Bias Reduction for Stereo based Motion Estimation with Applications to Large Scale Visual Odometry

Gijs Dubbelman
TNO Defence, Security and Safety
Oude Waalsdorperweg 63, The Hague, NL
gijs.dubbelman@tno.nl

Frans C.A. Groen
University of Amsterdam
Kruislaan 403, Amsterdam, NL
F.C.A.Groen@uva.nl

Abstract

This contribution addresses the problem of bias in stereo based motion estimation. Using a biased estimator within a visual-odometry system will cause significant drift on large trajectories. This drift is often minimized by exploiting auxiliary sensors, (semi-)global optimization or loop-closing. In this paper it is shown that bias in the motion estimates can be caused by incorrect modeling of the uncertainties in landmark locations. Furthermore, there exists a relation between the bias, the true motion and the distribution of landmarks in space. Guided by these observations, a novel bias reduction technique has been developed. The core of the proposed method is computing the difference between motion estimates obtained using dissimilar heteroscedastic landmark uncertainty models. This approach is accurate, efficient and does not rely on auxiliary sensors, (semi-)global optimization or loop-closing. To show the real-world applicability of the proposed method, it has been tested on several data-sets including a challenging 5 km urban trajectory. The gain in performance is clearly noticeable.

1. Introduction

In this article, the focus is on accurate ego-motion estimation of a moving stereo-camera. If the camera is mounted on a vehicle this is also known as stereo-based visual-odometry. Stereo-processing allows estimation of the three dimensional (3D) location and associated uncertainty of landmarks observed by a stereo-camera. Subsequently, 3D point clouds can be obtained for each stereo-frame. By establishing correspondences between visual landmarks, the point clouds of two successive stereo-frames, i.e. from $t-1$ to t, can be related to each other. From these two corresponding point clouds the pose at t relative to the pose at $t-1$ can be estimated. The position and orientation of the stereo-rig in the global coordinate frame can be tracked by integrating all the relative-pose estimates. In the past decades several methods have been proposed to estimate motion parameters from corresponding 3D point patterns [1, 7, 11, 23, 29]. The uncertainty of stereo-reconstruction is heteroscedastic i.e. inhomogeneous (not the same for each point) and anisotropic (different in each dimension). For this type of noise the Heteroscedastic Error-In-Variables (HEIV) estimator was developed by Matei and Meer [18, 19]. By taking heteroscedasticity into account the HEIV estimator attains higher accuracy than estimators which wrongly assume independent identically distributed noise. The approaches mentioned so far directly minimize a 3D error. An alternative is minimizing an error in image space. While these approaches can also be used for accurate stereo-based visual-odometry [4], they are not the focus of this work.

In general, vision based approaches to motion estimation are susceptible to outlier landmarks. Sources of outlier landmarks range from sensor noise and correspondence errors to independently moving objects such as cars or people that are visible in the camera views. Robust estimation techniques, such as RANSAC [9], are therefore frequently applied [21, 22, 24]. Recently, Dubbelman et al. [6] proposed using Expectation Maximization directly on the motion space SE(3). This technique depends on the theory of Riemannian Geometry. In the case of visual-odometry it has advantages in terms of accuracy and efficiency.

The integration of relative-pose estimates to track the global-pose is sensitive to error-propagation, i.e. small frame-to-frame motion errors eventually cause large errors in the estimated trajectory. In the literature several vision and non-vision based approaches can be found to minimize this drift. For example, techniques such as (semi-)global optimization like (sliding-window) bundle adjustment [27,28], loop-closing [15,30] or exploiting auxiliary sensors such as an IMU [14,24] are frequently used. One of the most popular approaches of the past decade is Simultaneous Localization and Mapping (SLAM) and many stereo-vision SLAM approaches exists [2,8,25,26]. The benefit of SLAM is that

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

it can combine all previous mentioned methods i.e. global optimization, loop-closing and using auxiliary sensors in one mathematical framework.

In this contribution, however, the focus is on achieving accurate visual-odometry results for loop-less trajectories without relying on auxiliary sensors or (semi-)global optimization. Therefore, it is investigated if, and how, drift in stereo-vision based relative-pose estimates can be related to structural errors i.e. bias in the optimization process. In sec. 2 and 3 HEIV based motion estimation using stereo-vision is analyzed. It will be shown that bias in the estimated motion is caused by approximations made on the noise distributions governing the data. The proposed bias reduction technique is then introduced in sec. 4. The performance gain due to bias reduction is presented for simulated data in sec. 5 and for real-world data in sec. 6. The conclusions and a discussion can be found in sec. 7.

2. HEIV motion estimation

All static landmarks observed by a stereo-camera which moves according to a 3D motion, described using a rotation matrix $\bar{\mathbf{R}}$ and translation vector $\bar{\mathbf{t}}$, obey $\bar{\mathbf{v}}_i = \bar{\mathbf{R}}\bar{\mathbf{u}}_i + \bar{\mathbf{t}}$. Here $\bar{\mathbf{v}}_i$ and $\bar{\mathbf{u}}_i$ are the noise free coordinates of a particular landmark observed relative to the coordinate frame of the moving sensor at respectively $t-1$ and t. Two corresponding landmarks $\bar{\mathbf{v}}_i$ and $\bar{\mathbf{u}}_i$ can be combined into the matrix

$$\bar{\mathbf{M}}_i = \begin{bmatrix} \bar{v}_x - \bar{u}_x & 0 & -\bar{v}_z - \bar{u}_z & \bar{v}_y - \bar{u}_y \\ \bar{v}_y - \bar{u}_y & \bar{v}_z + \bar{u}_z & 0 & -\bar{v}_x - \bar{u}_x \\ \bar{v}_z - \bar{u}_z & -\bar{v}_y - \bar{u}_y & \bar{v}_x + \bar{u}_x & 0 \end{bmatrix} . \tag{1}$$

Then the motion constraint between $\bar{\mathbf{v}}_i$ and $\bar{\mathbf{u}}_i$ can also be expressed as

$$\bar{\mathbf{M}}_i\bar{\mathbf{q}} + \bar{\mathbf{Q}}\bar{\mathbf{t}} = \mathbf{0} \ , \tag{2}$$

Ohta and Kanatani [23] where $\bar{\mathbf{q}} = [q, q_i, q_j, q_k]^\top$ is the quaternion expressing the rotation $\mathbf{R}$ and

$$\bar{\mathbf{Q}} = \begin{bmatrix} -q & -q_k & q_j \\ q_k & -q & -q_i \\ -q_j & q_i & -q \end{bmatrix} . \tag{3}$$

Clearly $\bar{\mathbf{v}}_i$ and $\bar{\mathbf{u}}_i$ are not observed directly. According to [18, 19] the noisy observations of $\bar{\mathbf{v}}_i$ and $\bar{\mathbf{u}}_i$ can be modeled with.

$$\mathbf{v}_i = \bar{\mathbf{v}}_i + \varepsilon_{\mathbf{v}_i} \ , \qquad \mathbf{u}_i = \bar{\mathbf{u}}_i + \varepsilon_{\mathbf{u}_i} \ , \tag{4}$$

where $\varepsilon_{\mathbf{v}_i}$ and $\varepsilon_{\mathbf{u}_i}$ are drawn from symmetric and independent distributions with zero mean and data dependent covariances $\mathcal{S}(0, \eta\boldsymbol{\Sigma}_{\mathbf{v}_i})$ and $\mathcal{S}(0, \eta\boldsymbol{\Sigma}_{\mathbf{u}_i})$ respectively. It is thus assumed that the noise distribution can be described using a Gaussian. Note that the covariances only need to be known up to a common scale factor η. Clearly the noise governing the observed data is modeled as heteroscedastic, i.e. anisotropic and inhomogeneous. Analogue to eq. 1 the observed landmarks can be combined into the matrix $\mathbf{M}$. From the matrices $\mathbf{M}_i$ and $\bar{\mathbf{M}}_i$ the vectors $\mathbf{w}_i = \left[\mathbf{m}_i^1, \mathbf{m}_i^2, \mathbf{m}_i^3\right]^\top$ and $\bar{\mathbf{w}}_i = \left[\bar{\mathbf{m}}_i^1, \bar{\mathbf{m}}_i^2, \bar{\mathbf{m}}_i^3\right]^\top$ can be constructed, where the superscript is used to index the rows of the $\mathbf{M}_i$ and $\bar{\mathbf{M}}_i$ matrices. The noise effecting $\mathbf{w}_i$ will be denoted as $\mathbf{C}_i$ and it can be computed from $\boldsymbol{\Sigma}_{\mathbf{z}_i}$ and $\boldsymbol{\Sigma}_{\mathbf{u}_i}$ [18]. The HEIV based motion estimator then minimizes the following objective function

$$[\mathbf{q}, \mathbf{t}] = \arg\min_{\{\mathbf{q},\mathbf{t},\bar{\mathbf{w}}\}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathbf{w}_\mathrm{i} - \bar{\mathbf{w}}_\mathrm{i})^\top \mathbf{C}_\mathrm{i} (\mathbf{w}_\mathrm{i} - \bar{\mathbf{w}}_\mathrm{i}) \quad . \tag{5}$$

under the constraint eq. 2. In [18] it is explained that a solution to this non-linear problem can be obtained by iteratively solving a generalized eigen problem. More detail on this approach can be found in [17, 19]. In the rest of the paper $\{\mathbf{R}, \mathbf{t}\} = \mathcal{HEIV}(\mathbf{v}, \boldsymbol{\Sigma}_\mathbf{v}, \mathbf{u}, \boldsymbol{\Sigma}_\mathbf{u})$ denotes the motion estimated on the landmarks $\mathbf{v}_i$ and $\mathbf{u}_i$ with the covariances $\boldsymbol{\Sigma}_{\mathbf{v}_i}$ and $\boldsymbol{\Sigma}_{\mathbf{u}_i}$ for $i = 1..n$.

In [19] it is shown that optimization approaches such as Generalized Total Least Squares (GTLS), Sampson method and the *renormalization* approach of Kanatani [13] can be derived from HEIV when simplifications are assumed. Furthermore, its accuracy is at least equal to other advanced optimization techniques such as the Fundamental Numerical Scheme [3] and Levenberg-Marquardt, whereas HEIV has better convergence and is less influenced by the initial parameters. The benefit of using HEIV has been noted for many computer vision problems such as motion estimation, camera calibration, tri-focal tensor estimation and structure from motion. The relevant references can be found in [19].

In the derivation of the algorithm an implicit assumption, apart from symmetry, is made on the error models governing the observations. Recall that the observations are modeled with an additive noise term $\varepsilon_{\mathbf{z}_i}$, drawn from $\mathcal{S}(0, \eta\boldsymbol{\Sigma}_{\mathbf{z}_i})$, on the true data i.e. $\mathbf{z}_i = \bar{\mathbf{z}}_i + \varepsilon_{\mathbf{z}_i}$. Here $\mathbf{z}_i$ is either $\mathbf{v}_i$ or $\mathbf{u}_i$. Clearly the noise governing the observations can not depend, in the physical sense, on those same observations that are being generated. When the error is modeled as additive on the true data the general heteroscedastic model is

$$\mathbf{v}_i = \bar{\mathbf{v}}_i + \varepsilon_{\bar{\mathbf{v}}_i} \ , \qquad \mathbf{u}_i = \bar{\mathbf{u}}_i + \varepsilon_{\bar{\mathbf{u}}_i} \tag{6}$$

and eq. 5 becomes

$$[\mathbf{q}, \mathbf{t}] = \arg\min_{\{\mathbf{q},\mathbf{t},\bar{\mathbf{w}}\}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathbf{w}_\mathrm{i} - \bar{\mathbf{w}}_\mathrm{i})^\top \bar{\mathbf{C}}_\mathrm{i} (\mathbf{w}_\mathrm{i} - \bar{\mathbf{w}}_\mathrm{i}) \quad , \tag{7}$$

where $\varepsilon_{\bar{\mathbf{v}}_i}$ and $\varepsilon_{\bar{\mathbf{u}}_i}$ are drawn from symmetric and independent distributions with zero mean and covariances dependent on the true data i.e. $\mathcal{S}(0, \eta\bar{\boldsymbol{\Sigma}}_{\bar{\mathbf{v}}_i})$ and $\mathcal{S}(0, \eta\bar{\boldsymbol{\Sigma}}_{\bar{\mathbf{u}}_i})$. The problem is that $\bar{\boldsymbol{\Sigma}}_{\bar{\mathbf{v}}_i}$ and $\bar{\boldsymbol{\Sigma}}_{\bar{\mathbf{u}}_i}$ can be unobservable or simply be unknown. In this case, they are often approximated with $\boldsymbol{\Sigma}_{\mathbf{v}_i}$ and $\boldsymbol{\Sigma}_{\mathbf{u}_i}$. This approximation can cause a small bias

in the estimate of the motion parameters. Since the absolute pose is the integration of possible thousands of relative-pose estimates, this small bias will eventually cause a significant drift.

3. Stereo-vision

To obtain the static landmarks needed for motion estimation a stereo based approach is used. This requires image feature correspondences between successive stereo-frames and between images in the stereo-frames themselves. To this purpose the Scale Invariant Feature Transform (SIFT), Lowe [16], is used. A threshold is applied on the distance between SIFT descriptors to ensure reliable matches between image features. Furthermore, the epipolar constraint, back-and-forth and left-to-right consistency are enforced. It is assumed that stereo images are rectified according to the epipolar geometry of the used stereo-rig, Hartley and Zisserman [10]. From an image point in the left image $\mathrm{z}_l = [x_l, y_l]'$ and its corresponding point in the right image $\mathrm{z}_r = [x_r, y_r]'$, the disparity can be obtained with sub-pixel accuracy $d = x_l - x_r$. Using the disparity d, the focal length f of the left camera and the stereo base line b, the 3D position of the landmark $\bar{\mathbf{z}}$ imaged by z_l and z_r, relative to the optical center, of the left camera can be estimated with

$$\mathbf{z} = \left[\frac{x_l b}{d}, \quad \frac{y_l b}{d}, \quad \frac{f b}{d}\right]^{\mathrm{T}}. \tag{8}$$

More advanced stereo reconstruction methods can be found in [10, 13]. Using them would be an interesting research direction. In this work, however, the estimated motion is utilized to find an improved estimate of the landmark positions.

3.1. Stereo-uncertainty modeling

The landmark $\bar{\mathbf{z}}$ is projected onto the images of a stereo camera resulting in the noise free image points $\bar{\mathrm{z}}_l$ and $\bar{\mathrm{z}}_r$. Due to noise in the sensing process only z_l and z_r are observed, where $\mathrm{z}_l = \bar{\mathrm{z}}_l + \varepsilon_l$ and $\mathrm{z}_r = \bar{\mathrm{z}}_r + \varepsilon_r$. Assuming that ε_l and ε_r are drawn from independent identically distributed (i.i.d.) Gaussian noise with isotropic covariance Σ, the regions around $\bar{\mathrm{z}}_l$ and $\bar{\mathrm{z}}_r$ that have a probability of α to contain z_l and z_r can be described using circles. The reconstruction based on z_l and z_r i.e. $\mathbf{z}$ has then a probability of α^2 to lie within the intersection of the two elliptical cones spanned by the circles and the camera optical centers, see fig 1.c and 1.e. Depending on the true position of the landmark $\bar{\mathbf{z}}$ the intersection volume around $\bar{\mathbf{z}}$ changes. Clearly the general heteroscedastic model of eq. 6 is appropriate. Note also, in fig. 1.d, how the skewness of the intersection volume increases with the depth of the landmark.

It is the intersection-volume, approximated with the symmetric distribution $\mathcal{S}(0, \eta\bar{\mathbf{\Sigma}}_{\bar{\mathbf{z}}})$, that should be used in the optimization. Unfortunately, enforcing symmetry is necessary within the HEIV optimization scheme. Important is the correct relative scale and orientation of $\bar{\mathbf{\Sigma}}_{\bar{\mathrm{z}}}$. Because the scale and orientation of the intersection volume depends on $\bar{\mathbf{z}}$,which is unobservable, it is not straightforward to obtain. Note that enforcing symmetry is an issue when minimizing 3D errors as well as when minimizing reprojection errors for most optimization strategies. This is a consequence of the fact that the reprojection of the uncertainty in the reconstructed landmarks position is not Gaussian either.

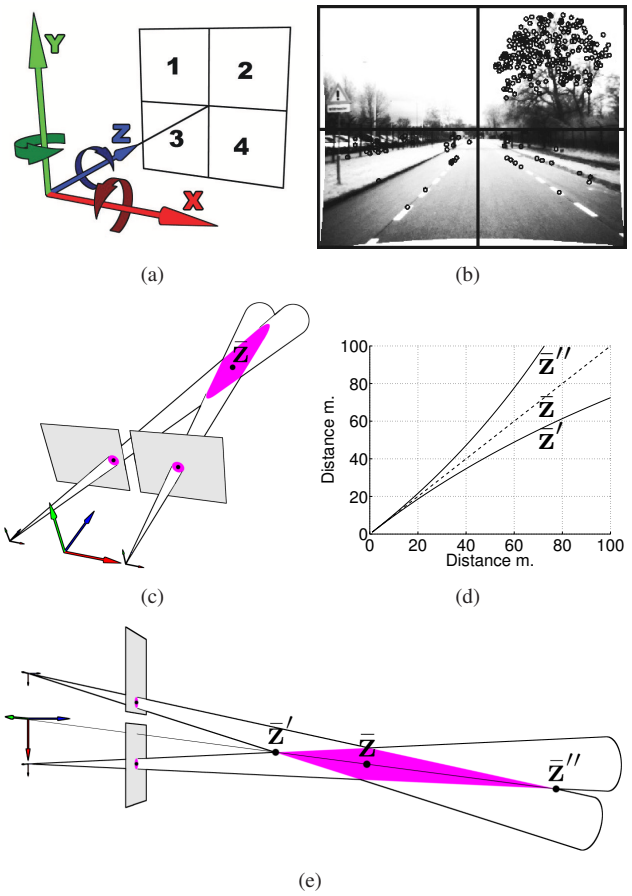

Figure 1. Coordinate axis and image quadrant convention (a). Example of inhomogeneously distributed landmarks (b). Impression of stereo-reconstruction uncertainty modeling (c), top-view (e). Skewness of stereo-reconstruction uncertainty (d).

In [18] the authors propose estimating the stereo-reconstruction uncertainty with a bootstrap approach using residual resampling. The residuals are added to the reprojection z of the estimated landmark position $\mathbf{z}$. As a direct consequence, $\mathbf{\Sigma}_{\mathrm{z}}$ is estimated instead of $\bar{\mathbf{\Sigma}}_{\bar{\mathrm{z}}}$. The stereo-reconstruction uncertainty can also be estimated using error-propagation of the image-feature position uncertainty Σ using the Jacobian $\mathbf{J}$ of the reconstruction function,

Matthies and Shafer [20],

$$\mathbf{\Sigma_z} = \mathbf{J}_z \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \Sigma \end{bmatrix} \mathbf{J}_z^\top, \tag{9}$$

$$\mathbf{J}_z = \begin{bmatrix} \frac{-x_l b}{d^2} + \frac{b}{d} & 0 & \frac{x_l b}{d^2} & 0 \\ \frac{-y_l b}{d^2} & \frac{b}{d} & \frac{y_l b}{d^2} & 0 \\ \frac{-fb}{d^2} & 0 & \frac{fb}{d^2} & 0 \end{bmatrix}. \tag{10}$$

Because the jacobian is calculated on the observed projection z, $\mathbf{\Sigma}_z$ is estimated instead of $\bar{\mathbf{\Sigma}}_{\bar{z}}$.

3.2. Improved stereo uncertainty modeling

To obtain improved estimates of the stereo-reconstruction uncertainties they are first approximated using eq. 9 and 10. Then by using the rotation $\hat{\mathbf{R}}$ and translation $\hat{\mathbf{t}}$ estimated with $\{\hat{\mathbf{R}}, \hat{\mathbf{t}}\} = \mathcal{H}EIV(\mathbf{v}, \Sigma_\mathbf{v}, \mathbf{u}, \Sigma_\mathbf{u})$, the observed points can be corrected. Firstly, they are transformed into the same coordinate frame with

$$\begin{aligned} \mathbf{u}'_i &= \hat{\mathbf{R}}\mathbf{u}_i + \hat{\mathbf{t}} \\ \mathbf{\Sigma}_{\mathbf{u}'_i} &= \hat{\mathbf{R}}\mathbf{\Sigma}_{\mathbf{u}_i}\hat{\mathbf{R}}^\top \end{aligned}. \tag{11}$$

In this coordinate frame the landmark positions can be fused according to their uncertainties with

$$\begin{aligned} \mathbf{K} &= \mathbf{\Sigma}_{\mathbf{v}_i}(\mathbf{\Sigma}_{\mathbf{v}_i} + \mathbf{\Sigma}_{\mathbf{u}'_i})^{-1} \\ \hat{\mathbf{v}}_i &= \mathbf{v}_i + \mathbf{K}(\mathbf{u}'_i - \mathbf{v}_i) \\ \hat{\mathbf{u}}_i &= \hat{\mathbf{R}}^\top(\hat{\mathbf{v}}_i - \hat{\mathbf{t}}) \end{aligned}, \tag{12}$$

Kalman [12]. Finally, a copy of the fused landmark positions is transformed according to the inverse of estimated motion. The process results in an improved estimate of the landmark positions which exactly correspond to the estimated motion. The real goal is an improved estimate of the landmark uncertainties. To obtain them, the new estimates $\hat{\mathbf{v}}_i$ and $\hat{\mathbf{u}}_i$ can be projected onto the imaging planes of a modeled stereo-camera. The appropriate parameters can be obtained by calibration of the actual stereo-camera being used. From these projections $\hat{\mathrm{v}}_i$ and $\hat{\mathrm{u}}_i$ an improved estimate of the covariances i.e. $\hat{\mathbf{\Sigma}}_{\hat{\mathbf{v}}_i}$ and $\hat{\mathbf{\Sigma}}_{\hat{\mathbf{u}}_i}$ can be obtained with eq. 9 and 10. This technique is preferred because it produces covariances with the correct orientation and relative scale given $\hat{\mathbf{v}}_i$ and $\hat{\mathbf{u}}_i$.

4. Estimating and compensating motion bias

A premise of the proposed bias reduction technique is the absence of landmark outliers. An initial robust estimate of the motion can be obtained using techniques explained in [6]. Given the robust estimate the improved location and uncertainty of the landmarks can be calculated with eq.11 and 12. Landmarks can then be discarded based on their Mahalanobis distance to the improved landmark positions

$$(\mathbf{v}_i - \hat{\mathbf{v}}_i)^\top \hat{\mathbf{\Sigma}}_{\hat{\mathbf{v}}_i}(\mathbf{v}_i - \hat{\mathbf{v}}_i) + (\mathbf{u}_i - \hat{\mathbf{u}}_i)^\top \hat{\mathbf{\Sigma}}_{\hat{\mathbf{u}}_i}(\mathbf{u}_i - \hat{\mathbf{u}}_i). \tag{13}$$

A new motion estimate is then calculated using all the inliers. The process can be iterated several times or until convergence.

From now on $\mathbf{v}_i$ and $\mathbf{u}_i$ and their covariances $\mathbf{\Sigma}_{\mathbf{v}_i}$ and $\mathbf{\Sigma}_{\mathbf{u}_i}$, obtained using eq. 9 and eq. 10, for $i = 1..n$ are assumed to be inliers only. The bias reduction technique then estimates the motion on these inliers

$$\left\{\hat{\mathbf{R}}, \hat{\mathbf{t}}\right\} = \mathcal{H}EIV(\mathbf{v}, \mathbf{\Sigma}_\mathbf{v}, \mathbf{u}, \mathbf{\Sigma}_\mathbf{u}). \tag{14}$$

Given $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$ the uncertainties are improved using the technique describe in sec. 3.2 resulting in $\hat{\mathbf{\Sigma}}_{\hat{\mathbf{v}}_i}$ and $\hat{\mathbf{\Sigma}}_{\hat{\mathbf{u}}_i}$. Another motion estimate, using the new covariances, is then generated

$$\left\{\hat{\hat{\mathbf{R}}}, \hat{\hat{\mathbf{t}}}\right\} = \mathcal{H}EIV(\mathbf{v}, \hat{\mathbf{\Sigma}}_{\hat{\mathbf{v}}}, \mathbf{u}, \hat{\mathbf{\Sigma}}_{\hat{\mathbf{u}}}). \tag{15}$$

The motion bias is then approximated using

$$\begin{aligned} \mathbf{t}_{bias} &= \begin{bmatrix} \omega_x & & \\ & \omega_y & \\ & & \omega_z \end{bmatrix} \hat{\hat{\mathbf{t}}} - \hat{\mathbf{t}} \\ \mathbf{R}_{bias} &= \mathcal{DCM}\left(\begin{bmatrix} \omega_p & & \\ & \omega_h & \\ & & \omega_r \end{bmatrix} \mathcal{A}(\hat{\mathbf{R}}^\top \hat{\hat{\mathbf{R}}}) \right) \end{aligned}. \tag{16}$$

Here ω_x, ω_y and ω_z are the appropriate gains that scale the estimated tendency of the translation bias to the correct magnitude. By using the gains ω_p, ω_h and ω_r the same is applied to the Euler angles (pitch,heading,roll) obtained with $\mathcal{A}$, of the rotation bias tendency. The function $\mathcal{DCM}$ transforms the scaled Euler angles back into a rotation matrix, [5]. Finally, an unbiased motion estimate is obtained with

$$\begin{aligned} \mathbf{R}_{unbiased} &= \hat{\mathbf{R}}\mathbf{R}_{bias} \\ \mathbf{t}_{unbiased} &= \hat{\mathbf{t}} + \mathbf{t}_{bias} \end{aligned}. \tag{17}$$

The need for the bias gains $(\omega_x, \omega_y, \omega_z, \omega_p, \omega_h, \omega_r)$ is a direct consequence of the fact that $\hat{\mathbf{\Sigma}}_{\hat{\mathbf{v}}}$ and $\hat{\mathbf{\Sigma}}_{\hat{\mathbf{u}}}$ are only, on average, improved estimates of the true landmark uncertainties $\bar{\mathbf{\Sigma}}_{\bar{\mathbf{v}}}$ and $\bar{\mathbf{\Sigma}}_{\bar{\mathbf{u}}}$. In reality, this improvement might even be very small. Nevertheless, the improvement reveals the bias tendency. The gains then amplify the estimated tendency to the correct magnitude.

5. Results on simulated data

Simulation is used to give insight into the relevance of the claims and methods introduced in this contribution. Most important is the claim that eq. 4 and 5 are essentially wrong and should be replaced with eq. 6 and 7. Furthermore, interesting observations regarding the dependency of

the bias on the landmark distribution are given in this section. Using the available groundtruth $\bar{\mathbf{R}}$ and $\bar{\mathbf{t}}$, the bias in the estimators is calculated as follows

$$\begin{aligned} \mathbf{Bias_t} &= \left(\tfrac{1}{m} \sum_{i=1}^{m} \hat{\mathbf{t}}_i \right) - \bar{\mathbf{t}} \ , \\ \mathbf{Bias_R} &= \tfrac{1}{m} \sum_{i=1}^{m} \mathcal{A}(\bar{\mathbf{R}}^\top \hat{\mathbf{R}}_i) \end{aligned} \quad . \tag{18}$$

5.1. Symmetric 'stereo-like' noise

For the first experiment only the bias due to approximating $\bar{\mathbf{\Sigma}}_{\bar{\mathbf{z}}}$ with $\mathbf{\Sigma_z}$ is of interest. The possible bias introduced by using a symmetric distribution for what in reality is an asymmetric distribution is neglected. The purpose is to show that the general heteroscedastic model of eq. 6 and 7 is to be preferred and will cause an unbiased HEIV estimate.

In order to generate noise that is symmetric and at the same time mimics stereo-reconstruction noise the following approach has been chosen. The artificial points $\bar{\mathbf{u}}_i.....\bar{\mathbf{u}}_{150}$ were generated homogenously within the space defined by the optical center of the left camera and the first image quadrant, see fig. 1.a. The distances of the generated landmarks ranged from 5 m to 150 m. The points $\bar{\mathbf{v}}_i.....\bar{\mathbf{v}}_{150}$ were then generated by transforming $\bar{\mathbf{u}}_i.....\bar{\mathbf{u}}_{150}$ with the groundtruth motion $\bar{\mathbf{R}}$ and $\bar{\mathbf{t}}$. These 3D points were projected onto the imaging planes of a simulated stereo-camera and $\bar{\mathbf{\Sigma}}_{\bar{\mathbf{v}}_i}$ and $\bar{\mathbf{\Sigma}}_{\bar{\mathbf{u}}_i}$ were calculated using eq. 9 and 10. For each point a random perturbation, drawn from either $\mathcal{N}(0, \bar{\mathbf{\Sigma}}_{\bar{\mathbf{v}}_i})$ or $\mathcal{N}(0, \bar{\mathbf{\Sigma}}_{\bar{\mathbf{u}}_i})$, was added to the true 3D landmark locations resulting in $\mathbf{v}_i$ and $\mathbf{u}_i$. The noisy landmark locations were then also projected onto the imaging planes of the stereo-camera and from these $\mathbf{\Sigma}_{\mathbf{v}_i}$ and $\mathbf{\Sigma}_{\mathbf{u}_i}$ were estimated using eq. 9 and 10. Then two motion estimates were obtained, one using $\mathcal{HEIV}(\mathbf{v}, \bar{\mathbf{\Sigma}}_{\bar{\mathbf{v}}}, \mathbf{u}, \bar{\mathbf{\Sigma}}_{\bar{\mathbf{u}}})$ and another one using $\mathcal{HEIV}(\mathbf{v}, \mathbf{\Sigma_v}, \mathbf{u}, \mathbf{\Sigma_u})$. The experiment was repeated one thousand times for each of nine different motions. The result are shown in fig. 2. It can clearly be seen that using the general heteroscedastic model of eq. 6 and 7 results in an unbiased motion estimate. In contrast to this, modeling $\bar{\mathbf{\Sigma}}_{\bar{\mathbf{z}}}$ with $\mathbf{\Sigma_z}$ introduces bias. As can be seen in fig. 2, this bias is relatively small. When many of these biased relative-pose estimates are integrated to track the absolute-pose, however, they will cause significant drift.

5.2. Simulated stereo-noise

In this section the stereo-reconstruction noise will be modeled more accurately. Furthermore, the effectiveness of the proposed bias reduction technique on simulated data will be presented.

The artificial landmarks $\bar{\mathbf{u}}_i.....\bar{\mathbf{u}}_{150}$ and $\bar{\mathbf{v}}_i.....\bar{\mathbf{v}}_{150}$ were generated similarly to sec. 5.1. For this experiment also

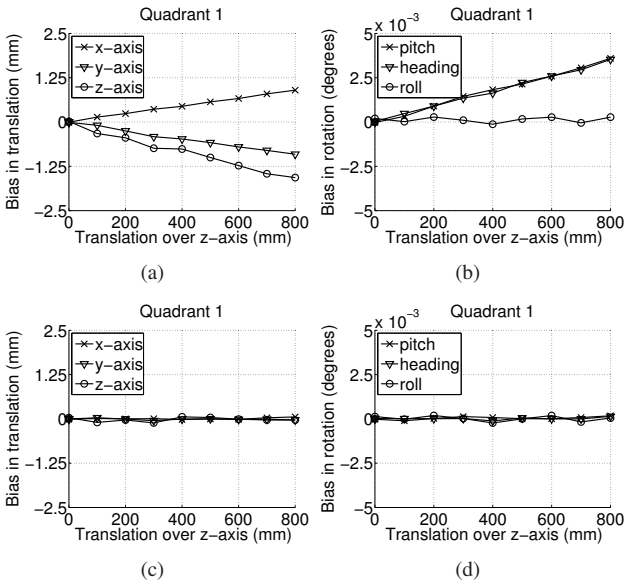

Figure 2. Bias in estimated parameters for motions with a constant heading of 1 degree and increasing translation over z-axis, using $\mathcal{HEIV}(\mathbf{v}, \mathbf{\Sigma_v}, \mathbf{u}, \mathbf{\Sigma_u})$ a,b and using $\mathcal{HEIV}(\mathbf{v}, \bar{\mathbf{\Sigma}}_{\bar{\mathbf{v}}}, \mathbf{u}, \bar{\mathbf{\Sigma}}_{\bar{\mathbf{u}}})$ c and d.

different image quadrants were used i.e. quadrant 2 and quadrant 3, see fig. 1.a. By doing so, the dependency of the bias on the landmark distribution can be visualized. A real-world example of a situation in which the landmarks are not homogenously distributed is shown in fig. 1.b. Again the landmarks were projected onto the imaging planes of a simulated stereo-camera. Now, however, isotropic i.i.d. gaussian noise (with standard deviation of 0.25 pixel) is added to the image projections. By using stereo-reconstruction, on the basis of these noisy image points, the landmark positions are estimated resulting in $\mathbf{u}_i.....\mathbf{u}_{150}$ and $\mathbf{v}_i.....\mathbf{v}_{150}$. Also $\mathbf{\Sigma}_{\mathbf{v}_i}$ and $\mathbf{\Sigma}_{\mathbf{u}_i}$ were estimated, using eq. 9 and eq. 10, from the noisy image points. Again a motion estimate is generated with $\mathcal{HEIV}(\mathbf{v}, \mathbf{\Sigma_v}, \mathbf{u}, \mathbf{\Sigma_u})$ and the experiment is repeated one thousand times for nine different motions. The results for different landmark distributions is shown in fig. 3. The result of applying the bias reduction technique of sec. 4 is shown in fig. 4. The used bias gains $(\omega_x, \omega_y, \omega_z, \omega_p, \omega_h, \omega_r)$ were all set to 0.8. The benefit of the proposed bias reduction technique is clearly visible. Important is the fact that the mean absolute error in motion parameters did not change by using the bias reduction technique. The error in translation was approximately x=1.0 mm, y=1.2 mm and z=3.0 mm and the error in rotation angles for pitch=$3 \cdot 10^{-3}\,^\circ$, heading=$2 \cdot 10^{-3}\,^\circ$ and roll=$7 \cdot 10^{-3}\,^\circ$ for all experiments. Furthermore, the graphs from fig. 3 visualize the effect of true motion and the landmark distribution on the bias. Interestingly, from fig. 3 and the image quadrant and axis conventions from fig. 1.a, it can be seen that the bias causes a rotation slightly towards

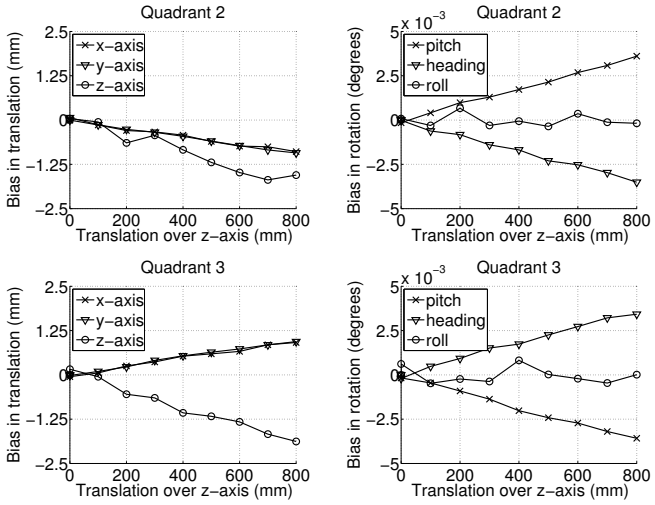

Figure 3. Effects of estimator bias for a motion with a constant heading of 1 degree and increasing translation over z-axis. Bias in translation (left images), bias in rotation (right images), landmarks only in image quadrant 2 (top images) and landmarks only in image quadrant 3 (bottom images).

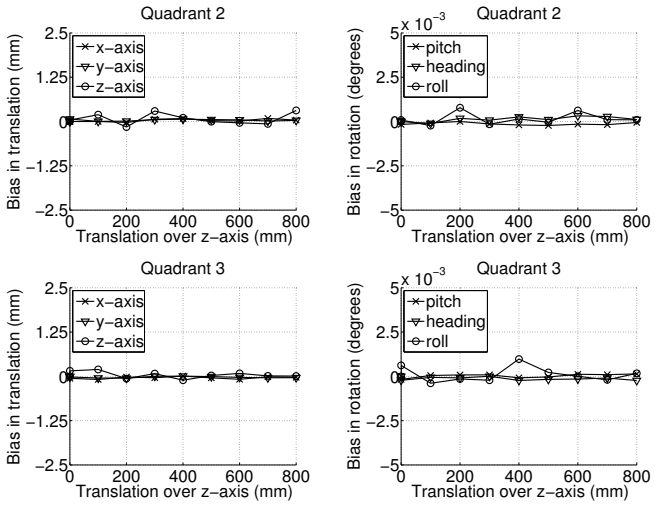

Figure 4. Results of the proposed bias reduction method for a motion with a constant heading of 1 degree and increasing translation over z-axis. Bias in translation (left images), bias in rotation (right images), landmarks only in image quadrant 2 (top images) and landmarks only in image quadrant 3 (bottom images).

the landmarks and a translation slightly away from the landmarks.

6. Results on real-world data

To show the applicability of the proposed bias reduction technique it has been tested on a challenging 5 km urban data-set. To our knowledge it is currently (one of) the largest urban data-sets used for relative-pose based visual-odometry research. Many possible sources for outlier landmarks, such as moving cars, trucks and pedestrian, are included in the data-set. An impression of the approximately 19000 images in the data-set is given in fig. 5. The data-

Figure 5. Impression of different images from the used 5 km long urban data-set.

set was recorded using a stereo-camera with a baseline of 40 cm and an image resolution of 640 by 480 pixels running at 30 Hz. The correct values for the real-world bias gains $(\omega_x, \omega_y, \omega_z, \omega_p, \omega_h, \omega_r)$ were obtained by manual selection, such that the loop in the calibration data-set, see fig. 6, was approximately closed in 3D. These exact bias reduction gains were then used for the 5 km trajectory. A minimal estimated distance of 30 cm is enforced on-line between frames. If two successive frames do not reach this distance, the latest of these frames is dropped. The process results in approximately 14500 relative-pose estimates for the 19000 images in the data-set. The driven trajectory is obtained by integrating all the relative pose estimates, the results are visualized in fig. 7. Probably, the most significant improvement is in the estimated height profile, see fig. 8. Due to systematic bias in the estimated roll angle the trajectory without bias reduction spirals downward. By compensation the bias in roll, using the proposed technique, this spiraling effect is significantly reduced. Due to these biased rotation estimates the error in the final position as percentage of the traveled distance, when not using the bias reduction technique, was approximately 20%. This reduced to 1% when the proposed bias reduction technique was used. The relative computation time of the most intensive processing stages were approximately, 45% for image-feature extraction and matching and 45% for obtaining the robust motion estimate. The relative computation time of the bias reduction technique was only 4%.

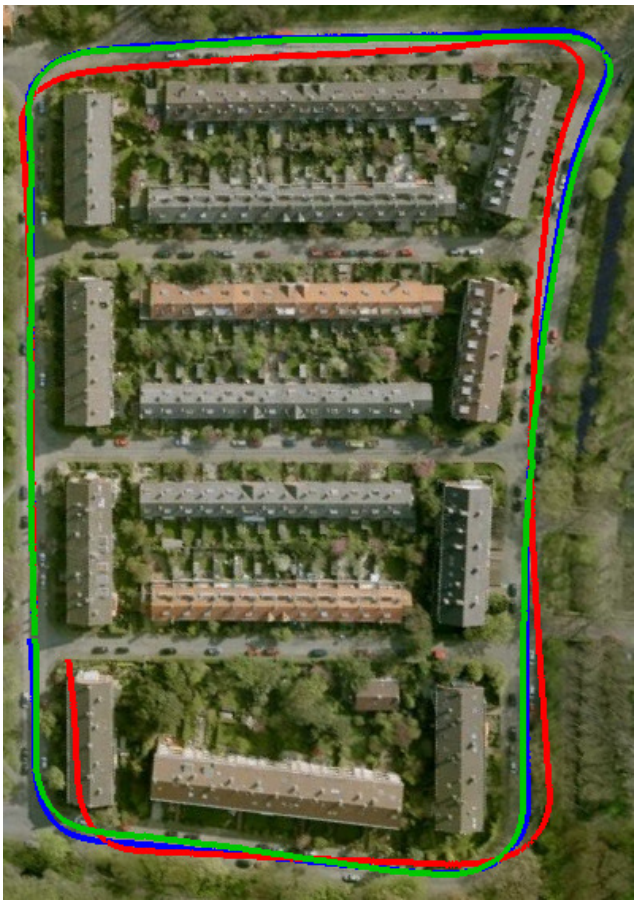

Figure 6. Calibration data-set, of approximately 800 m, used to obtain the appropriate bias gains. DGPS based groundtruth (green), HEIV on inlier landmarks (red) and HEIV using the manually tuned bias reduction technique (blue).

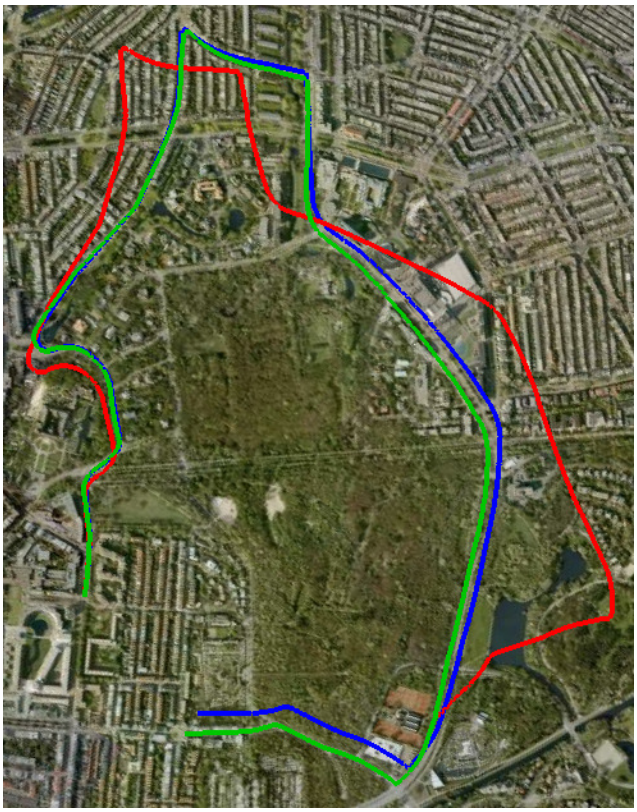

Figure 7. Results on the 5 km urban trajectory. DGPS based groundtruth (green), HEIV on inlier landmarks (red) and HEIV using the calibrated bias reduction technique (blue).

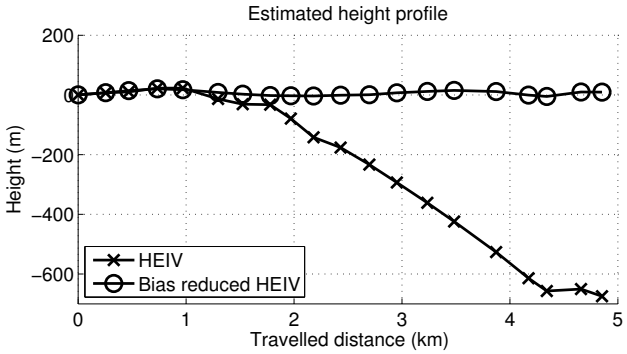

Figure 8. Estimated height profile of the 5 km. urban trajectory on relatively flat terrain.

7. Conclusion

This contribution addresses possible sources of structural errors i.e. bias in stereo based relative-pose estimation. Bias in the relative pose estimates can lead to significant drift when they are integrated to track the absolute-pose of the stereo camera. In relation with the Heteroscedastic Error-In-Variables estimator two sources of bias are identified. Firstly, approximating the noise distribution of the stereo reconstructed landmarks using the observations instead of the true landmark locations. Secondly, neglecting the skewness of the landmark reconstruction uncertainty. For the first source a novel bias reduction technique is presented that significantly reduces the structural error in stereo-vision based motion estimation. The benefit of this approach is most apparent when the relative-pose estimates are integrated to track the absolute-pose of the camera, as is the case with visual-odometry. The work presented in this contribution mainly identifies the existence of bias and, most importantly, the sources of bias for stereo based relative-pose estimation. The solution proposed must be seen as a first attempt to overcome the bias problem. Clearly, in future work we will investigate other possible solutions to reduce bias in the motion estimates. Especially, providing a solution that does not depend on the bias gains has our priority.

References

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700, September 1987.

[2] T. Barfoot. Online visual motion estimation using fastslam with sift features. In *IEEE/RSJ Intelligent Robots and Systems*, pages 579–585, 2005.

[3] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley. From fns to heiv: A link between two vision parameter estimation methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):264–268, 2004.

[4] A. Comport, E. Malis, and P. Rives. Accurate quadrifocal tracking for robust 3d visual odometry. In *IEEE International Conference on Robotics and Automation*, pages 40–45, April 2007.

[5] J. Diebel. Representing attitude: Euler angles, unit quaternions, and rotation vectors. Technical report, Stanford University, Stanford, California 94301-9010, October 2006.

[6] G. Dubbelman, W. van der Mark, and F. C. A. Groen. Accurate and robust ego-motion estimation using expectation maximization. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3914–3920, September 2008.

[7] D. Eggert, A. Lorusso, and R. Fisher. Estimating 3-d rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications*, 9(5-6):272–290, March 1997.

[8] P. Elinas, R. Sim, and J. J. Little. σslam: Stereo vision slam using the rao-blackwellised particle filter and a novel mixture proposal distribution. In *IEEE International Conference on Robotics and Automation*, pages 1564–1570, 2006.

[9] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381 – 395, June 1981.

[10] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[11] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, 4:629, 1987.

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

[13] K. Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Dover Publications, ISBN: 0486443086, 1996.

[14] K. Konolige, M. Agrawal, and J. Solà. Large scale visual odometry for rough terrain. In *International Symposium on Research in Robotics*, 2007.

[15] A. Levin and R. Szeliski. Visual odometry and map correlation. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 611 – 618, June 2004.

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

[17] B. C. Matei. *Heteroscedastic errors-in-variables models in computer vision*. PhD thesis, The State University of New Jersey, May 2001.

[18] B. C. Matei and P. Meer. Optimal rigid motion estimation and performance evaluation with bootstrap. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, page 345, 1999.

[19] B. C. Matei and P. Meer. Estimation of nonlinear errors-in-variables models for computer vision applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1537 – 1552, October 2006.

[20] L. Matthies and S. A. Shafer. Error modeling in stereo navigation. *IEEE Journal of Robotics and Automation*, 3(3):239 – 248, June 1988.

[21] D. Nistér, O. Naroditsky, and J. Bergen. Visual odometry. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 652–659, 2004.

[22] D. Nistér, O. Naroditsky, and J. Bergen. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23(1):3–20, January 2006.

[23] N. Ohta and K. Kanatani. *Optimal estimation of three-dimensional rotation and reliability evaluation*, volume 1406 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 1998.

[24] C. F. Olson, L. H. Matthies, M. Schoppers, and M. W. Maimoneb. Rover navigation using stereo ego-motion. *Robotics and Autonomous Systems*, 43(4):215–229, June 2003.

[25] J. M. Saez and F. Escolano. 6dof entropy minimization slam. In *IEEE International Conference on Robotics and Automation*, pages 1548–1555, 2006.

[26] S. Se, D. Lowe, and J. Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *The International Journal of Robotics Research*, 21(8):735–758, August 2002.

[27] N. Sunderhauf, K. Konolige, S. Lacroix, and P. Protzel. Visual odometry using sparse bundle adjustment on an autonomous outdoor vehicle. *Tagungsband Autonome Mobile Systeme*, pages 157–163, 2005.

[28] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *International Workshop on Vision Algorithms: Theory and Practice*, pages 298 – 372, 1999.

[29] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376 – 380, 1991.

[30] Z. Zhu, T. Oskiper, S. Samarasekera, R. Kumar, and H. Sawhney. Ten-fold improvement in visual odometry using landmark matching. In *IEEE International Conference on Computer Vision*, pages 1 – 8, October 2007.