Robotics: Science and Systems 2006
Philadelphia, PA, USA, August 16-19, 2006

The Iterated Sigma Point Kalman Filter with Applications to Long Range Stereo

Gabe Sibley[†,‡], Gaurav Sukhatme[†] and Larry Matthies[‡]

[†]Robotic Embedded Systems Laboratory,
University of Southern California,
Los Angeles, CA 90089
gsibley|gaurav@usc.edu

[‡]Jet Propulsion Laboratory,
California Institute of Technology,
Pasadena, CA 91109
lhm@helios.jpl.nasa.gov

***Abstract*— This paper investigates the use of *statistical linearization* to improve *iterative* non-linear least squares estimators. In particular, we look at improving long range stereo by filtering feature tracks from sequences of stereo pairs. A novel filter called the Iterated Sigma Point Kalman Filter (ISPKF) is developed from first principles; this filter is shown to achieve superior performance in terms of efficiency and accuracy when compared to the Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), and Gauss-Newton filter. We also compare the ISPKF to the optimal Batch filter and to a Gauss-Newton Smoothing filter. For the long range stereo problem the ISPKF comes closest to matching the performance of the full batch MLE estimator. Further, the ISPKF is demonstrated on real data in the context of modeling environment structure from long range stereo data.**

I. Introduction

This paper investigates the *Iterated* Sigma Point Kalman Filter (ISPKF), which is a principled extension to statistical linearization methods such as the Unscented Kalman Filter (UKF), Central Difference Kalman Filter (CDKF), and Sigma Point Kalman Filter (SPKF) [6, 8, 18, 19]. In the seminal work [18] the Iterated Sigma Point Kalman Filter was discussed as future work as an "ad-hoc improvement" to non-iterated methods. In contrast, we highlight the fundamental importance of iteration in solving non-linear least squares problems. To this end we develop the ISPKF directly from the underlying probability density function, grounding it in non-linear optimization theory and Newton's Method. For non-linear problems we emphasize that the measurement update for methods that do not iterate, such as the EKF, UKF or SPKF, cannot be expected to achieve the Maximum Likelihood solution, whereas iterative methods are *provably* convergent and have a rich convergence theory [3].

After deriving the ISPKF, we compare it to a number of techniques including an optimal batch non-linear least squares solution, a Gauss-Newton filter, an Extended Kalman Filter, (as in [13]) a Gauss-Newton Smoother, and an Unscented Kalman Filter. We find that the ISPKF both converges faster and with greater accuracy when compared to the other estimators.

We are interested in improving the range resolution of stereo to the point where we can accurately estimate depth from disparities on the order of one pixel or less. Our approach is to filter measurement sequences of image features over time. Temporal filtering of such measurements can improve depth estimation, which is important for mobile robots. For instance, better stereo range resolution increases a robot's ability to perform tasks such as navigation, long range path planning, obstacle avoidance, mapping and localization and high-speed driving.

Stereo is an interesting problem on which to apply the ISPKF because it is non-linear, suffers from inherent statistical bias, and because the true range probability distribution is non-Gaussian. When using traditional filters, such as the Extended Kalman Filter, these issues can cause *apparent divergence* (e.g. convergence to the wrong value). While the difficulties with bias in stereo can be largely overcome without recourse to statistical linearization, [13] we show that using statistical linearization results in significant improvements. Further, we show that the iterative nature of the new estimator greatly improves upon results of non-iterative statistical linearization methods like the UKF.

Finally, we demonstrate the ISPKF's capabilities in the real-world context of building accurate environment models from long range stereo data.

II. The Iterated Sigma Point Kalman Filter

In this section we derive the measurement update equations for the Iterated Sigma Point Kalman Filter. For completeness we start from first principles with the joint probability density function describing our problem, work through Bayes's rule, Newton's method, Gauss-Newton iterated non-linear least squares, the Iterated Extended Kalman Filter, linearized error propagation, statistically linearized error propagation, and finally arrive at the ISPKF.

We are interested in estimating the value of an unknown parameter vector $\mathbf{x}$ from noisy measurements of related quantities $\mathbf{z}$. We consider the estimate vector $\hat{\mathbf{x}}$ and observation vector $\mathbf{z}$ to be independent realizations of multivariate normal distributions

$$\mathbf{z} \sim N(h(\mathbf{x}),\, \mathbf{R})$$

$$\hat{\mathbf{x}} \sim N(\mathbf{x}, \mathbf{P})$$

where $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a non-linear measurement function relating the state to the measurements; and $\mathbf{R}$ and $\mathbf{P}$

are the measurement and state error covariance matrices, respectively.

From the Bayesian perspective we have $P(\mathbf{x}|\mathbf{z}) = \eta P(\mathbf{z}|\mathbf{x})P(\mathbf{x})$ where $\eta = P(\mathbf{z})^{-1}$ amounts to a normalizing factor, $P(\mathbf{x})$ is the prior, and $P(\mathbf{z}|\mathbf{x})$ is the likelihood. Note that in this paper we ignore any system dynamics in the process model that might be acting on the prior and focus on the *distinct* problem of the measurement update. The task of *maximum a posterior* (MAP) estimation is to find the $\mathbf{x}$ that maximizes the scalar quantity $P(\mathbf{z}|\mathbf{x})P(\mathbf{x})$. Recall that the distributions on $\mathbf{x}$ and $\mathbf{z}$ are normal,

$$P(\mathbf{z}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m|\mathbf{R}|}} \exp\left(-\frac{1}{2}(\mathbf{z}-h(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{z}-h(\mathbf{x}))\right)$$

$$P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{P}|}} \exp\left(-\frac{1}{2}(\hat{\mathbf{x}}-\mathbf{x})^T\mathbf{P}^{-1}(\hat{\mathbf{x}}-\mathbf{x})\right)$$

where $|\cdot|$ is the determinant. The solution that maximizes $P(\mathbf{z}|\mathbf{x})P(\mathbf{x})$ is equivalent to minimizing its negative log, which reduces to the quadratic,

$$\ell(\mathbf{x}) = \frac{1}{2}\Big[(\mathbf{z}-h(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{z}-h(\mathbf{x})) + (\hat{\mathbf{x}}-\mathbf{x})^T\mathbf{P}^{-1}(\hat{\mathbf{x}}-\mathbf{x})\Big] + k$$

where k is a constant which may be dropped. An algebraically equivalent way to look at MAP is to consider the prior estimate as a pseudo-observation and then to write a new observation vector and function,

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z} \\ \hat{\mathbf{x}} \end{bmatrix}, \quad g(\mathbf{x}) = \begin{bmatrix} h(\mathbf{x}) \\ \mathbf{x} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{R} & 0 \\ 0 & \mathbf{P} \end{bmatrix}$$

which gives (ignoring k)

$$\ell(\mathbf{x}) \approx \frac{1}{2}[(\mathbf{Z}-g(\mathbf{x}))\mathbf{C}^{-1}(\mathbf{Z}-g(\mathbf{x}))]. \tag{1}$$

If we let $\mathbf{S}^T\mathbf{S} = \mathbf{C}^{-1}$ and

$$f(\mathbf{x}) = \mathbf{S}(\mathbf{Z}-g(\mathbf{x})) \tag{2}$$

then (1) is clearly a non-linear least squares problem of the form

$$\ell(\mathbf{x}) = \frac{1}{2}||f(\mathbf{x})||^2.$$

Newton's solution to such optimization problems is the iterative sequence

$$\mathbf{x}_{i+1} = \mathbf{x}_i - (\nabla^2\ell(\mathbf{x}_i))^{-1}\nabla\ell(\mathbf{x}_i). \tag{3}$$

For small residual problems a useful approximation to (3) is the Gauss-Newton method, which approximates the Hessian $\nabla^2\ell(\mathbf{x}_i)$ by $f'(\mathbf{x}_i)^Tf'(\mathbf{x}_i)$. Thus, since the gradient of (2) is $\nabla\ell(\mathbf{x}_i) = f'(\mathbf{x}_i)^Tf(\mathbf{x}_i)$, the Gauss-Newton method defines the sequence of iterates [3]

$$\mathbf{x}_{i+1} = \mathbf{x}_i - (f'(\mathbf{x}_i)^Tf'(\mathbf{x}_i))^{-1}f'(\mathbf{x}_i)^Tf(\mathbf{x}_i) \tag{4}$$

where $f'(\mathbf{x}_i)$ is the Jacobian of (2). Noting that $f'(\mathbf{x}_i) = -\mathbf{S}\mathbf{G}_i$ where $\mathbf{G}_i$ is the Jacobian of $g(\mathbf{x}_i)$, (4) becomes

$$\mathbf{x}_{i+1} = (\mathbf{G}_i^T\mathbf{C}^{-1}\mathbf{G}_i)^{-1}\mathbf{G}_i^T\mathbf{C}^{-1}(\mathbf{Z}-g(\mathbf{x}_i)+\mathbf{G}_i\mathbf{x}_i) \tag{5}$$

which (when $g(\mathbf{x})$ is first approximated by its first order expansion) is also the *normal equation* solution to (1). This sequence is iterated to convergence. After convergence the updated covariance $\hat{\mathbf{P}}_{k+1}$ is approximated using

$$\hat{\mathbf{P}}_{k+1} = (\mathbf{G}_i^T\mathbf{C}^{-1}\mathbf{G}_i)^{-1} = (\mathbf{H}_i^T\mathbf{R}^{-1}\mathbf{H}_i + \mathbf{P}^{-1})^{-1} \tag{6}$$

where $\mathbf{H}_i$ is the Jacobian of $h(\mathbf{x}_i)$ (for notational simplicity the k is dropped on the current covariance $\mathbf{P}$). At this point we have derived the Gauss-Newton measurement update and made explicit the connection to Newton's method. In contrast to one shot linearization methods like the Extended Kalman Filter or the Unscented Kalman Filter, the Gauss-Newton method is locally convergent to the MAP estimate for near zero-residual problems [3]. In fact the discrete EKF is algebraically equivalent to a **single** iteration of the Gauss-Newton method [1]. The Gauss-Newton method is simply on the other side of the *matrix inversion lemma*. To see this we will need two forms of the matrix inversion lemma: 1)

$$(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})^{-1}\mathbf{H}^T\mathbf{R}^{-1} = \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}$$

and 2)

$$(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})^{-1} = \mathbf{P} - \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}$$

Expanding (5) and making use of the matrix inversion lemma we obtain, $\mathbf{x}_{i+1}$

$$= \mathbf{x}_i + (\mathbf{H}_i^T\mathbf{R}^{-1}\mathbf{H}_i + \mathbf{P}^{-1})^{-1}\big[\mathbf{H}_i^T\mathbf{R}^{-1}(\mathbf{z}-h(\mathbf{x}_i)) + \mathbf{P}^{-1}(\hat{\mathbf{x}}_k - \mathbf{x}_i)\big]$$

$$= \hat{\mathbf{x}}_k + \mathbf{P}\mathbf{H}_i^T(\mathbf{H}_i\mathbf{P}\mathbf{H}_i^T + \mathbf{R})^{-1}(\mathbf{z}-h(\mathbf{x}_i)) - \mathbf{P}\mathbf{H}_i^T(\mathbf{H}_i\mathbf{P}\mathbf{H}_i^T + \mathbf{R})^{-1}\mathbf{H}_i\mathbf{P}\mathbf{P}^{-1}(\hat{\mathbf{x}}_k - \mathbf{x}_i) \tag{7}$$

$$= \hat{\mathbf{x}}_k + \mathbf{P}\mathbf{H}_i^T(\mathbf{H}_i\mathbf{P}\mathbf{H}_i^T + \mathbf{R})^{-1}\big(\mathbf{z} - h(\mathbf{x}_i) - \mathbf{H}_i(\hat{\mathbf{x}}_k - \mathbf{x}_i)\big)$$

$$= \hat{\mathbf{x}}_k + \mathbf{K}\big(\mathbf{z} - h(\mathbf{x}_i) - \mathbf{H}_i(\hat{\mathbf{x}}_k - \mathbf{x}_i)\big) \tag{8}$$

which is the Iterated Extended Kalman Filter measurement update; $\mathbf{K} = \mathbf{P}\mathbf{H}_i^T(\mathbf{H}_i\mathbf{P}\mathbf{H}_i^T + \mathbf{R})^{-1}$ is the Kalman gain. Once the sequence over i is iterated to convergence the next time estimate $\hat{\mathbf{x}}_{k+1}$ is set to $\mathbf{x}_i$. Similarly, after convergence, the covariance update (6) can be manipulated into the familiar EKF form $\hat{\mathbf{P}}_{k+1} = (\mathbf{I} - \mathbf{K}\mathbf{H}_i)\mathbf{P}$. On the first iteration $\mathbf{x}_i = \hat{\mathbf{x}}$, $i = 1$, and (8) reduces to the Extended Kalman Filter. The point here is that iteration is not just an extension to the EKF; on the contrary, for near zero-residual problems, the EKF is a sub-optimal approximation of the provably convergent Gauss-Newton method [3].

A. Linear Error Propagation

Notice that in (8) there are numerous linearized error propagation terms like $\mathbf{P}\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}\mathbf{H}^T$. If we have a linear function $\mathbf{z} = \mathbf{H}\mathbf{x}$ like the linearized equations above,

then using the linearity of expectation we can propagate covariance in $\mathbf{x}$ through to $\mathbf{z}$

$$\begin{aligned}\mathrm{cov}(\mathbf{z}) &= E[(\hat{\mathbf{z}}-\mathbf{z})(\hat{\mathbf{z}}-\mathbf{z})^T] \\ &= E[(\mathbf{H}\hat{\mathbf{x}}-\mathbf{H}\mathbf{x})(\mathbf{H}\hat{\mathbf{x}}-\mathbf{H}\mathbf{x})^T] \\ &= \mathbf{H}\mathrm{cov}(\mathbf{x})\mathbf{H}^T \\ &= \mathbf{H}\mathbf{P}\mathbf{H}^T\end{aligned}$$

and the cross-covariance is

$$\begin{aligned}\mathrm{cov}(\mathbf{x},\mathbf{z}) &= E[(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{z}}-\mathbf{z})^T] \\ &= E[(\hat{\mathbf{x}}-\mathbf{x})(\mathbf{H}\hat{\mathbf{x}}-\mathbf{H}\mathbf{x})^T] \\ &= \mathrm{cov}(\mathbf{x})\mathbf{H}^T \\ &= \mathbf{P}\mathbf{H}^T\end{aligned}$$

Note also that $\mathrm{cov}(\mathbf{z},\mathbf{x}) = \mathrm{cov}(\mathbf{x},\mathbf{z})^T = \mathbf{H}\mathrm{cov}(\mathbf{x}) = \mathbf{H}\mathbf{P}$. With these in mind the IEKF update in (7) and (8) is

$$\begin{aligned}\mathbf{x}_{i+1} = \hat{\mathbf{x}} &+ \mathrm{cov}(\mathbf{x},\mathbf{z})\big(\mathrm{cov}(\mathbf{z})+\mathbf{R}\big)^{-1} \\ &\left(\mathbf{z}-h(\mathbf{x}_i)-\mathrm{cov}(\mathbf{x},\mathbf{z})^T\mathbf{P}^{-1}(\hat{\mathbf{x}}-\mathbf{x}_i)\right) \quad (9)\end{aligned}$$

This equation is based on linearization of the observation function h, and linearized error propagation of the random variables $\hat{\mathbf{x}}$ and $\mathbf{z}$.

A more accurate method of error propagation could be employed, such as statistically linearized error propagation [5]. There are various methods that use this technique, such as the UKF, SPKF, CDKF and LRKF [6, 8, 9, 18, 19]. However, these methods do not iterate the sequence to convergence, and typically end their analysis with comparison to the Extended Kalman Filter. This practice misses the direct and intuitive relationship between the Iterated Extended Kalman Filter and Newton's method. One would never iterate the Newton method just once. Hence, it is important to extend the statistical linearized filters so that iteration is possible.

B. Statistically Linearized Error Propagation

Statistical linear error propagation is generally more accurate than error propagation via first order Taylor series expansion [5, 7, 18]. The idea behind statistical linear error propagation is simple: select from the distribution on $\mathbf{x}$ a set of regression points $\mathcal{X}$, such that they maintain certain properties of the input distribution (such as the mean and covariance). These points are then mapped through h

$$\mathcal{Z}_i = h(\mathcal{X}_i)$$

creating a set of transformed regression points, from which the mean and covariance are computed. Recent research has led to a number of filters that employ statistical linearization [6, 7, 19], which can all be understood as examples of the so-called Sigma Point approach [18]. In this method the Sigma Points (regression points) are selected to lie on the principle component axes of the input covariance, plus one extra point for the mean of the distribution

$$\begin{aligned}&\mathcal{X}_0 = \hat{\mathbf{x}} \\ &\mathcal{X}_i = \hat{\mathbf{x}} + (\sqrt{(L+\lambda)\mathbf{P}})_i \quad i = 1,\cdots,L \\ &\mathcal{X}_i = \hat{\mathbf{x}} - (\sqrt{(L+\lambda)\mathbf{P}})_i \quad i = L+1,\cdots,2L\end{aligned}$$

$$\begin{aligned}&w_0^{(m)} = \tfrac{\lambda}{L+\lambda} \\ &w_0^{(c)} = \tfrac{\lambda}{L+\lambda} + (1-\alpha^2+\beta) \\ &w_i^{(m)} = w_i^{(c)} = \tfrac{\lambda}{L+\lambda} \qquad i = 1,\cdots,2L\end{aligned} \quad (10)$$

where $\lambda = \alpha^2(L+\kappa) - L$, L is the dimension of the state space, and the parameters α, β, and κ are tuning parameters[1].

The mean and covariance are computed from the Sigma points $\mathcal{X}$ and transformed Sigma points $\mathcal{Z}$ using

$$\begin{aligned}\bar{\mathcal{Z}} &= \sum_{i=0}^{2L} w_i^{(m)}\mathcal{Z}_i \\ \mathrm{cov}(\mathcal{Z}) &= \sum_{i=0}^{2L} w_i^{(c)}(\mathcal{Z}_i-\bar{\mathcal{Z}})(\mathcal{Z}_i-\bar{\mathcal{Z}})^T \\ \mathrm{cov}(\mathcal{X},\mathcal{Z}) &= \sum_{i=0}^{2L} w_i^{(c)}(\mathcal{X}_i-\hat{\mathbf{x}})(\mathcal{Z}_i-\bar{\mathcal{Z}})^T \quad (11)\end{aligned}$$

Selecting the regression points in this way is the key insight behind the Sigma Point approach. After pushing the Sigma points through h, we compute the first and second order statistics of the transformed points.

C. The Iterated Sigma Point Kalman Filter

From (7) we see that in (9) we have expressed the Iterated Extended Kalman Filter equations so that they do not depend on the sensor model Jacobian, but are instead expressed in terms of propagated error terms. This allows replacing the linear error propagation terms with statistical error propagation terms. By replacing the linearized error propagation terms in (9) with statistically linearized error propagation terms (11) we get

$$\begin{aligned}\mathbf{x}_{i+1} = \hat{\mathbf{x}} &+ \mathrm{cov}(\mathcal{X},\mathcal{Z})\big(\mathrm{cov}(\mathcal{Z})+\mathbf{R}\big)^{-1} \\ &\left(\mathbf{z}-h(\mathbf{x}_i)-\mathrm{cov}(\mathcal{X},\mathcal{Z})^T\mathbf{P}^{-1}(\hat{\mathbf{x}}_k-\mathbf{x}_i)\right) \quad (12)\end{aligned}$$

$$\mathbf{P}_{k+1} = \mathbf{P}_k - \mathrm{cov}(\mathcal{X},\mathcal{Z})\big(\mathrm{cov}(\mathcal{Z})+\mathbf{R}\big)^{-1}\mathrm{cov}(\mathcal{X},\mathcal{Z})^T \quad (13)$$

The Iterated Sigma Point Kalman Filter measurement update is defined by (10), (11), (12) and (13), which completes the derivation. This filter utilizes the benefit of statistical linearization *and* iteration.

III. Developed Example: Long Range Stereo

We have applied the ISPKF to the problem of filtering a sequence of feature measurements from a Stereo Vision system. Stereo is an interesting problem because it exposes

[1]For more on setting these see [18]. In the examples here $\alpha = 10^{-4}$, $\beta = 2$, and $\kappa = 0$, which is appropriate for Gaussian distributions [19]. The parameter α controls the distance of the Sigma Points away from $\hat{\mathbf{x}}$ along the principle components of $\mathbf{P}$; β can be used to include information about higher order moments; and $\kappa \geq 0$ helps guarantee the positive-semi-definiteness of the covariance matrix.

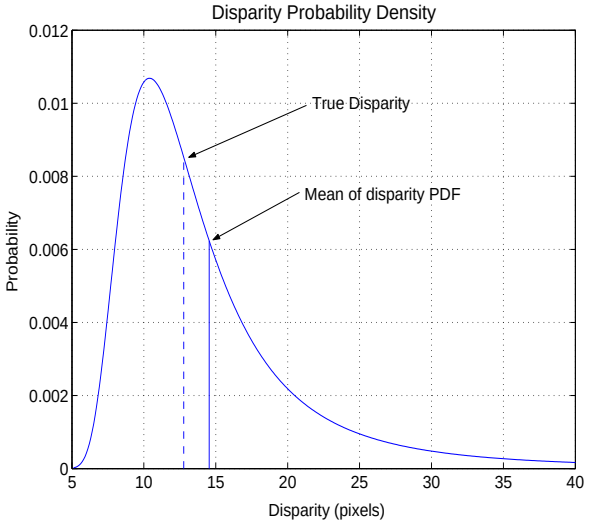

Fig. 1. Probability density function for disparity derived from a 3D landmark at 4m with a range standard deviation of 1.5m. The p.d.f. is clearly non-Gaussian, non-symmetric, and has a heavy tail. The difference between the p.d.f. mean and the true mean can lead to bias and apparent divergence when filtering stereo measurements.

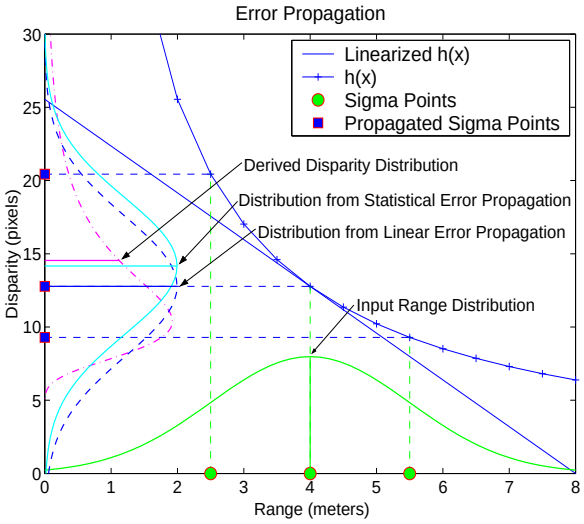

Fig. 2. Error propagation from the state space to the measurement space for stereo ranging. Propagation of this type takes place when we compute $\bar{\mathcal{Z}}$ and $\mathrm{cov}(\mathcal{Z})$ in (11) or $\mathrm{cov}(\mathbf{z})$ via $\mathbf{HPH}^T$. Statistical error propagation captures the mean of the analytical p.d.f. more accurately, which is biased long due to the heavy tail.

the inability of linear error propagation to faithfully transform a probability density through a non-linear function. This is apparent in Fig. 1 which shows the shape of the *analytically propagated* disparity probability density function – it is not Gaussian.

Consider a standard camera model that maps a 3D point $\mathbf{x} = [x,\ y,\ z]^T$ to pixel locations in an axis aligned stereo rig. Standard perspective projection gives

$$\mathbf{z} = \begin{bmatrix} u_l \\ v_l \\ u_r \\ v_r \end{bmatrix} = h(\mathbf{x}) = \begin{bmatrix} fx/z \\ fy/z \\ f(b-x)/z \\ fy/z \end{bmatrix}$$

where $[u_l,\ v_l,\ u_r,\ v_r]^T$ is the measurement of the pixels in the left and right images, b is the baseline, and f is the focal length[2]. For rectified imagery, the disparity, $d = u_l - u_r$, is inversely related to range, $d = s(z) = bf/z$, and clearly range is inversely related to disparity $z = s(z)^{-1} = bf/d$. If both the left pixel $[u_l,\ v_l]^T$ and right pixel $[u_r,\ v_r]^T$ are realizations of normal distributions, then the measured disparity will also be normally distributed. Analogously, notice that in (8) that the covariance of the measurement distribution, derived from the state covariance $\mathbf{P}$, is represented by the linear approximation $\mathbf{H}_i\mathbf{P}\mathbf{H}_i^T$. To understand the impact of this approximated measurement distribution better we use the range p.d.f., $f_z(z)$, which we have modeled as Gaussian, to derive analytically the disparity p.d.f., $f_d(d)$. This gives the *derived distribution* [2],

[2]Throughout this portion of the paper we use linear camera models with a resolution of 512×384 pixels, a horizontal FOV of 68.12° and vertical FOV of 51.37°. Unless otherwise stated, we use uncorrelated image measurement noise with a standard deviation of 0.25 pixels.

$$f_d(d) = f_z(s(d)^{-1})\left|\frac{\partial s(d)^{-1}}{\partial d}\right|$$

which expands to

$$f_d(d) = \frac{fb}{\sqrt{2\pi}\sigma_z d^2} exp\left(\frac{-(fb/d - \mu_z)^2}{2\sigma_z^2}\right) \qquad (14)$$

Where σ_z is the range standard deviation and μ_z is the range mean. This non-Gaussian p.d.f. is shown in Fig. 1 and on the y-axis of Fig. 2. Clearly, this predicted measurement disparity p.d.f. is non-symmetric and has a long tail (which causes bias in stereo). Using statistical linearization to compute these distributions is generally more accurate that linearized error propagation when the function h is non-linear [5, 18]. This is apparent in Fig. 2 which shows that statistical error propagation captures the mean of the disparity probability density function more accurately than linear error propagation.

A. Comparison of methods

In this section we derive and compare a number of methods that estimate the 3D position of a distant feature given stereo measurements. Consider some general stereo triangulation function $r : \mathbb{R}^4 \rightarrow \mathbb{R}^3$

$$r(\mathbf{z}) = \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad (15)$$

All the filters we compare are initialized with $r(\mathbf{z}_1)$, the triangulated stereo value from the first measurement. The initial state error covariance matrix $\mathbf{P}$ is found via error propagation of image errors

$$\mathbf{P} = \frac{\partial r}{\partial \mathbf{z}}\mathbf{R}\frac{\partial r}{\partial \mathbf{z}}^T \quad , \quad \mathbf{R} = \begin{bmatrix} \mathbf{\Sigma}_l & 0 \\ 0 & \mathbf{\Sigma}_r \end{bmatrix} \qquad (16)$$

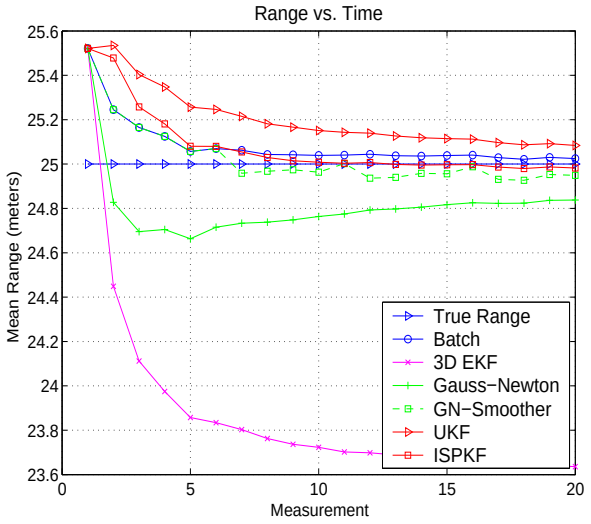

Fig. 3. Filter convergence for a sequence of 20 measurements of a feature at 25m averaged over 1000 trials. Image noise is modeled as uncorrelated standard deviation of 0.25 pixels.

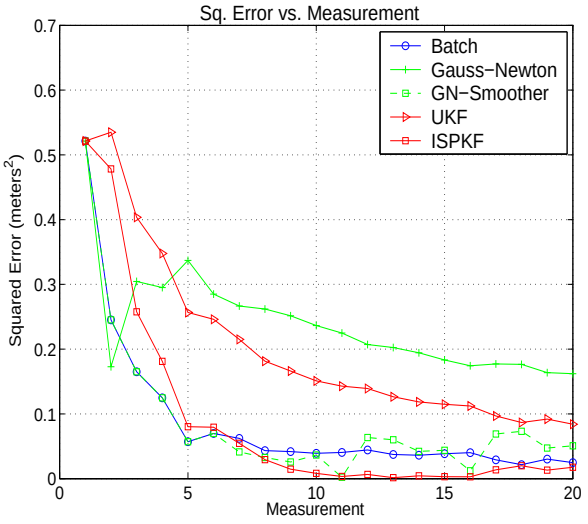

Fig. 4. Squared Filter Error vs. Measurement number for a sequence of 20 measurements of a feature at 25m averaged over 1000 trials. The improvement of the UKF over the Gauss-Newton method can be attributed to statistical error propagation. The improvement of the ISPKF over the UKF can be attributed to Iteration. Notice that filters incorporating prior information modeled as Gaussian (all but the batch filter) suffer from initialization problems – either over confidence or under confidence. This is due to modeling range uncertainty as Gaussian, which is incorrect.

where $\mathbf{\Sigma}_l$ and $\mathbf{\Sigma}_r$ are the left and right 2×2 measurement error covariance matrices, respectively.

We compare the following methods: the optimal Batch non-linear least squares solution (III-A.1), an Iterated EKF (III-A.3), a Gauss-Newton Smoother (III-A.2), an EKF with 3D measurements (III-A.4), the UKF (III-A.5), and finally the Iterated Sigma Point Kalman Filter (III-A.6).

It is well known that stereo triangulation suffers from a statistical bias, and recently solutions for bias removal have been proposed [12, 13] However, in the context of temporal filtering bias removal is not enough – we must also make sure the underlying estimation machinery is robust to the non-linear nature of the problem at hand, or else the estimator may exhibit apparent divergence, as we will see with the "3D EKF" described below.

1) Batch Least Squares: If we have measurements $\mathbf{z}_j$ for $j = 1, 2, ..., k$ up to time k then we can lump all the measurements together

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_k \end{bmatrix}, h = \begin{bmatrix} h_1(\mathbf{x}) \\ h_2(\mathbf{x}) \\ \vdots \\ h_k(\mathbf{x}) \end{bmatrix},$$

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & 0 & \dots & 0 \\ \vdots & \mathbf{R}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{R}_k \end{bmatrix}$$

Where the j^{th} observation function h_j is

$$h_j(\mathbf{x}) = \begin{bmatrix} h_{left}(\mathbf{x}) \\ h_{right}(\mathbf{x}) \end{bmatrix} \tag{17}$$

and $h_{left} : \mathbb{R}^3 \to \mathbb{R}^2$ and $h_{right} : \mathbb{R}^3 \to \mathbb{R}^2$ are the left and right camera projection functions. Depending on the camera models in use, h_{left} and h_{right} can be formulated in a variety of ways [17, 20]. We only require that these functions and their first derivatives are available, and otherwise leave them unspecified.

From this we can write a large batch least squares problem

$$\ell_{\mathbf{z}}(\mathbf{x}) \approx \frac{1}{2}(\mathbf{z} - h(\mathbf{x}))\mathbf{R}^{-1}(\mathbf{z} - h(\mathbf{x})) \tag{18}$$

the solution to which is the *maximum likelihood estimate* (MLE) given all measurements at all times. The Gauss-Newton Iteration for this problem is

$$\mathbf{x}_{i+1} = (\mathbf{H}_i^T\mathbf{R}^{-1}\mathbf{H}_i)^{-1}\mathbf{H}_i^T\mathbf{R}^{-1}\big(\mathbf{z} - h(\mathbf{x}_i) + \mathbf{H}_i\mathbf{x}_i\big).$$

As usual after convergence the covariance is approximated as

$$\hat{\mathbf{P}}_{k+1} = (\mathbf{H}_i^T\mathbf{R}^{-1}\mathbf{H}_i)^{-1}.$$

Notice that for identical and independently distributed Gaussian measurements this is exactly the inverse Fisher Information matrix, and hence the batch estimator matches the Cramer-Rao lower bound. In the maximum likelihood sense this is the *best* unbiased estimate possible. In Fig. 3 and Fig. 4 we expect this estimator to perform the best.

2) Gauss-Newton Smoother: Including a prior in the batch estimator results in a formulation identical to the Gauss-Newton recursive filter in (5) except that the measurement vector $\mathbf{z}$ is much larger (it contains all measurements from all times). This batch/recursive filter can use any number of past measurements, such as a sliding window over the previous m measurements. The longer the window the closer the results will approximate the MLE batch solution. A history window of $m = 1$ is equivalent to the Gauss-Newton recursive filter (5); a window with all measurements is equivalent to the batch method in Sec. (III-A.1). Results for smoothing with a window of

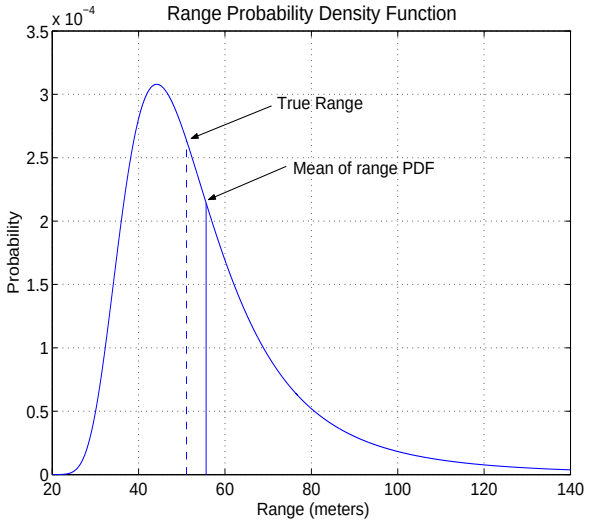

Fig. 5. Range probability density function analytically derived through the stereo triangulation equations (inverse of the sensor model). Propagating measurements that are normally distributed results in a non-Gaussian range distribution. Estimators that model the feature state as a 3D Gaussian will suffer from this mis-representation. The filter will likely exhibit either bias or divergence. Both iteration and statistical linearization can help overcome these effects. *All* the filters except the batch-optimal filter in Fig. 3 suffer from this mis-representation so some extent.

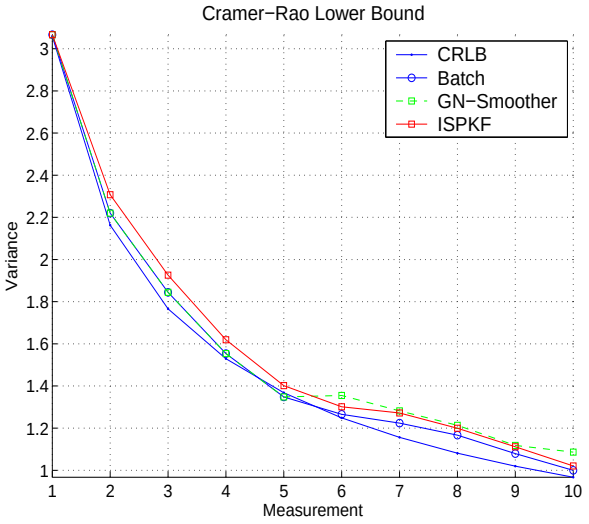

Fig. 6. Cramer Rao Lower Bound for the ISPKF, the Gauss-Newton Smoother, and the optimal Batch filter. As expected the Batch filter tracks the CRLB, and the Gauss-Newton Smoother tracks the CRLB for the first 5 steps. However, after 5 steps the the Smoother starts rolling old measurements into a prior, and the ISPKF starts to out perform the Smoother. As a recursive filter the ISPKF is nearly as efficient as the Batch filter.

$m = 5$ are shown in Fig. 3 and Fig. 4. As expected, the Smoother matches the Batch estimator up until the time window starts dropping older measurements at $m = 5$.

3) Iterated Extended Kalman Filter: An Iterated Extended Kalman Filter is algebraically equivalent to a Gauss-Newton Smoother with a time window of $m = 1$. The results for this filter are plotted in Fig. 3 and Fig. 4. Initially the filter exhibits apparent divergence along the lines of the 3D EKF. This can be attributed to the inaccuracy of modeling range uncertainty as a Gaussian, when the true range distribution is biased long, as shown in Fig. 5.

4) EKF with 3D measurements: In order to see the deleterious effects linearized error propagation can have on a filter, consider a Kalman Filter with 3D observations, $\mathbf{z}_{3D}$, as output from stereo triangulation. The state estimate and observation are independent realizations of multivariate normal distributions: $\mathbf{z}_{3D} \sim N(r(\mathbf{z}), \mathbf{R}_{3D})$ and $\hat{\mathbf{x}} \sim N(\mathbf{x}, \hat{\mathbf{P}})$ where $\mathbf{R}_{3D}$ and $\hat{\mathbf{P}}$ are the measurement and state error covariance matrices, respectively. The measurement error covariance matrix $\mathbf{R}_{3D}$ is found via error propagation of image errors just like (16).

Notice that using this formulation introduces an extra linear error propagation into the equations. This has serious implications for filter performance – indeed from Fig. 3 we see that the 3D EKF converges to a biased value. Filtering sequences of 3-D measurements from stereo leads to biased range estimates when the uncertainty of each 3-D measurement is modeled by standard linear error propagation techniques; this emphasizes the fact that linear error propagation does not faithfully represent the transformed distribution.

5) Unscented Kalman Filter: The Unscented Kalman Filter is equivalent to the Iterated Sigma Point Filter where only one iteration is carried out during the measurement update. This filter is shown in Fig. 3 and Fig. 4.

6) Iterated Sigma Point Kalman Filter: Applying the Iterated Sigma Point Filter developed in Sec. III to the stereo problem is straightforward. Like the other filters the state is initialized via stereo triangulation and error propagation. The iterated measurement update typically converges within 9 iterations. As can be seen in Fig. 3 and Fig. 4 the ISPKF comes closest to matching the batch solution, followed by the Gauss-Newton Smoother.

IV. Estimator Efficiency

It is also important to address the efficiency of the ISPKF; that is, how well it approximates a minimal variance estimate of the parameters. The information inequality, $\mathrm{cov}_{\mathbf{x}}(\mathbf{x}) \geq \mathcal{I}_{\mathbf{z}}(\mathbf{x})^{-1}$, defines such a bound, which is called Cramer-Rao lower bound [2]. Here the Fisher information matrix $\mathcal{I}_{\mathbf{z}}(\mathbf{x})$ is given by the symmetric matrix whose i^{th},j^{th} element is the covariance between first partial derivatives of the log-likelihood function (18),

$$\mathcal{I}_{\mathbf{z}}(\mathbf{x})_{,i,j} = \mathrm{cov}_{\mathbf{x}}\left(\frac{\partial \ell_{\mathbf{z}}}{\partial \mathbf{x}_i}, \frac{\partial \ell_{\mathbf{z}}}{\partial \mathbf{x}_j}\right) \tag{19}$$

which, for a multivariate normal distribution, reduces to [15, 21]

$$\mathcal{I}_{\mathbf{z}}(\mathbf{x})_{,i,j} = \frac{\partial h}{\partial \mathbf{x}_i}^T \mathbf{R}^{-1} \frac{\partial h}{\partial \mathbf{x}_j}.$$

For n independent identically distributed measurements the Fisher information is simply $n\mathcal{I}$. Qualitatively, an estimator that comes close to the CRLB is efficient. Fig. 6 shows comparison between the ISPKF, the Gauss-Newton smoother and the CRLB, and demonstrates that the ISPKF

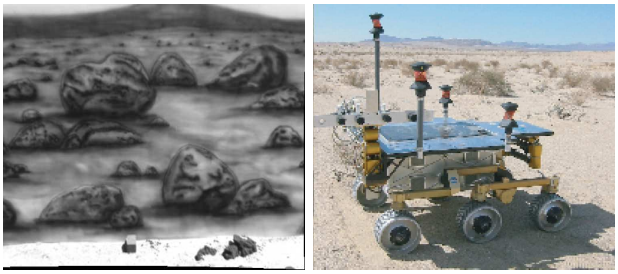

Fig. 7. Sample image (left) from experiments using the FIDO Mars Rover (right). The wall is approximately 6m from the robot. Over 20 frames there were 8005 stereo measurements of 899 different SIFT features. FIDO uses rectified camera models with a resolution of 512×384 pixels, a horizontal FOV of 37.22°, a vertical FOV of 29.53° and a baseline of ∼20cm.

Fig. 8. Sample image (left) from experiments using a stereo setup configured analogously to the LAGR robot (right). The wall is approximately 20m from the robot. This sequence uses rectified camera models with a resolution of 512×384 pixels, a horizontal FOV of 54.08°, a vertical FOV of 40.87° and baseline of ∼18cm.

is efficient. Note that this is also an indication of consistency, since the estimator is not over-confident (e.g. it does not dip below the CRLB).

V. Long Range Stereo Experiments

To verify that the ISPKF is indeed converging to a reasonable estimate we have performed the following experiments. We took a sequence of 20 stereo images of a highly textured large flat wall, triangulated and tracked SIFT features on this wall [10]. We then fit a plane to all of the triangulated 3D points from all of the frames using RANSAC with an inlier threshold of 5cm [4]. If the stereo rig is well calibrated then the plane fit should give a reasonable 'ground truth' against which to measure error. Fig. 7 and 8 show sample images from two of the robots used in two different experiments. Fig. 9 shows a sample image from a sequence of 733 frames of a large checkerboard pattern measured from approximately 12m. In this case features are extracted with sub-pixel accuracy using a saddle point least squares fit [11]. This experimental setup is meant to ensure that features will follow a normal distribution, and for purposes of clarity we report results for this sequence.

The error plot in Fig. 10 for the checker wall sequence shows that filtering with the ISPKF improves the planar-error – e.g. the structure estimate is improved with filtering. The Gauss-Newton and Batch estimators are also plotted.

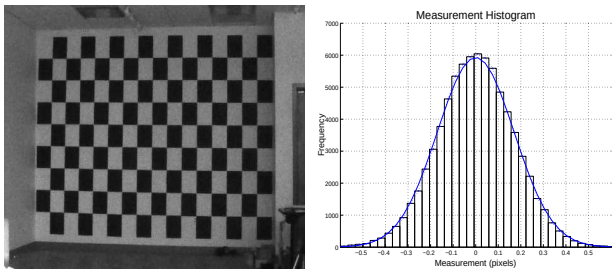

Fig. 9. Sample image of the checker wall sequence. The wall is approximately 12m in front of the robot. For scale, each checker is a US Letter size piece of paper (8.5 in by 11 in). Over 733 frames 105 corner features were tracked from the center portion of the pattern. The lighting and camera shutter speed were deliberately set to induce image noise so that the feature tracker (a sub-pixel saddle point corner fit in this case [11]) would produce high measurement noise. This is apparent in the measurement histogram on the right, which is a histogram of 77,049 measurements for real data from the checker wall sequence (for the horizontal pixel dimension). This sequence uses rectified camera models with a resolution of 1024x768 pixels, a horizontal FOV of 54.08°, a vertical FOV of 40.87° and baseline of ∼4cm. This small baseline setup was deliberately chosen to to challenge the estimators in this paper with noisy data and small disparities.

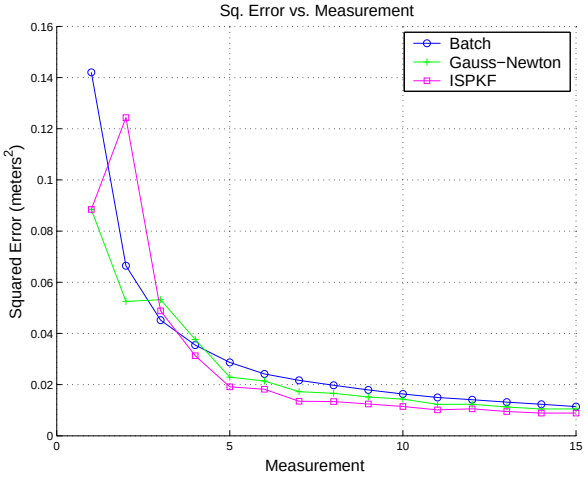

Fig. 10. Results from real data showing error measured against RANSAC ground truth plane for the Batch, ISPKF and Gauss-Newton filter for the checker wall sequence.

While all the estimators show improvement, it is difficult to distinguish difference in their performance.

VI. Future Work

In the near future we have plans to use the estimation machinery developed in this paper to improve the structure of maps built from moving robot platforms. This leads naturally into Structure from Motion (SFM) and Simultaneous Localization and Mapping (SLAM) research – problems that we are actively working on. Another interesting direction to consider is filtering dense stereo range images like the one shown in Fig. 11.

Generally, the problem with using simple parametric distributions (like Gaussians) is that error-propagation through non-linear functions can transform the p.d.f. into highly non-Gaussian shapes. For instance, in stereo, the input p.d.f. *is* quite nearly Gaussian (see for instance the

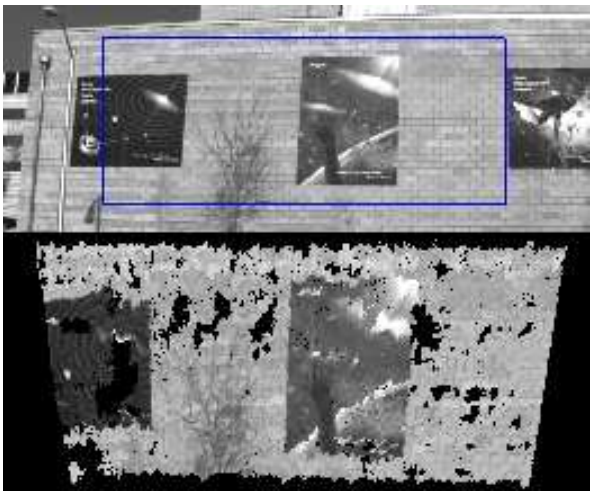

Fig. 11. Screen shot of stereo range image from the LAGR wall sequence. The bottom portion shows that there is range data for most of the wall. We use a robust plane fit to all of the triangulated 3D points from all of the frames to establish a reference against which to compare the various estimators developed in this paper. Each range point can potentially be filtered to yield more accurate structure.

measurement histogram in Fig. 9) while the transformed range p.d.f. is not. Since our state space representation assumes the normal distribution, we are likely to face difficulties. The underlying problem is that in state space the distribution we are tracking is not Gaussian - even if our sensor measurements are. There are many solutions, such as non-parametric density estimation [14], or tracking higher order moments to estimate more accurately the a posteriori density [16], to name just two. In general comparison to Monte-Carlo methods should give more insight into the ISPKF's performance.

VII. Conclusion

This paper develops the Iterated Sigma Point Kalman Filter and emphasizes the importance of iteration in solving non-linear least squares problems. We develop the ISPKF directly from first principles, grounding it in non-linear optimization theory and Newton's Method. This derivation shows that the *Iterated* Sigma Point Kalman Filter (ISPKF) is more than just an extension to non-iterated statistical linearization methods.

To establish the usefulness of the ISPKF we compare against a number of methods in the context of filtering long range stereo range measurements. Stereo is an interesting problem on which to apply the ISPKF because triangulation is an inherently biased, non-linear problem. We compare the ISPKF to the optimal batch non-linear least squares solution, a Gauss-Newton smoother, the Iterated Extended Kalman Filter, an Extended Kalman Filter with 3D measurements, and the Unscented Kalman Filter. This comparison shows that the ISPKF out performs all but the full batch solution. For experimental validation we have also applied ISPKF to filtering long range stereo data using three stereo rigs.

VIII. Acknowledgements

This work is supported in part by Caltech/JPL under contract 1277958, and by NSF grants IIS-0133947 and CCR-0120778.

References

[1] F.W. Bell, B.M. Cathey. The iterated Kalman filter update as a Gauss-Newton method. *IEEE Transactions on Automatic Control*, 38(2):294–297, Feb 1993.

[2] M. H. DeGroot and M. J. Schervish. *Probability and Statistics*. Addison Wesley, 2001.

[3] Jr Dennis J.E. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Soc for Industrial & Applied Math, 1996.

[4] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381 – 395, 1981.

[5] A. Gelb. *Applied Optimal Estimation*. MIT Press, Cambridge, MA, 1974.

[6] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45(5):910–927, May 2000.

[7] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45:477–482, 2000.

[8] S. J. Julier and J. K. Uhlmann. Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations. In *Proceedings of the 2002 American Control Conference*, 2002.

[9] T. Lefebvre, H. Bruyninckx, and J. De Schutter. Kalman filters for non-linear systems: a comparison of performance. *International Journal of Control*, 77(7):639–653, May 2004.

[10] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[11] L. Lucchese and S.K. Mitra. Using saddle points for subpixel feature detection in camera calibration targets. In *Proceedings of the 2002 Asia Pacific Conference on Circuits and Systems*, volume 2, pages 191– 195, 2002.

[12] A. Roy-Chowdhury and R. Chellappa. Statistical error propagation in 3d modeling from monocular video. In *2003 Conference on Computer Vision and Pattern Recognition Workshop*, volume 8, pages 89 –, Madison, Wisconsin, June 2003.

[13] G. Sibley, L. Matthies, and G. Sukhatme. Bias reduction and filter convergence for long range stereo. In *International Symposium of Robotics Research*, 2005.

[14] A. F. M. Smith and A. E. Gelfand. Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, 46 (2):84–88, May 1992.

[15] H. W. Sorenson. *Parameter Estimation: Principles and Problems*. Marcel Drekker, Inc., 1980.

[16] H. W. Sorenson and A. R. Stubberud. Non-linear filtering by approximation of the posteriori density. *International Journal of Control*, 8(1):33–51, 1968.

[17] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):pp 323–344, 1987.

[18] R. van der Merwe. *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*. PhD thesis, Oregon Health & Science University, OGI School of Science & Engineering, 2004.

[19] E. A. Wan and R. van der Merwe. The unscented kalman filter for nonlinear estimation. In *Symposium 2000 on Adaptive Systems for Signal Processing*, 2000.

[20] Y. Yakimovsky and R. Cunningham. A system for extracting three-dimensional measurements from a stereo pair of tv cameras. In *Computer Graphics and Image Processing*, volume 7, pages 1995–2010, 1978.

[21] G.S. Young and R. Chellappa. Statistical analysis of inherent ambiguities in recovering 3-d motion from a noisy flow field. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 14(10): 995–1013, 1992.