Exact Bias Correction and Covariance Estimation for Stereo Vision

Charles Freundlich
Duke University
charles.freundlich@duke.edu

Michael Zavlanos
Duke University
michael.zavlanos@duke.edu

Philippos Mordohai
Stevens Institute of Technology
mordohai@cs.stevens.edu

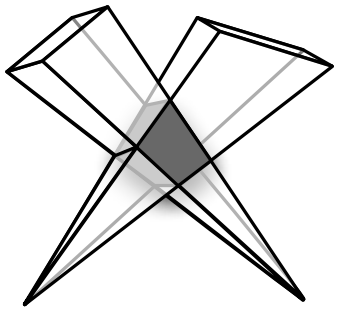
Abstract

We present an approach for correcting the bias in 3D reconstruction of points imaged by a calibrated stereo rig. Our analysis is based on the observation that, due to quantization error, a 3D point reconstructed by triangulation essentially represents an entire region in space. The true location of the world point that generated the triangulated point could be anywhere in this region. We argue that the reconstructed point, if it is to represent this region in space without bias, should be located at the centroid of this region, which is not what has been done in the literature. We derive the exact geometry of these regions in space, which we call 3D cells, and we show how they can be viewed as uniform distributions of possible pre-images of the pair of corresponding pixels. By assuming a uniform distribution of points in 3D, as opposed to a uniform distribution of the projections of these 3D points on the images, we arrive at a fast and exact computation of the triangulation bias in each cell. In addition, we derive the exact covariance matrices of the 3D cells. We validate our approach in a variety of simulations ranging from 3D reconstruction to camera localization and relative motion estimation. In all cases, we are able to demonstrate a marked improvement compared to conventional techniques for small disparity values, for which bias is significant and the required corrections are large.

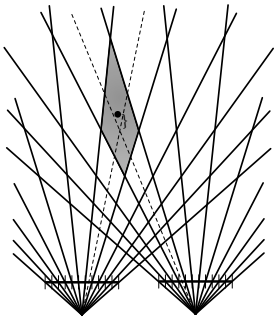
1. Introduction

Pixelation, or quantization, error has long been acknowledged as a limiting factor for stereo vision systems [1, 10, 14, 2]. Since digital camera pairs map points from the continuous world to a discrete set of pixel pairs, their observations are subject to pixelation error. As a result, distinct world points become indistinguishable after projection and reconstruction by triangulation. Pixelation effectively groups sets of points in the world, which we hereafter refer to as *3D cells*, by assigning to each group a single 3D location: the *reconstructed point*. In classical stereo vision the reconstructed point is the intersection of two rays that pass through the two camera centers and the centers of the two corresponding pixels. It represents an entire region formed by the intersection of two infinitely long pyramids created by the camera centers and the pixels, shown in Fig. 1(a). The sides of each pyramid are formed by the planes defined by the camera centers and edges of the pixels. 3D cells are larger, more elongated and asymmetric the further away from the cameras they are. This means that points far from the cameras are often subject to large error.

Assuming that points in the 3D cell are uniformly distributed in space, then the expected value of the reconstructed point is the first moment, or the centroid, of the



(a) Illustration of a 3D cell



(b) Cross-section of the cell

Figure 1. Illustrations of quantization error in stereo vision. (a) a 3D cell created by the intersection of the viewing cones (pyramids) emanating from corresponding pixels in the left and right camera. (b) Cross-section of the 3D cell. The dark dot is the centroid of the cell and should be used as the reconstructed point, instead of the intersection of the rays that is currently used.

3D cell. Conventional stereo vision, however, does not use the centroid as the reconstructed point, causing systematic error. Several authors [1, 14, 16] have reported the bias in long range stereo vision, but to the best of our knowledge, the treatment and correction to the systematic error proposed here is novel. Our approach differs from previous work in that it is an exact and computationally simple solution that is able to remove the bias. An illustration of the proposed solution can be seen in Fig. 1(b) which shows the intersection of the rays (dashed lines) and the centroid of the 3D cell (dark dot). The distance between the two is the bias of conventional reconstruction. The bias can be computed as shown in the remainder of the paper and the coordinates of the reconstructed point can be corrected.

Bias has largely been neglected in stereo vision because, for close ranges (small disparities), the 3D cells are almost symmetric and the difference between the classical reconstructed point and the centroid of the 3D cell is negligible. For long ranges, on the other hand, the variance of the reconstruction errors is very large, both for classical reconstruction methods and for the correction proposed in this paper. Our approach does not guarantee that a single reconstructed point can be converted from an essentially useless outlier to an inlier. Instead, it ensures that the overall accuracy of a stereo system relying on measurement of distant points can be improved significantly on average. We show several examples of such improvements in Section 4.

The second contribution of this paper is a novel way to derive the covariance matrix of a reconstructed point by computing the second moments of a uniform distribution in the corresponding 3D cell. This estimate is exact and more accurate than common approximations that propagate uncertainty from the image plane to the reconstructed points under Gaussian assumptions [10]. Note that higher order moments of these distributions exist, but we ignore them in this work. We propose a test for the validity of our new covariance estimation method and show that it is indeed superior to conventional covariance propagation [10].

In summary, the contributions of this paper are:

- an exact bias correction method for 3D reconstruction, and
- an exact estimation of the covariance matrix of a reconstructed 3D point.

These contributions are validated in simulations of triangulation, localization of a stereo rig given 3D landmarks using Horn's absolute orientation algorithm [7], and relative motion estimation for a stereo rig using Horn's algorithm on reconstructed 3D points. **The only assumption is that points are uniformly distributed in 3D.** We make no assumptions on how the projections of these points or the disparity are distributed in the images.

2. Related Work

While several authors have observed the bias in 3D reconstruction dating back to at least 30 years ago [11], none of them have presented a closed-from, effective correction for it. In most cases, the covariance of the reconstructed point is either only modeled in the Z (range) direction, or is approximated by some form of covariance propagation. In this section, we briefly review methods that have made observations relevant to the problem we address in this paper.

A common assumption that appears reasonable, but is incorrect, is to model the points as generated from uniform distributions on the two image planes. This is equivalent to assuming a uniform distribution on the image plane of the reference camera and a uniform distribution for the disparity. Both formulations lead to non-uniform distributions of points in 3D, since depth and disparity are inversely proportional. We argue that 3D points have equal probability of appearing anywhere in space and start our analysis by modeling the distribution of points in 3D as uniform.

McVey and Lee [11] are arguably the first to publish a study on the error of stereo vision as a function of camera parameters. Later, Blostein and Huang [1] investigated the errors in stereo reconstruction under the assumption that correct correspondences across the two images had been found. They also assumed that the exact unknown 3D point is uniformly distributed in the world. (We make the same assumptions, up to this point.) Blostein and Huang further assume that the reconstructed point is restricted to a quadrilateral in 3D and derive the probability of the errors in all direction being less than specified tolerance values. The major findings are that these probabilities are functions of the disparity and that errors in range dominate errors in other directions. Rodriguez and Aggarwal [14] derived a probability density function for the error in range as a function of the parameters of the imaging system. They assumed that the x coordinates of the corresponding pixels were corrupted independently by uniform quantization noise. Under these assumptions, they were able to compute the expected value of the relative range error given the system parameters. A similar analysis was carried out for convergent, non-parallel stereo by Chang et al. [2].

Matthies and Shafer [10] proposed the use of 3D Gaussian distributions to model triangulation errors, in contrast to previous research that used scalar error models for the range only. They point out that quantization noise on the images follows a uniform distribution and that the resulting distribution in 3D is skewed. For convenience, however, they approximate the uniform image noise by a 2D Gaussian distribution, which they propagate to 3D to obtain the final error distribution. They acknowledge that the approximation does not capture the bias or the tails of the distribution, which are significant for small disparity values. Kriegman et al. [8] also modeled the uncertainty in

image coordinates as Gaussian. Due to the nonlinearity of the triangulation equation, the resulting uncertainty in 3D is not Gaussian, but can be approximated as such locally by linearization and uncertainty propagation.

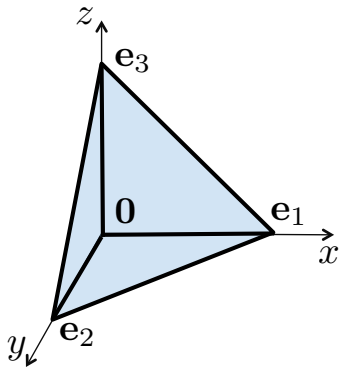
The publication that is more closely related to ours is that of Sibley et al. [16] who proposed a bias reduction technique for long range stereo. Unlike the above approaches, as well as ours, Sibley et al. detected the features with subpixel accuracy and experimentally verified that localization error on the images is normally distributed around the true subpixel locations. Under this assumption, range estimation is biased towards overestimating the true range. Since an analytical solution for the bias appears to be impossible within this framework, the authors resort to an approximate bias reduction technique based on a Taylor series expansion. Due to different assumptions, our results are not directly comparable with those of Sibley et al. It should be noted, however, that our bias correction is closed-form and exact, while we also provide an exact estimate of the covariance of the coordinates of the reconstructed points.

Recently, Fooladgar et al. [4] analyzed the localization error in stereo based on a similar observation to ours: entire regions of 3D points are mapped to the same pixels. This causes uncertainty, which can be modeled by approximating the volume of the intersection of two cones emanating from one pixel each in the left and right camera, assuming that pixels are circular. Due to this choice the shape and volume of the intersection can only be approximated. Besides using the volume as a measure of uncertainty and evaluating the sensitivity with respect to various parameters, no attempt to correct the errors is made.

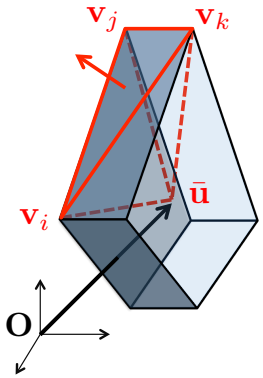
Triangulation [6] also addresses the estimation of the 3D coordinates of a point given the image coordinates of at least two of its projections. Despite this apparent similarity, however, the emphasis of the triangulation literature is on noise due to miscalibration and perturbation in image coordinates. Instead, here, we assume perfect knowledge of calibration. There is also a relationship between our work and inverse depth parametrization [3]. The latter has been proposed to address issues that arise due to the high uncertainty of 3D points with very small disparity values. In our terminology, these difficulties are due to the large size of cells that correspond to small disparity values. We consider structure from motion to be out of the scope of this paper and defer this discussion.

3. Bias Correction and Covariance Estimation

In this section, we compute the first and second moments of a 3D cell that corresponds to a given pair of pixels in the two cameras. The first moments are used for bias correction and the second moments for covariance estimation. We rely on the approach of Lien and Kajiya [9] for integration over polyhedral domains.



(a) The orthogonal unit tetrahedron



(b) The decomposition of a cell

Figure 2. The orthogonal unit tetrahedron has a vertex at the origin and the three unit length edges aligned with the axes. The cells can be decomposed into 12 tetrahedra, each of which can be mapped to the orthogonal unit tetrahedron by an appropriate transformation.

3.1. Bias Correction

Denote by $\mathcal{S}$ any 3D cell as described in Section 1. Since any world-point in $\mathcal{S}$ corresponds to the same pixel pair, the logical choice for the reconstructed point, which we denote by $\boldsymbol{\mu}$, is the expected value of a uniform distribution on $\mathcal{S}$. This expected value is given by

$$\boldsymbol{\mu} = \frac{1}{V_{\mathcal{S}}} \int_{\mathcal{S}} \mathbf{u} \, dV_{\mathbf{u}}, \tag{1}$$

where $V_{\mathcal{S}}$ is the volume of the 3D cell $\mathcal{S}$, $1/V_{\mathcal{S}}$ is its density, and $dV_{\mathbf{u}} = d\mathbf{u}_x d\mathbf{u}_y d\mathbf{u}_z$ is the volume element.

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_8$ denote the eight vertices of the hexahedral cell $\mathcal{S}$, and divide each one of the six quadrilateral faces of $\mathcal{S}$ into two triangles, so that the boundary of $\mathcal{S}$ consists of twelve triangular faces. Let $\mathbf{v}_0$ denote any point in space. Then $\mathcal{S}$ can be decomposed into twelve possibly overlapping (depending on whether $\mathbf{v}_0$ is an interior or exterior point of $\mathcal{S}$) tetrahedra $\mathcal{T}_\ell$, for $\ell = 1, \ldots, 12$, formed by the twelve triangular faces of $\mathcal{S}$ and the common point $\mathbf{v}_0$; see Fig. 2(b). In what follows, we find the expected value $\boldsymbol{\mu}$ of the uniformly distributed cell $\mathcal{S}$ in terms of the geometric

centers $\bar{\mathbf{u}}_\ell$ of the 12 constituent tetrahedra $\mathcal{T}_\ell$.

To find the center $\bar{\mathbf{u}}_\ell$ of the tetrahedron $\mathcal{T}_\ell$, we first express $\mathcal{T}_\ell$ as a linear transformation of the orthogonal unit tetrahedron $\mathcal{T}_o = \text{conv}\{\mathbf{0}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, where $conv\{\cdot\}$ denotes the convex hull and $\mathbf{e}_i$ are the vectors of the standard basis in $\mathbb{R}^3$, as shown in Fig. 2(a). In particular, let

$$\mathbf{u} = A_\ell \mathbf{x} \tag{2}$$

with

$$A_\ell = \left[\begin{array}{c|c|c} \mathbf{v}_i - \mathbf{v}_0 & \mathbf{v}_j - \mathbf{v}_0 & \mathbf{v}_k - \mathbf{v}_0 \end{array} \right], \tag{3}$$

where $\mathbf{v}_0$, $\mathbf{v}_i$, $\mathbf{v}_j$, and $\mathbf{v}_k$ are the four vertices that define the tetrahedron $\mathcal{T}_\ell$. The order of the vertices $\mathbf{v}_i$, $\mathbf{v}_j$, and $\mathbf{v}_k$ is specified clockwise so that the normal vector to the $\mathbf{v}_i\mathbf{v}_j\mathbf{v}_k$ face points away from the tetrahedron; see Fig. 2(b). Then,

$$\begin{aligned} \mathbf{v}_i - \mathbf{v}_0 &= A_\ell \mathbf{e}_1, \\ \mathbf{v}_j - \mathbf{v}_0 &= A_\ell \mathbf{e}_2, \\ \mathbf{v}_k - \mathbf{v}_0 &= A_\ell \mathbf{e}_3, \end{aligned}$$

so A_ℓ maps the orthogonal unit tetrahedron $\mathcal{T}_o = \text{conv}\{\mathbf{0}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to the tetrahedron $\mathcal{T}_\ell^* = \text{conv}\{\mathbf{0}, \mathbf{v}_i - \mathbf{v}_0, \mathbf{v}_j - \mathbf{v}_0, \mathbf{v}_k - \mathbf{v}_0\}$, that is essentially the desired tetrahedron $\mathcal{T}_\ell = \text{conv}\{\mathbf{v}_0, \mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_k\}$ translated by $\mathbf{v}_0$, i.e., by abuse of notation $\mathcal{T}_\ell^* = \mathcal{T}_\ell - \mathbf{v}_0$. Let $\bar{\mathbf{u}}_\ell^*$ denote the center of mass of the tetrahedron $\mathcal{T}_\ell^*$, which is equal to the mean of its four vertices:

$$\bar{\mathbf{u}}_\ell^* = |\det(A_\ell)| A_\ell \bar{\mathbf{x}}_o = \frac{1}{4} V_{\mathcal{T}_\ell} (\mathbf{v}_i + \mathbf{v}_j + \mathbf{v}_k - 3\mathbf{v}_0).$$

The center of mass of the orthogonal unit tetrahedron is $\bar{\mathbf{x}}_o = [1/4 \;\; 1/4 \;\; 1/4]^T$.

To compute the volume of a tetrahedron, we need the relationship between the volume element $dV_\mathbf{x}$ in the $\mathbf{x}$ coordinates and the volume element $dV_\mathbf{u}$ in the $\mathbf{u}$ coordinates, which is

$$dV_\mathbf{u} = \left| \det \left(\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right) \right| dV_\mathbf{x},$$

where $|\cdot|$ denotes the absolute value and the matrix $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ denotes the Jacobian of the transformation (2) defined as

$$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \frac{\partial A_\ell \mathbf{x}}{\partial \mathbf{x}} = A_\ell.$$

Therefore, the term $|\det(A_\ell)|$ captures the amount by which the transformation A_ℓ distorts the volume element. The volume of the tetrahedron $\mathcal{T}_\ell$ is equal to that of $\mathcal{T}_\ell^*$. Given that $V_{\mathcal{T}_o} = \int_{\mathcal{T}_o} dV_\mathbf{x} = \frac{1}{6}$ is the volume of the orthogonal unit tetrahedron $\mathcal{T}_o$,

$$V_{\mathcal{T}_\ell} = V_{\mathcal{T}_\ell^*} = \int_{\mathcal{T}_\ell^*} dV_\mathbf{u} = \det(A_\ell) \int_{\mathcal{T}_o} dV_\mathbf{x} = \frac{|\det(A_\ell)|}{6} \tag{4}$$

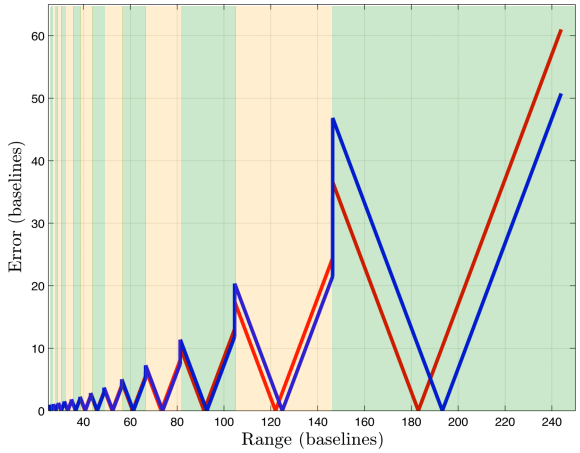


Figure 3. Plotting the distance from world points placed along the optical axis to their representative points using the classical notion of reconstructed points (red) and the proposed correction (blue). The x-axis is range (depth) and is divided in blocks corresponding to disparity levels. Due to quantization, all points within a block are represented by the same point after reconstruction. The error is 0 when the original point is identical to the reconstructed point and increases as the original point moves away from the reconstructed point. Uncorrected reconstruction (red curve) underestimates depth leading to larger errors in the far range, especially for small disparities.

Using the above decomposition of $\mathcal{S}$ in the twelve tetrahedra $\mathcal{T}_\ell$, (1) becomes

$$\boldsymbol{\mu} = \frac{1}{V_\mathcal{S}} \sum_{\ell=1}^{12} \int_{\mathcal{T}_\ell} \mathbf{u} \, dV_\mathbf{u} = \frac{1}{V_\mathcal{S}} \sum_{\ell=1}^{12} V_{\mathcal{T}_\ell} \bar{\mathbf{u}}_\ell, \tag{5}$$

where

$$V_\mathcal{S} = \sum_{\ell=1}^{12} V_{\mathcal{T}_\ell} \tag{6}$$

is the volume of the 3D cell $\mathcal{S}$. A tetrahedron can have positive or negative contribution to the integral over $\mathcal{T}_\ell$ depending on the sign of $\det(A_\ell)$ in (4). In particular, $\text{sign}(\det(A_\ell)) = +1$ if $\mathcal{S}$ and $\mathcal{T}_\ell$ occupy the same half space defined by the common face shared by $\mathcal{S}$ and $\mathcal{T}_\ell$, and $\text{sign}(\det(A_\ell)) = -1$ if $\mathcal{S}$ and $\mathcal{T}_\ell$ occupy different half spaces; see Lien and Kajiya [9]. If $\mathbf{v}_0$ is an interior point of $\mathcal{S}$, as illustrated in Fig. 2(b), then $\text{sign}(\det(A_\ell)) = +1$ for all $\ell = 1, \ldots, 12$. On the other hand, if $\mathbf{v}_0$ is an exterior point of $\mathcal{S}$, then $\text{sign}(\det(A_\ell))$ will be positive for some ℓ and negative for others.

The point $\boldsymbol{\mu}$ in (5) is not equal to the classical notion of a reconstructed point in stereo vision, which is the intersection of the rays that pass through the two camera centers and the two pixel centers. Essentially, Section 3.1 proposes a new set of points with which a stereo rig can represent the world. To illustrate the comparison, consider a sequence of points on a straight line along the viewing direction starting midway between the left and right cameras in a stereo rig

and extending to infinity. Using the classical and corrected methods, the distances from each point on the line to the respective representative points are plotted in Fig. 3. World points that lie on the optical axis fall in 3D cells symmetric about this axis, so the representative point for each method will also lie on the optical axis. This is why we see the error go to zero for each method once in each disparity region and increase linearly in either direction as points on the line get further away from the representative point. Note also in Fig. 3 that the classical reconstructed point typically underestimates the true range, but with the proposed correction, overestimates and underestimates occur almost equally often. In fact, when we sample points in the world uniformly in all three directions in Section 4, we see that reconstruction with the correction is, on average, completely unbiased.

3.2. Covariance Estimation for a 3D Cell

The proposed correction also provides us with a fast and accurate way of computing the covariance of the reconstructed points via the second central moment of the 3D cells. The second central moment of any region $\mathcal{S}$ with uniform density $1/V_{\mathcal{S}}$ is given by

$$\mathbf{C} = \frac{1}{V_{\mathcal{S}}} \int_{\mathcal{S}} (\mathbf{u} - \boldsymbol{\mu})(\mathbf{u} - \boldsymbol{\mu})^T dV_{\mathbf{u}}, \tag{7}$$

where $\boldsymbol{\mu}$ is the centroid of $\mathcal{S}$ determined by (5) and $V_{\mathcal{S}}$ is its volume obtained in (6). As in Section 3.1, we compute the covariance $\mathbf{C}$ of the 3D cell $\mathcal{S}$ by decomposing it into twelve tetrahedra $\mathcal{T}_\ell$, computing their non-central moments, and combining these moments to compute the second central moment of the cell $\mathcal{S}$. In this section, we choose the common vertex $\mathbf{v}_0$ used to form the 12 tetrahedra to be the mean value $\boldsymbol{\mu}$ of $\mathcal{S}$. This choice is convenient because it is equivalent to translating the 3D cell so that its mean value is at the origin, making its central and non-central moments coincide. The other three vertices of each tetrahedron $\mathcal{T}_\ell$ are the same as in Section 3.1 and maintain the positive orientation of the tetrahedra. Summing the non-central moments of the 12 constituent tetrahedra will result in the central moments of the cell $\mathcal{S}$.

In particular, let

$$\mathbf{u} = B_\ell \mathbf{x} \tag{8}$$

with

$$B_\ell = \left[\begin{array}{c|c|c} \mathbf{v}_i - \boldsymbol{\mu} & \mathbf{v}_j - \boldsymbol{\mu} & \mathbf{v}_k - \boldsymbol{\mu} \end{array} \right]. \tag{9}$$

As in Section 3.1, B_ℓ maps the orthogonal unit tetrahedron $\mathcal{T}_o = \text{conv}\{\mathbf{0}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to the tetrahedron $\mathcal{T}_\ell^* = \text{conv}\{\mathbf{0}, \mathbf{v}_i - \boldsymbol{\mu}, \mathbf{v}_j - \boldsymbol{\mu}, \mathbf{v}_k - \boldsymbol{\mu}\}$, that is essentially the desired tetrahedron $\mathcal{T}_\ell = \text{conv}\{\boldsymbol{\mu}, \mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_k\}$ translated by $\boldsymbol{\mu}$, i.e., $\mathcal{T}_\ell^* = \mathcal{T}_\ell - \boldsymbol{\mu}$. Let $C_{\mathcal{T}_\ell}^*$ denote the second moment of the tetrahedron $\mathcal{T}_\ell^*$. Then, applying the change of coordinates (8) we get

$$\begin{aligned} C_{\mathcal{T}_\ell}^* &= \int_{\mathcal{T}_\ell^*} \mathbf{u}\mathbf{u}^T dV_{\mathbf{u}} = \int_{\mathcal{T}_o} B_\ell \mathbf{x}\mathbf{x}^T B_\ell^T \det(B_\ell) dV_{\mathbf{x}} \\ &= \det(B_\ell) B_\ell \left(\int_{\mathcal{T}_o} \mathbf{x}\mathbf{x}^T dV_{\mathbf{x}} \right) B_\ell^T \\ &= \det(B_\ell) B_\ell C_{\mathcal{T}_o} B_\ell^T, \end{aligned} \tag{10}$$

where

$$\begin{aligned} C_{\mathcal{T}_o} &= \int_{x=0}^{1} \int_{y=0}^{1-x} \int_{z=0}^{1-x-y} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}^T dx dy dz \\ &= \begin{bmatrix} 1/60 & 1/120 & 1/120 \\ 1/120 & 1/60 & 1/120 \\ 1/120 & 1/120 & 1/60 \end{bmatrix}. \end{aligned} \tag{11}$$

As in Section 3.1, after decomposing $\mathcal{S}$ into the twelve tetrahedra $\mathcal{T}_\ell$, the covariance in (7) becomes

$$\begin{aligned} \mathbf{C} &= \frac{1}{V_{\mathcal{S}}} \sum_{\ell=1}^{12} \int_{\mathcal{T}_\ell} (\mathbf{u} - \boldsymbol{\mu})(\mathbf{u} - \boldsymbol{\mu})^T dV_{\mathbf{u}} \\ &= \frac{1}{V_{\mathcal{S}}} \sum_{\ell=1}^{12} \int_{\mathcal{T}_\ell^*} \mathbf{u}\mathbf{u}^T dV_{\mathbf{u}} = \frac{1}{V_{\mathcal{S}}} \sum_{\ell=1}^{12} C_{\mathcal{T}_\ell}^* \end{aligned} \tag{12}$$

Since the point $\boldsymbol{\mu}$ chosen to form the twelve tetrahedra $\mathcal{T}_\ell$ is an interior point of $\mathcal{S}$, we have that $\text{sign}(\det(B_\ell)) = +1$ for all $\ell = 1, \ldots, 12$.

Note that for each tetrahedron $\mathcal{T}_\ell$, the vertices of the face that does not contain $\boldsymbol{\mu}$ must have a clockwise orientation so that the normal vector to this face points away from the polyhedron $\mathcal{S}$.

4. Simulation Results

We begin this section by presenting the method we used to evaluate the effectiveness of our contributions. In all simulations, the image resolution is 1025×1025 pixels and the focal length f is 731.93 pixels, corresponding to a field of view of $70°$. The unit of length is always equal to the baseline between the cameras in the stereo pair.

4.1. The *Uncorrected* Method

Throughout these simulations, we will compare our method, which will be referred to as the *corrected method*, with an approximation, which will be referred to as the *uncorrected method*. The uncorrected method does not address bias and approximates the covariance by propagating the covariance of the image noise to 3D, similarly to the work of Matthies and Shafer [10]. Let the image coordinates for a 3D target be x_L, x_R, y denoting the x coordinate in the left and right image and the common y coordinate in

both images, respectively. We will assume that these coordinates are corrupted by independent, uniform quantization noise, which we will approximate as Gaussian. Then the covariance of the image observations is

$$Q \approx \text{diag}\left[\sigma_L^2 \quad \sigma_R^2 \quad \sigma_y^2\right], \tag{13}$$

where σ_L^2, σ_R^2, and σ_y^2 denote the variance of the corresponding observation. In the absence of other information, we assume that all three variances are equal. The coordinates of a reconstructed 3D point $\mathbf{p}_i$ in a coordinate frame anchored to the left camera center, without correction, based on these observations are

$$\mathbf{p}_i \triangleq \mathbf{p}(x_{Li}, x_{Ri}, y_i) = \begin{bmatrix} \frac{bx_{Li}}{x_{Li}-x_{Ri}} \\ \frac{by_i}{x_{Li}-x_{Ri}} \\ \frac{bf}{x_{Li}-x_{Ri}} \end{bmatrix}, \tag{14}$$

where f denotes the focal length of the camera and b is the baseline of the stereo pair. The Jacobian of $\mathbf{p}_i$ is given by

$$J_i = \frac{1}{(x_{Li}-x_{Ri})^2} \begin{bmatrix} -bx_{Ri} & bx_{Li} & 0 \\ -by_i & by_i & b(x_{Li}-x_{Ri}) \\ -bf & bf & 0 \end{bmatrix}, \tag{15}$$

and the covariance of $\mathbf{p}_i$ in the camera coordinate system can be approximated by

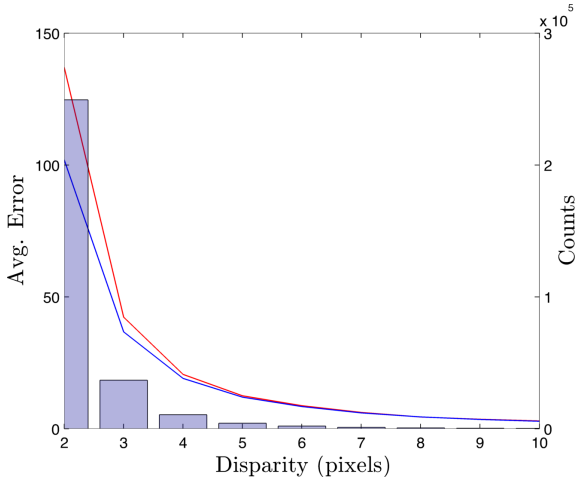
$$U_i = \text{cov}[\mathbf{p}(x_{Li}, x_{Ri}, y_i)] \approx J_i Q J_i^T. \tag{16}$$

4.2. Single-frame Bias Correction

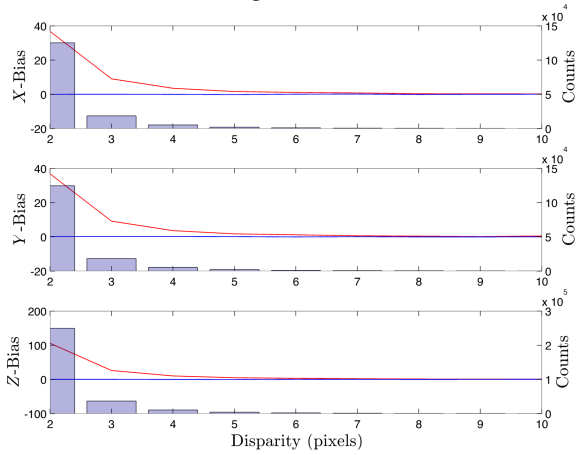
We uniformly generated ten million sample points in front of a stereo camera pair, projected them on the images and triangulated the location of each point with and without the proposed correction. The sample space was a cube, and it extended in all directions up to the range at disparity 1. We discard any points that were not visible to both cameras or that corresponded to cells with disparity 1, which have extremely large volume. Due to the elongation of the cells as one moves away from the cameras, the samples are dominated by those with small disparities.

The left axis in Fig. 4(a) plots the average absolute errors, defined as the Euclidean distance between each reconstructed 3D point and the ground truth, for each disparity region that contained at least 200 sample points, and the right axis measures the height of the bars – the amount of samples at each disparity used to determine the averages. Both methods suffer from severe errors at small disparities in Fig. 4(a). This error is unavoidable due to pixelation, even though the *corrected* method suffers less on average.

Figure 4(b) verifies the zero bias claims of Section 3.1, plotting the average biases for each method at each disparity on the left axes, along with bar graphs to show the number of samples used to compute the each average bias on the right vertical axes. Samples were symmetric about the optical (Z) axis, so in order to observe bias orthogonal to the viewing direction, which would otherwise average out, we partitioned the sample space into half spaces. In the first two boxes of Fig. 4(b), we observe the average bias at each disparity for samples in the $+X$ and $+Y$ half spaces. We do not plot the results for samples that fell in the $-X$ and $-Y$ half spaces, since their average biases were equal in magnitude and opposite in direction compared to the biases for $+X$ and $+Y$.



(a) Average absolute error



(b) Bias per axis

Figure 4. Average absolute error (top) and bias (bottom) versus disparity in units of baselines. Red and blue curves are the average values for the *uncorrected* and *corrected* methods. The bars, measured by the right vertical axes, indicate the number of samples at each disparity.

4.3. Verification of Covariance Estimation

We assessed the accuracy of the proposed covariance estimation method by testing the distribution of the squared Mahalanobis distances from each input 3D point to its rep-

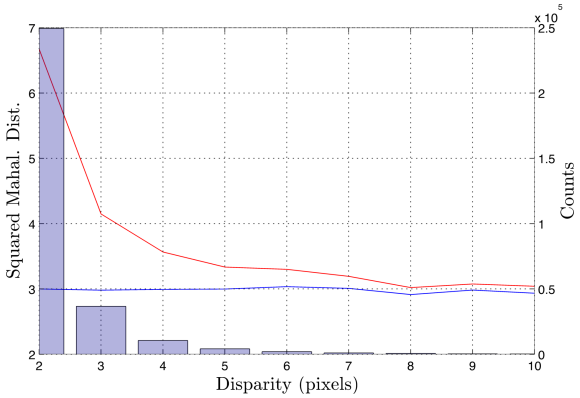


Figure 5. The average squared Mahalanobis distance from the samples to their representative points in units of baselines. Red and blue curves are the average values for the *uncorrected* and *corrected* methods. The bar graph, measured on the right vertical axis, depicts the number of samples at each disparity

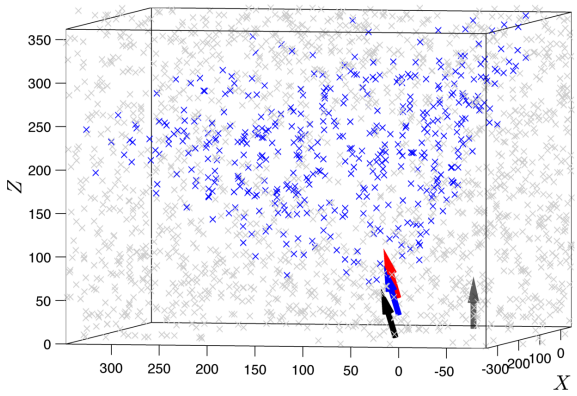


Figure 6. An example of the camera localization simulation. The black arrow is ground truth, the red and blue arrows are the *uncorrected* and *corrected* estimates, and the grey arrow is the coordinate origin. Units are in baselines

resentative point, i.e., the ray intersection if we use the *uncorrected* method and the centroid if we use the *corrected* method. If $\boldsymbol{\mu}_i$ and $\mathbf{C}_i$ are the representative point and predicted covariance of cell $\mathcal{S}_i$, respectively, then the squared Mahalanobis distance from a 3D point $\mathbf{p}_j \in \mathcal{S}_i$ to $\boldsymbol{\mu}_i$ is

$$d^2 = (\mathbf{p}_j - \boldsymbol{\mu}_i)^T \mathbf{C}_i^{-1} (\mathbf{p}_j - \boldsymbol{\mu}_i). \tag{17}$$

Setting

$$\mathbf{q}_j = \mathbf{C}_i^{-\frac{1}{2}} (\mathbf{p}_j - \boldsymbol{\mu}_i), \tag{18}$$

we obtain a set of new 3D vectors $\mathbf{q}_j$ whose coordinates $(\mathbf{q}_j)_k$ are i.i.d. white Gaussian variables. The squared Mahalanobis distance can be written as

$$d^2 = \mathbf{q}_j^T \mathbf{q}_j = \sum_{k=1}^{3} (\mathbf{q}_j)_k^2. \tag{19}$$

But the sum of squares of n i.i.d. $N(0,1)$ random variables is distributed according to the χ^2 distribution with parameter n, where n is also equal to the mean. Thus, in our settings, the mean of the squared Mahalanobis distance d^2 has to be equal to 3, if the distribution mean and covariance have been estimated accurately.

We computed the average of d^2 from (17) among samples to their representative point for every disparity for each method using the data from Section 4.2. Figure 5 plots these averages. Note that, while the *corrected* method relies on the geometry of the 3D cells to predict covariance, the *uncorrected* method requires parameters σ_L^2, σ_R^2, and σ_y^2 from (13), which we set to $1/12$, the variance of uniform noise on the unit interval. While the resulting *uncorrected* covariance is a reasonable approximation for large disparities where bias is small, the squared Mahalanbois distance deviates from the ideal value at small disparities, where the majority of world-points are found. The corrected method obtains the theoretically expected vale of d^2 for all disparities.

4.4. Camera Localization

In this simulation, we attempt to estimate the location of a stereo rig that observes a set of known 3D landmarks. This is accomplished by reconstructing the landmarks based on their image projections and by using Horn's absolute orientation algorithm [7] to estimate the camera's pose. We uniformly generated 100 sets of 5,000 points in a cubic region. For each set of 5,000 points, we placed the stereo cameras in a uniformly random location within the cube. The cameras were always oriented toward the center of the cube to ensure that some landmarks would be visible. Since the difference between the *corrected* and *uncorrected* method is only significant for long ranges, we needed to prevent the rig from using only nearby points for localization, so only samples that projected to pixels with disparity between 3 and 10 were used. This restriction, along with the lower sample density than in Section 4.2, meant that only about 223 landmarks were available in each trial. One such scenario is shown in Fig. 6.

Statistics for 100 camera localization simulations are presented in Table 1. The *corrected* method dominates in terms of position error. The difference is smaller in terms of orientation because distant landmarks serve as reasonable bearing constraints for both methods.

4.5. Relative Pose Estimation

This set of simulations is similar to Section 4.4, but now the coordinates of the landmarks are unknown. The stereo pair makes one observation from an unknown location using the same pose generation parameters as Section 4.4. Then, the cameras undergo another motion that is unknown to them and make an observation from the new pose. The

	Error Statistics	Camera Localization	
		Uncorrected	*Corrected*
Position	Mean	20.20	6.25
	Median	19.96	5.65
Orientation	Mean	1.21	1.16
	Median	1.13	1.08

Table 1. Position and orientation errors for the camera self localization simulations (Section 4.4) in baselines and degrees

goal is to estimate the relative motion between the two observations by registering the two reconstructed sets of landmarks. The main difference between this simulation and the one of Section 4.4 is that now both point sets are noisy.

In each relative pose simulation, we generated 12,000 features because the two random poses may not have large overlapping fields of view, and we wanted to ensure that at least 150 landmarks would be mutually visible. Using this density of samples, and still restricting to disparities 3 through 10, the average number of mutually visible samples was 168. Statistics for 100 relative pose simulations are presented in Table 2, which shows that the correction is effective in this case as well despite the overall increase in noise.

	Error Statistics	Relative Pose Estimation	
		Uncorrected	*Corrected*
Position	Mean	51.49	34.08
	Median	47.34	23.71
Orientation	Mean	7.89	6.88
	Median	5.87	5.15

Table 2. Position and orientation errors for the relative pose estimation simulations (Section 4.5) in baselines and degrees

5. Conclusions

We presented a new method for 3D reconstruction that does not suffer from bias in range or in horizontal displacement. This is accomplished by analytically deriving the bias of conventional triangulation and applying the appropriate corrections to the reconstructed points. We verified that there is no remaining bias in a wide variety of simulations. We also derived the covariance matrix of the expected error of a reconstructed point. These covariance matrices can be calculated easily and accurately using elementary matrix operations. We proved in simulation that this covariance is correct. Since the proposed correction has minimal overhead, we are optimistic that it will be adopted by the research community.

Our findings are directly applicable to the majority of current stereo matching algorithms that treat disparity as a discrete variable. Most of the top performing methods participating in the evaluations hosted by Middlebury [15] and KITTI [5], including those using Markov Random Fields or Semi-Global Matching for optimization, treat disparity as a discrete variable. When needed, subpixel disparity estimates are typically obtained by "cost refinement" as defined by Szeliski and Scharstein [17]. The most common implementation of cost refinement is fitting a parabola around the detected minimum. This does not remove the bias directly since the interpolation is based on cost computations at integer disparity values. The mapping from these disparity values to depth is biased and interpolation adjusts the reconstructed 3D point ignoring the bias. We claim that if the triangulated points corresponding to the integer disparities are corrected using our method, cost refinement will lead to more accurate approximation of the continuous 3D coordinates of the reconstructed point.

Szeliski and Scharstein [17] recommend upsampling the images via cubic interpolation before computing the matching costs. This would reduce the size of the 3D cells, thus reducing bias, but to the best of our knowledge, this recommendation has been ignored by the research community. This may be due to the use of percentage of "bad pixels", with disparity errors above a certain threshold, as the most popular error metric by the benchmarks [15, 5]. We leave the analysis of matching methods that interpolate the input images, as well as of methods that treat disparity as a continuous variable by fitting planes to image patches or by relying on variational techniques, for future work. A possible approach would be to build upon the work of Robinson and Milanfar [13] on performance limits in image registration. We refer interested readers to recent work by Pinggera et al. [12] that evaluates the subpixel accuracy of stereo matching methods, using discrete and continuous disparity representations, on the problem of estimating the disparity and velocity of a single plane at large distances from the cameras.

Acknowledgements: This research has been supported in part by National Science Foundation awards DGE-1068871 and IIS-1217797.

References

[1] S. D. Blostein and T. S. Huang. Error analysis in stereo determination of 3-d point positions. *PAMI*, 9(6):752–766, 1987. 1, 2

[2] C. C. Chang, S. Chatterjee, and P. R. Kube. A quantization error analysis for convergent stereo. In *ICIP*, pages II: 735–739, 1994. 1, 2

[3] J. Civera, A. Davison, and J. Montiel. Inverse depth parametrization for monocular slam. *Robotics, IEEE Transactions on*, 24(5):932–945, 2008. 3

[4] F. Fooladgar, S. Samavi, S. Soroushmehr, and S. Shirani. Geometrical analysis of localization error in stereo vision systems. *IEEE Sensors Journal*, 13(11):4236–4246, 2013. 3

[5] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *CVPR*, pages 3354–3361, 2012. 8

[6] R. Hartley and P. Sturm. Triangulation. *CVIU*, 68(2):146–157, 1997. 3

[7] B. Horn. Closed form solutions of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America A*, 5(7):1127–1135, 1987. 2, 7

[8] D. Kriegman, E. Triendl, and T. O. Binford. Stereo vision and navigation in buildings for mobile robots. *IEEE Transactions on Robotics and Automation*, 5(6):792–803, 1989. 2

[9] S. Lien and J. T. Kajiya. A symbolic method for calculating the integral properties of arbitrary nonconvex polyhedra. *IEEE Computer Graphics and Applications*, 4(10):35–42, 1984. 3, 4

[10] L. H. Matthies and S. A. Shafer. Error modelling in stereo navigation. *IEEE Journal of Robotics and Automation*, 3(3):239–250, 1987. 1, 2, 5

[11] E. S. McVey and J. W. Lee. Some accuracy and resolution aspects of computer vision distance measurements. *PAMI*, 4(6):646–649, 1982. 2

[12] P. Pinggera, D. Pfeiffer, U. Franke, and R. Mester. Know your limits: Accuracy of long range stereoscopic object measurements in practice. In *ECCV*, pages 96–111. Springer, 2014. 8

[13] D. Robinson and P. Milanfar. Fundamental performance limits in image registration. *IEEE Trans. on Image Processing*, 13(9):1185–1199, 2004. 8

[14] J. J. Rodriguez and J. K. Aggarwal. Stochastic analysis of stereo quantization error. *PAMI*, 12(5):467–470, 1990. 1, 2

[15] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1-3):7–42, 2002. 8

[16] G. Sibley, L. Matthies, and G. Sukhatme. Bias reduction and filter convergence for long range stereo. In *Robotics Research*, volume 28, pages 285–294, 2007. 2, 3

[17] R. Szeliski and D. Scharstein. Sampling the disparity space image. *PAMI*, 26(3):419–425, 2004. 8