THE IGMARP DATA FUSION ALGORITHM

Andrew R. Runnalls

Notes:

1. A version of this paper, radically abridged to meet a four-page limit, was presented at the IEEE Nonlinear Statistical Signal Processing at Cambridge, UK in September 2006 (NSSPW2006).

2. In earlier publications by the author and coworkers, the IGMARP algorithm is referred to simply as IGMAP. Both acronyms refer to the same algorithm.

1. INTRODUCTION

IGMARP (Iterative Gaussian Mixture Approximation of the Reduced-Dimension Posterior) is a data fusion algorithm for handling non-linear measurements, particularly ambiguous measurements (i.e. measurements for which the likelihood function may be multimodal), in conjunction with a linear or linearisable system model. It is particularly well suited to system models of high dimensionality, and applications where it is desired to interoperate with existing approaches using a Kalman Filter or multi-hypothesis Kalman Filter.

The algorithm has been developed over the period since 2001 as a means of integrating data from terrain-referenced navigation systems into a multiway integrated navigation solution also comprising an inertial navigation system (INS) and GPS. An example of a terrain-referenced navigation system is terrain-contour navigation (TCN), in which an air vehicle uses a radio altimeter or similar sensor to take measurements of the height above sea level of the terrain being overflown.

The paper describes the mathematical foundations of the algorithm, and illustrates its application to an integrated TCN/INS system. Sec. 2 introduces the motivating application, TCN. Sec 3 reviews the measurement update equations for the multi-hypothesis Kalman filter (MHKF), which represents an application of Bayes' Theorem to the case in which the prior distribution is a Gaussian mixture, and the likelihood function also has the form of a (slightly generalised) Gaussian mixture. Sec. 4 then discusses how the likelihood function can be computed for TCN, and gives the flavour of

the resulting functions, which are by no means of a Gaussian mixture form; this motivates Sec. 5, which discusses how the MHKF approach can be adapted to handle more general likelihood functions, and introduces the key theorems on which the IGMARP method depends. Then Sec. 6 describes the algorithm itself, and Sec. 7 illustrates the results of applying the algorithm to TCN/INS flight data. Finally Sec. 8 discusses conclusions and possible further work.

2. TERRAIN-CONTOUR NAVIGATION

An inertial navigation system (INS) operates continuously and provides a high bandwidth (>50 Hz) navigational output with low short term noise. It also provides effective attitude, angular rate and acceleration measurements as well as position and velocity. However, its navigation accuracy degrades with time as the noise and biases on its inertial instrument outputs are mathematically integrated through the navigation equations that generate the final output.

The Global Positioning System (GPS), and similar satellite navigation systems, provide a high accuracy position solution—of the order of 10 m (2σ) in each axis—that does not degrade with time. The GPS navigation solution is noisier than that of an INS, has a lower bandwidth ($\sim$1 Hz) and does not normally include attitude. GPS and INS are thus complementary. Consequently, many aircraft and guided weapons use an integrated INS/GPS navigation system. The INS provides the core navigation solution, whilst the GPS measurements are used to correct and calibrate the INS via an integration algorithm.

However, satellite navigation signals are extremely vulnerable to interference. Unintentional interference sources include broadcast television, mobile satellite services, ultrawide-band communications, over-the-horizon radar and cellular telephones [1]. In military applications, deliberate jamming is highly likely, and must be planned for. Interference can be mitigated using a controlled reception pattern antenna (CRPA) system (e.g. [2, 3, 4]), together with advanced INS/GPS integration techniques. These techniques enable satellite navigation signals to be tracked under higher levels of interference. However, they do not eliminate the effects of jamming and interference completely. The cost and complexity of jamming technology that can defeat them is signif-

icantly less than that of the CRPA systems themselves, and this technology is being communicated across the internet.

As soon as GPS measurements are lost, the INS begins to drift out of calibration. Aircraft-grade INS can maintain a horizontal position accuracy within 100 m through GPS outages of more than 10 minutes. However, the lower cost INS common in guided weapons, unmanned air vehicles and general aviation (private) aircraft can only maintain this accuracy for 2 to 3 minutes. To attain robust navigation in a GPS jamming environment, reversionary navigation systems are required.

Terrain-referenced navigation (TRN) techniques offer a solution. The most established TRN technique, terrain contour navigation (TCN), began to be developed in the 1970s and a number of systems have been produced commercially over the years. Conventionally, such systems estimate the height of the terrain below the air vehicle by subtracting radio altimeter height from INS or barometric/INS altitude. Measurements are typically taken around once a second. These are then compared with a terrain height database, such as Digital Terrain Elevation Data (DTED) [5]. Performance may be enhanced by using a laser range-finder as the sensor [6, 7].

A range of different processing techniques have been developed to obtain position fixes from the comparisons of measured and database terrain heights [8]. These may be divided into two broad categories: batch and sequential.

2.1. Batch processing

In batch processing, a series of terrain height measurements, known as a transect, are processed together; this was the approach taken in the original TERCOM system [9], and in early work of the present author [10]. The transect is compared to the terrain height database at different offsets in latitude, longitude and height from the current estimated aircraft position. The residuals of each comparison are used to calculate what, in the context of the present paper, we can consider to be the likelihood function.

If the transect is sufficiently long—spanning several kilometres, say—then the observed terrain profile over long distances will usually be effectively unique within the area of navigational uncertainty. Consequently the resulting likelihood function will have a single sharp peak: early TCN systems would simply take the position of this peak as the basis of a position fix to be input to a Kalman filter.

The disadvantage with this 'long transect' approach is there there is a considerable delay before the navigational estimate is updated. If this delay could be reduced, there would be less time for residual navigational drifts to accumulate before the TCN update becomes available. This in turn may allow the air vehicle to fly at lower altitude, since the position of terrain features and obstacles in relation to the aircraft will be better known. This consideration is particularly strong for applications such as unmanned air vehicles and guided weapons, where the use of lower grade INS will result in high rates of navigational drift.

2.2. Sequential processing

In sequential processing, each measurement is processed separately. The difference between the radalt generated and database indicated terrain height is input as a measurement to a data fusion algorithm. Typically this data fusion algorithm is an extended Kalman filter (EKF): the terrain gradient at the current best estimate of position is used to attribute the observed height difference to a linear combination of the latitude, longitude and height components of the INS position error. Sequential processing is well established in commercial TCN systems such as BAE Systems' TERPROM [11] and the American SITAN [12, 13, 14].

The principal advantage of the EKF sequential approach is relative simplicity and comparatively low processor load. However, it relies on accurate knowledge of the terrain gradient below the aircraft, which is a demanding requirement on existing low resolution, low accuracy databases like DTED. To a certain extent, the limitations of terrain height databases may be overcome by using sophisticated linearisation algorithms [15]. However, a fundamental problem remains in that the gradient is calculated below the aircraft's position *as indicated by the navigation system*, not its true position. Thus, if the horizontal position error exceeds about 250 m, mainstream sequential processing does not work and a 'recovery' mode must be instigated, for example batch processing.

Particle-filtering approaches [16, 8] also take a sequential approach, but without the linear/Gaussian assumptions of the extended Kalman filter. There are, however, some disadvantages with MCMC (including particle filter) approaches to integrated navigation:

- For relatively straightforward integrated navigation tasks, such as INS/GPS integration, MCMC methods are an overkill: as compared with a Kalman filter, they require much greater computational loads, but offer little or no performance advantage.

- Consequently, for multiway integration, e.g. INS/GPS/TRN, it would seem to be desirable for the overall integration to be carried out using Kalman filtering techniques, leaving MCMC approaches to deal with the non-linearities and ambiguities of TRN, and somehow feeding the MCMC output into the overall Kalman filter. Unfortunately, the output of MCMC methods is not in a form which readily lends itself to use as a Kalman filter input.

- Typical INS models often have 15 or more elements in the state vector. Such high dimensionality may cause problems for MCMC approaches.

2.3. IGMARP objectives

The IGMARP algorithm described in this paper arose from a need for a data fusion algorithm suited to multiway integrated navigation with terrain-referenced navigation as one (or more) of the inputs. There were three primary objectives in developing the algorithm:

1. The algorithm should be capable of updating the system's navigation solution quickly—e.g. within a few seconds—in response to newly-gathered terrain data.

2. The algorithm should be capable of handling the position ambiguities that often arise using terrain-referenced navigation, in such a way as to make maximum and timely use of the terrain data.

3. The algorithm should be directly compatible with Kalman filter approaches to multiway integrated navigation, including in particular the multi-hypothesis Kalman filter (MHKF).

3. THE MHKF MEASUREMENT UPDATE

3.1. Prior distribution and likelihood

Let $g(\mathbf{x}; \boldsymbol{\mu}, P)$ represent the multivariate Gaussian density for a d-dimensional r.v. $\mathbf{x}$, with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix P. In general, if P—adorned perhaps with subscripts and other diacritical marks—represents a covariance matrix, we shall use $\mathcal{I}$—similarly adorned—to represent the corresponding information matrix $\mathcal{I} = P^{-1}$.

Suppose that a linear dynamical system has a d-dimensional state vector $\mathbf{x}$, and that our information about $\mathbf{x}$ can be represented as a Gaussian mixture distribution with n components, as follows:

$$
f(\mathbf{x}) = \sum_{i=1}^{n} w_i g(\mathbf{x}; \boldsymbol{\mu}_i, P_i)
$$
$$
= \sum_{i=1}^{n} w_i \sqrt{\frac{\det \mathcal{I}_i}{(2\pi)^d}} \exp -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \mathcal{I}_i (\mathbf{x} - \boldsymbol{\mu}_i) \quad (1)
$$

where $\sum_{i=1}^{n} w_i = 1$ and all of the information matrices $\mathcal{I}_i$ are positive definite.

Suppose further that a measurement $\mathbf{z}$ becomes available, where the distribution of $\mathbf{z}$ depends on $\mathbf{x}$ only via $H\mathbf{x}$, where H is $d' \times d$ matrix of full rank, with $d' \leq d$. In many cases, d' is much smaller than d, and it is to these cases that the IGMARP algorithm is particularly applicable.

In a Bayesian formulation, the information provided by $\mathbf{z}$ about $\mathbf{x}$ is entirely encapsulated within the likelihood function $L(\mathbf{x}|\mathbf{z})$, which we assume to have the form of a linear combination of m d'-dimensional components as follows:

$$
L(\mathbf{x}|\mathbf{z}) = \sum_{j=1}^{m} \lambda_j \exp -\frac{1}{2} (H\mathbf{x} - \boldsymbol{\zeta}_j)^T \Lambda_j (H\mathbf{x} - \boldsymbol{\zeta}_j) \quad (2)
$$

where $\boldsymbol{\zeta}_j$ defines the location of the jth component, λ_j is its weight, and Λ_j is a $d' \times d'$ symmetric matrix describing its shape. For the time being we shall assume that Λ_j is non-negative definite, but we shall relax this condition at the end of Sec. 3.

By Bayes' Theorem, the posterior density $f(\mathbf{x}|\mathbf{z})$ is proportional to the product of the prior distribution (1) and the likelihood function (2).

3.2. Terms of the posterior distribution

In this section we examine the form of a typical term of the posterior distribution. Let:

$$
\phi = \sqrt{\frac{\det \mathcal{I}}{(2\pi)^d}} \exp -\frac{1}{2} \Big[(\mathbf{x} - \boldsymbol{\mu})^T \mathcal{I} (\mathbf{x} - \boldsymbol{\mu})
$$
$$
+ (H\mathbf{x} - \boldsymbol{\zeta})^T \Lambda (H\mathbf{x} - \boldsymbol{\zeta}) \Big] \quad (3)
$$

Consider the expression in square brackets. We have:

$$
(\mathbf{x} - \boldsymbol{\mu})^T \mathcal{I} (\mathbf{x} - \boldsymbol{\mu}) + (H\mathbf{x} - \boldsymbol{\zeta})^T \Lambda (H\mathbf{x} - \boldsymbol{\zeta})
$$
$$
= \mathbf{x}^T (\mathcal{I} + H^T \Lambda H) \mathbf{x} - \mathbf{x}^T (\mathcal{I}\boldsymbol{\mu} + H^T \Lambda \boldsymbol{\zeta})
$$
$$
- (\boldsymbol{\mu}^T \mathcal{I} + \boldsymbol{\zeta}^T \Lambda H) \mathbf{x} + (\boldsymbol{\mu}^T \mathcal{I}\boldsymbol{\mu} + \boldsymbol{\zeta}^T \Lambda \boldsymbol{\zeta}) \quad (4)
$$

Now define

$$
\mathcal{I}' = \mathcal{I} + H^T \Lambda H \quad (5)
$$

Since $\mathcal{I}$ is positive definite, $\mathcal{I}'$ is also, and therefore has an inverse. Further define:

$$
\boldsymbol{\mu}' = \mathcal{I}'^{-1} (\mathcal{I}\boldsymbol{\mu} + H^T \Lambda \boldsymbol{\zeta}) \quad (6)
$$
$$
= \boldsymbol{\mu} + \mathcal{I}'^{-1} H^T \Lambda (\boldsymbol{\zeta} - H\boldsymbol{\mu}) \quad (7)
$$

On substituting from (5) and (6) into (4), the square-bracketed expression in (3) becomes:

$$
\mathbf{x}^T \mathcal{I}' \mathbf{x} - \mathbf{x}^T \mathcal{I}' \boldsymbol{\mu}' - \boldsymbol{\mu}'^T \mathcal{I}' \mathbf{x} + (\boldsymbol{\mu}^T \mathcal{I}\boldsymbol{\mu} + \boldsymbol{\zeta}^T \Lambda \boldsymbol{\zeta})
$$
$$
= (\mathbf{x} - \boldsymbol{\mu}')^T \mathcal{I}' (\mathbf{x} - \boldsymbol{\mu}') + (\boldsymbol{\mu}^T \mathcal{I}\boldsymbol{\mu} + \boldsymbol{\zeta}^T \Lambda \boldsymbol{\zeta} - \boldsymbol{\mu}'^T \mathcal{I}' \boldsymbol{\mu}')
$$

Note that the last term, $(\boldsymbol{\mu}^T \mathcal{I}\boldsymbol{\mu} + \boldsymbol{\zeta}^T \Lambda \boldsymbol{\zeta} - \boldsymbol{\mu}'^T \mathcal{I}' \boldsymbol{\mu}')$, does not depend on $\mathbf{x}$.

If we now substitute this result back into (3), we get:

$$
\phi = \sqrt{\frac{\det \mathcal{I}}{\det \mathcal{I}'}} \exp -\frac{1}{2} (\boldsymbol{\mu}^T \mathcal{I}\boldsymbol{\mu} + \boldsymbol{\zeta}^T \Lambda \boldsymbol{\zeta} - \boldsymbol{\mu}'^T \mathcal{I}' \boldsymbol{\mu}')
$$
$$
\times g(\mathbf{x}; \boldsymbol{\mu}', \mathcal{I}'^{-1}) \quad (8)
$$

In (8) it is possible to represent the subexpression

$$
S = \boldsymbol{\mu}^T \mathcal{I}\boldsymbol{\mu} + \boldsymbol{\zeta}^T \Lambda \boldsymbol{\zeta} - \boldsymbol{\mu}'^T \mathcal{I}' \boldsymbol{\mu}' \quad (9)
$$

as a quadratic form: by substituting from (7) into (9) it is straightforward to show that

$$
S = (\boldsymbol{\zeta} - H\boldsymbol{\mu})^T (\Lambda - \Lambda H \mathcal{I}'^{-1} H^T \Lambda)(\boldsymbol{\zeta} - H\boldsymbol{\mu}) \quad (10)
$$

3.3. Posterior distribution

When the prior density has the Gaussian mixture form of (1), the corresponding posterior density can be expressed in the form:

$$f(\mathbf{x}|\mathbf{z}) \propto \sum_{i=1}^{n} w_i \phi_i(\mathbf{x}|\mathbf{z}) \tag{11}$$

where

$$\phi_i(\mathbf{x}|\mathbf{z}) = g(\mathbf{x}; \boldsymbol{\mu}_i, P_i) L(\mathbf{x}|\mathbf{z}) \tag{12}$$

and the constant of proportionality in (11) is such that the left-hand side integrates to unity. We shall refer to the positive measures defined by the density functions $\phi_i(\mathbf{x}|\mathbf{z})$ as the **unnormalised posterior components (UPCs)**.

In the case where the likelihood function is itself of the Gaussian mixture form (2), it follows from (8) and (10) that the UPCs are given by:

$$\phi_i(\mathbf{x}|\mathbf{z}) = \sum_{j=1}^{m} \lambda_j c_{ij} g(\mathbf{x}; \boldsymbol{\mu}'_{ij}, P'_{ij}) \tag{13}$$

where

$$P'_{ij} = (\mathcal{I}'_{ij})^{-1} \tag{13a}$$

$$\mathcal{I}'_{ij} = \mathcal{I}_i + H^T \Lambda_j H \tag{13b}$$

$$\boldsymbol{\mu}'_{ij} = \boldsymbol{\mu}_i + K_{ij}(\boldsymbol{\zeta}_j - H\boldsymbol{\mu}_i) \tag{13c}$$

$$K_{ij} = P'_{ij} H^T \Lambda_j \tag{13d}$$

$$c_{ij} = \sqrt{\frac{\det \mathcal{I}_i}{\det \mathcal{I}'_{ij}}} \times \exp{-\frac{1}{2}(\boldsymbol{\zeta}_j - H\mu_i)^T \Lambda'_{ij}(\boldsymbol{\zeta}_j - H\mu_i)} \tag{13e}$$

$$\Lambda'_{ij} = \Lambda_j - \Lambda_j H P'_{ij} H^T \Lambda_j \tag{13f}$$

We now observe that for (11) to represent a well-defined Gaussian mixture of positive-definite components, it is not strictly necessary for the matrices Λ_j to be non-negative definite: all that is required is that $\mathcal{I}'_{ij}$ given by (13b) is positive definite for each i and j.

4. THE TCN LIKELIHOOD FUNCTION

As applied to TCN, IGMARP has been implemented as a batch processing algorithm, processing short transects of TCN data spanning a few hundred metres of terrain with the radio altimeter sampled at 1–2 Hz, so as to achieve a horizontal separation of about 100 m between the samples. Closer sampling than this would increase the processing load but yield little accuracy benefit, owing to the limited resolution both of the radio altimeter and the DEM.

Using short transects avoids the latency problems noted above for long transects, and helps to meet the first design objective. (It is straightforward to configure the algorithm so that these 'batches' each comprise a single radio-altimeter reading, thus turning it into a fully sequential algorithm, but experience to date indicates that this fully-sequential approach does not offer best performance, and it substantially increases the computational load.)

To motivate the discussion in the following sections, we now take a look at the characteristics of the likelihood function that arise from processing a short transect of TCN data. The transect data are processed by subtracting the height above ground measured by the radio altimeter from the aircraft height measured by the INS (as corrected by the integrated navigation system) to yield a **sensed terrain profile**. Fig. 1(a) shows the sensed terrain profile measured during a transect of 4 seconds' duration, with the radio altimeter sampled at 2 Hz, yielding 9 samples in all. Notice that sensed terrain profile plotted here will be offset from the true terrain profile—the section of terrain profile that was actually being overflown while the transect data were gathered—because of residual position errors in the INS, both horizontal and vertical. In the initial stages of the operation of the algorithm, this *absolute* offset may be of the order of many hundreds of metres horizontally, and many tens of metres vertically. However, the *relative* positions of the points along the sensed terrain profile will normally be in much better agreement with the relative positions of the points along the true terrain profile, although they will still be affected by residual velocity errors in the INS, radio altimeter errors, and errors in the digital elevation map (DEM).

The next stage is therefore to search for horizontal and vertical position offsets that will bring the sensed profile into good agreement with the terrain profile given by the DEM. Fig. 1(b) shows the form of the terrain surface within the area of navigational uncertainty, as given by the DEM. This search proceeds by working through a series of hypotheses about the true position of the aircraft, both vertically and horizontally, at the time of the **mid-transect point**: the time when the fifth of the nine radio altimeter readings forming the transect was sampled. (Working from the middle of the transect helps to minimise the effects of residual INS velocity errors.) For each hypothesis about the mid-transect point, we examine the DEM to determine the terrain profile that would have been overflown during the transect if that hypothesis were true, and compare that DEM terrain profile with the (appropriately offset) sensed terrain profile. This comparison is performed using a statistical model characterising the errors arising from the radio altimeter, and from inaccuracies in the DEM itself, and yields the likelihood of the mid-transect point hypothesis.

The resulting likelihood function (reduced for presentational clarity to two dimensions) is shown in Fig. 1(c). The reader will observe that there are three areas of relatively high likelihood towards the north of the area shown, and further such regions to the south and to the west of the centre of the region. There is generally rather poor agreement elsewhere, for example towards the south-west corner. (In fact, the true position of the aircraft lies within one of the peaks to the west
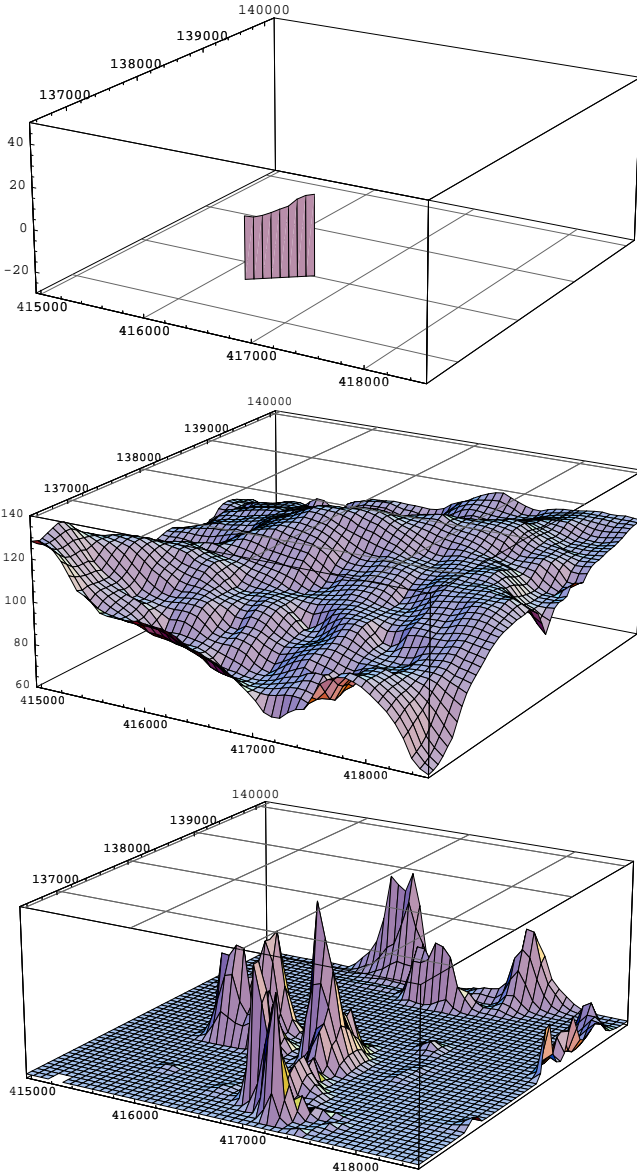
of the region's centre.)

Fig. 1. (*a*) The sensed terrain profile observed during a transect; the aircraft was flying approximately south-west while the transect data were gathered. (*b*) The terrain topography as given by the Digital Elevation Map (DEM). (*c*) The resulting likelihood function. The horizontal graticule comprises 1 km squares, and in panels (*a*) and (*b*), heights are in metres. Note that the vertical scale is exaggerated by a factor of about 17.5 in comparison with the horizontal scale.

5. HANDLING MORE GENERAL LIKELIHOODS

We now consider how the results of Sec. 3 might be adapted to handle measurements $\mathbf{z}$ which give rise to likelihood functions that do not conform to the Gaussian mixture pattern of (2), for example the likelihood function seen in Fig. 1(c). We shall however assume that the likelihood function is bounded above, i.e. that there exists ℓ such that $L(\mathbf{x}|\mathbf{z}) < \ell$ for all $\mathbf{x}$.

One approach would be directly to approximate the likelihood function as a mixture of Gaussians. This is not the approach we shall take. This is partly because, in the application for which IGMARP was developed, the likelihood value is not well localised, and may have substantial peaks well away from the region of the state space where the prior distribution is concentrated. This is true, for example, of the case illustrated in Fig. 1(c), where the peaks are by no means confined to the horizontal area included in the figure.

Indeed, the likelihood function may well be bounded *below* by a positive value: this would for example rule out the use of the Redner-Walker algorithm [17] for Gaussian mixture approximation, since the required integrals would not converge.

When faced with poorly localised likelihood functions such as these, we should perhaps think not about approximating the likelihood function *as a whole*, but only within the region where the prior density is non-negligible. Indeed, we may think about approximating the likelihood function *weighted* by the prior density. But of course, the likelihood weighted by the prior is (but for a normalisation constant) exactly the same thing as the posterior density.

This points to the approach we shall take, which is to approximate the posterior distribution as a Gaussian mixture. As an intermediate step, we shall approximate the UPCs $\phi_i(\mathbf{x}|\mathbf{z})$ as Gaussian mixtures. However, these UPCs (like the full posterior itself) span $\mathbb{R}^d$, where d is the full dimension of the state space. Fortunately, as we shall now show, the numerically intensive part of the approximation can be carried out in merely d' dimensions.

Let $\breve{\phi}_i(\breve{\mathbf{x}}|\mathbf{z})$ be the (unnormalised) density function over $\mathbb{R}^{d'}$ induced by $\phi_i(\mathbf{x}|\mathbf{z})$ under the mapping H. Here we have introduced the convention of using a breve (superposed $\breve{}$) to denote a quantity defined over the d'-dimensional space spanned by the columns of H; $\breve{\mathbf{x}} (= H\mathbf{x})$ is a variable ranging over this space. We shall refer to the density functions $\breve{\phi}_i(\breve{\mathbf{x}}|\mathbf{z})$ as the **projected UPCs**.

In the case where a UPC has the form given in (13), the corresponding projected UPC is clearly given by:

$$\breve{\phi}_i(\breve{\mathbf{x}}|\mathbf{z}) = \sum_{j=1}^{m} \lambda_j c_{ij} g(\breve{\mathbf{x}}; \breve{\boldsymbol{\mu}}'_{ij}, \breve{P}'_{ij}) \qquad (14)$$

where

$$\breve{\boldsymbol{\mu}}'_{ij} = H\boldsymbol{\mu}'_{ij} \qquad (14a)$$

$$\breve{P}'_{ij} = HP_{ij}H^T \qquad (14b)$$

Here we have derived the parameters of the projected UPC from the parameters of the unprojected UPC. However, it is possible to convert in the opposite direction, provided that we know the parameters of the corresponding component of the prior distribution. This is shown by the following theorem:

Theorem 1 *Suppose that for a particular i and j (which we shall leave implicit in what follows, by omitting the subscripts), P', $\mathcal{I}'$, $\boldsymbol{\mu}'$, $\breve{\boldsymbol{\mu}}'$ and $\breve{P}'$ are given by* (13a), (13b), (13c), (14a) *and* (14b) *respectively. Then:*

$$\boldsymbol{\mu}' = C(\breve{\boldsymbol{\mu}}' - \breve{\boldsymbol{\mu}}) \qquad (15)$$

$$P' = P + PH^T(\breve{P}^{-1}\breve{P}'\breve{P}^{-1} - \breve{P}^{-1})HP \qquad (16)$$

where

$$C = PH^T\breve{P}'^{-1} \qquad (17)$$

and $\breve{\boldsymbol{\mu}}$ and $\breve{P}$ are the parameters of the projected prior component, as follows:

$$\breve{\boldsymbol{\mu}} = H\boldsymbol{\mu}$$

$$\breve{P} = HPH^T$$

Proof: From (13c) and (13d) we have

$$\boldsymbol{\mu}' = \boldsymbol{\mu} + P'H^T\Lambda(\boldsymbol{\zeta} - H\boldsymbol{\mu}) \qquad (18)$$

and therefore

$$\breve{\boldsymbol{\mu}}' = \breve{\boldsymbol{\mu}} + \breve{P}'\Lambda(\boldsymbol{\zeta} - H\boldsymbol{\mu})$$

Since $\mathcal{I}'$ is positive definite and H is of full rank, P' and $\breve{P}'$ will also be positive definite **(check this!)** and therefore nonsingular. Consequently:

$$\Lambda(\boldsymbol{\zeta} - H\boldsymbol{\mu}) = \breve{P}'^{-1}(\breve{\boldsymbol{\mu}}' - \breve{\boldsymbol{\mu}})$$

Substituting this back into (18) establishes (15).

Now let us turn our attention to (16). If $\Lambda = \mathbf{0}$ then $\mathcal{I}' = \mathcal{I}$ and (16) follows trivially. Otherwise, Λ will have $d'' \geq 1$ non-zero eigenvalues, and Λ can **(check this!)** be expressed in the form $G^T DG$, where D is a $d'' \times d''$ diagonal matrix comprising the non-zero eigenvalues, and G is a $d'' \times d'$ matrix.

Using (13b) and the Woodbury matrix identity, we now have:

$$P' = (\mathcal{I} + H^T G^T DGH)^{-1}$$
$$= P - PH^T G^T (D^{-1} + G\breve{P}G^T)^{-1}GHP \qquad (19)$$

Consequently

$$\breve{P}' = \breve{P} - \breve{P}G^T(D^{-1} + G\breve{P}G^T)^{-1}G\breve{P}$$

which can be rearranged to yield:

$$G^T(D^{-1} + G\breve{P}G^T)^{-1} = \breve{P}^{-1} - \breve{P}^{-1}\breve{P}'\breve{P}^{-1}$$

Substituting this back into (19) yields (16). ∎

Computationally, it is preferable to recast (16) in the following equivalent form:

$$P' = (I - CH)P(I - CH)^T + C\breve{P}'C^T \qquad (20)$$

since it avoids subtracting one positive definite matrix from another and relying on the result being positive definite. (20) is comparable to (and was inspired by) the Joseph form of the standard Kalman filter covariance update equation (cf. [18, p. 305]).

Note that the proof of Theorem 1 did not assume that Λ is non-negative definite: only that is symmetric. How can we be sure that the matrix P' given by (16) is positive definite? Here we are assisted by the following theorem:

Theorem 2 *Suppose that P and $\breve{P}'$ are positive definite. Then the matrix P' given by* (16) *is also positive definite.*

Proof: By inspection of (20) it is clear that P' is at least non-negative definite.

First recast (16) in the following form:

$$P' = (P - PH^T\breve{P}^{-1}HP) + PH^T\breve{P}^{-1}\breve{P}'\breve{P}^{-1}HP \qquad (21)$$

and note, by comparison with (20), that both terms on the right-hand side are at least non-negative definite.

Now suppose (contrary to what we wish to establish) that there exists a non-zero d-vector $\mathbf{x}$ such that $\mathbf{x}^T P'\mathbf{x} = 0$. This implies:

$$\mathbf{x}^T(P - PH^T\breve{P}^{-1}HP)\mathbf{x} = 0 \qquad (22)$$

$$\mathbf{x}^T PH^T\breve{P}^{-1}\breve{P}'\breve{P}^{-1}HP\mathbf{x} = 0 \qquad (23)$$

since neither of these terms can be negative. Now, since $\breve{P}^{-1}\breve{P}'\breve{P}^{-1}$ is positive definite, the only way in which (23) can be satisfied is if $HP\mathbf{x} = 0$. But in that case (22) reduces to $\mathbf{x}^T P\mathbf{x} = 0$, which is impossible since P is positive definite. The supposition that P' is *not* positive definite therefore leads to a contradiction. ∎

So, to recap, P' will be positive definite provided $\breve{P}'$ is. This leads to some possibilities not embraced by the standard Kalman filter formulation: in particular, we can have 'negative updates' in which some of the diagonal elements of P' are *larger* than the corresponding elements of P.

6. THE ALGORITHM

A schematic diagram of the IGMARP method is given in Fig. 2, which is cross-referenced to the description that follows. The IGMARP method works by iterating through the following steps:

1. For each component i, $i = 1, \ldots n$, of the prior distribution (1):

   (a) Determine the projection of the distribution component $g(\mathbf{x}|\boldsymbol{\mu}_i, P_i)$ under the mapping H. This will be the p-dimensional Gaussian distribution $\breve{g}(\breve{\mathbf{x}}|\breve{\boldsymbol{\mu}}_i, \breve{P}_i)$, with mean

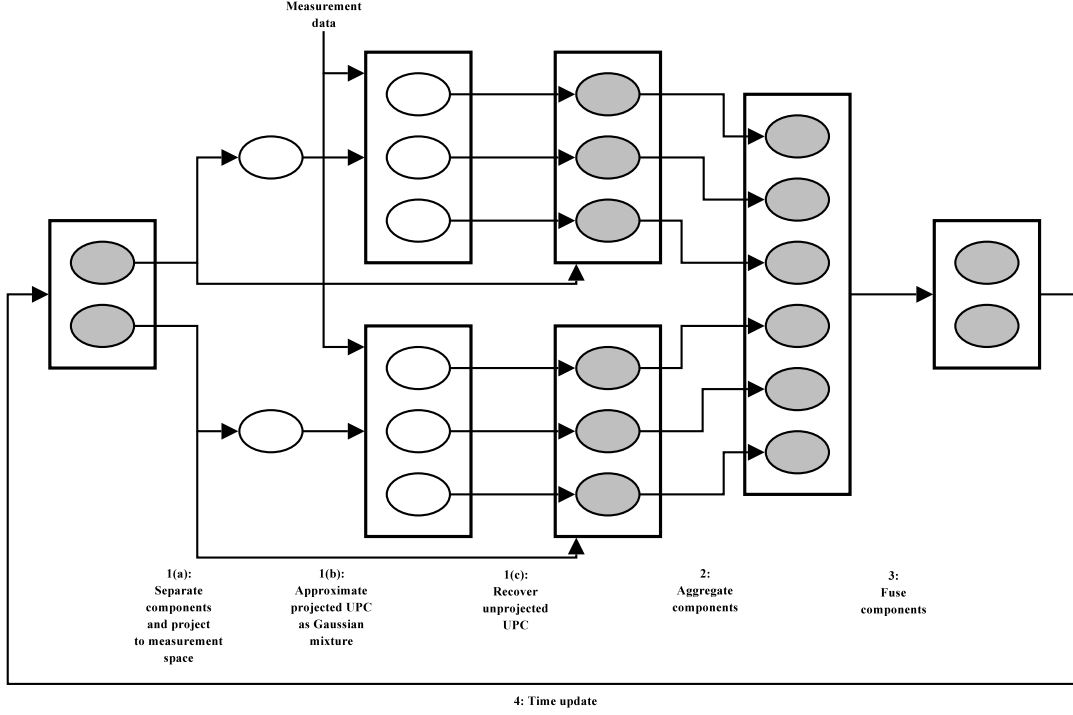 $$\breve{\boldsymbol{\mu}}_i = H\boldsymbol{\mu}_i$$

Fig. 2. Schematic diagram of the IGMARP method, for $n = 2$ and $m = 3$. Gaussian components of full dimension (d) are represented as shaded ellipses, while Gaussian components projected to the measurement space are represented as hollow ellipses.

and covariance matrix

$$\breve{P}_i = HP_iH^T$$

(b) Approximate the projected UPC given by

$$\breve{\phi}'_i(\breve{\mathbf{x}}) = f(\mathbf{z}|\breve{\mathbf{x}})\breve{g}(\breve{\mathbf{x}}|\breve{\boldsymbol{\mu}}_i, \breve{P}_i) \qquad (24)$$

as a mixture of some number m of Gaussian components:

$$\breve{\phi}'_i(\breve{\mathbf{x}}) = \sum_{j=1}^{m} w_{ij}\breve{g}(\breve{\mathbf{x}}; \breve{\boldsymbol{\mu}}'_{ij}, \breve{P}'_{ij})$$

This approximation can be accomplished using the now well-known Redner-Walker algorithm [17] (often known as the EM algorithm). Notice that this will involve numerical integrations in only d' dimensions, rather than the full dimensionality d of the state vector.

(c) For each j, $j = 1, \ldots m$, use Theorem 1 to determine the parameters $\boldsymbol{\mu}'_{ij}$, P'_{ij} of the corresponding unprojected UPC $\phi'_i(\breve{\mathbf{x}})$.

2. Assemble together all the Gaussian components resulting from Step 1c and normalise the result, to obtain our first-round Gaussian mixture approximation to the posterior:

$$f'(\mathbf{x}) \approx \sum_{i=1}^{n}\sum_{j=1}^{m} w'_{ij}g(\mathbf{x}; \boldsymbol{\mu}'_{ij}, P'_{ij}) \qquad (25)$$

where

$$w'_{ij} = kw_iw_{ij}$$

with the constant k being chosen so that:

$$\sum_{i=1}^{m}\sum_{j=1}^{m} w'_{ij} = 1$$

3. Unfortunately, the mixture in (25) contains nm components, rather than the n components we started with. Obviously, if this procedure is to be used recursively, we cannot permit the number of components to increase on each iteration, so it is necessary to bring the number of components back down to n.

Various methods have been discussed in the literature for performing Gaussian mixture reduction, i.e. approximating a Gaussian mixture with another mixture with fewer components: see for example [19, 20, 21].

However the method favoured for IGMARP is that described in [22], which is specially adapted to merging mixtures of high dimensionality.

The method works by repeatedly choosing a pair of components and **fusing** them, i.e. replacing them by a single component whose weight is the sum of the weights of the two fused components, and whose first and second moments are equal to the (joint) first and second moments of the components being fused. This pairwise fusing continues until the number of components is reduced back down to n.

[22] puts forward a criterion, based on Kullback-Leibler divergence, for choosing which pair of components to fuse at each stage. This criterion has a strong tendency to select low-weighted components as candidates for fusing, either with other low-weighted components, or with more heavily weighted nearby components. This means that the 'dynamic range' of the weights in the eventual n-component mixture is likely to be considerably smaller than that in the original nm-component mixture, which is probably desirable.

4. Allow for the passage of time before the next measurement data become available by applying the Kalman filter time update equations to each of the new n components, leaving their weights unchanged. Alternatively, if the system dynamics are appreciably nonlinear, each component can by time-updated using the method of the unscented Kalman filter [23].

7. THE IGMARP ALGORITHM IN ACTION

A previous paper [24] used simulation to compare a number of different techniques for integrating TCN with INS, and for the three-way integration of TCN with INS and GPS. The techniques considered for integrating TCN data included a best-fix (effectively, maximum likelihood) method, a probabilistic data association filter (PDAF), and IGMARP. It concluded that a weighted fix integration technique—of which PDAF and IGMARP are examples—makes the navigation solution more robust against false TCN fixes than a simple best-fix integration. The simulation results obtained indicated that IGMARP performs sufficiently better than the PDAF algorithm to justify the greater complexity and processor load that it entails.

In this section, we describe the results, first reported in [25], of applying IGMARP not to simulated data, but to real flight data recorded during a sortie of a QinetiQ Tornado GR1 aircraft over southern Britain. The equipment for the trial included a Honeywell H764G Embedded GPS/INS (EGI) and a BAE Systems AD1990 radio altimeter, both mounted in a pod fitted under the fuselage of the aircraft. The H764G incorporates a GPS receiver and a ring-laser gyro INS, and provides
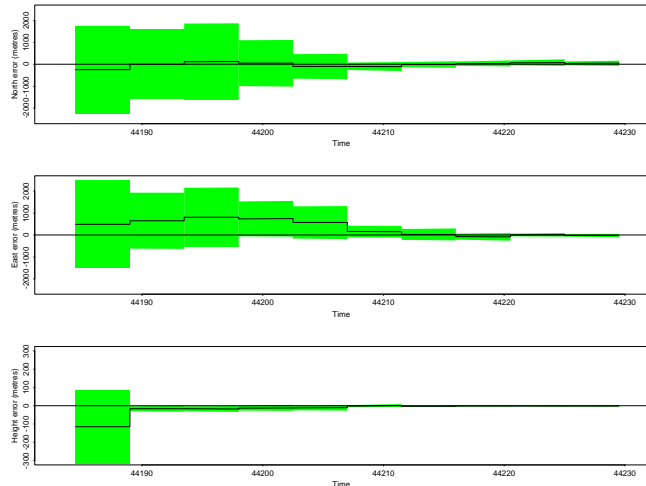

Fig. 3. Position errors during the first 45 seconds after terrain-aided navigation begins. (Time is measured as seconds after midnight.)

both a blended GPS/INS navigation output and a pure inertial output (with barometric damping of the vertical channel). GPS signals were obtained via a typical fixed reception pattern antenna (FRPA). The IGMARP algorithm was applied to combine data from the AD1990 with the pure inertial output from the H764G, using DTED Level 1 [5] (compiled *circa* 1986) as the digital elevation map. The blended INS/GPS output from the H764G was used as the 'truth' measure with which the IGMARP output is compared.

Although the INS in this trial was of aircraft grade, the model of the INS incorporated in the Kalman filter system model was very pessimistic, particularly as regards the initial conditions: the initial position was assumed to be accurate only to within ± 2 km (2σ) in each horizontal axis, and to within ± 200 m (2σ) in height. The initial velocity was assumed to be accurate only to within ± 20 m/s (2σ) in each horizontal axis, and ± 2 m/s (2σ) vertically. Although the H764G was barometrically aided, the height channel in the Kalman filter model does not assume this.

The data from this sortie, of duration just over $1\frac{1}{2}$ hours, were analysed using the IGMARP algorithm, with the system state estimated as a mixture of $n = 4$ Gaussian components. (In fact, Fig. 1 is based on data from the third transect of this sortie.)

First consider the initial capture phase, immediately after terrain data becomes available. The position errors during the first 45 seconds are shown in Fig. 3. In the figure the black lines represent the components of position error, based on comparing the overall mean of the 4-component Gaussian mixture with the 'true' position given by the INS/GPS blended data. The green band represents a 2σ tolerance band based on the overall standard deviation of the Gaussian mix-

ture.

It will be noted that already after the first 4 second transect has been analysed, there has been a substantial reduction the horizontal position uncertainty, particularly in the east-west axis, as well as in the vertical position uncertainty. After just 22.5 seconds (five transects), the position uncertainty is of the order of ±220 m (2σ) in each horizontal axis, ±6 m (2σ) vertically. After 45 seconds these uncertainties are further reduced to ±48 m and ±2.6 m respectively.

Now let us consider the sortie as a whole. A characteristic of the sortie was the high level of manoeuvre, with the aircraft making numerous sharp turns and climbs and dives. Consequently, for a large proportion of the flight, the radio altimeter could not provide usable data, either because the aircraft's height above ground was too great (i.e. above about 500 m), or its bank angle too great (i.e. greater than about $30°$). Fig. 4, which shows the components of position error throughout the sortie, indicates these parts of the sortie by colouring the 2σ tolerance bands red. Yellow coloration indicates that the radio altimeter was usable, but that the terrain roughness was no greater than 1%. (Terrain roughness is here measured as the RMS terrain gradient along the true path of the aircraft during a transect, with the true path measured by INS/GPS.) The remaining periods are when the radio altimeter was usable and the terrain roughness exceeded 1%; these periods together accounted for about a quarter of the sortie duration, and are indicated by green colour.

Obviously, during the 'red' periods, the horizontal and vertical position uncertainties increase continuously as a result of INS drift, though the rate of drift decreases as the sortie progresses because the INS becomes better calibrated, particularly the velocity errors. During the yellow periods, horizontal uncertainties normally continue to grow, but vertical errors are kept in check. Only during the green periods is the horizontal navigation materially assisted by the TCN data.

Fig. 4 clearly indicates the ability of the IGMARP algorithm quickly to recover accurate navigation once 'green' data comes along following a period of drift. Taking all the green periods together, the radial horizontal position error had a median value of just under 28 m.

The RMS height error during these periods works out as 5.4 m; however, it is noticeable in Fig. 4 that the height error appears to have a slowly varying bias: this is believed to be due to GPS errors influencing the measurement of 'true' height. Consequently the height accuracy is probably better than this.

8. CONCLUSIONS AND PROSPECTS

In this paper we have described the design objectives that led to the development of the IGMARP algorithm: in a nutshell, this was to have a data fusion algorithm that could cope with the ambiguities inherent in terrain-referenced navigation data, but at the same time make efficient use of the available data, in a manner compatible with the use of a Kalman filter (or MHKF) architecture for multiway integrated navigation. The paper has described the foundations of the IGMARP algorithm, and described its performance when applied to recorded data from a fast-jet sortie with some challenging characteristics. The results illustrated the algorithm's ability rapidly to acquire and reacquire accurate navigation from high initial position and velocity uncertainties.

The following are some areas for possible exploration in the future:

- The studies of the use of IGMARP with medium- and low-grade IN systems reported in [24] were based on simulated data. It would be desirable to corroborate its conclusions using real data, particularly real radio altimeter data, or—better still—laser rangefinder data.

- All studies of IGMARP applied to TRN have so far been based on terrain-contour navigation (TCN): it would be interesting to apply it to an imaging-based technique such as continuous visual navigation (CVN) [26].

- IGMARP may also be applicable to some target-tracking problems.

9. REFERENCES

[1] James Carroll et al. Vulnerability assessment of the transportation infrastructure relying on the global positioning system. Technical report, Volpe National Transportation Systems Center, August 2001. Report for US Department of Transportation.

[2] John I. R. Owen and Michael Wells. An advanced digital antenna control unit for GPS. In *Proceedings of the 2001 ION National Technical Meeting*, pages 402–7, 2001.

[3] Michael Wells and John Owen. A digital adaptive array demonstrator for GPS. In *Proceedings of NATO SET Symposium 'Emerging Military Capabilities Enabled by Advances in Navigation Sensors'*, October 2002.

[4] Neil J. Boasman and Peter Briggs. The development of an anechoic GPS test facility. In *Proceedings of the ION 58th Annual Meeting and CIGTF 21st Guidance Test Symposium*, 2002.

[5] Defense Mapping Agency. *Product Specifications for the Digital Landmass System (DLMS) Data Base*, April 1983.

[6] Robin J. Handley, Paul D. Groves, Paul McNeil, and Lawrence Dack. Future terrain referenced navigation techniques exploiting sensor synergy. In *Proceedings of GNSS 2003, The European Navigation Conference*, April 2003.
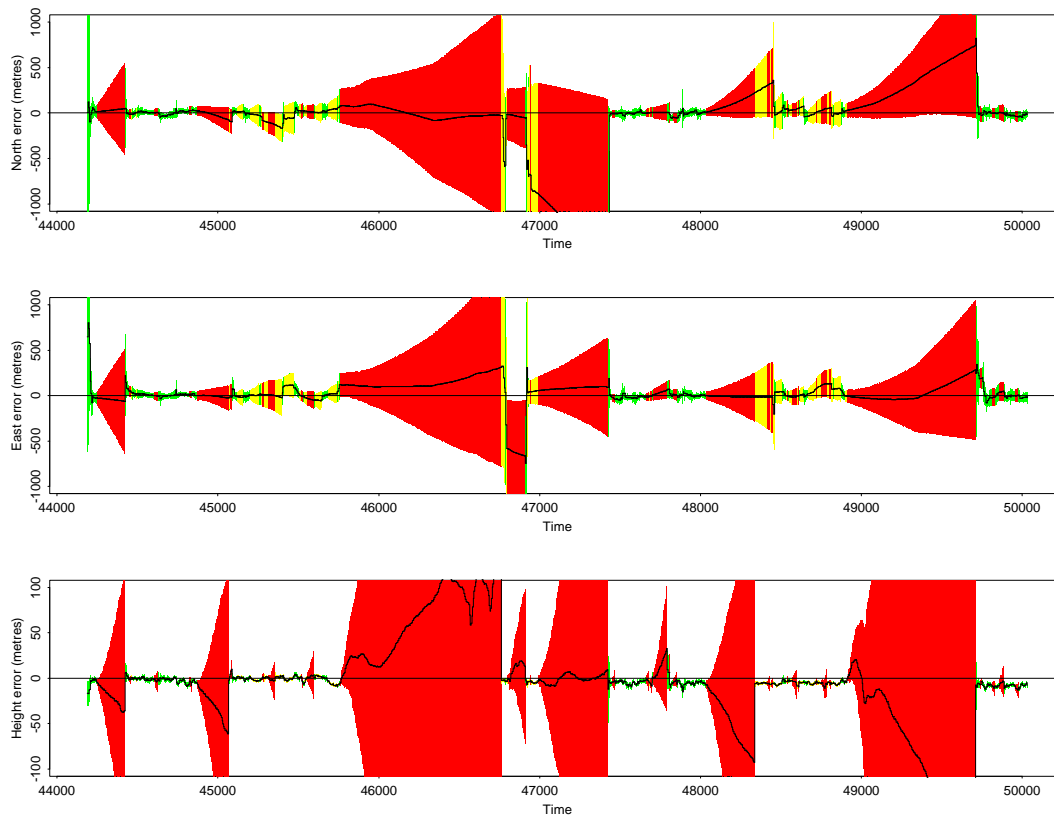
Fig. 4. Position errors throughout the sortie. Timescale is in seconds. Red colour indicates periods in which the radio altimeter was unusable. Yellow colour indicates terrain roughness $\leq 1\%$; green indicates terrain roughness $> 1\%$.

[7] Jacob Campbell, Maarten Uijt de Haag, and Frank van Graas. Terrain reference navigation using Airborne LAser SCAnner (ALASCA)—preliminary flight test results. In *ION 60th Annual Meeting*, pages 671–8, June 2004.

[8] Jürgen Metzger, Jan Wendel, and Gert F. Trommer. Comparison of different terrain referenced navigation techniques. In *Proceedings of ION 58th Annual Meeting*, pages 132–41, 2002.

[9] J. P. Golden. Terrain contour matching (TERCOM): A cruise missile guidance aid. *Proc. Int'l Soc. for Optical Eng. (SPIE)*, 238, July/Aug 1980.

[10] A. R. Runnalls. A Bayesian approach to terrain contour navigation. In *Guidance and Control Panel, 40th Symposium*. NATO Advisory Group for Aerospace Research and Development, May 1985.

[11] Alan Robins. Recent developments in the 'TERPROM' integrated navigation system. In *Proceeding of the ION 44th Annual Meeting*, June 1998.

[12] Larry D. Hostetler and Richard C. Beckman. Expanding the region of convergence for SITAN through improved modelling of terrain nonlinearities. In *NAECON*, pages 1023–30, 1979.

[13] L. D. Hostetler and R. D. Andreas. Non-linear Kalman filtering techniques for terrain-aided navigation. *IEEE Transactions on Automatic Control*, AC-28(3):315–23, March 1983.

[14] Jeff Hollowell. HELI/SITAN: a terrain referenced navigation algorithm, for helicopters. In *Proceedings of the IEEE Position Location and Navigation Symposium (PLANS)*, 1990.

[15] Pei-jun Yu, Zhe Chen, and James C. Hung. Performance evaluation of six terrain stochastic linearisation techniques for TAN. In *Proceedings of the IEEE National Aerospace and Electronics Conference (NAECON)*, 1991.

[16] Niclas Bergman. *Recursive Bayesian Estimation: Navigation and Tracking Applications*. PhD thesis, Linköping University, Sweden, 1999.

[17] R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26:195–239, 1984.

[18] Arthur Gelb, editor. *Applied Optimal Estimation*. M. I. T. Press, 1974.

[19] D. J. Salmond. Mixture reduction algorithms for target tracking in clutter. In *Signal and Data Processing of Small Targets 1990*, pages 434–45, 1990. Proceedings of SPIE Vol. 1305.

[20] Jason L. Williams. Gaussian mixture reduction for tracking multiple maneuvering targets in clutter. Master's thesis, M.S.E.E. Thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, 2003.

[21] Jason L. Williams and Peter S. Maybeck. Cost-function-based Gaussian mixture reduction. In *Sixth International Conference on Information Fusion*. ISIF, 2003.

[22] Andrew R. Runnalls. A Kullback-Leibler approach to Gaussian mixture reduction. Submitted to IEEE Transactions on Aerospace and Electronic Systems, 2006.

[23] S. Julier, J. Uhlmann, and H. Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of the American Control Conference*, pages 1628–32, 1995.

[24] Paul D. Groves, Robin J. Handley, and Andrew R. Runnalls. Optimising the integration of terrain-referenced navigation with INS and GPS. In *Proceedings of Institute of Navigation National GNSS2004*, pages 1048–59, September 2004. Long Beach, California, USA.

[25] A. R. Runnalls, P. D. Groves, and R. J. Handley. Terrain-referenced navigation using the IGMAP data fusion algorithm. In *Institute of Navigation 61st Annual Meeting*, pages 976–87, June 2005.

[26] Robin J. Handley, Lawrence Dack, and Paul McNeil. Flight trials of the continuous visual navigation system. In *Proceedings of the ION National Technical Meeting*, pages 185–92, 2001.