On the Importance of Modeling Camera Calibration Uncertainty in Visual SLAM

Paul Ozog and Ryan M. Eustice

***Abstract*— This paper reports on methods for incorporating camera calibration uncertainty into a two-view sparse bundle adjustment (SBA) framework. The co-registration of two images is useful in mobile robotics for determining motion over time. These camera measurements can constrain a robot's relative poses so that the trajectory and map can be estimated in a technique known as simultaneous localization and mapping (SLAM). Here, we comment on the importance of propagating uncertainty in both feature extraction and camera calibration in visual pose-graph SLAM. We derive an improved pose covariance estimate that leverages the Unscented Transform, and compare its performance to previous methods in both simulated and experimental trials. The two experiments reported here involve data from a camera mounted on a KUKA robotic arm (where a precise ground-truth trajectory is available) and a Hovering Autonomous Underwater Vehicle (HAUV) for large-scale autonomous ship hull inspection.**

I. Introduction

The spatial relationship between two images of the same scene, taken from different poses, must be estimated to leverage the camera as a viable sensor for real-time, high-precision simultaneous localization and mapping (SLAM). The problem of extracting 3D pose and structure given only an image sequence is a problem widely referred to as structure-from-motion (SFM). Estimating the structure of a scene and motion of cameras given only an image sequence can be done with an iterative technique called sparse bundle adjustment (SBA) [1]. By minimizing the reprojection error using sparse optimization techniques, SBA quickly solves for the relative-pose between camera frames together with a sparse 3D model of the scene with either known or unknown camera internal parameters. The egomotion of the camera as determined by SBA is commonly used in SLAM to constrain the trajectory of a robot [2]–[5]. Moreover, to achieve good qualitative and quantitative results in SLAM, the covariance of the spatial constraints must also be accurately determined. Accurate covariance estimation techniques for camera constraints using keyframe-based SBA is the focus of our work.

SLAM can be represented as a pose-graph optimization problem [6]–[8]. A pose-graph representation encodes the poses of a robot together with spatial constraints between poses, as depicted in Fig. 1. These constraints are typically assumed to be Gaussian with some mean and covariance.

This work was supported by the Office of Naval Research under award N00014-12-1-0092.

P. Ozog is with the Department of Electrical Engineering & Computer Science, University of Michigan, Ann Arbor, MI 48109, USA paulozog@umich.edu

R. Eustice is with the Department of Naval Architecture & Marine Engineering, University of Michigan, Ann Arbor, MI 48109, USA eustice@umich.edu

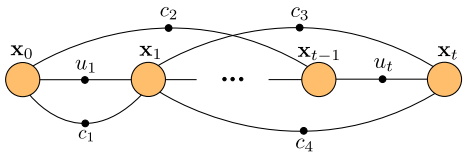


Fig. 1. Simple example of pose-graph visual SLAM with camera constraints, c, odometry constraints, u, and pose nodes, $\mathbf{x}_0 \dots \mathbf{x}_t$. Each constraint is composed of a mean measurement and covariance estimate. Corresponding visual features (not shown) are fed into a pairwise bundle adjustment framework to compute the camera constraints.

Under this assumption, the posterior trajectory and map can be optimized with nonlinear optimization techniques in a back-end SLAM framework, such as incremental smoothing and mapping (iSAM) [9] or Tree-based netwORk Optimizer (TORO) [10]. SLAM back-ends are frequently tasked with recovering marginal posterior confidence regions of each pose in the optimized trajectory. Overconfidence and under-confidence are of particular concern in mobile robotics—overconfidence leads to navigating hazardous areas dangerously [11], while under-confidence can lead to poor performance in data association [12]. This can be corrected when using accurate covariances for pose-graph constraints.

A. Related Work

Because high-quality cameras are relatively affordable and data-rich compared to other sensors like LIDAR, there has been significant research involving the use of cameras on robotic platforms. In some instances, cameras act to augment or even replace odometry sensors like wheel encoders or inertial measurement units (IMUs) altogether [13], [14].

There have been many efforts to model SBA under the presence of uncertainty in feature points and camera calibration values. When the camera calibration is known, Hartley and Zisserman [15] solve sparse bundle adjustment as a nonlinear least square problem using the Levenberg-Marquardt (LM) algorithm. This method computes first-order covariance estimates assuming Gaussian pixel noise, which is either estimated or tuned by the user. However, in general bundle adjustment, calibration parameters can be included as variables to be optimized and there has been some effort to evaluate the effect of these in the pose estimation. Grossman and Santes [16] derived an estimate of the pose, 3D structure, and internal parameters by applying a first-order Taylor expansion of the reprojection error. They consider maximum likelihood estimation (MLE) and maximum *a posteriori* (MAP) estimation of the camera motion, and examined the effect of intrinsic parameters on

the uncertainty of the estimated motion. However, they do not consider lens distortion uncertainty, nor do they consider the case of reconstruction from only two keyframes. In [17], [18] the effect of the calibration matrix uncertainty on the estimated motion and structure was examined. In an effort to evaluate the influence of a poor calibration to the resulting motion, they computed the first-order covariance matrix and showed that the uncertainty propagation is also affected by the geometry of the scene and motion of the cameras.

Uncertainty in lens distortion models is typically not accounted for in visual SLAM research because images are usually preprocessed with distortion removal filters before performing SBA. Though there is some work for pairwise appearance-based matching of visual features in the presence of radial distortion [19], there is little work being done for incorporating lens distortion uncertainty in pairwise egomotion estimation. When the camera calibration is poor, common practice would be to inflate the feature covariance to add additional uncertainty to camera measurements. Our work aims to provide accurate and efficient estimation of relative-pose covariance by modeling feature uncertainty, intrinsic calibration parameter uncertainty, and lens distortion uncertainty.

B. Outline

In §II we begin with a derivation of a first-order covariance estimate that extends previous methods by taking into account uncertainty in the lens distortion parameters. From this derivation, an improved estimate leveraging the Unscented Transform (UT) can be developed. In §III we evaluate the different covariance estimates by comparing each one to a Monte-Carlo distribution inferred from many trials of SBA. In §IV, we apply these techniques to two real-world experimental platforms: a industrial KUKA 7-axis arm and a Hovering Autonomous Underwater Vehicle (HAUV) for autonomous hull inspection. The KUKA arm experiment provides a precise, drift-free ground-truth in a small-scale visual SLAM experiment, while the HAUV provides a large-scale dataset on which to test our methods. Finally, in §V we discuss our results and offer some concluding remarks.

II. Uncertainty Analysis of Sparse Bundle Adjustment

A. Probabilistic Model

The projection of points from a 3D scene onto an image plane is commonly modeled by a camera with a set of internal parameters comprised of focal lengths, f_x and f_y, center of projection, c_x and c_y, and radial distortion coefficients k_1, k_2, k_3, p_1, and p_2. The distortion model used for our work is taken from Brown's model described in [20]. Though the MAP-based estimators from [16], [17], [21] subject these internal parameters to optimization, we treat them as fixed and not subject to change after the one-time initial calibration. This ensures that for every 3D reconstruction estimated from pairwise images, the camera calibration values are the same. This approach is convenient for pose-graphs like the one shown in Fig. 1 because every camera constraint, c, is estimated from the same set of calibration parameters.

For two-view 3D reconstruction, the features in a 3D scene are projected onto the 2D image of each camera. These locations are denoted by $\mathbf{u}^{i1}, i = 1 \dots N$ for camera 1 and $\mathbf{u}^{i2}, i = 1 \dots N$ for camera 2. By stacking these locations in a vector, we define the stacked vector, $\mathbf{X}_u \in \mathbb{R}^{4N}$, containing the feature locations in both cameras. Since our work treats camera 1 and 2 as two camera poses induced by motion, the calibration values for the two cameras are the same. We define the camera calibration values, $\mathbf{X}_c \in \mathbb{R}^9$, as $\mathbf{X}_c = [f_x, f_y, c_x, c_y, k_1, k_2, k_3, p_1, p_2]$. We combine these two vectors into a single measurement vector $\mathbf{X} = [\mathbf{X}_u^\top, \mathbf{X}_c^\top]^\top \in \mathbb{R}^{4N+9}$. We assume this measurement vector to be Gaussian with mean and covariance as in

$$\mathbf{X} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\mathbf{X}_u} \\ \mu_{\mathbf{X}_c} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{X}_u} & 0 \\ 0 & \Sigma_{\mathbf{X}_c} \end{bmatrix} \right). \tag{1}$$

B. Camera Projection Models

1) Feature Correspondence from Essential Matrix: When projecting a general (non-planar) 3D scene onto an image plane, the projective camera matrices $\mathrm{K}[\mathrm{I}|0]$ and $\mathrm{K}[\mathrm{R}|\mathbf{t}]$ for cameras 1 and 2, respectively, project a scene point $\mathbf{P}^i$ to an undistorted location on each image plane. Thus, the pose of camera 1 is at the origin and the pose of camera 2 is described by the transformation $[\mathrm{R}|\mathbf{t}]$. After applying the function $r(\,\cdot\,)$ that dehomogenizes the projected point and applies radial distortion, we have the predicted feature measurements for both cameras:

$$\begin{bmatrix} \widehat{\mathbf{u}}^{i1} \\ \widehat{\mathbf{u}}^{i2} \end{bmatrix} = \begin{bmatrix} r(\mathrm{K}[\mathrm{I}|0]\mathbf{P}^i) \\ r(\mathrm{K}[\mathrm{R}|\mathbf{t}]\mathbf{P}^i) \end{bmatrix}, \tag{2}$$

where K is the camera internal matrix. K and $r(\,\cdot\,)$ are both parametrized by the elements in $\mathbf{X}_c$.

2) Feature Correspondence from Plane-Induced Homography: For two images of a planar scene, homographies relate undistorted features in camera 1 to undistorted features in camera 2 through the plane-induced homography matrix, H. Taking radial distortion into account, the relation between features in camera 1 and camera 2 is given by:

$$\widehat{\mathbf{u}}^{i2} = r\left(\mathrm{H}\; r^{-1}\left(\widehat{\mathbf{u}}^{i1}\right)\right), \tag{3}$$

where $\mathrm{H} = \mathrm{K}\left(\mathrm{R} + \frac{\mathbf{t}\mathbf{n}^\top}{d}\right)\mathrm{K}^{-1}$ is the plane-induced homography. The plane itself, $\pi = [\mathbf{n}, d]^\top$, consists of the normal, $\mathbf{n}$, and perpendicular distance from the camera to the ground plane, d. Like the essential matrix model, K and r are both parametrized by the elements of $\mathbf{X}_c$.

C. Two-view Bundle Adjustment

In general, bundle adjustment minimizes the weighted squared error of measurements and those predicted by some nonlinear projection model, f, as in

$$\widehat{\mathbf{\Theta}} = \underset{\mathbf{\Theta}}{\operatorname{argmin}} \|\mathbf{X} - f(\mathbf{\Theta})\|^2_{\Sigma_{\mathbf{X}}}, \tag{4}$$

where $\|\cdot\|^2_\Sigma$ denotes the squared Mahalanobis distance according to the covariance Σ. $\mathbf{\Theta}$ is a vector of M unknown

parameters containing the relative-pose and 3D structure. This optimization problem forms the MLE of $\mathbf{\Theta}$, assuming the measurements, $\mathbf{X}$, are corrupted by additive Gaussian noise. If the measurements are taken from two cameras, and if the Jacobian of f with respect to $\mathbf{\Theta}$ is sparse, this optimization problem is known as two-view SBA.

For the camera projection models described in §II-B, we adapt this optimization so that $f(\cdot)$ is parametrized by calibration values in $\mathbf{X}_c$ as follows:

$$\widehat{\mathbf{\Theta}} = \underset{\mathbf{\Theta}}{\operatorname{argmin}} \|\mathbf{X}_u - f(\mathbf{\Theta}; \mathbf{X}_c)\|^2_{\Sigma_{\mathbf{X}_u}}. \tag{5}$$

1) Essential Matrix and SBA: For the essential matrix model, $f(\cdot)$ is the stacked projected points given by (2). Thus, the parameter vector consists of the 5-degree of freedom (DOF) relative-pose and sparse 3D scene points, $\mathbf{P}_i$. The relative-pose from camera 2 to camera 1 is modeled as the azimuth, α_{21}, and elevation, β_{21}, of the baseline direction of motion, and the relative Euler orientations, ϕ_{21}, θ_{21}, ψ_{21}, i.e.,

$$\mathbf{z}_{21} = \begin{bmatrix} \alpha_{21} & \beta_{21} & \phi_{21} & \theta_{21} & \psi_{21} \end{bmatrix}^\top.$$

If the baseline direction of motion is assumed to be unit length, then R and $\mathbf{t}$ used in (2) can be easily extracted from $\mathbf{z}_{21}$. Thus, for the essential matrix model, our parameter vector is $\mathbf{\Theta}^\top = [\mathbf{z}_{21}, \mathbf{P}^i \dots \mathbf{P}^N]$.

2) Plane-Induced Homography and SBA: For the homography-based projection model, $f(\cdot)$ is the stacked projected points given by (3). In this case, the parameter vector simply consists of the relative-pose, $\mathbf{z}_{21}$, world plane, π, and the predicted distorted 2D feature locations in camera 1's image plane: $\mathbf{\Theta}^\top = [\mathbf{z}_{21}, \pi, \widehat{\mathbf{u}}^{11}, \dots, \widehat{\mathbf{u}}^{N1}]$.

D. Derivation of First-Order Covariance Estimate

The Levenberg-Marquardt (LM) algorithm solves optimization problems in the form of (4) by linearizing $f(\cdot)$ around the current estimate of $\mathbf{\Theta}$, applying a damped Gauss-Newton iteration to update $\mathbf{\Theta}$, and repeating the process until convergence. Applying this to (5) and using Haralick's framework from [22], we have a scalar-valued cost function that can be written as

$$F(\mathbf{X}, \mathbf{\Theta}) = \|\mathbf{X}_u - \mathrm{J}\mathbf{\Theta}\|^2_{\Sigma_{\mathbf{X}_u}}, \tag{6}$$

where $\mathrm{J} = \frac{\partial f}{\partial \mathbf{\Theta}}$ is the Jacobian of $f(\cdot)$ (which is parametrized by $\mathbf{X}_c$). Then, to first-order,

$$\Sigma_{\mathbf{\Theta}} = \left(\frac{\partial g}{\partial \mathbf{\Theta}}\right)^{-1} \frac{\partial g}{\partial \mathbf{X}}^\top \begin{bmatrix} \Sigma_{\mathbf{X}_u} & 0 \\ 0 & \Sigma_{\mathbf{X}_c} \end{bmatrix} \frac{\partial g}{\partial \mathbf{X}} \left(\frac{\partial g}{\partial \mathbf{\Theta}}\right)^{-1}, \tag{7}$$

where $g(\mathbf{X}, \mathbf{\Theta}) = \frac{\partial F}{\partial \mathbf{\Theta}}$. Applying this technique to (6) we have

$$g(\mathbf{X}, \mathbf{\Theta}) = 2\mathrm{J}^\top \Sigma^{-1}_{\mathbf{X}_u} \mathrm{J}\mathbf{\Theta} - 2\mathrm{J}^\top \Sigma^{-1}_{\mathbf{X}_u} \mathbf{X}_u,$$

with partials

$$\frac{\partial g}{\partial \mathbf{\Theta}} = 2\mathrm{J}^\top \Sigma^{-1}_{\mathbf{X}_u} \mathrm{J},$$
$$\frac{\partial g}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial g}{\partial \mathbf{X}_u} \\ \frac{\partial g}{\partial \mathbf{X}_c} \end{bmatrix} = \begin{bmatrix} -2\Sigma^{-1}_{\mathbf{X}_u} \mathrm{J} \\ \mathrm{A} \end{bmatrix},$$

where $\mathrm{A} \in \mathbb{R}^{9 \times M}$ is a dense matrix with no easily computed closed-form expression. Thus, we compute A using numerical differentiation. After substituting these values into (7), it is not difficult to simplify the covariance estimate to

$$\begin{aligned} \Sigma_{\mathbf{\Theta}} &= \left(\mathrm{J}^\top \Sigma^{-1}_{\mathbf{X}_u} \mathrm{J}\right)^{-1} + \mathrm{B}\Sigma_{\mathbf{X}_c}\mathrm{B}^\top \\ &= \Sigma_{\text{HZ}} \quad + \quad \Sigma_{\text{Fwd}} \end{aligned} \tag{8}$$

where $\mathrm{B} = \frac{1}{2}\left(\mathrm{J}^\top \Sigma^{-1}_{\mathbf{X}_u} \mathrm{J}\right)^{-1} \mathrm{A}^\top$. This covariance estimate may be thought of as the addition of the classic first-order backward propagation of feature covariance, Σ_{HZ}, from Hartley/Zisserman [15] and a forward propagation of calibration covariance, Σ_{Fwd}. For the remainder of this paper, we will refer to (8) as Haralick's method.

E. Extension to Unscented Transform

There are some concerns for using Haralick's method for relative-pose uncertainty. First, the elements of A must be individually computed, which is computationally costly when numerically differentiating Brown's distortion model. Further, because A's elements are second-order derivatives, the error from finite differencing can be large and the differentiation step size needs to be tuned depending on the scene and relative-pose. Finally, linearization error from differentiating the nonlinear lens distortion model can be a significant source of inaccuracy in the covariance estimate.

The form of (8) suggests that instead the Unscented Transform can be used to model the forward propagation of calibration uncertainty, rather than linearizing the camera projection models from §II-B [23]. Essentially, we are replacing the $\mathrm{B}\Sigma_{\mathbf{X}_c}\mathrm{B}^\top$ term from (8) with one computed from the UT by approximating the distribution of camera calibration values with $2\times 9+1 = 19$ sigma points, $\boldsymbol{\mathcal{X}}_i$ (with corresponding weights W_i). By doing so, this method avoids error created by linearizing the camera projection models from §II-B. Our proposed estimate takes the form

$$\Sigma_{\mathbf{\Theta}} = \Sigma_{\text{HZ}} + \Sigma_{\text{UT}}. \tag{9}$$

Like (8), this partitions the relative-pose covariance into two additive terms: a first-order backward propagation of feature covariance, Σ_{HZ}, and a UT-based model of the forward-propagation of camera uncertainty, Σ_{UT}. This method is described in more detail in Algorithm 1.

In the following, we empirically verify our extension to the UT using our simulation and experimentation frameworks. In particular, as $\Sigma_{\mathbf{X}_u} \to 0$, it is clear that the first-order propagation of uncertainty, Σ_{Fwd}, closely approximates a UT-based approximation, Σ_{UT}.

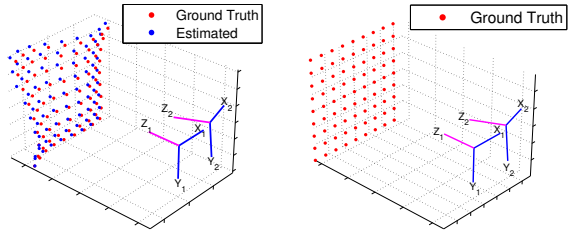
III. Simulated Trials

A. Overview

In our simulation shown in Fig. 2, we fix the two cameras in a single relative-pose that is representative of the typical baseline distance and orientation of two keyframes from the robots we use in experimentation. Sparse ground-truth 3D scene points are also generated, and projected into the image planes of the two cameras. Finally, five hundred realizations

Algorithm 1 Proposed covariance estimate for two-view SBA

Input: Feature locations $\mathbf{X}_u$, camera calibration values $\mathbf{X}_c$, initial guess $\mathbf{\Theta}_0$, covariance matrices $\Sigma_{\mathbf{X}_u}, \Sigma_{\mathbf{X}_c}$
$\widehat{\mathbf{\Theta}} \leftarrow \text{SBA}(f, \mathbf{X}_u, \Sigma_{\mathbf{X}_u}, \mathbf{X}_c, \mathbf{\Theta}_0)$
$\mathrm{J} \leftarrow \frac{\partial f}{\partial \mathbf{\Theta}}|_{\mathbf{\Theta}=\widehat{\mathbf{\Theta}}}$
$\Sigma_{\text{HZ}} \leftarrow \left(\mathrm{J}^\top \Sigma_{\mathbf{X}_u}^{-1} \mathrm{J}\right)^{-1}$
$(\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_{19}, \mathbf{W}) \leftarrow \text{UNSCENTEDXFM}(\mathbf{X}_c, \Sigma_{\mathbf{X}_c})$
for $i = 1 : 19$ **do**
    $\boldsymbol{\mathcal{Y}}_i \leftarrow \text{SBA}(f, \mathbf{X}_u, \Sigma_{\mathbf{X}_u}, \boldsymbol{\mathcal{X}}_i, \widehat{\mathbf{\Theta}})$
end for
$\bar{\mathbf{y}} \leftarrow \sum_{i=1}^{p} W_i \boldsymbol{\mathcal{Y}}_i$
$\Sigma_{\text{UT}} = \sum_{i=1}^{p} W_i \left(\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}}\right)\left(\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}}\right)^\top$
Output: $\Sigma_{\text{HZ}} + \Sigma_{\text{UT}}$



(a) Essential Matrix (b) Homography

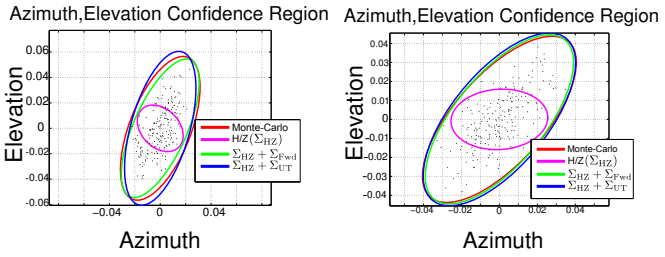
Fig. 2. The virtual scene used for estimating relative-pose with SBA is representative of the robots and cameras used in §IV. For non-planar scenes, shown in (a), we used the essential matrix model from §II-C.1. When the scene is planar, as in (b), we used the plane-induced homography model from §II-C.2.

of the measurement vector, $\mathbf{X}$, were generated from (1) and fed into a SBA estimator, yielding five hundred independent estimates of the parameter vector, $\mathbf{\Theta}$. From these estimates, a Monte-Carlo distribution on the estimated parameter vector was inferred. This process was done for both the essential matrix model and plane-induced homography models presented in §II-B.1 and §II-B.2.

The scene points in the measurement vector where corrupted with unity pixel variance, and the calibration values, though synthetic, were taken such that the calibration quality was quite good by real-world standards. The relationship between feature and calibration uncertainty and its effect on the covariance estimates will be shown in more detail in Fig. 5.

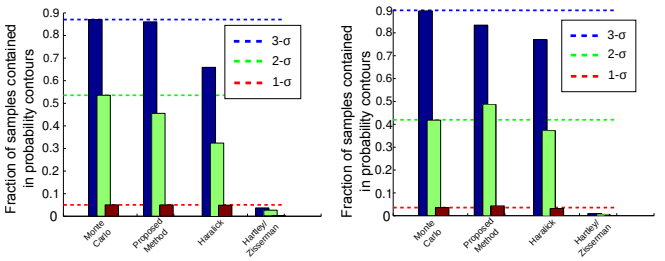
B. Evaluation

Once the optimal parameter vector $\widehat{\mathbf{\Theta}}$ was estimated, the 5-DOF pose covariance estimates were computed and evaluated using: (*i*) a Monte-Carlo (ground-truth) covariance inferred from the 500 independent realizations of egomotion estimates, (*ii*) Hartley and Zisserman's backward-propagation of feature covariance, Σ_{HZ} from (8), (*iii*) Haralick's method, $\Sigma_{\text{HZ}} + \Sigma_{\text{Fwd}}$ (8), and (*iv*) our proposed method, $\Sigma_{\text{HZ}} + \Sigma_{\text{UT}}$ (9). A qualitative comparison of these covariance estimates is shown in Fig. 3. Though Σ_{Fwd} is a slightly better approximation than Σ_{UT} for this plot, this comparison neglects the off diagonal terms in the full 5×5 covariance matrix. Those interested in a full comparison of the covariance estimates should consult Fig. 4.



(a) Essential Matrix (b) Homography

Fig. 3. Marginal 2-dimensional confidence region for the azimuth and elevation of the baseline direction of motion estimated from the simulated scene in Fig. 2. For these plots, the projected 3D scene points were corrupted with unity pixel variance. In red is the Monte-Carlo covariance inferred from the black dots, which are independent estimates of the relative-pose. The covariance estimates which take calibration uncertainty into account approximate the size and orientation of the Monte-Carlo ellipse much more closely than from the Hartley/Zisserman estimate. Because (8) and (9) share the same Σ_{HZ} term, this demonstrates that Σ_{Fwd} and Σ_{UT} are indeed fulfilling a similar role in the overall covariance estimate.



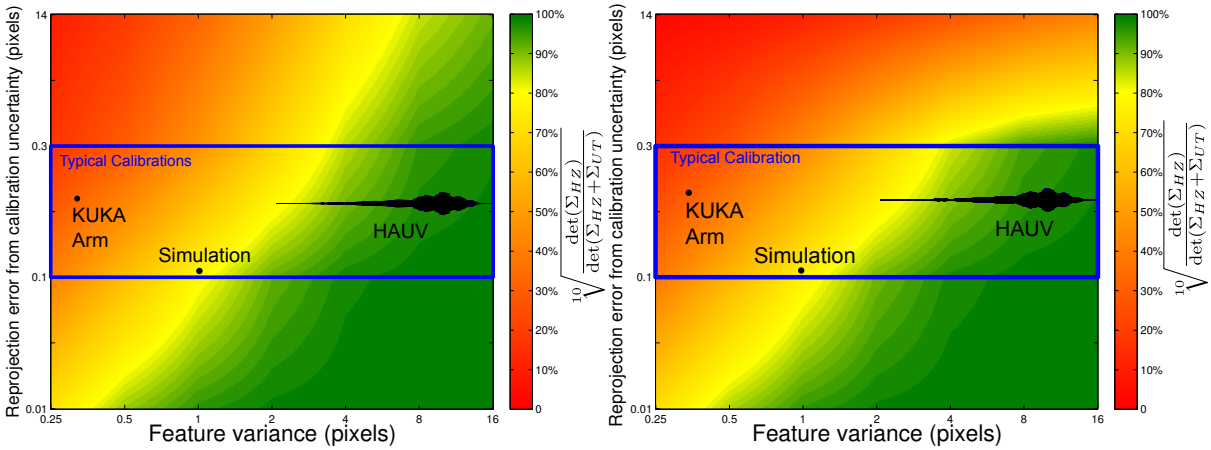
(a) Essential Matrix (b) Homography

Fig. 4. Fraction of SBA-derived relative-poses contained in each of the covariance estimates' 5-dimensional 1, 2, and 3-σ probability contours (red, green, and blue, respectively). Horizontal dotted lines denote Monte-Carlo probability contours. These plots were generated from the same covariance estimates whose 2-dimensional marginal ellipses are shown in Fig. 3. Clearly, our proposed estimate tends to be more representative of the Monte-Carlo probability contours.

The covariances were quantitatively evaluated by counting the number of samples from the true distribution (black dots in Fig. 3) that lie within the 1, 2, and 3-σ probability contours for each of the different covariance estimates (Fig. 4). Ideally, this fraction is approximately equal to the fraction of samples contained in the Monte-Carlo probability contours. For our proposed estimate, the probability contours approximate the Monte-Carlo covariance significantly better than the Hartley/Zisserman and Haralick covariance estimates. These results are described in Fig. 4.

C. Discussion

From these simulation results, we can develop simple criteria to model when it is important to model calibration uncertainty for visual SLAM. To do this, we examine the relative size of Hartley/Zisserman's covariance estimate, Σ_{HZ}, with one that takes calibration uncertainty into account, $\Sigma_{\text{HZ}} + \Sigma_{\text{UT}}$. We model the change in overall size of the covariance estimates by first computing the characteristic lengths of the matrices (taken to be the $2n^{\text{th}}$ root of the determinant). For a 5-DOF pose measurement, the covariance's



(a) Essential Matrix (b) Homography

Fig. 5. Relative sizes of Hartley/Zisserman and proposed covariances for varying levels of feature noise and calibration uncertainty. From this figure, the question of whether or not to model camera calibration uncertainty can be considered as a function of feature noise and calibration uncertainty. Green regions denote regimes where it is not necessary, yellow denotes situations where it is important, and orange denotes situations where it is critically important. In the regions outlined in blue are typical camera calibration values. The calibration uncertainty used in the simulation results from Figs. 3 and 4 are shown at the bottom of these regions. We also experimentally consider two cases: a KUKA arm with very low feature detection error, and an underwater robot that uses SIFT. The thickness for the line labeled "HAUV" encodes the pdf of the feature variance over many images, showing that the majority of feature variances fall in the region where modeling calibration uncertainty is not necessary.

characteristic length will be the 10^{th} root of the determinant, which will have units of radians. We then take the ratio of characteristic lengths,

$$\sqrt[10]{\frac{\det(\Sigma_{\text{HZ}})}{\det(\Sigma_{\text{HZ}}+\Sigma_{\text{UT}})}},$$

to express what percentage of relative-pose uncertainty is captured by modeling just feature noise versus modeling feature noise plus calibration uncertainty. This value is plotted against feature noise and calibration uncertainty in Fig. 5, holding the scene and relative-pose fixed.

It is important to note that Fig. 5 will change with relative-pose, the scene itself, and the location of corresponding features on the image plane. It is infeasible to generalize this information for every application, however, similar simulation analysis can be performed on an application-specific basis. For our purposes, the simulation environment is representative of the kinds of imagery we see on the KUKA 7-axis arm and HAUV experiments, to be shown in §IV.

IV. Experimental Trials

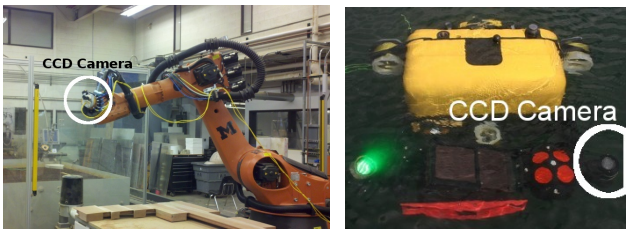
A. KUKA Robotic Arm Experiment

To obtain a ground-truth trajectory for the experimental trials, an industrial-grade 7-Axis KUKA robotic arm (Fig. 6(a)) was used to move and trigger a digital camera pointed toward a calibration target. The robot is able to place itself within a millimeter of a commanded pose anywhere in its surrounding work area—providing very accurate ground-truth for visual SLAM experiments. The calibration target was used to solve for the rigid-body transform (RBT) between the robotic arm end-effector and the camera center of projection; it also provided good features to perform SBA (Fig. 7). The quality of the RBT was assessed by compounding it with the end-effector pose and comparing the result to the camera pose computed from calibration. The translational and rotational errors were minuscule compared to the pose covariances computed from a least squares SLAM backend so their effect on this experiment was ignored.

The uncertainty of the feature detector was also known with high precision, having a measured standard deviation of 0.237 pixels; we assumed it to be isotropic noise as in [24]. The covariance of the calibration values, $\Sigma_{\mathbf{x}_c}$, was given by performing camera calibration over 500 trials with 50 independently-chosen images per trial. The distribution of the calibration values was reasonably Gaussian.

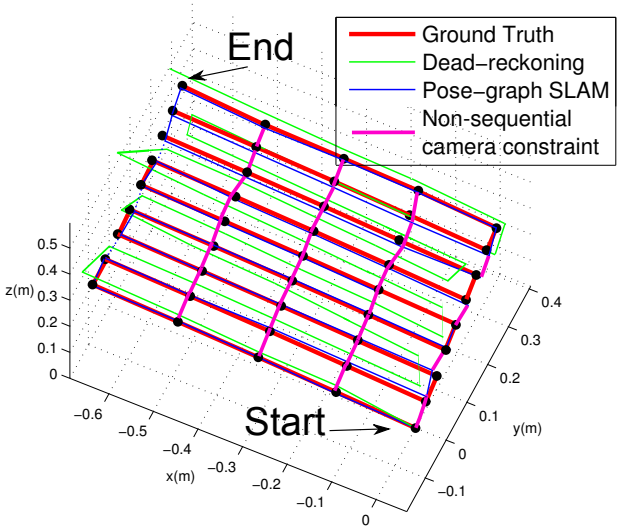
For the 5-dimensional relative-pose estimated from SBA, $\widehat{\mathbf{z}}_{21}$, the distribution of the squared Mahalanobis distances to the true pose will ideally follow a chi-squared distribution with 5 degrees of freedom,

$$\|\widehat{\mathbf{z}}_{21} - \boldsymbol{\mu}_{\mathbf{z}_{21}}\|^2_{\Sigma_{\widehat{\mathbf{z}}_{21}}} \sim \chi^2(5), \tag{10}$$

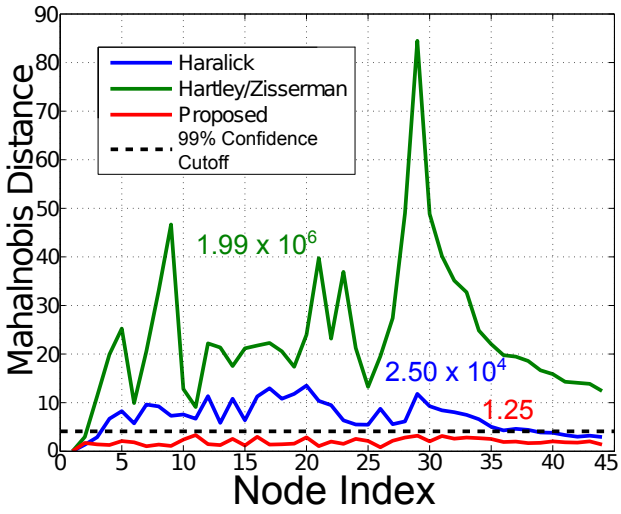


(a) KUKA 7-axis arm (b) Bluefin HAUV

Fig. 6. The robots used for testing the accuracy of various covariance estimates. (a) The industrial-grade 7-axis robotic arm provides precise ground-truth in small-scale visual SLAM experiment. (b) For large-scale visual SLAM, we use data from a HAUV for autonomous hull inspection.



(a) SLAM trajectory



(b) Mahalanobis error

Fig. 8. The results of small-scale visual SLAM on the robotic arm data is greatly improved using our camera measurement covariance estimate. The trajectory using the proposed covariance technique is shown in (a). Using the covariance derived from Haralick and Hartley/Zisserman (not shown), the posterior diverges heavily from ground-truth, suggesting camera-derived relative-pose covariance estimates are over-confident. In (b) is the Mahalanobis distance of each node to ground-truth with the dotted line denoting the 99.0% confidence cutoff. Clearly, our method produces much more probabilistically reasonable confidence regions than Haralick or Hartley/Zisserman. This is further verified with normalized chi-squared error (denoted by the overlain colored text), which is reasonable with our covariance estimate.

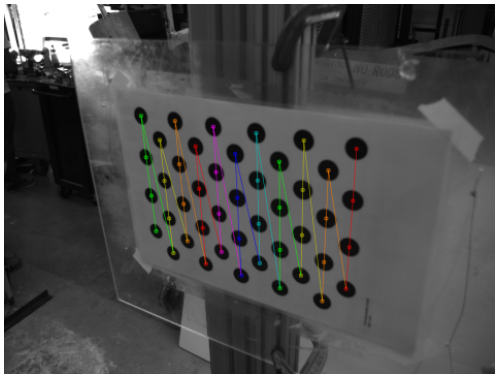


Fig. 7. A CCD camera was affixed to the 7-Axis robot end-effector, shown in Fig. 6(a). The target allowed us to solve for both the RBT from the robot frame to the camera frame and the feature noise, $\Sigma_{\mathbf{X}_u}$. An asymmetric circles pattern was used as in [25].

TABLE I

KUKA DATASET: DISTRIBUTION OF NORMALIZED ERROR TO GROUND TRUTH FOR 28,280 IMAGE PAIRS FOR DIFFERENT COVARIANCE ESTIMATES

	KLD of Normalized Error from $\chi^2(5)$	
Covariance estimate	Essential Matrix	Homography
Hartley/Zisserman ($\Sigma_{\mathbf{\Theta}} = \Sigma_{\text{HZ}}$)	93.172	66.075
Haralick's Method ($\Sigma_{\mathbf{\Theta}} = \Sigma_{\text{HZ}} + \Sigma_{\text{Fwd}}$)	1.036	0.898
Proposed Method ($\Sigma_{\mathbf{\Theta}} = \Sigma_{\text{HZ}} + \Sigma_{\text{UT}}$)	0.251	0.142

where $\boldsymbol{\mu}_{\mathbf{z}_{21}}$ is the ground-truth relative-pose, and $\Sigma_{\hat{\mathbf{z}}_{21}}$ is the upper 5×5 block of the estimate of $\Sigma_{\mathbf{\Theta}}$, which is determined from either Hartley/Zisserman, Haralick's method, or our proposed method.

By computing (10) over many pairs of images taken from the KUKA arm, we are able to verify greatly improved accuracy when taking into account camera calibration uncertainty. As shown in Table I, the distribution of (10) when using our covariance estimate resembles a chi-squared distribution with 5 degrees of freedom more closely because it has a much smaller Kullback-Leibler Divergence (KLD). When discarding the calibration uncertainty, the covariance estimate is much too overconfident and its lack of resemblance to the chi-squared distribution is extreme. This behavior is observed for both the essential matrix and homography registration models.

The high-precision ground-truth from the 7-Axis KUKA robotic arm also allowed us to accurately analyze the performance of small-scale visual SLAM. One such trajectory is shown in Fig. 8(a), which consists of 45 nodes with 127 camera constraints. Three pose-graphs were created for the three covariance estimation techniques: Hartley/Zisserman, Haralick, and our proposed method from §II-E. Odometry was generated by corrupting the ground-truth relative-poses with Gaussian noise. Each pose-graph was solved using the iSAM backend, which computed the posterior covariances used for computing the Mahalanobis distance of the posterior mean to the ground-truth [26]. Using both the Hartley/Zisserman and Haralick covariance estimates, the posterior did not capture the ground-truth within a reasonable confidence region. The proposed method, on the other hand,

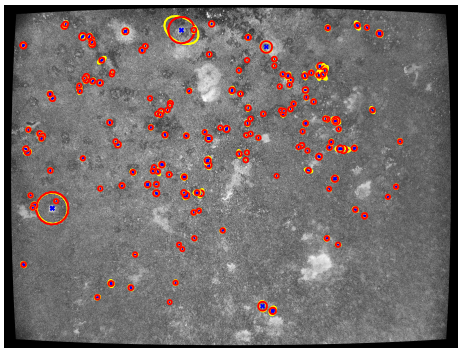


Fig. 9. Our method to estimate SIFT detection error covariance is an approximation of [28]. For typical underwater imagery of ship hulls, our approximation is sufficient. The feature points, denoted with blue marks, have exact covariance estimates in yellow and our approximation in red.

produced confidence intervals that capture the ground-truth with reasonable probability as shown in Fig. 8.

B. HAUV Experiment

We applied the methods in this paper to determine the effect of modeling calibration uncertainty for a HAUV (Fig. 6(b)) performing visually-augmented SLAM on a 183 m-long container ship for autonomous hull inspection. The robot uses monocular camera imagery in a SBA framework to provide sequential measurements and loop-closures to iSAM. A visual overview of the vehicle and the survey is shown in Fig. 10(a).

Feature correspondence in this application uses the SIFT [27] descriptor for its performance in appearance-based matching in underwater images, and because methods exist which estimate the covariance of a SIFT feature point location [28]. For this work, we simplify [28] using an approximation of that estimate so that the covariance of the ith feature, $\Sigma_{\mathbf{X}_{u_i}}$, is given by

$$\Sigma_{\mathbf{X}_{u_i}} = \sigma_i^2 \mathrm{I}_{2\times 2},$$

where σ_i is the scale of the ith feature and $\mathrm{I}_{2\times 2}$ is the 2×2 identity matrix. As shown in Fig. 9, the approximation is acceptable for the underwater imagery captured by the HAUV.

By constructing a pose-graph for each type of camera covariance estimate discussed in this paper, we were able to assess basic qualitative and quantitative performance of each method for visual SLAM on the HAUV. We used four different covariance estimates for these camera constraints: (*i*) Hartley/Zisserman with feature covariance taken to be identity, (*ii*) Hartley/Zisserman with an inflated feature covariance (5 times unity pixel variance), (*iii*) Hartley/Zisserman with feature covariance taken from our approximation of the method discussed in [28], and (*iv*) our UT-based method that uses the feature covariance from (*iii*), but also takes into account calibration uncertainty. Haralick's method was not included in this section because its performance was far too slow for large datasets, particularly for keyframe pairs with many corresponding visual features.

TABLE II

HAUV NORMALIZED CHI-SQUARED ERROR FOR A VARIOUS COVARIANCE ESTIMATION TECHNIQUES (1.0 IS IDEAL)

Covariance estimate	Normalized chi-square error
Σ_{HZ} w/ univariate i.i.d. feature noise	12.864
Σ_{HZ} w/ inflated i.i.d. feature noise	2.477
Σ_{HZ} w/ SIFT covariance estimate	1.743
$\Sigma_{\mathrm{HZ}} + \Sigma_{\mathrm{UT}}$ w/ SIFT, calibration	1.336

As shown in Table II, incorporating camera calibration uncertainty in two-view SBA can offer better chi-squared error results. In particular, the normalized chi-squared error is closer to the expected value of 1.0 when using our technique. However, from a qualitative standpoint, the results are very similar to modeling SIFT covariance only, which agrees with our simulation conclusion, shown earlier in Fig. 5.

V. CONCLUSION

This work presented and experimentally evaluated a method for partitioning the covariance estimate of two-view relative-pose into two additive terms: a first-order backward propagation of feature covariance and a UT that does a forward-propagation of camera uncertainty. This proposed technique has superior accuracy to other methods that rely on differentiating nonlinear transformations. This technique offers improved SBA-derived relative-pose covariances at the cost of extra computation. We establish a simple relationship between feature noise and calibration uncertainty as a basis for determining whether this extra effort is necessary in pose-graph visual SLAM.

We showed two visual SLAM experiments with varying regimes of feature detection accuracy: a camera-equipped KUKA arm using aid of a fiducial marker for feature detection, and a HAUV that uses SIFT to detect visual features on a ship hull. We experimentally verified that modeling camera calibration uncertainty is necessary for such situations where the feature detection error is small. On the other hand, using this technique for the HAUV only offers minor improvements in the chi-squared error and qualitative accuracy of the posterior trajectory.

ACKNOWLEDGMENTS

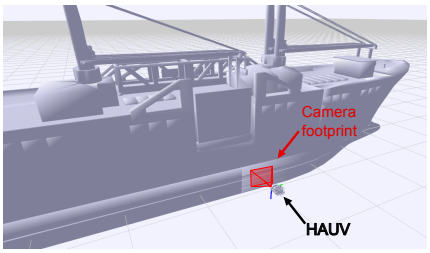
We are grateful to Wes McGee at the Digital Fab Lab at the University of Michigan Taubman College for letting us use the lab's robotic workcell for our experiments.

REFERENCES

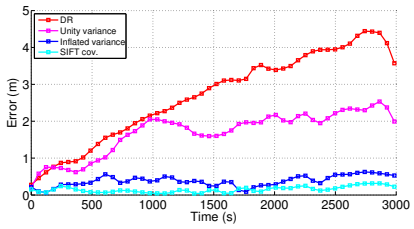
[1] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *Proc. Intl. Workshop on Vision Algorithms: Theory and Practice*, ser. ICCV '99. London, UK: Springer-Verlag, 2000, pp. 298–372.

[2] A. Kim and R. M. Eustice, "Real-time visual SLAM for autonomous underwater hull inspection using visual saliency," *IEEE Trans. Robot.*, 2013, in Press.

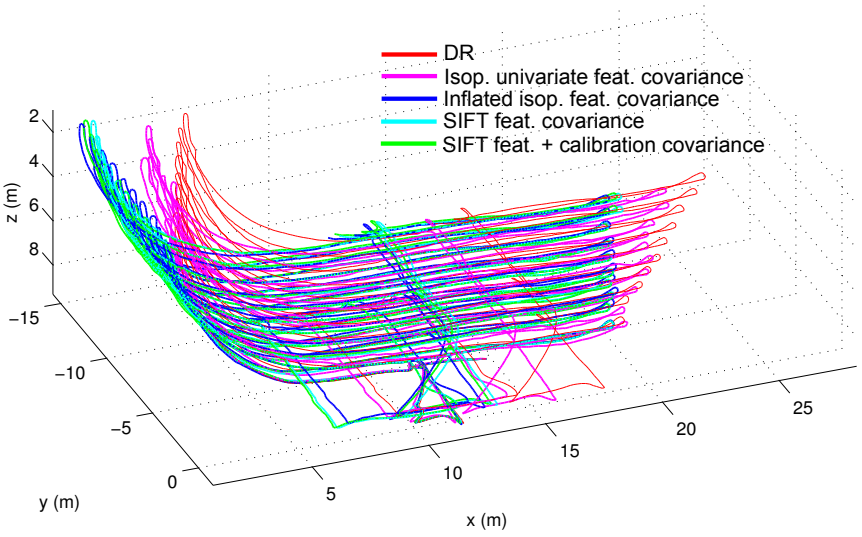
[3] M. Kaess and F. Dellaert, "Probabilistic structure matching for visual SLAM with a multi-camera rig," *Comput. Vis. Image Understanding*, vol. 114, pp. 286–296, Feb 2010.



(a) HAUV visualization



(b) Error with respect to proposed covariance estimate (green curve from (c))



(c) Posterior trajectory

Fig. 10. The HAUV is deployed alongside large ship hulls, similar to the 3D model in (a). When using the commonly-assumed isotropic unity variance pixel noise, the posterior SLAM trajectory is strongly warped due to overconfident camera measurements in the pose-graph. When modeling the feature uncertainty with SIFT covariance [28], the trajectory more closely follows the shape of a ship hull. By looking at the x, y error in (b), it's clear that modeling camera calibration adds no noticeable qualitative effect on the posterior. As discussed in Fig. 5, this is because the feature covariance, $\Sigma_{\mathbf{x}_u}$, dominates.

[4] A. Stewart and P. Newman, "Laps—localisation using appearance of prior structure: 6-dof monocular camera localisation using prior pointclouds," in *Proc. IEEE Int. Conf. Robot. and Automation*, Minnesota, USA, May 2012, pp. 2625–2632.

[5] C. Beall, F. Dellaert, I. Mahon, and S. Williams, "Bundle adjustment in large-scale 3d reconstructions based on underwater robotic surveys," in *Proc. IEEE/MTS OCEANS Conf. Exhib.*, June 2011, pp. 1–6.

[6] F. Dellaert and M. Kaess, "Square root SAM: Simultaneous localization and mapping via square root information smoothing," *Int. J. Robot. Res.*, vol. 25, no. 12, pp. 1181–1203, 2006.

[7] R. M. Eustice, H. Singh, and J. J. Leonard, "Exactly sparse delayed-state filters for view-based SLAM," *IEEE Transactions on Robotics*, vol. 22, no. 6, pp. 1100–1114, Dec. 2006.

[8] E. Olson, J. Leonard, and S. Teller, "Fast iterative alignment of pose graphs with poor initial estimates," in *Proc. IEEE Int. Conf. Robot. and Automation*, 2006, pp. 2262–2269.

[9] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Trans. Robot.*, vol. 24, no. 6, pp. 1365–1378, Dec. 2008.

[10] G. Grisetti, D. Lodi Rizzini, C. Stachniss, E. Olson, and W. Burgard, "Online constraint network optimization for efficient maximum likelihood map learning," in *Proc. IEEE Int. Conf. Robot. and Automation*, Pasadena, CA, May 2008, pp. 1880–1885.

[11] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. van Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, A. Nefian, and P. Mahoney, "Stanley: The robot that won the DARPA Grand Challenge," *J. Field Robot.*, vol. 23, no. 9, pp. 661–692, 2006.

[12] J. Neira and J. Tardos, "Data association in stochastic mapping using the joint compatibility test," *IEEE Trans. Robot. Autom.*, vol. 17, no. 6, pp. 890–897, Dec. 2001.

[13] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2003, pp. 1403–1410.

[14] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 1, 2004, pp. 652–659.

[15] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[16] E. Grossmann and J. Santos-Victor, "Uncertainty analysis of 3d reconstruction from uncalibrated views," *Image and Vision Computing*, vol. 18, pp. 686–696, 2000.

[17] M. Zucchelli and J. Košecká, "Motion bias and structure distortion induced by intrinsic calibration errors," *Image Vision Comput.*, vol. 26, pp. 639–646, May 2008.

[18] T. Svoboda and P. Sturm, "A badly calibrated camera in ego-motion estimation-propagation of uncertainty," in *Computer Analysis of Images and Patterns*. Springer, 1997, pp. 183–190.

[19] M. Lourenco, J. P. Barreto, and A. Malti, "Feature detection and matching in images with radial distortion," in *Proc. IEEE Int. Conf. Robot. and Automation*, Anchorage, AK, May 2010, pp. 1028–1034.

[20] D. Brown, "Decentering distortion of lenses," *Photometric Engineering*, vol. 32, no. 3, pp. 444–462, 1966.

[21] T. Svoboda and P. Sturm, "What can be done with a badly calibrated camera in ego-motion estimation?" Czech Technical University, Tech. Rep., 1996.

[22] R. Haralick, "Propagating covariance in computer vision," in *Proc. Int. Conf. Pattern Recog.*, vol. 1, Jerusalem, Israel, Oct. 1994, pp. 493–498.

[23] S. Julier, "The scaled unscented transformation," in *Proc. Amer. Control Conf.*, vol. 6, Anchorage, AK, May 2002, pp. 4555–4559.

[24] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proc. IEEE Int. Conf. Comput. Vis.*, Corfu, Greece, Sept. 1999, pp. 666–673.

[25] G. Mateos, "A camera calibration technique using targets of circular features," in *5th Ibero-America Symposium On Pattern Recognition (SIARP)*, 2000.

[26] M. Kaess and F. Dellaert, "Covariance recovery from a square root information matrix for data association," *Robot. and Auton. Syst.*, vol. 57, pp. 1198–1210, Dec. 2009.

[27] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[28] B. Zeisl, P. Georgel, F. Schweiger, E. Steinbach, and N. Navab, "Estimation of location uncertainty for scale invariant feature points," in *Proc. British Mach. Vis. Conf.*, 2009.