E. Shannon. *A Mathematical Theory of Communication*. Bell System Technical Journal, Vol. 27, pp. 379-423, 623-656, 1948.

[3] B. K. Skagerstam. *On the Notions of Entropy and Information*. Journal of Statistical Physics, Vol. 12, No. 6, pp. 449-462, 1975.

[4] K. J. Hintz, and E. S. McVey. *Multi-Process Constrained Estimation*. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 21, No. 1, pp. 237-244, 1991.

[5] J. Manyika, and H. Durrant-Whyte, *Data Fusion and Sensor Management: A Decentralized Information-Theoretic Appraoch*, Ellis Horwood, New York, 1994.

[6] F. Zhao, J. Shin, and J. Reich. *Information-Driven Dynamic Sensor Collaboration*. Signal Processing Magazine, IEEE, Vol. 19, No. 2, pp. 61-72, 2002.

[7] E. Ertin, J. W. Fisher, and L. C. Potter, *Maximum Mutual Information Principle for Dynamic Sensor Query Problems*, 2nd International Workshop on Information Processing in Sensor Networks (2003), Palo Alto, CA, April 2003.

[8] J. Liu, J. Reich, and F. Zhao. *Collaborative in-Network Processing for Target Tracking*. Eurasip Journal on Applied Signal Processing, Vol. 2003, No. 4, pp. 378-391, 2003.

[9] H. F. Durrant-Whyte, B. Y. S. Rao, and H. Hu, *Toward a Fully Decentralized Architecture for Multi-Sensor Data Fusion* (1990), Cincinnati, OH, USA, pp. 1331-1336.

[10] H. Wang, K. Yao, and D. Estrin. *Information-Theoretic Approaches for Sensor Selection and Placement in Sensor Networks for Target Localization and Tracking*. Journal of Communications and Networks, Vol. 7, No. 4, pp. 438-449, 2005.

[11] Y. Bar-Shalom, *Tracking and Data Association*, Academic Press, Boston, 1988.