Bias Reduction and Filter Convergence for Long Range Stereo

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

†Robotic Embedded Systems Laboratory, University of Southern California, Los Angeles, CA 90089
‡Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109

Abstract. We are concerned here with improving long range stereo by filtering image sequences. Traditionally, measurement errors from stereo camera systems have been approximated as 3-D Gaussians, where the mean is derived by triangulation and the covariance by linearized error propagation. However, there are two problems that arise when filtering such 3-D measurements. First, stereo triangulation suffers from a range dependent statistical bias; when filtering this leads to over-estimating the true range. Second, filtering 3-D measurements derived via linearized error propagation leads to apparent filter divergence; the estimator is biased to under-estimate range. To address the first issue, we examine the statistical behavior of stereo triangulation and show how to remove the bias by series expansion. The solution to the second problem is to filter with image coordinates as measurements instead of triangulated 3-D coordinates. Compared to the traditional approach, we show that bias is reduced by more than an order of magnitude, and that the variance of the estimator approaches the Cramer-Rao lower bound.

1 Introduction

This paper details our efforts to enhance long range depth estimation in stereo systems by filtering feature measurements from image sequences. We would like to accurately estimate the depth of distant objects from disparities on the order of 1 pixel. Improving depth estimation in stereo systems is an important pursuit. For instance, better stereo range resolution will enhance a robot's ability to perform tasks such as navigation, long range path planning, obstacle avoidance, mapping and localization and high-speed driving. Unbiased sensing is a prerequisite for these algorithms to perform well.

In the balance of this paper we will encounter two problems with traditional stereo error modeling and in turn describe their solutions. First, because of the non-linearity in stereo triangulation, we will see that range estimates produced by standard triangulation methods are statistically biased. While bias in stereo is a known phenomenon, previous research focused on how range bias is induced from uncertain camera positions[3, 15, 19], or dismissed it as insignificant [12]. Second, 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 linearized error propagation techniques; this stems from the fact that the uncertainty model is biased.

For the former issue, analyzing the statistical behavior of stereo triangulation leads us to new triangulation equations based on series expansion; this new bias-corrected formulation is shown to be an improvement over traditional stereo triangulation by more than order of magnitude. For the latter biased filter problem, we find that formulating the filter with image coordinates as measurements leads to efficient and unbiased estimation of range. Lastly, using the Fisher information inequality we show that the combination of bias-corrected stereo and a Gauss-Newton recursive filter yield estimates that closely approach the minimum variance Cramer-Rao lower bound.

2 Statistical Bias in Stereo

Consider some general stereo triangulation function $s : \mathbb{R}^4 \mapsto \mathbb{R}^3$

$$s(\mathbf{z}) = \mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} \tag{1}$$

where the current observation, $\mathbf{z}$ is the vector of pixel coordinates $[u_1, v_1, u_2, v_2]^T$, and the pixels $[u_1, v_1]^T$ and $[u_2, v_2]^T$ are projections of $\mathbf{x}$ into the two camera image planes. Let x_2 be the range component of $\mathbf{x}$ - i.e. x_2 is aligned with the optical axis of the cameras.

We are interested in how a particular model of pixel measurement uncertainty will translate into range uncertainty. Before we address the issue of bias in more detail we first need to establish an appropriate observation probability density function.

2.1 Measurement Distribution

A common approximation is that many measurements of a stationary feature point, such corner features [6, 8], follow a normal distribution [10, 12, 13, 14]. To establish how features are actually distributed we have performed the following experiment: we took a sequence of images from a stationary camera of a stationary checker board and tracked the corners over time with sub-pixel accuracy [11]; we then re-centered each feature track about zero by subtracting its mean.

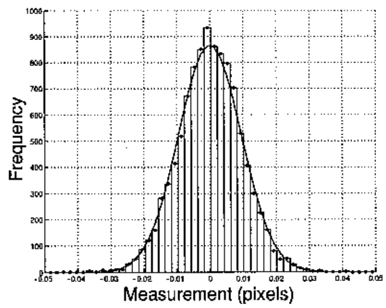

Figure 1: Feature measurement histogram of 10,360 measurements.

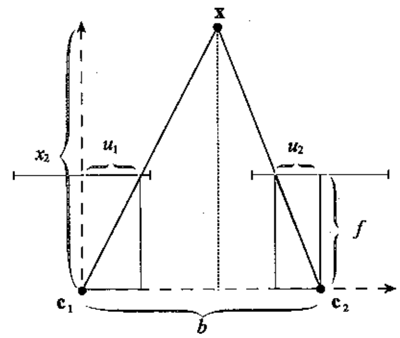

Figure 2: Standard model of linear perspective projection for stereo triangulation with axis aligned cameras.

For each pixel dimension a histogram of all the measurements is then plotted. This histogram approximates the true distribution we should expect in measurements. Qualitatively, the histogram in Fig. 1 indicates that the distribution is close to Gaussian.

2.2 Derived Range Distribution

Recall the fronto-parallel configuration, whose geometry is shown in Fig. 2. We will derive the range p.d.f. using this simple geometry, though the methods and results used here apply to other camera models as well. Using Fig. 2, the stereo equations are

$$s(\mathbf{z}) = \mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} u_1\, b/d) \\ v_1\, b/d \\ (b\ f)/d \end{bmatrix} \tag{2}$$

where the last element of $\mathbf{x}$ is the range component, and $d = (u_1 - u_2)$ is the disparity. Monte-Carlo simulation using these equations indicates that if image feature positions, and hence disparity, are normally distributed, then the expected range will be biased toward over estimating the true value[1]. The bias is empirically visible in Fig. 3. Analytically, the bias can be seen by deriving the range p.d.f., $f_{x_2}(x_2)$, from the disparity p.d.f., $f_d(d)$ [4, 9, 12]. From (2) we have $x_2 = s_2(d) = k/d$, where $k = bf$. Since s_2 and s_2^{-1} are continuously differentiable, then

$$f_{x_2}(x_2) = f_d(d) \left| \frac{\partial f_d(d)}{\partial x_2} \right| = f_d[s_2^{-1}(x_2)] \left| \frac{\partial s_2^{-1}(x_2)}{\partial x_2} \right|$$

[1] Throughout this paper we use linear camera models with a resolution of 512x384 pixels, a horizontal FOV of 68.12 degrees and vertical FOV 51.3662 of degrees

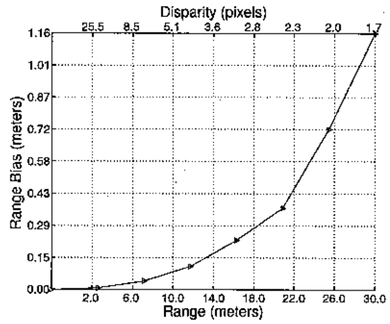

Figure 3: Range vs. Bias at 8 different ranges averaged over 10,000 trials. Clearly, bias is a strong function of range. Pixel standard deviation is 0.3 pixels in each pixel dimension, with no covariance.

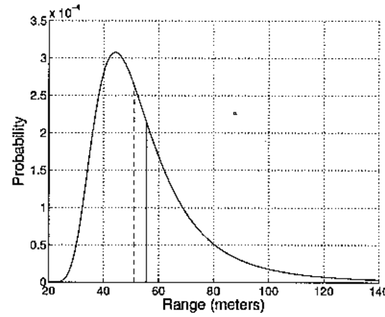

Figure 4: Range p.d.f. for $\sigma_d = 0.3$ pixels. True range is 51.08m = 1 pixel disparity. Note that because of the tail, the mean is at 55.63m, which is a bias of almost 10%.

where $|\cdot|$ denotes absolute value of the Jacobian determinant and $s_2^{-1} = d = k/x_2$. Thus, since $f_d(d)$ is modeled as Gaussian

$$f_{x_2}(x_2) = \frac{k}{\sqrt{2\pi}\sigma_d x_2^2} exp\left(\frac{-(k/x_2 - \mu_d)^2}{2\sigma_d^2}\right) \tag{3}$$

where μ_d and σ_d are the disparity mean and variance. The mean of (3) is

$$\mu_{x_2} = E[x_2] = \int_{-\infty}^{\infty} x_2 f_{x_2}(x_2) dx_2$$

which unfortunately does not appear to have an analytical solution, so we resort to numerical integration. Plots of $f_{x_2}(x_2)$ are shown in Fig. 4; clearly, for distant features the p.d.f. is non-Gaussian, non-symmetric and exhibits a long tail. The tail shifts the mean away from the true range and hence we see the source of bias in stereo.

2.3 Bias Reduction

Naturally, we would like an unbiased method for calculating range. Recall that the distribution on $\hat{d}$ is approximately Gaussian, the mean of which we take to approximate some true underlying state, d. If the true value of d was known, then the true unbiased range could be calculated with $x_2 = s_2(d)$, but due to the variation in $\hat{d}$, $s_2(\hat{d})$ is, as we have seen, slightly biased. However, if the variation of $\hat{d}$ around d is small, then a Taylor series expansion of s_2 may provide a better estimate [2],

$$s_2(d) \approx s(d) + \left.\frac{\partial s_2}{\partial d}\right|_d (\hat{d} - d) + \frac{1}{2}(\hat{d} - d)^2 \left.\frac{\partial^2 s_2}{\partial d}\right|_d$$

Taking the expectation, noting that $E[\hat{d}-d]=0$, that $E[(\hat{d}-d)^2]$ is the definition of variance, and replacing d with $\hat{d}$ we get

$$\tilde{s}_2(\hat{d}) \approx s_2(\hat{d}) + \frac{1}{2}\mathrm{var}(\hat{d})\frac{\partial^2 s_2}{\partial d}\bigg|_{\hat{d}} \tag{4}$$

which is the new range equation that we use to correct for bias. Note that this formulation requires accurate knowledge of the measurement variance; which is reasonable. Looking at Fig. 5 the improvement is immediately visible for small disparities; in fact, for the ranges shown, bias is reduced by more than an order of magnitude. Note that higher order series approximation, which should theoretically provide a better estimate, will depend on higher order moments, $E[(\hat{d}-d)^n], n > 2$. But if the input distribution has negligible higher order moments, then the second term in (4) makes use of all the available information. By considering the variance and how it impacts the range distribution, this bias-correction method largely removes the bias from long range stereo.

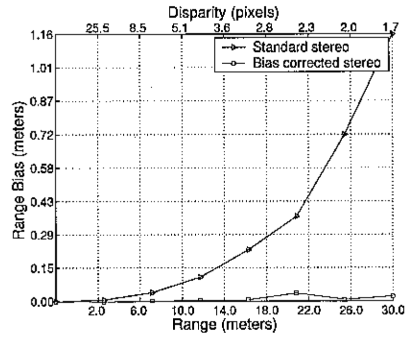

Figure 5: Bias-corrected stereo compared to traditional stereo.

3 3-D Estimation

In this section we uncover another type of bias that results from filtering a sequence of 3-D estimates produced by triangulation and linearized error propagation. To alleviate this we develop a non-linear Gauss-Newton iterative measurement update using image space measurements instead of 3-D measurements. Finally, the statistical efficiency of the 3-D measurement update and the Gauss-Newton update are compared to the Cramer-Rao lower bound.

3.1 3-D Measurement Update

Let $\mathbf{x} \in \mathbb{R}^3$, $\hat{\mathbf{x}} \in \mathbb{R}^3$, $\mathbf{z}_{3D} \in \mathbb{R}^3$ denote the current state, current state estimate and the current 3-D observation, respectively. For the case at hand, the current observation, $\mathbf{z}_{3D}$, is the vector found via bias-corrected stereo. The state estimate and observation are independent realizations of multivariate Gaussian distributions: $\mathbf{z}_{3D} \sim N(s(\mathbf{z}), \mathbf{R}_{3D})$ 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

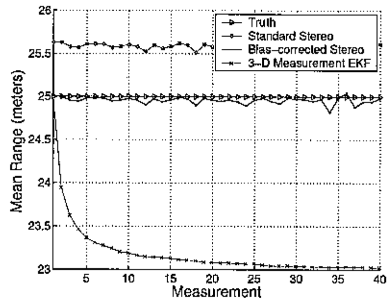

Figure 6: Kalman Filter of a sequence of 40 3-D stereo measurements averaged over 10,000 trials.

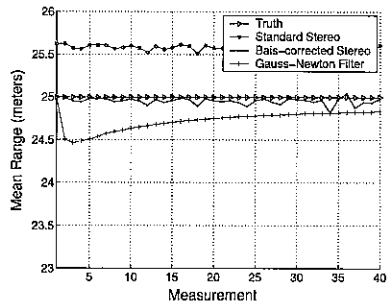

Figure 7: Gauss-Newton filter of a sequence of 40 image measurements averaged over 10,000 trials.

error covariance matrix $\mathbf{R}_{3D}$ is found via error propagation of image errors

$$\mathbf{R}_{3D} = \tfrac{\partial s}{\partial \mathbf{z}} \mathbf{R} \tfrac{\partial s}{\partial \mathbf{z}}^T, \quad \mathbf{R} = \begin{bmatrix} \mathbf{\Sigma}_l & 0 \\ 0 & \mathbf{\Sigma}_r \end{bmatrix} \tag{5}$$

where s is the bias corrected stereo equation and $\mathbf{\Sigma}_l$ and $\mathbf{\Sigma}_r$ are 2x2 error covariance matrices from the left and right images, respectively.

In this scenario the sensor model $h_{3D} : \mathbb{R}^3 \mapsto \mathbb{R}^3$ is the vector function that returns a predicted measurement for $\mathbf{z}_{3D}$ given $\mathbf{x}$. If $\mathbf{x}_{ws}$ is the global position of the stereo head with orientation matrix $\mathbf{R}_{ws}$ then the generative sensor model is, $h_{3D}(x) = \mathbf{R}_{ws}^T x - \mathbf{R}_{ws}^T x_{ws}$. Note that while we focus here on the stationary case and solving issues of bias and filter convergence, this forumlation extends to the moible sensor case (an issue we are actively working on that is beyond the scope of this paper). The Kalman Filter update equations for this system are

$$\begin{aligned} \hat{\mathbf{x}}_{k+1}^{+} &= \hat{\mathbf{x}}_k^{-} + \mathbf{K}(\mathbf{z}_{3D} - h_{3D}(\hat{\mathbf{x}}_k^{-})) \\ \hat{\mathbf{P}}_{k+1}^{+} &= (\mathbf{I} - \mathbf{K}\mathbf{H}_k)\hat{\mathbf{P}}_k^{-} \\ \mathbf{K} &= \hat{\mathbf{P}}_k^{-}\mathbf{H}_k^T(\mathbf{H}_k\hat{\mathbf{P}}_k^{-}\mathbf{H}_k^T - \mathbf{R}_{3D})^{-1} \end{aligned} \tag{6}$$

where $\mathbf{H}_k$ is the Jacobian of h_{3D}. Filtering with this setup leads to the situation depicted in Fig. 6, which clearly shows what is called *apparent* divergence - i.e. convergence to the wrong result[7]. This can be explained by the fact that linearized error propagation in (5) gives quadratically larger range variance for more distant features. This means that the weighted average in (6) will always place more confidence in closer measurements and $\hat{\mathbf{x}}^+$ will be biased toward the short measurements. In essence, linearized error propagation leads to over confidence for shorter measurements, which in turn leads to serious filter bias.

3.2 Gauss-Newton Measurement Update

Instead of using the triangulated point $\mathbf{z}_{3D}$ as the observation, let the observation again be the vector of pixel coordinates $\mathbf{z} = [u_1, v_1, u_2, v_2]^T$. Thus our sensor model $h : \mathbb{R}^3 \mapsto \mathbb{R}^4$ is the vector function that projects $\mathbf{x}$ into the left and the right images,

$$h(\mathbf{x}) = \begin{bmatrix} h_l(\mathbf{x}) \\ h_r(\mathbf{x}) \end{bmatrix} \tag{7}$$

where $h_l : \mathbb{R}^3 \mapsto \mathbb{R}^2$ and $h_r : \mathbb{R}^3 \mapsto \mathbb{R}^2$ are the left and right camera projection functions. Depending on the camera models in use, h_l and h_r can be formulated in a variety of ways [17, 18]. We only require that these functions and their first derivatives are available, and otherwise leave them unspecified.

For convenience we choose to formulate the measurement update as an iterative Gauss-Newton method, which is equivalent to an iterated Extended Kalman Filter [1]. To integrate prior information about $\hat{\mathbf{x}}$ we write the current state estimate and and current observation as a single measurement vector

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z} \\ \hat{\mathbf{x}} \end{bmatrix}, \quad g(\mathbf{x}) = \begin{bmatrix} h(\mathbf{x}) \\ \mathbf{x} \end{bmatrix}$$

For the first measurement, the filter is initialized with $\hat{\mathbf{x}} = \mathbf{z}_{3D}$ and $\hat{\mathbf{P}} = \mathbf{R}_{3D}$ which are calculated by bias-corrected triangulation and linearized error propagation in as in section 3.1. Since the current observation and state estimate are realizations of independent normal distributions we have $\mathbf{Z} \sim N(g(\mathbf{x}), \mathbf{C})$ where $\mathbf{C} = \left[\begin{smallmatrix} \mathbf{R} & 0 \\ 0 & \mathbf{P} \end{smallmatrix}\right]$. Given the measurement $\mathbf{Z}$, we can write the likelihood function

$$\mathcal{L}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^7 |\mathbf{C}|}} exp\left(-\frac{1}{2}(\mathbf{Z} - g(\mathbf{x}))\mathbf{C}^{-1}(\mathbf{Z} - g(\mathbf{x}))\right) \tag{8}$$

where $|\cdot|$ is the determinant. The maximum likelihood estimate for this expression is $\hat{\mathbf{x}}^+ = argmax_{\mathbf{x}}\ \mathcal{L}(\mathbf{x})$, whose solution is equivalent the solution minimizing the negative log-likelihood, $argmin_{\mathbf{x}}\ \ell(\mathbf{x})$,

$$\ell(\mathbf{x}) = \frac{1}{2}(\mathbf{Z} - g(\mathbf{x}))\mathbf{C}^{-1}(\mathbf{Z} - g(\mathbf{x})) + k \tag{9}$$

where k is a constant.If we let $\mathbf{S}^T\mathbf{S} = \mathbf{C}^{-1}$ and

$$r(\mathbf{x}) = \mathbf{S}(\mathbf{Z} - g(\mathbf{x})) \tag{10}$$

then (9) is a non-linear least squares problem to minimize $r(\mathbf{x})^T r(\mathbf{x})$. The Gauss-Newton method to solve non-linear problems of this form is the sequence of iterates [5]

$$\mathbf{x}_{i+1} = \mathbf{x}_i - (\mathbf{J}(\mathbf{x}_i)^T\mathbf{J}(\mathbf{x}_i))^{-1}\mathbf{J}(\mathbf{x}_i)^T r(\mathbf{x}_i) \tag{11}$$

where $\mathbf{J}$ is the Jacobian of (10). Noting that $\mathbf{J} = -\mathbf{S}\mathbf{G}_i$ where $\mathbf{G}_i$ is the Jacobian of $g(\mathbf{x}_i)$, (11) becomes

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

which is the familiar normal equation solution. Once iterated to convergence the covariance $\hat{\mathbf{P}}^+$ can then be approximated using $\hat{\mathbf{P}}^+ = (\mathbf{G}^T\mathbf{C}^{-1}\mathbf{G})^{-1}$. As noted in [1], this is equivalent to the iterated Extended Kalman Filter measurement update.

Filtering with this setup leads to the situation depicted in Fig 7. Typically, the measurement update converges after 3 to 4 iterations. Compared to the 3D measurement update, the fact that the Gauss-Newton (IEKF) method converges without bias is not surprising considering that we avoid the intermediate stereo triangulation and linearized error propagation for calculating the 3-D error covariance matrix.

3.3 Estimator Efficiency

Having derived a bias-corrected estimator it is important to address its efficiency, that is, how well it approximates a minimal variance estimate of the parameters. The information inequality, $\text{cov}_\mathbf{x}(\mathbf{x}) \geq \mathcal{I}_\mathbf{z}(\mathbf{x})^{-1}$, defines such a bound, which is called Cramer-Rao lower bound[4]. Here the Fisher information matrix $\mathcal{I}_\mathbf{z}(\mathbf{x})$ is given by the symmetric matrix whose ij^{th} element is the covariance between first partial derivatives of the *measurement* log-likelihood function,

$$\mathcal{I}_\mathbf{z}(\mathbf{x})_{,i,j} = \text{cov}_\mathbf{x}\left(\frac{\partial \ell_\mathbf{z}}{\partial \mathbf{x}_i}, \frac{\partial \ell_\mathbf{z}}{\partial \mathbf{x}_j}\right) \tag{12}$$

The measurement log-likelihood function is $\ell_\mathbf{z}(\mathbf{x}) = \frac{1}{2}(\mathbf{z}-h(\mathbf{x}))\mathbf{R}^{-1}(\mathbf{z}-h(\mathbf{x})) + k$. For a multivariate normal distribution (12) reduces to[16, 19]

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

For n i.i.d. measurements the Fisher information is simply $n\mathcal{I}$. An estimator that achieves the CRLB is said to be efficient. Fig. 8 shows range variance convergence for the Gauss-Newton estimator; this demonstrates that the Gauss-Newton stereo estimator is efficient.

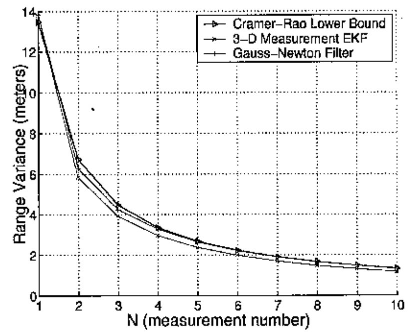

Figure 8: Gauss-Newton and 3-D EKF estimator efficiency compared against the Cramer-Rao lower bound over a sequence of 10 measurements of a feature 25m away. At each step estimator variance is found via Monte-Carlo simulation over 10,000 trials.

4 Conclusion

In our efforts to improve long range stereo by filtering image sequences we have come across two problems: the first is that stereo triangulated range estimates are statistically biased. To address this we have re-expressed the stereo triangulation equations using a second order series expansion. This new formulation reduces bias in stereo triangulation by more than an order of magnitude. The second problem is that temporal filtering of 3D stereo measurements also leads to biased estimates. The solution to this problem is to filter with image coordinates as measurements instead of triangulated 3D coordinates. Finally, using the Fisher information inequality we show that the bias-corrected Gauss-Newton stereo estimator approaches the minimum variance Cramer-Rao lower bound. While the scope of this paper is constrained to address stereo bias and estimator efficiency, our ultimate goal is to filter feature points from a moving platform. This is a task that requires a solid solution to the simultaneous localization and mapping problem, which we are actively exploring.

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] D.R. Cox and D.V. Hinkley. *Theoretical Statistics*. Chapman and Hall, 1979.

[3] Kostas Daniilidis and Minas E. Spetsakis. *Understanding noise sensitivity in structure from motion*, chapter 4, pages 61–88. Lawrence Erlbaum Associates, 1996.

[4] Morris H. DeGroot and Mark J. Schervish. *Probability and Statistics*. Addison Wesley, 2001.

[5] Jr Dennis J.E. and Robert B. Schnabel. *Numerical Methods for Uncon-*

strained Optimization and Nonlinear Equations. Soc for Industrial & Applied Math, 1996.

[6] W. Forstner. A feature based correspondence algorithm for image matching. *International Archives of Photogrammetry & Remote Sensing*, 26(3):150–166, 1986.

[7] Arthur Gelb. *Applied Optimal Estimation*. MIT Press, Cambridge, MA, 1974. Good reference for various forms of the Kalman Filter.

[8] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Proceedings of The Fourth Alvey Vision Conference*, Manchester, 1988.

[9] Andrew H Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.

[10] Il-Kyun Jung and Simon Lacroix. Simultaneous localization and mapping with sterevision. In *International Symposium on Robotics Research*, 2003.

[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*, 2002.

[12] L. Matthies and P. Grandjean. Stochastic performance modeling and evaluation of obstacle detectability with imaging range sensors. *IEEE Transactions on Robotics and Automation*, Vol.10(6):783–792, Dec 1994.

[13] L. Matthies and S. Shafer. Error modelling in stereo navigation. *IEEE Journal of Robotics and Automation*, 3(3):239–248, 1987.

[14] Clark F. Olson, Larry H. Matthies, Marcel Schoppers, and Mark W. Maimone. Stereo ego-motion improvements for robust rover navigation. In *In Proceedings IEEE Conference on Robotics and Automation (ICRA)*, 2001.

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

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

[17] Roger 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):323–344, 1987.

[18] 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*, 1978.

[19] 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.