The Cramer-Rao Lower Bound for 3-D State Estimation from Rectified Stereo Cameras

Daniel E. Clark
Joint Research Institute in Signal and Image Processing
Heriot-Watt University
Edinburgh, UK.
d.e.clark@hw.ac.uk

Špela Ivekovič
Lessells Scholar
Royal Society of Edinburgh
Edinburgh, UK
spela.ivekovic@gmail.com

Abstract – *It is well known that any 3-D state estimate computed from stereo camera measurements is corrupted by heteroscedastic noise due to the nature of the perspective projection. It is also well understood that the image measurements used to estimate the 3-D state are inherently noisy. Despite the wealth of research in this area, the accurate statistical characterisation of the uncertainty for any 3-D state estimation from stereo algorithm is less well understood. This paper presents the Cramer-Rao Lower Bound (CRLB) for 3-dimensional state estimation from a rectified stereo pair of cameras. The paper also presents a method for efficient stereo estimation via Bayesian triangulation that achieves the CRLB. These results provide a basis for 3-D statistical estimation for camera-based sensor measurements.*

Keywords: Cramer-Rao Lower Bound, Bayesian Triangulation, 3-D State Estimation, Rectified Stereo Cameras

1 Introduction

In statistical stereo estimation, the hidden state of a three-dimensional system is estimated from image measurements generated by a stereo pair of cameras. The majority of stereo estimation algorithms estimate the hidden state in 3-D Euclidean space, since we are generally interested in knowing the state of the object, such as position and velocity, in the world co-ordinate system. Unfortunately, the projection from the 3-D space (the state space) onto the camera image planes (the observation space) is non-linear, violating the linear/Gaussian assumptions and making it necessary to resort to nonlinear approximations instead. A further difficulty with statistical estimation in 3-D is that the noise in the state estimate is heteroscedastic, that is, the variance of the state estimate changes with the position of the system. The heteroscedasticity is a direct consequence of the nature of stereo reconstruction, as shown in Figure 1. These difficulties can seriously impede the reliability of statistical estimation algorithms. To address these, we propose a fundamentally different approach by performing recursive Bayesian stereo estimation in disparity space as an intermediate step to 3-D object state estimation.

Stereo-estimation in disparity space has two key advantages over 3-D Euclidean space: (i) the noise in the state estimate is homoscedastic, and (ii) the projections into the observation space (the two image planes) are linear. The immediate consequence of this is that the hidden state can now be estimated with the linear Gaussian assumptions and hence optimally and in closed form. The linear and Gaussian assumptions in disparity space allow for a straightforward error analysis such as the computation of the Fisher information and Cramer-Rao Lower Bound (CRLB). Once the CRLB in disparity space is known, one can derive the corresponding CRLB in 3-D by means of reparametrization, thereby characterising the minimum variance of any 3-D estimator from a rectified stereo pair of cameras.

We illustrate the stereo estimation in disparity space with Bayesian stereo triangulation and provide the corresponding 3-D estimates and their variance as well as compare the accuracy of the estimator with the derived CRLB.

This paper is organised as follows. In Section 2, we describe the related work in this area. In Section 3, we describe disparity space and its relation to the two image planes and 3-D. In section 4 we explain the concept of statistical estimation in disparity space. We provide the statistical error analysis for disparity space and 3-D in Section 5 and illustrate the approach with Bayesian triangulation in Section 6. In Section 7 we test the Bayesian triangulation on simulated data and compare the performance to the CRLB derived in Section 5. We conclude in Section 8.

2 Background

It is well known that 3-D points reconstructed from stereo images are corrupted by heteroscedastic noise and hence using stereo-reconstructed data as an input to further estimation algorithms requires use of specific estimation methods which explicitly model the error covariance [1, 2, 3, 4]. We show that a proxy estimation problem can be formulated in disparity space [5, 6, 7, 8], where the noise is reasonably assumed to be homoscedastic and the usual linear/Gaussian assumptions hold. The results of linear/Gaussian estimation in disparity space can then be mapped into 3-D via

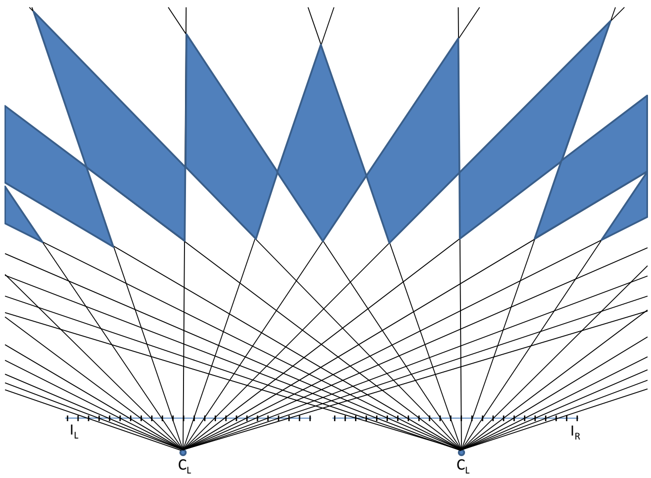

Figure 1: Schematic illustration of the heteroscedastic noise property in stereo-reconstructed 3-D state estimates. The image planes have been split into discrete pixels and the corresponding image rays are shown. As the object moves further away from the sensor, its position uncertainty increases, as illustrated by the quadrilateral formed by the intersecting image rays. The shaded quadrilaterals illustrate the uncertainty in the reconstructed state estimate for a disparity value $d = 3$.

reparametrization of the posterior distribution which completely characterises the estimate in 3-D.

Using the concept of disparity space, we also perform an error analysis of the 3-D from stereo estimation. Performance estimation in computer vision is an actively debated problem [9] and in the area of 3-D reconstruction from stereo, several papers address the error propagation due to noisy image observations, for example [10, 11, 12, 13, 14], however, little is known about the efficiency of these estimators. Sibley *et al.* [15] attempted to derive the CRLB for estimation from stereo, though the work did not provide an analytic form for the CRLB and assumed that the estimation in depth was independent of co-ordinates parallel to the image plane.

We derive the Cramer-Rao Lower Bound (CRLB) for 3-D state estimation from rectified stereo images, which provides the theoretical lower bound on the variance of any 3-D estimation from rectified stereo and can as such be used as a performance metric in addition to the ground-truth data used, for example, in [16] to evaluate the performance of stereo reconstruction algorithms. An unbiased estimator which meets the CRLB with equality is said to be efficient, see, for example [17, 18]. We provide an example of such an efficient estimator in the form of Bayesian triangulation in Section 6. A stereo multi-object filtering extension of this method was presented recently [19], which can be used for automatic Bayesian feature correspondence in the presence of false alarms and missed detections.

3 Disparity space

Let us assume that a point $\mathbf{W} = (X, Y, Z)^T \in \mathbb{R}^3$ is viewed by two distinct cameras, left camera with projection matrix P_l and right camera with projection matrix P_r. We denote the homogeneous equivalent of the point $\mathbf{W}$ as $\bar{\mathbf{W}} = (X, Y, Z, 1)^T \in \mathbb{P}^3(\mathbb{R}^3)$. The images of the point $\mathbf{W}$ are defined as $\mathbf{w}_l = (u_l, v_l, 1)^T \simeq P_l\bar{\mathbf{W}}$ and $\mathbf{w}_r = (u_r, v_r, 1)^T \simeq P_r\bar{\mathbf{W}}$ in the left and right camera's image plane, respectively, where "$\simeq$" denotes equality up to a scale factor. The corresponding points $\mathbf{w}_l$ and $\mathbf{w}_r$ are related by a *disparity* which, in the general case, is defined as:

$$d(\mathbf{w}_l, \mathbf{w}_r) = (u_r - u_l, v_r - v_l). \tag{1}$$

In the case of rectified images, the two corresponding points lie on the same scanline, and the disparity simplifies to a displacement along that scanline:

$$d(\mathbf{w}_l, \mathbf{w}_r) = u_r - u_l. \tag{2}$$

For a rectified stereo pair of images, the disparity space is then defined as a three-dimensional space $\mathbb{D}^3 = \{u, v, d\}$. The so-defined disparity space is a projective space. This can be shown by deriving a projective transformation P_D between $\mathbb{P}^3(\mathbb{R}^3)$ and $\mathbb{P}^3(\mathbb{D}^3)$, as follows.
We assume, without a loss of generality, a specific form of the rectified projection matrices [20]. Let the left and right rectified camera projection matrices, $\tilde{P}_l$ and $\tilde{P}_r$, be written as:

$$\tilde{P}_l = \begin{pmatrix} p^l_{11} & p^l_{12} & p^l_{13} & p^l_{14} \\ p^l_{21} & p^l_{22} & p^l_{23} & p^l_{24} \\ p^l_{31} & p^l_{32} & p^l_{33} & p^l_{34} \end{pmatrix} \tag{3}$$

$$\tilde{P}_r = \begin{pmatrix} p^r_{11} & p^r_{12} & p^r_{13} & p^r_{14} \\ p^l_{21} & p^l_{22} & p^l_{23} & p^l_{24} \\ p^l_{31} & p^l_{32} & p^l_{33} & p^l_{34} \end{pmatrix}. \tag{4}$$

Projecting a point $\bar{\mathbf{W}} = (X, Y, Z, 1)^T$ into left and right view gives the left and right image point, $\mathbf{w}_l$ and $\mathbf{w}_r$:

$$\mathbf{w}_l = \begin{pmatrix} u_l \\ v_l \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{p^l_{11}X+p^l_{12}Y+p^l_{13}Z+p^l_{14}}{p^l_{31}X+p^l_{32}Y+p^l_{33}Z+p^l_{34}} \\ \frac{p^l_{21}X+p^l_{22}Y+p^l_{23}Z+p^l_{24}}{p^l_{31}X+p^l_{32}Y+p^l_{33}Z+p^l_{34}} \\ 1 \end{pmatrix} \simeq \tilde{P}_l\bar{\mathbf{W}} \tag{5}$$

$$\mathbf{w}_r = \begin{pmatrix} u_r \\ v_r \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{p^r_{11}X+p^r_{12}Y+p^r_{13}Z+p^r_{14}}{p^l_{31}X+p^l_{32}Y+p^l_{33}Z+p^l_{34}} \\ \frac{p^l_{21}X+p^l_{22}Y+p^l_{23}Z+p^l_{24}}{p^l_{31}X+p^l_{32}Y+p^l_{33}Z+p^l_{34}} \\ 1 \end{pmatrix} \simeq \tilde{P}_r\bar{\mathbf{W}} \tag{6}$$

A point $\boldsymbol{s} \in \mathbb{P}^3(\mathbb{D}^3)$, is defined as:

$$\boldsymbol{s} = \begin{pmatrix} u_l \\ v_l \\ u_r - u_l \\ 1 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{p^l_{11}X+p^l_{12}Y+p^l_{13}Z+p^l_{14}}{p^l_{31}X+p^l_{32}Y+p^l_{33}Z+p^l_{34}} \\ \frac{p^l_{21}X+p^l_{22}Y+p^l_{23}Z+p^l_{24}}{p^l_{31}X+p^l_{32}Y+p^l_{33}Z+p^l_{34}} \\ \frac{p^r_{11}X+p^r_{12}Y+p^r_{13}Z+p^r_{14}}{p^l_{31}X+p^l_{32}Y+p^l_{33}Z+p^l_{34}} - \frac{p^l_{11}X+p^l_{12}Y+p^l_{13}Z+p^l_{14}}{p^l_{31}X+p^l_{32}Y+p^l_{33}Z+p^l_{34}} \\ 1 \end{pmatrix}, \tag{7}$$

and the linear transformation P_D, for which

$$\mathbf{s} \simeq P_D\bar{\mathbf{W}}, \quad \mathbf{s}, \bar{\mathbf{W}} \in \mathbb{P}^3 \tag{8}$$

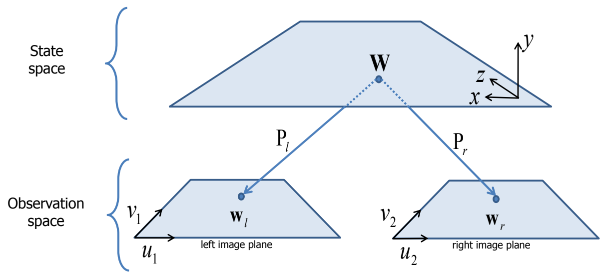

Figure 2: The nonlinear nature of 3-D estimation from stereo. The mapping from the state space onto the observation space is specified by the camera matrices P_l and P_r and is nonlinear in Euclidean coordinates.

as

$$P_D = \begin{pmatrix} p_{11}^l & p_{12}^l & p_{13}^l & p_{14}^l \\ p_{21}^l & p_{22}^l & p_{23}^l & p_{24}^l \\ p_{11}^r - p_{11}^l & p_{12}^r - p_{12}^l & p_{13}^r - p_{13}^l & p_{14}^r - p_{14}^l \\ p_{31}^l & p_{32}^l & p_{33}^l & p_{34}^l \end{pmatrix} \tag{9}$$

The transformation P_D is the link between $\mathbb{R}^3$ and $\mathbb{D}^3$ which allows us to map the disparity space points into 3-D and vice-versa.

4 Statistical Estimation in Disparity Space

Statistical estimation in 3-D from observations generated by a stereo pair of cameras must account for the nonlinearity in the observation model as shown in the Figure 2. For example, in recursive Bayesian estimation, the object's state transition model in 3-D can be modelled as linear, however, the observation model is clearly nonlinear. As a consequence, estimating the hidden state in 3-D with observations given in the image plane does not have a closed form linear/Gaussian solution and approximations, such as particle filters, must be used instead. Frequently, the nonlinearity in the observation model is avoided by mapping the observations into 3-D via triangulation, however, as pointed out by Sibley *et al.* [15], this inevitably leads to filter divergence.

The nonlinear observation model can be overcome in a principled manner by estimating in disparity space instead. Disparity space is linked to the 3-D space, in which the object really exists, via the projective transformation P_D given in Equation (9). If we know the hidden state estimates in disparity space and have a calibrated and rectified stereo pair of cameras, we are able to compute the corresponding hidden state estimates in 3-D space. The trick is in realising that the transformation between disparity space and 3-D does not need to happen in the Bayesian recursion, where the nonlinearity of the observation model complicates things. Instead, the recursive Bayesian estimation is done in disparity space and the resulting posterior distribution is transformed into 3-D afterwards.

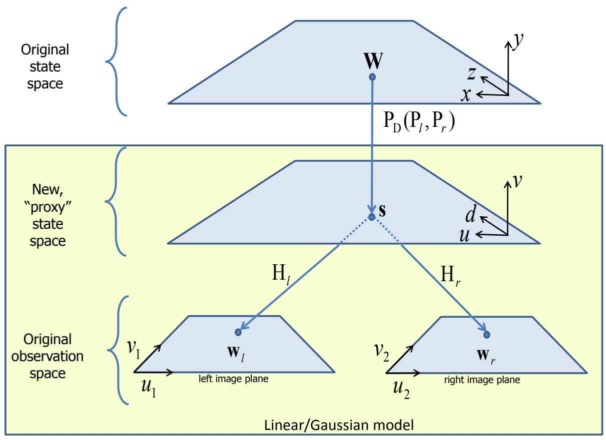

Figure 3: Linear estimation in disparity space. Instead of approximating the non-linear transformation from $\mathbb{R}^3$ to the two image planes, linear estimation can be done in disparity space and the resulting state estimates transformed into $\mathbb{R}^3$ using P_D from Equation (9).

When estimating in disparity space, the hidden state now becomes

$$s = (u, v, d)^T, \tag{10}$$

where u and v designate image column and row dimensions, respectively, with respect to the left image plane, and d denotes the disparity between the point (u_l, v_l) in the left image and the corresponding point (u_r, v_r) in the right image. We assume a rectified camera setup, which means that the corresponding points are in fact (u_l, v_l) and (u_r, v_l), and the disparity becomes a scalar value, $d = u_r - u_l$.

Contrary to estimation in 3-D, we now do not have to work with homogeneous coordinates anymore because the observation model becomes a simple orthographic projection, one for each of the cameras:

$$H^{[1]} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \tag{11}$$

$$H^{[2]} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{12}$$

and hence the observation model is linear:

$$(u, v)^T = H^{[1]}\mathbf{s} \tag{13}$$
$$(u + d, v)^T = H^{[2]}\mathbf{s}, \tag{14}$$

where $H^[1]$, $H^[2]$ are the projections for left and right camera, respectively.

The proposed framework is shown in Figure 3. An additional benefit of estimating in disparity space is that the noise in the hidden-state estimates is homoscedastic. This is not true for 3-D space, where the variance of the state estimate increases with the distance from the cameras. This in turn means that we can formulate the recursive Bayesian estimation in disparity space as a linear/Gaussian problem and provide an optimal, closed-form solution. The linear/Gaussian model in disparity space lends itself to a straightforward error analysis and provides a perfect platform for statistical error analysis of 3-D estimation from stereo, based on the observations in the image plane, which we provide in the next section.

5 Error Analysis for Estimation from Stereo

In this section we perform a formal analysis of the estimation error from a rectified stereo pair of cameras. We provide the statistical characterisation of the estimation error in 3-D using the Fisher Information and the Cramer-Rao Lower Bound (CRLB) [17].

5.1 Cramer-Rao Lower Bound in Disparity Space

Let $\mathbf{s} = [u, v, d]^T$ be a state in disparity space. Let us then suppose that we sample n independent, identically distributed (i.i.d.) random variables $\mathbf{z}_1^{[1]}, \ldots, \mathbf{z}_n^{[1]}$ from the first camera likelihood function $g_{\mathbf{s}}^{[1]}(\mathbf{z}|\mathbf{s})$ and m i.i.d. random variables $\mathbf{z}_1^{[2]}, \ldots, \mathbf{z}_m^{[2]}$ from the second camera likelihood function $g_{\mathbf{s}}^{[2]}(\mathbf{z}|\mathbf{s})$. We assume that the observations in the left camera are independent from the observations in the right camera.

The variance $\mathrm{var}(\hat{\mathbf{s}})$ of any unbiased estimator $\hat{\mathbf{s}}$ of the state $\mathbf{s} \in \mathbb{D}^3$ is bounded by the inverse of the Fisher Information $I(\mathbf{s})$:

$$\mathrm{var}(\hat{\mathbf{s}}) \geq I(\mathbf{s})^{-1}, \tag{15}$$

where

$$\begin{aligned} I(\mathbf{s}) &= E_{\mathbf{s}}\left[- \bigtriangledown^2 \ln \left(\prod_{i=1}^{n} g_{\mathbf{s}}^{[1]}(\mathbf{z}_i^{[1]}|\mathbf{s}) \prod_{j=1}^{m} g_{\mathbf{s}}^{[2]}(\mathbf{z}_j^{[2]}|\mathbf{s}) \right) \right] \qquad (16) \\ &= \sum_{i=1}^{n} E_{\mathbf{s}}\left[- \bigtriangledown^2 \ln \left(g_{\mathbf{s}}^{[1]}(\mathbf{z}_i^{[1]}|\mathbf{s}) \right) \right] \\ &+ \sum_{j=1}^{m} E_{\mathbf{s}}\left[- \bigtriangledown^2 \ln \left(g_{\mathbf{s}}^{[2]}(\mathbf{z}_j^{[2]}|\mathbf{s}) \right) \right] \end{aligned}$$

Now suppose that the camera likelihood functions $g_{\mathbf{s}}^{[1]}(\mathbf{z}^{[1]}|\mathbf{s})$ and $g_{\mathbf{s}}^{[2]}(\mathbf{z}^{[2]}|\mathbf{s})$ are Gaussian, i.e.,

$$g_{\mathbf{s}}^{[1]}(\mathbf{z}^{[1]}|\mathbf{s}) = \mathcal{N}(\mathbf{z}^{[1]}; H^{[1]}\mathbf{s}, R^{[1]}), \tag{17}$$

$$g_{\mathbf{s}}^{[2]}(\mathbf{z}^{[2]}|\mathbf{s}) = \mathcal{N}(\mathbf{z}^{[2]}; H^{[2]}\mathbf{s}, R^{[2]}), \tag{18}$$

where the observation noise covariances $R^{[1]}$ and $R^{[2]}$ are assumed to be diagonal. The (i, j)-th element of the Fisher Information Matrix for a Gaussian distribution is

$$I_{i,j} = \frac{\partial(\mathbf{z} - H\mathbf{s})^T}{\partial \mathbf{s}_i} R^{-1} \frac{\partial(\mathbf{z} - H\mathbf{s})}{\partial \mathbf{s}_j}, \tag{19}$$

where H is the projection from $\mathbb{D}^3$ onto the image plane, R is the measurement noise covariance, and $\mathbf{s}_i$ ($\mathbf{s}_j$) is the i-th (j-th) element of the parameter vector $\mathbf{s}$.

Hence the Fisher Information Matrices for the camera likelihoods are

$$I^{[1]}(\mathbf{s}) = \begin{bmatrix} (R_{11}^{[1]})^{-1} & 0 & 0 \\ 0 & (R_{22}^{[1]})^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{20}$$

$$I^{[2]}(\mathbf{s}) = \begin{bmatrix} (R_{11}^{[2]})^{-1} & 0 & (R_{11}^{[2]})^{-1} \\ 0 & (R_{22}^{[2]})^{-1} & 0 \\ (R_{11}^{[2]})^{-1} & 0 & (R_{11}^{[2]})^{-1} \end{bmatrix} \tag{21}$$

For the joint likelihood:

$$\begin{aligned} & g_{\mathbf{s}}\left(\mathbf{z}_1^{[1]}, \ldots, \mathbf{z}_n^{[1]}, \mathbf{z}_1^{[2]}, \ \ldots, \mathbf{z}_m^{[2]}|\mathbf{s}\right) \qquad (22) \\ & \qquad = \prod_{i=1}^{n} g_{\mathbf{s}}^{[1]}(\mathbf{z}_i^{[1]}|\mathbf{s}) \prod_{j=1}^{m} g_{\mathbf{s}}^{[2]}(\mathbf{z}_j^{[2]}|\mathbf{s}) \end{aligned}$$

the Fisher Information Matrix is then:

$$I(\mathbf{s}) = n \cdot I^{[1]}(\mathbf{s}) + m \cdot I^{[2]}(\mathbf{s}), \tag{23}$$

which is a constant for any given number of measurements n and m.

Hence

$$I(\mathbf{s}) = \begin{bmatrix} I_{11}^s & 0 & I_{13}^s \\ 0 & I_{22}^s & 0 \\ I_{13}^s & 0 & I_{13}^s \end{bmatrix} \tag{24}$$

where

$$I_{11}^s = n(R_{11}^{[1]})^{-1} + m(R_{11}^{[2]})^{-1}, \tag{25}$$

$$I_{13}^s = m(R_{11}^{[2]})^{-1}, \tag{26}$$

$$I_{22}^s = n(R_{22}^{[1]})^{-1} + m(R_{22}^{[2]})^{-1}. \tag{27}$$

The Cramer-Rao Lower Bound (CRLB) is then found by taking the inverse of (24).

$$\mathrm{var}(\mathbf{s}) = \begin{bmatrix} \left(I_{11}^s - I_{13}^s\right)^{-1} & 0 & -\left(I_{11}^s - I_{13}^s\right)^{-1} \\ 0 & {I_{22}^s}^{-1} & 0 \\ -\left(I_{11}^s - I_{13}^s\right)^{-1} & 0 & I_{11}^s / \left(I_{13}^s \left(I_{11}^s - I_{13}^s\right)\right) \end{bmatrix} \tag{28}$$

5.2 Cramer-Rao Lower Bound in 3-D

The estimation in disparity space described so far does not require any knowledge of the actual camera parameters involved. The only assumption is that the cameras are rectified. In order to establish the connection between estimates in disparity space and estimates in 3-D, the actual cameras must then be specified. In the following we assume, without loss of generality, a canonical rectified stereo pair of cameras, where the left camera is at the world origin and the right camera is displaced along the x-axis by a baseline B:

$$\tilde{P}_l = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{29}$$

$$\tilde{P}_r = \begin{pmatrix} 1 & 0 & 0 & B \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{30}$$

The corresponding P_D is then

$$P_D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & B \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{31}$$

and the resulting link between the disparity space and 3-D in Euclidean coordinates is as follows:

$$(\frac{X}{Z}, \frac{Y}{Z}, \frac{B}{Z}, 1)^T \equiv (X, Y, B, Z)^T \simeq P_D(X, Y, Z, 1)^T, \tag{32}$$

where the homogeneous 4-vector with the last coordinate equal to 1 can then be interpreted as a Euclidean 3-vector and hence

$$(u, v, d)^T = (\frac{X}{Z}, \frac{Y}{Z}, \frac{B}{Z})^T. \tag{33}$$

Let therefore $\mathbf{s} = [u, v, d]^T$ be a state in disparity space as before, $\mathbf{s} \in \mathbb{D}^3$, and $\mathbf{W} = [X, Y, Z]^T$ be the corresponding state in 3-D, so that $(u, v, d, 1)^T \simeq P_D(X, Y, Z, 1)^T$, where P_D is now given in Equation (31).

The Fisher information is dependent on the parametrisation of the state space chosen. By the chain rule, we can use the Fisher Information Matrix in disparity space, (24), to find the Fisher Information Matrix in 3-D,

$$I(\mathbf{W}) = J^T I(\mathbf{s}) J, \tag{34}$$

where J is the Jacobian matrix $J = \frac{\partial \mathbf{s}}{\partial \mathbf{W}}$ with components $J_{ij} = \frac{\partial \mathbf{s}_i}{\partial \mathbf{W}_j}$.

Hence, the Jacobian matrix J is

$$\begin{bmatrix} Z^{-1} & 0 & -\frac{X}{Z^2} \\ 0 & Z^{-1} & -\frac{Y}{Z^2} \\ 0 & 0 & -\frac{B}{Z^2} \end{bmatrix} \tag{35}$$

Let I^s_{ij} refer to the individual elements in $I(\mathbf{s})$. Then the Fisher Information Matrix for 3-D becomes

$$I(\mathbf{W}) = \frac{1}{Z^2} \begin{bmatrix} I^s_{11} & 0 & K_{13} \\ 0 & I^s_{22} & -\frac{1}{Z}\left(I^s_{22} \cdot Y\right) \\ K_{13} & -\frac{1}{Z}\left(I^s_{22} \cdot Y\right) & K_{33} \end{bmatrix}, \tag{36}$$

where

$$K_{13} = -\frac{1}{Z}\left(I^s_{11} \cdot X + I^s_{13} \cdot B\right), \tag{37}$$

$$K_{33} = \frac{1}{Z^2}\left(I^s_{11} \cdot X^2 + I^s_{13} \cdot (2X + B)B + I^s_{22} \cdot Y^2\right). \tag{38}$$

The inverse of this gives us the minimum variance on the estimator for 3-D. The information about any estimate in X or Y degrades with $\frac{1}{Z^2}$, and the information about Z degrades with $\frac{1}{Z^4}$. It should be stressed that $I(\mathbf{W})$ has no dependence on disparity space and that the result above is the Fisher Information for a pair of rectified cameras. The inverse of $I(\mathbf{W})$ (CRLB) gives the minimum measurement error covariance for estimation in 3-D.

$$\text{var}(\mathbf{W}) = \frac{Z^2}{B^2} \cdot \frac{1}{I^s_{13}\left(I^s_{11} - I^s_{13}\right)} \begin{bmatrix} L_{11} & L_{12} & L_{13} \\ L_{12} & L_{22} & I^s_{11} Y Z \\ L_{13} & I^s_{11} Y Z & I^s_{11} Z^2 \end{bmatrix} \tag{39}$$

where

$$L_{11} = \left(I^s_{11} X^2 + I^s_{13}(2\,X + B)B\right), \tag{40}$$

$$L_{12} = \left(I^s_{11} X + I^s_{13} B\right) Y, \tag{41}$$

$$L_{13} = \left(I^s_{11} X + I^s_{13} B\right) Z, \tag{42}$$

$$L_{22} = \frac{I_{11} I_{22} Y^2 + I_{13} B^2 \left(I_{11} - I_{13}\right)}{I_{22}}. \tag{43}$$

6 Bayesian Triangulation

In this section, we illustrate the statistical 3-D estimation process with a Bayesian approach to stereo triangulation from a rectified pair of cameras and compare the performance of the estimator to the Cramer-Rao Lower Bound derived in Section 5.2.

6.1 Bayes Update in Disparity Space

We select a prior distribution $\rho_{\mathbf{s}}(\mathbf{s})$ in disparity space and apply Bayes' rule with the likelihood given in Equation (22). Hence the posterior distribution updated with n measurements from the first sensor and m measurements from the second sensor is

$$p_{\mathbf{s}}\left(\mathbf{s}|\mathbf{z}^{[1]}_1, \ldots, \mathbf{z}^{[1]}_n, \mathbf{z}^{[2]}_1, \ldots, \mathbf{z}^{[2]}_m\right) = \frac{\rho_{\mathbf{s}}(\mathbf{s}) g_{\mathbf{s}}\left(\mathbf{z}^{[1]}_1, \ldots, \mathbf{z}^{[1]}_n, \mathbf{z}^{[2]}_1, \ldots, \mathbf{z}^{[2]}_m|\mathbf{s}\right)}{\int \rho_{\mathbf{s}}(\mathbf{s}') g_{\mathbf{s}}\left(\mathbf{z}^{[1]}_1, \ldots, \mathbf{z}^{[1]}_n, \mathbf{z}^{[2]}_1, \ldots, \mathbf{z}^{[2]}_m|\mathbf{s}'\right) d\mathbf{s}'} \tag{44}$$

In order to obtain a closed form solution for the posterior distribution in equation (44), we select a Gaussian prior $\rho_{\mathbf{s}}(\mathbf{s})$ in disparity space. The posterior distribution is then obtained through the recursive application of the Kalman filter update [21, 22].

Suppose that the prior distribution in disparity space is the Gaussian $\mathcal{N}(\mathbf{s}; \mathbf{m}, P)$. Then, for each measurement $\mathbf{z}^{[i]}$ from the camera i, we update the prior with the Gaussian likelihood $\mathcal{N}(\mathbf{z}^{[i]}; H^{[i]}\mathbf{s}, R^{[i]})$ to obtain the updated posterior Gaussian $\mathcal{N}(\mathbf{s}; \hat{\mathbf{m}}(\mathbf{z}^{[i]}), \hat{P})$ as follows.

$$\mathcal{N}(\mathbf{z}^{[i]}; H^{[i]}\mathbf{s}, R^{[i]})\mathcal{N}(\mathbf{s}; \mathbf{m}, P) = q(\mathbf{z}^{[i]})\mathcal{N}(\mathbf{s}; \hat{\mathbf{m}}(\mathbf{z}^{[i]}), \hat{P}) \tag{45}$$

where the updated mean and covariance are calculated with

$$\hat{\mathbf{m}}(\mathbf{z}^{[i]}) = \mathbf{m} + K(\mathbf{z}^{[i]} - \hat{\mathbf{z}}), \tag{46}$$

$$\hat{P} = (I - KH^{[i]})P, \tag{47}$$

and where

$$q(\mathbf{z}^{[i]}) = \mathcal{N}(\mathbf{z}^{[i]}; \hat{\mathbf{z}}, S), \tag{48}$$

$$\hat{\mathbf{z}} = H^{[i]}\mathbf{m}, \tag{49}$$

$$K = P(H^{[i]})^T S^{-1}, \tag{50}$$

$$S = H^{[i]}P(H^{[i]})^T + R^{[i]}. \tag{51}$$

The normalising factor for the Bayes update with a single measurement is $q(\mathbf{z}^{[i]})$, so this term cancels and we are left with the Gaussian $\mathcal{N}(\mathbf{s}; \hat{\mathbf{m}}(\mathbf{z}^{[i]}), \hat{P})$. The update described above can be applied recursively for all of the measurements received from both of the cameras. Hence the resulting posterior is Gaussian in disparity space.

6.2 Reparametrization in to 3-D

In order to obtain the state in 3-D, we need to reparametrize the posterior distribution from disparity space to the 3-D co-ordinate system.

Theorem Suppose that the posterior distribution in disparity space is given by the Gaussian distribution

$$p_{\mathbf{s}}\left(\mathbf{s}|\mathbf{z}_1^{[1]}, \ldots, \mathbf{z}_n^{[1]}, \mathbf{z}_1^{[2]}, \ldots, \mathbf{z}_m^{[2]}\right) = \mathcal{N}(\mathbf{s}; \hat{\mathbf{m}}, \hat{P}). \tag{52}$$

Then the posterior distribution in 3-D is given by

$$\begin{aligned} p_{\mathbf{W}}&\left(\mathbf{W}|\mathbf{z}_1^{[1]}, \ldots, \mathbf{z}_n^{[1]}, \mathbf{z}_1^{[2]}, \ldots, \mathbf{z}_m^{[2]}\right) \\ &= \frac{B}{Z^4} \cdot \mathcal{N}(P_D(\mathbf{W}); \hat{\mathbf{m}}, \hat{P}), \end{aligned} \tag{53}$$

where $P_D(\mathbf{W})$ denotes the disparity space projection of the 3-D point $\mathbf{W} = (X, Y, Z)^T \in \mathbb{R}^3$ in Euclidean coordinates. For example, for a canonical set of rectified cameras with a baseline B, $(X/Z, Y/Z, B/Z)^T = P_D(\mathbf{W})$.

Proof In order to simplify the notation, we write

$$p_{\mathbf{s}}(\mathbf{s}) := p_{\mathbf{s}}\left(\mathbf{s}|\mathbf{z}_1^{[1]}, \ldots, \mathbf{z}_n^{[1]}, \mathbf{z}_1^{[2]}, \ldots, \mathbf{z}_m^{[2]}\right) \tag{54}$$

$$p_{\mathbf{W}}(\mathbf{W}) := p_{\mathbf{W}}\left(\mathbf{W}|\mathbf{z}_1^{[1]}, \ldots, \mathbf{z}_n^{[1]}, \mathbf{z}_1^{[2]}, \ldots, \mathbf{z}_m^{[2]}\right). \tag{55}$$

Then, if we reparameterize the posterior into 3-D, we have [23]:

$$p_{\mathbf{W}}(\mathbf{W}) = p_{\mathbf{s}}\left(P_D(\mathbf{W})\right) \cdot |J_{\mathbf{W}}(\mathbf{P_D}(\mathbf{W}))| \tag{56}$$

where $|J_{\mathbf{W}}(P_D(\mathbf{W}))|$ is the absolute value of the Jacobian determinant, with the Jacobian matrix given in Equation (35). The Jacobian determinant equals B/Z^4 and hence the posterior in (54) follows.

7 Simulations

In this section, we present some simulated results for the Bayesian triangulation and compare with the Cramer-Rao Lower Bound for stereo estimation derived in section 5. The results compare the diagonal elements of the CRLB with the covariance matrix from Bayesian triangulation that has been reparametrized into 3-D. The covariance is found by reparametrising the with

$$I(\mathbf{W}) = J^T P^{-1} J, \tag{57}$$

where P is the covariance estimate determined in disparity space through the Kalman filter update and J is the Jacobian in section 5.2.

The mean in X, Y and Z are found with

$$E[X] = \int_0^\infty X \cdot p_{\mathbf{W}}(\mathbf{W}) d\mathbf{W} \tag{58}$$

$$E[Y] = \int_0^\infty Y \cdot p_{\mathbf{W}}(\mathbf{W}) d\mathbf{W} \tag{59}$$

$$E[Z] = \int_0^\infty Z \cdot p_{\mathbf{W}}(\mathbf{W}) d\mathbf{W} \tag{60}$$

respectively. in order to compute $E[X]$, $E[Y]$ and $E[Z]$.

The simulations results were taken over 100 Monte Carlo runs. The baseline was chosen as $B = 100m$, and the point in 3-D space was $[50, 0, 25]^T$, which gives a point in disparity space of $[2, 0, 2]^T$. The observation covariances R_1 and R_2 were the same for each camera,

$$R_1 = R_2 = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{61}$$

with observation noise standard deviation $\sigma = 0.3$. The filter is initialised with a random mean and large covariance matrix to represent an uninformative prior.

Figure 4 compares the diagonal elements in the CRLB and Bayes updated covariance matrix. Note that the results of the Bayesian triangulation correspond with the CRLB which indicates that the estimator is efficient. This should be expected since the estimator is Bayes-optimal in disparity space and finding the mean in the reparametrised space is therefore also Bayes-optimal. Figure 5 compares the sample variance from 100 Monte Carlo runs using numerical integration in Z. The sample variance is actually below the CRLB in these simulation runs which indicates a possible bias in the experiments. This bias could be introduced in the numerical integration step since we are required to define a finite region in which to integrate.

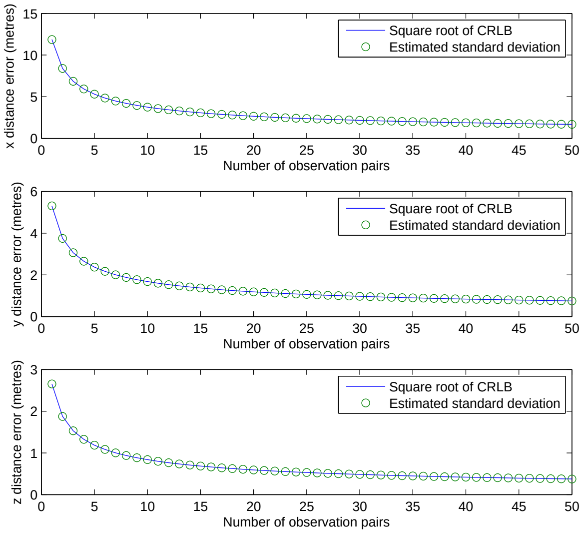

Figure 4: Comparison of the diagonal elements in the CRLB and Bayes updated covariance matrix.

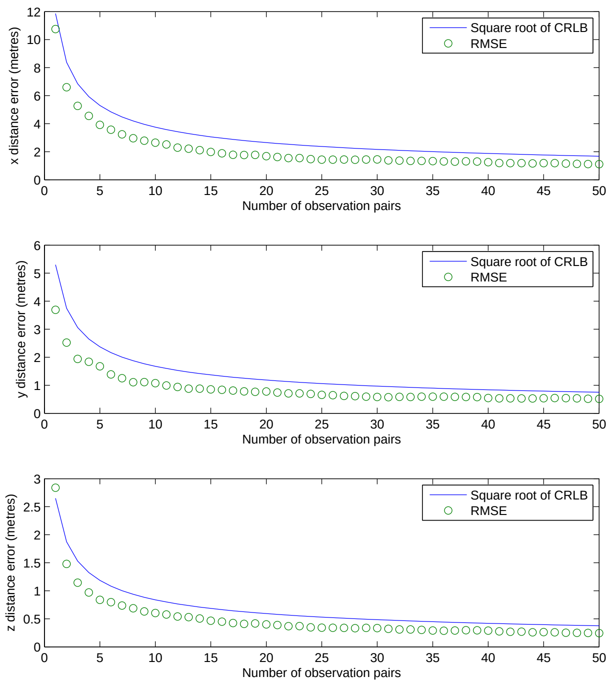

Figure 5: Comparison of the diagonal elements in the CRLB and the RMSE in X, Y and Z for 100 Monte Carlo runs.

8 Conclusion

In this paper, we derived the Cramer-Rao Lower Bound for 3-D state estimation from rectified stereo. Using disparity space as an intermediate space between 3-D and the measurement spaces (the image planes), we derived an analytic form for the Fisher Information matrix and the Cramer-Rao Lower Bound. We illustrated the 3-D estimation with a Bayesian approach to stereo triangulation in disparity space and analysed the performance of the algorithm with respect to the derived bound. The results presented are an essential performance measure for the reliable assessment of object estimation and 3-D reconstruction algorithms from stereo.

Acknowledgements

Dr Clark is a Royal Academy of Engineering/ EPSRC Research Fellow. Dr Ivekovic is a Lessells Scholar of the Royal Society of Edinburgh.

References

[1] B. Matei and P. Meer. A general method for errors-in-variables problems in computer vision. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 18 – 25, 2000.

[2] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley. On the fitting of surfaces to data with covariances. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 22(11):1294–1303, November 2000.

[3] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley. Fns, cfns and heiv: A unifying approach. *Journal of Mathematical Imaging and Vision*, 23(2):175–183, September 2005.

[4] P. Chen and D. Suter. A bilinear approach to the parameter estimation of a general heteroscedastic linear system, with application to conic fitting. *Journal of Mathematical Imaging and Vision*, 28(3):191–208, July 2007.

[5] D. Demirdjian and T. Darrell. Using multiple-hypothesis disparity maps and image velocity for 3-d motion estimation. *International Journal of Computer Vision*, 47(1/2/3):219–228, 2002.

[6] M. Agrawal and K. Konolige. Real-time localization in outdoor environments using stereo vision and inexpensive gps. In *International Conference on Pattern Recognition*, pages 1063 – 1068, 2006.

[7] K.G. Derpanis and P. Chang. Closed-form linear solution to motion estimation in disparity space. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 268 – 275, 2006.

[8] S. Ivekovic and E. Trucco. Articulated 3-d modelling in a wide-baseline disparity space. In *Proceedings of the 4th European Conference on Visual Media Production*, pages 1 – 10, 2007.

[9] N. A. Thacker, A. F. Clark, J. L. Barron, J. R. Beveridge, P. Courtney, W. R. Crum, V. Ramesh, and C. Clark. Performance characterization in computer vision: A guide to best practices. *Computer Vision and Image Understanding*, 109(3):305–334, March 2008.

[10] S. D. Blostein and T. S. Huang. Error analysis in stereo determination of 3-d point positions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(6):752–765, November 1987.

[11] G. Kamberova and R. Bajczy. *Empirical Evaluation Techniques in Computer Vision*, chapter Sensor Errors and the Uncertainties in Stereo Reconstruction, pages 96–116. Wiley-IEEE Computer Society Press, June 1998.

[12] R. M. Haralick. *Performance Characterization in Computer Vision*, chapter Propagating Covariance in Computer Vision, pages 95–114. Computational Imaging and Vision. Kluwer Academic Publishers, 2000. ISBN-13: 978-0792363743, ISBN-10: 0792363744.

[13] M. Farenzena, A. Busti, A. Fusiello, and A. Benedetti. Rigorous accuracy bounds for calibrated stereo reconstruction. In *Proceedings of 17th International Conference on Pattern Recognition*, volume 4, pages 288–292. IEEE Computer Society, August 2004.

[14] A. Bier and L. Luchowski. Error analysis of stereo calibration and reconstruction. In A Gagalowicz and W. Philips, editors, *Proceedings of Mirage 2009*, volume LCNS 5496 of *Lecture Notes in Computer Science*, pages 230–241. Springer Verlag, 2009.

[15] G. Sibley, L. Matthies, and G. Sukhatme. *Robotics Research*, chapter Bias Reduction and Filter Convergence for Long Range Stereo, pages 285–294. Springer Tracts in Advanced Robotics. Springer Verlag, 2007.

[16] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 519 – 526, 2006.

[17] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, 1991.

[18] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter, Particle Filters for Tracking Applications*. Signal Processing. Artech House Publishers, 2004.

[19] S. Ivekovic and D. E. Clark. Multi-object stereo filtering in disparity space. *COGIS (COGnitive systems with Interactive Sensors), Paris. 16-18 November,*, 2009.

[20] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000.

[21] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[22] Y. C. Ho and R. C. K. Lee. A Bayesian approach to problems in stochastic estimation and control. *IEEE Trans. AC*, AC-9:333–339, 1964.

[23] A. Jazwinski. Stochastic processes and filtering theory. *Academic Press*, 1970.