Online Continuous Stereo Extrinsic Parameter Estimation

Peter Hansen
Computer Science Department
Carnegie Mellon University in Qatar
Doha, Qatar
phansen@qatar.cmu.edu

Hatem Alismail, Peter Rander, Brett Browning
National Robotics Engineering Center
Robotics Institute, Carnegie Mellon University
Pittsburgh PA, USA
{halismai,rander,brettb}@cs.cmu.edu

Abstract

Stereo visual odometry and dense scene reconstruction depend critically on accurate calibration of the extrinsic (relative) stereo camera poses. We present an algorithm for continuous, online stereo extrinsic re-calibration operating only on sparse stereo correspondences on a per-frame basis. We obtain the 5 degree of freedom extrinsic pose for each frame, with a fixed baseline, making it possible to model time-dependent variations. The initial extrinsic estimates are found by minimizing epipolar errors, and are refined via a Kalman Filter (KF). Observation covariances are derived from the Crámer-Rao lower bound of the solution uncertainty. The algorithm operates at frame rate with unoptimized Matlab code with over 1000 correspondences per frame. We validate its performance using a variety of real stereo datasets and simulations.

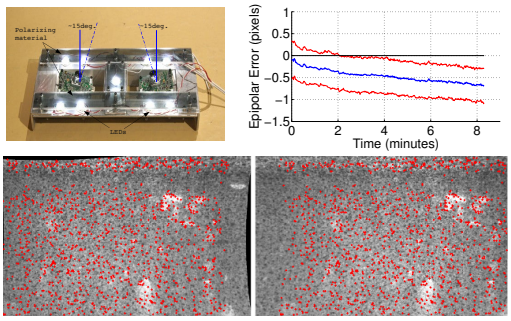
1. Introduction

Stereo vision is core to many $3D$ vision methods including visual odometry and dense scene reconstruction. Good calibration, both intrinsic and extrinsic, is essential to achieving high accuracy as it impacts image rectification, stereo correspondence search, and triangulation. Intrinsic calibration models image formation for each camera (e.g. [3]), while extrinsic calibration models the 6 degree of freedom (DOF) pose between the cameras. For real systems, extrinsic calibration errors occur more frequently due to larger exposure to shock, vibration, thermal variation and cycling. For visual odometry in particular, such errors lead to biased results. We propose a method to *recalibrate* extrinsic parameters online to correct drift or bias. Fig. 1 shows epipolar errors for a range of stereo heads. For 1b and 1c there is a near constant bias, while 1a drifts possibly caused by thermal expansion from the lighting assembly.

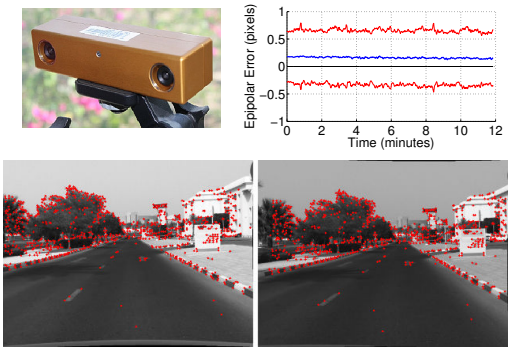
Online calibration remains an active area of research. Online intrinsic calibration (auto or self calibration) estimates intrinsic parameters using scene point correspondences from multiple views (e.g. [18, 17, 8, 11]). However, the results are generally less accurate than offline methods [8] using known relative Euclidean control points (e.g. [16]). Here, we focus on correcting drifting extrinsic calibration. Carrera *et al*. [2] calibrated multi-camera extrinsics using monocular visual SLAM maps for each camera [6], not necessarily with overlapping fields of view. However, the extrinsic estimates were assumed to be stable over time and monocular SLAM limits real-time performance in large environments. In contrast, *continuous* methods output a unique extrinsic pose for each stereo pair (per time step). In [1], a linear essential matrix estimate is used to find relative pose, followed by non-linear refinement incorporating depth ordering constraints. Some constraints were placed on the extrinsic pose DOF, and experimental testing was restricted to small indoor sequences with a stationary camera.

Dang *et al*. [5, 4] developed an approach that estimates the extrinsics using three error metrics incorporated into an iterative Extended Kalman Filter (EKF). The error metrics are derived from bundle adjustment (BA), epipolar constraints, and trilinear constraints. Comparisons were made via scene reconstruction accuracy, and they found that using epipolar constraints (epipolar reprojection errors) only to be inferior to using all three metrics. The number of correspondences was limited (< 50), and using more is likely to significantly impact real-time performance. Interestingly, there were several advantages to using epipolar errors only. These include the ability to obtain strictly per-frame estimates without needing temporal correspondences and the invariance to non-rigid scenes, which is important for operations in dynamic environments.

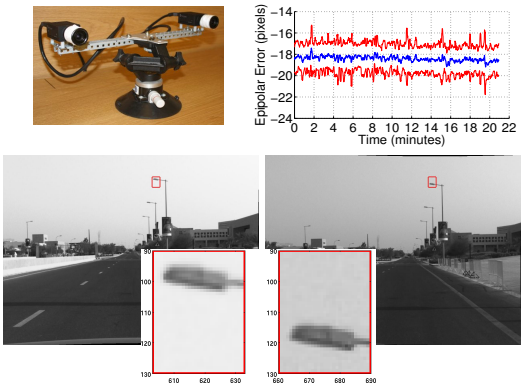
In this paper, we contribute a continuous, online, extrinsic re-calibration algorithm that operates in real-time using only sparse stereo correspondences and no temporal constraints. The initial extrinsic estimates are obtained by minimizing epipolar errors, and a Kalman Filter (KF) is used to limit over-fitting. The unique extrinsics estimated for each stereo pair enable temporal drift to be modeled and we



(a) Pipe dataset: $f = 1203pix$, $1200 \times 768pix^2$ images.



(b) Outdoor dataset 1: $f = 811pix$, $640 \times 480pix^2$ images.



(c) Outdoor dataset 2: $f = 1781pix$, $1024 \times 768pix^2$ images.

Figure 1: Mean (blue) and $\pm 3\sigma$ standard deviation (red) epipolar errors for sparse correspondences for different stereo data. The supplied calibration (Pointgrey Bumblebee2) was used for rectification in (b).

show that with enough correspondences (e.g. 1000), epipolar errors alone are sufficient for good re-calibration. Moreover, the approach is trivial to extend to multiple frames by combining correspondences. We validate the approach in simulation and on real stereo datasets by comparing visual odometry estimates with and without re-calibration, and reconstruction errors compared to offline calibration with a known target. We show the limitations of re-calibrating the baseline length, and suggest methods to partially address these.

2. Stereo Geometry and Error Metric

2.1. Stereo Pose and Epipolar Constraints

The stereo extrinsics $S = [R|\mathbf{t}]$, is composed of a rotation $R \in SO(3)$ and translation $\mathbf{t} \in \mathbb{R}^3$. It defines the projection of a scene point $\mathbf{X}_l = (X, Y, Z)^T$ in the left camera, to $\mathbf{X}_r$ in the right: $\mathbf{X}_r = R\,\mathbf{X}_l + \mathbf{t}$.

Our re-calibration algorithm uses image coordinates and errors in the left and right stereo rectified images. Let $\tilde{\mathbf{u}}_l \leftrightarrow \tilde{\mathbf{u}}_r$ be a set of homogeneous scene point correspondences in a pair of rectified images, which are related to the scene points coordinates $\mathbf{X}_l, \mathbf{X}_r$ by

$$\tilde{\mathbf{u}}_l \simeq K_l\,\tilde{R}_l\,\mathbf{X}_l = K_l\,\tilde{\mathbf{X}}_l \tag{1}$$

$$\tilde{\mathbf{u}}_r \simeq K_r\,\tilde{R}_r\,\mathbf{X}_r = K_r\,\tilde{\mathbf{X}}_r, \tag{2}$$

where $\tilde{R}_l, \tilde{R}_r$ are rotations applied to each camera, and K_l, K_r are pinhole projection matrices with zero skew and equal focal lengths f. For convenience we assume that

$$K_l = K_r = \begin{bmatrix} f & 0 & u_o \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

$\tilde{R}_l$ and $\tilde{R}_r$ are selected to produce a *rectified* extrinsic pose $\tilde{S} = [I_{3\times3}\,|\,(-b, 0, 0)^T]$, where $b = ||\mathbf{t}||$ is the original baseline, such that $\tilde{\mathbf{X}}_r = \tilde{\mathbf{X}}_l + (-b, 0, 0)^T$ (e.g. [12]).

The rectified coordinates are related by

$$(\tilde{u}_l, \tilde{v}_l, 1)^T = (\tilde{u}_r + d, \tilde{v}_r, 1)^T, \quad d = \frac{bf}{\tilde{Z}}, \tag{4}$$

where d is the *disparity* and $\tilde{Z}$ the depth of a scene point. The stereo rectified epipolar constraint is simply $\tilde{v}_l = \tilde{v}_r$, which is independent of the depth and baseline. This can also be derived from the monocular essential matrix $\tilde{\mathbf{u}}_l\,E\,\tilde{\mathbf{u}}_r = 0$ [12].

2.2. Calibration Error Metric

For re-calibration, we decompose each rotation, $\tilde{R}_l$ and $\tilde{R}_r$, as the product of two independent rotations:

$$\tilde{R}_l = \check{R}_l^T R'_l, \quad \tilde{R}_r = \check{R}_r^T R'_r. \tag{5}$$

They are the rotations R'_l and R'_r from the original stereo extrinsics S, and a rotation correction $\check{R}_l$ and $\check{R}_r$. We start with a set of correspondences $\mathbf{u}'_l \leftrightarrow \mathbf{u}'_r$ detected in imagery rectified with R'_l and R'_r. They are related to the correct rectified coordinates $\tilde{\mathbf{u}}_l \leftrightarrow \tilde{\mathbf{u}}_r$, satisfying the epipolar constraint by

$$\tilde{\mathbf{u}}_l \simeq K_l \check{R}_l^T K_l^{-1} \mathbf{u}'_l, \quad \tilde{\mathbf{u}}_r \simeq K_r \check{R}_r^T K_r^{-1} \mathbf{u}'_r. \tag{6}$$

For an estimate of $\check{R}_l$ and $\check{R}_r$, the epipolar error ϵ_i is

$$\epsilon_i = f\frac{\check{R}_{l[\mathbf{2}]}^T K_l^{-1} \mathbf{u}'_{l_i}}{\check{R}_{l[\mathbf{3}]}^T K_l^{-1} \mathbf{u}'_{l_i}} - f\frac{\check{R}_{r[\mathbf{2}]}^T K_r^{-1} \mathbf{u}'_{r_i}}{\check{R}_{r[\mathbf{3}]}^T K_r^{-1} \mathbf{u}'_{r_i}}, \tag{7}$$

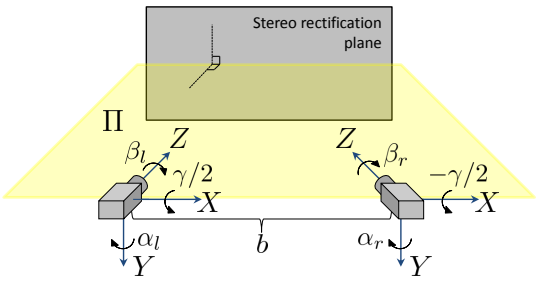


Figure 2: The parameterization of the rotation angles Φ. In the rectified pose, the cameras principal axes are parallel, and lie in the plane Π. The rectified coordinates are projected to a plane orthogonal to Π.

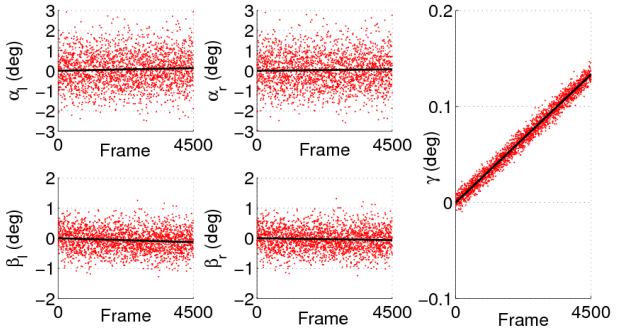


Figure 3: Ground truth simulated change in angles Φ' (black line), and the initial optimized estimates Φ (red dots).

where $R^T_{a[\mathbf{b}]}$ means row b of matrix R^T_a. The re-calibration objective function is the sum of squared epipolar errors ϵ:

$$\underset{\check{R}_l,\check{R}_r}{\operatorname{argmin}} \sum_{i=1}^{N} \epsilon_i^2, \tag{8}$$

giving the maximum likelihood estimate of $\check{R}_l$ and $\check{R}_r$, from which the new $\hat{S}$ stereo extrinsics can be recovered:

$$\hat{S} = \langle \hat{Q}_r, \hat{Q}_l^* \rangle, \tag{9}$$

$$\hat{Q}_l = \left[\left(\check{R}_l^T \, R_l' \right)^T | \mathbf{0} \right] \rightarrow \hat{Q}_l^* = \left[\check{R}_l^T \, R_l' | \mathbf{0} \right], \tag{10}$$

$$\hat{Q}_r = \left[\left(\check{R}_r^T \, R_r' \right)^T | \left(\check{R}_r^T \, R_r' \right)^T (-b, 0, 0)^T \right], \tag{11}$$

where $\langle \hat{Q}_r, \hat{Q}_l^* \rangle$ is the projection $\hat{Q}_l^*$ followed by $\hat{Q}_r$.

As we can only use epipolar constraints, there is no means for correcting the stereo baseline estimate b. We introduce a method to partially address this in section 4.3. We restrict the optimized extrinsic pose by 1 DOF as a result and instead optimize the 5 DOF vector of Euler angles $\Phi = [\alpha_l, \beta_l, \alpha_r, \beta_r, \gamma]^T$ by minimizing (8). Referring to Fig. 2, the rotations $\check{R}_l$ and $\check{R}_r$ are

$$\check{R}_l = R_X(\gamma/2) \, R_Z(\beta_l) \, R_Y(\alpha_l) \tag{12}$$

$$\check{R}_r = R_X(-\gamma/2) \, R_Z(\beta_r) \, R_Y(\alpha_r), \tag{13}$$

where R_A is the right-handed rotation about the axis A. Euler angles are a suitable parameterization as the initial extrinsic estimate is assumed to be near the solution, and the expected changes in angles are small.

3. Solution Covariance and Over Fitting

In practice, the correspondences $\mathbf{u}_l' \leftrightarrow \mathbf{u}_r'$ will be corrupted with noise and the ability to accurately estimate Φ from these is dependent on many factors. These include: the focal lengths, baseline, number of correspondences, spatial distribution of correspondences, and the depth of the scene points. Small rotation angles Φ make over-fitting a concern.

To test this, we simulated a time dependent change in the extrinsic pose of a $b = 150mm$ baseline, 640×480 resolution ($f = 1000pix$) stereo camera. For each stereo pair, 1000 random correspondences were generated, and uncorrelated Gaussian noise ($\sigma = 0.5pix$) added. The disparity values ranged between 1 and $25pix$, or equivalently depths Z between 3 and $150m$. Fig. 3 shows the simulated angular changes (black), and the noisy estimates of Φ (red).

3.1. Solution Covariance

Assuming that Φ is an unbiased estimate of the solution Φ', with expected error covariance $\mathcal{C} = \mathcal{E}\left[(\Phi - \Phi') (\Phi - \Phi')^T \right]$, the Cramér-Rao lower bound $\mathcal{C}$ is greater than or equal to the inverse of the Fisher information matrix F, which is the score variance at the solution [15]:

$$\mathcal{C} = \mathcal{E}\left[(\Phi - \Phi') (\Phi - \Phi')^T \right] \geq F^{-1} \tag{14}$$

$$F = \mathcal{E}\left[\left(\frac{\partial \ln p(\epsilon|\Phi)}{\partial \Phi} \right)^T \left(\frac{\partial \ln p(\epsilon|\Phi)}{\partial \Phi} \right) \right]. \tag{15}$$

Where $p(\epsilon|\Phi)$ is the conditional error probability. If the measurement errors of the imaged points are zero-mean Gaussian, then we can assume that $\epsilon \sim \mathcal{N}(0, \sigma)$ at the solution, and (15) can be written as

$$F = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(\frac{\partial \epsilon_i}{\partial \Phi} \right)^T \left(\frac{\partial \epsilon_i}{\partial \Phi} \right). \tag{16}$$

The summation in (16) is taken over all n correspondences, and the Jacobian $\frac{\partial \epsilon_i}{\partial \Phi}$ is the change in error with respect to the change in parameters Φ at the solution:

$$J_i = \left[\begin{array}{ccccc} \frac{\partial \epsilon_i}{\partial \alpha_l} & \frac{\partial \epsilon_i}{\partial \beta_l} & \frac{\partial \epsilon_i}{\partial \alpha_r} & \frac{\partial \epsilon_i}{\partial \beta_l} & \frac{\partial \epsilon_i}{\partial \gamma} \end{array} \right], \tag{17}$$

which, for the simple case where $\Phi = \mathbf{0}^T$ is

$$J_i|_{\mathbf{0}^T} = \left[\begin{array}{cccc} \frac{-x_{l_i} y_{l_i}}{f} & -x_{l_i} & \frac{x_{r_i} y_{r_i}}{f} & x_{r_i} & \frac{f^2 + y_{l_i}^2 + y_{r_i}^2}{2f} \end{array} \right]. \tag{18}$$

$\mathcal{C}$	α_l	β_l	α_r	β_r	γ
α_l	0.040	0.031	-0.010	0.030	0.019
β_l	0.031	3.142	0.070	3.127	1.969
α_r	-0.010	0.070	0.041	0.070	0.044
β_r	0.030	3.127	0.070	3.117	1.961
γ	0.019	1.969	0.044	1.961	1.235

(a) Pipe dataset (see Fig. 1a). All scene points are within $300mm$ of the camera. $\det(\mathcal{C}) = 8.452 \times 10^{-42}$.

$\mathcal{C}$	α_l	β_l	α_r	β_r	γ
α_l	178.414	2.562	178.884	2.737	-0.013
β_l	2.562	0.967	2.710	0.979	0.007
α_r	178.884	2.710	180.958	2.979	-0.013
β_r	2.737	0.979	2.979	1.002	0.007
γ	-0.013	0.007	-0.013	0.007	0.001

(b) Outdoor dataset 1 (see Fig. 1b). Many scene points are $> 10m$ from the camera. $\det(\mathcal{C}) = 4.602 \times 10^{-37}$.

Table 1: Covariance matrices for the correspondences in (a) Fig. 1a and (b) Fig. 1b. The units are deg^2/pix^2, and all values have been scale by 1.0×10^3 for display purposes.

From (6), $(x_l, y_l)^T = (\tilde{u}_l - u_0, \tilde{v}_l - v_0)^T$ and $(x_r, y_r)^T = (\tilde{u}_r - u_0, \tilde{v}_r - v_0)^T$. Due to its complexity we omit here the full Jacobian. For most perspective cameras with average fields of view the component $\frac{\partial \epsilon}{\partial \gamma}$ dominates the magnitude of J, suggesting that γ will be the most reliable estimate.

Table 1 shows the covariance matrices for the sets of correspondences in Fig. 1a and Fig. 1b. The variances of the angles (leading diagonal) differ significantly in the examples, and although the number of correspondences used was similar, the determinant of $\mathcal{C}$ for the pipe example is several orders of magnitude smaller than the outdoor 1 example. For the outdoor 1 example, the majority of the scene points are distant, and there is a large covariance between the α angles (α_l and α_r, highlighted in blue), as well as the β angles (β_l and β_r, highlighted in red)[1]. This shows that it is primarily the *relative* angles $\delta\alpha = \alpha_l - \alpha_r$ and $\delta\beta = \beta_l - \beta_r$ being estimated (see Fig. 5). For example, if points at an infinite distance are observed in a perfectly rectified stereo pair, such that $\mathbf{u}'_l = \mathbf{u}'_r$, the epipolar errors $\sum \epsilon_i^2$ will be zero for any rotations where $\beta_l = \beta_r$ ($\delta\beta = 0$). In effect this is attempting to estimate a small translation using points at infinity (Fig. 4). It is only when $\beta_l \neq \beta_r$ that $\sum \epsilon_i^2 > 0$.

4. Kalman Filter Re-Calibration

Given the noisy estimates Φ of the extrinsic pose obtained from the non-linear minimization of the epipolar errors, we use a KF [13] to produce a smoothed estimate $\hat{\Phi}$. We use a stationary process model so that we have at time k $\hat{\Phi}_k = \hat{\Phi}_{k-1}$, although more complex models could be used.

[1]For any point at infinity, $u'_l = u'_r$, so $\frac{\partial \epsilon}{\partial \alpha_l} = \frac{\partial \epsilon}{\partial \alpha_r}$ and $\frac{\partial \epsilon}{\partial \beta_l} = \frac{\partial \epsilon}{\partial \beta_r}$.

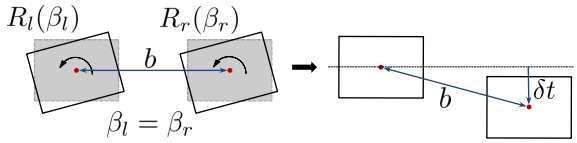


Figure 4: For a point at infinity, only relative angles can be estimated, for example $\delta\beta = \beta_l - \beta_r$. Rotating the cameras by the same angle $\beta_l = \beta_r$ ($\delta b = 0$) is approximately equivalent to adding a small translation change δt, and estimating small translations with distal points is problematic.

The lower bound $\mathcal{C}_k$ evaluated at time k is used as the measurement noise covariance. The process noise covariance $\mathcal{Q}$ is set to

$$\mathcal{Q} = \left(\frac{\pi}{180}\right)^2 \left(\frac{\tau}{60 \times fps}\right)^2 \text{Diag}(1, 1, 1, 1, 0.25), \quad (19)$$

where fps is frames per second, and τ is the selected angular rate of the process noise with units of degrees per minute.

4.1. Update Equations

The time update predictions for the camera state $\hat{\Phi}_k^-$, error covariance $\mathcal{P}_k^-$, and Kalman gain $\mathcal{K}_k$ are

$$\hat{\Phi}_k^- = \hat{\Phi}_{k-1} \quad (20)$$

$$\mathcal{P}_k^- = \mathcal{P}_{k-1} + \mathcal{Q} \quad (21)$$

$$\mathcal{K}_k = \mathcal{P}_k^- \left(\mathcal{P}_k^- + \mathcal{C}_k\right)^{-1}, \quad (22)$$

from which the updated estimate of the camera state $\hat{\Phi}_k$ and error covariance $\mathcal{P}_k$ are evaluated as

$$\hat{\Phi}_k = \hat{\Phi}_k^- + \mathcal{K}_k \left(\Phi - \hat{\Phi}_k^-\right) \quad (23)$$

$$\mathcal{P}_k = \left(I_{5\times5} - \mathcal{K}_k\right) \mathcal{P}_k^-. \quad (24)$$

4.2. Initializing the State Covariance

We estimate the initial state covariance $\mathcal{P}_{k=0}$ by generating 50 perfectly rectified frames of checkerboard scene points (120 points per frame). Random poses of the cameras with respect to the checkerboard target are simulated. Gaussian noise is then added to each image coordinate with $\sigma = 0.25pix$. The reprojection errors are defined as a function of the Euler angles (6) — the y error component is (7). The initial estimate $\mathcal{P}_{k=0}$ is calculated from the lower bound of the solution uncertainty.

Figure 5 shows the KF results $\hat{\Phi}$ obtained from the original optimized estimates Φ in the example in Fig. 3 using the process noise rate $\tau = 1e^{-3}$. It is clear from Fig. 5 that the KF estimates of the individual angles $\alpha_l, \alpha_r, \beta_l, \beta_r$ do not accurately estimate the simulated angles. However, the *differential* angles $\delta\alpha = \alpha_l - \alpha_r$ and $\delta\beta = \beta_l - \beta_r$ shown in the same figure are close approximations of the simulated differential angles. Note that γ is also a differential angle, and its filter estimate is very close to the simulated values.

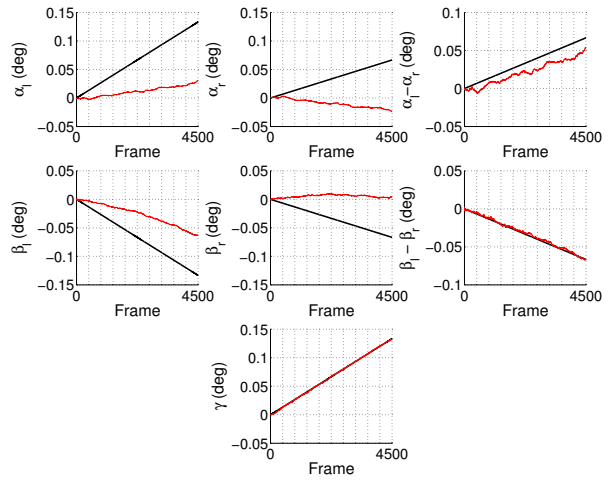


Figure 5: Ground truth angles Φ (black) and KF estimates $\hat{\Phi}$ (red) – original estimates Φ shown in Fig. 3. The differential angles $\delta\alpha = \alpha_l - \alpha_r$, $\delta\beta = \beta_l - \beta_r$ are also shown.

4.3. Baseline Estimation

The true baseline distance cannot be measured from stereo correspondences, however, it may be estimated using additional information. Examples include inertial or wheel odometry, fixed reference fiduciary markers, or structured light measurement observable in both images. Here, we used the following per-frame method to obtain the results in section 5. We assume that triangulated distance to a scene point $\mathbf{X}_i$ should be the same using both the original and re-calibrated extrinsics. We denote these l_i and l'_i, respectively. Since distances are proportional to the triangulated depths (see 4), we estimate the new baseline $\hat{b}$ as

$$\hat{b} = \frac{b}{n} \sum_{i=1}^{n} \frac{l_i}{l'_i}. \tag{25}$$

The summation is only taken over the nearest $n = 5$ stereo correspondences each frame as the nearest points are the most suitable for resolving translation magnitudes.

5. Experiments and Results

To evaluate the approach, we present a range of experimental online re-calibration results including visual odometry for the datasets in Fig. 1 (see table 2), and scene reconstruction using the dataset described in Sect. 5.

For all datasets, Harris corners [10] were detected in image pairs rectified using the original extrinsics. Sparse stereo correspondences were found by thresholding the cosine similarity between SIFT descriptors [14] for each feature. Although sub-pixel accuracy Harris corners were found, Zero-Normalized Cross Correlation (ZNCC) was used to refine the correspondences and improve accuracy.

	Pipe	Outoor 1	Outdoor 2
Camera	Assembled	Commercial	Assembled
Resolution	1202x768	640x480	1033x768
#images	971	8278	7567
fps	7.5	15	7.5
f (pix)	1203	811	1781
b (mm)	156	120	342
# stereo	1236	947	885
Length (m)	7.1	5477	6247

Table 2: Summary of the visual odometry datasets (see also Fig. 1). The notation # stereo is the mean number of stereo correspondences found per frame. The camera parameters are given for the stereo rectified images.

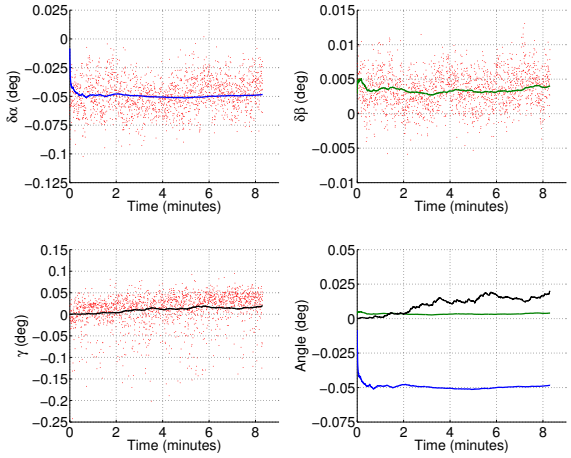
Importantly, we constrain the right stereo feature to an epipolar box and not a line.

For the visual odometry results, temporal correspondences between adjacent stereo pairs were found by thresholding the ambiguity ratio [14] between SIFT descriptors. Visual odometry estimates were computed using both the original and the re-calibrated stereo extrinsic pose. The 6 DOF change in pose Q between the left camera frames was estimated using Perspective-n-Points (PnP) and RANSAC [7], followed by non-linear minimization of the image reprojection errors. The KF process noise was set to $\tau = 0.001$ for each dataset, and $\mathcal{P}_{k=0}$ estimated using the method in Sect. 4.2.

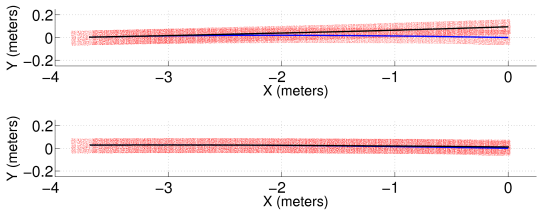
Pipe Dataset The stereo camera, original epipolar errors, and sample rectified imagery for the pipe dataset are shown in Fig.1a. As described in [9], the camera observed the upper surface of a $400mm$ diameter steel pipe as it moved forwards and then in reverse through the pipe. Lighting via nine LEDs was mounted to the camera housing, which raised the temperature of the camera housing from $25 - 30^oC$ ambient at the start to $27 - 38^oC$ at the end. We attribute the time dependent change in epipolar errors to thermal expansion.

The KF estimates of the camera rotation angles, visual odometry estimates, and 3D point clouds with original and re-calibration extrinsics are shown in Fig.6a, 6b, and 6c. Although GPS ground truth is unavailable, all scene points belong to the same curved surface, so the reconstructions in both directions should align. There is a large misalignment using the original extrinsic calibration, which is improved significantly using the online re-calibration estimates.

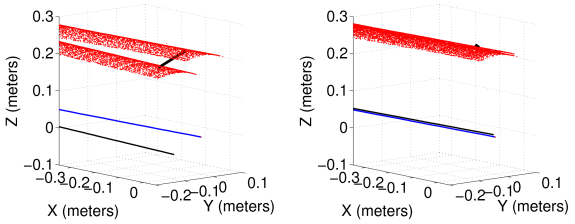
Outdoor Dataset (Camera 1) The first outdoor dataset (Fig.1b) includes imagery from a short baseline Pointgrey Bumblebee2 stereo camera. The rectified imagery was created using the supplied calibration data. The KF estimates of the extrinsics are provided in Fig.7a, and the compari-



(a) KF estimates of the rotations angles.



(b) VO result with pipe axis in X direction: original (top) and re-calibrated (bottom).

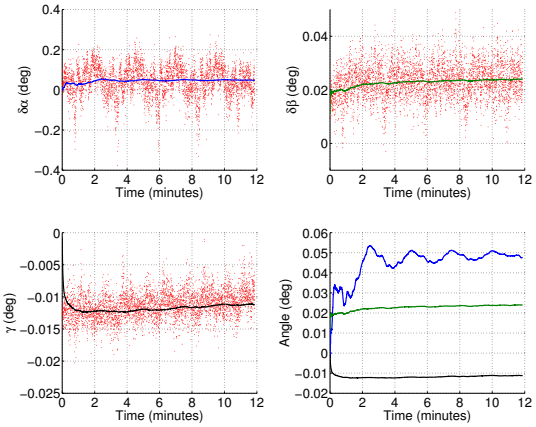


(c) VO result at the start/end: original (left) and re-calibrated (right). The points all belong to the same surface.

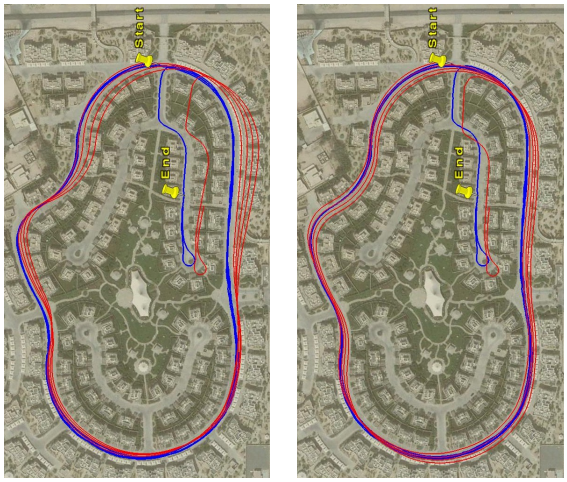
Figure 6: Results for the pipe dataset. The black line near the surface points in (c) connects the same ground truth marker, reconstructed at the start and end of the dataset. The Euclidean errors in the reconstructed coordinate are: 100.1mm for original calibration, 15.1mm for re-calibrated.

son of the visual odometry estimates using the original and re-calibrated extrinsic pose are shown in Fig.7b. The $5Hz$ GPS (non-RTK) measurements collected are included as ground truth. The visual odometry position estimates were linearly interpolated at the time stamps for each of the 1671 GPS readings[2], and then aligned with the GPS by minimizing the sum of squared distances. The average absolute distance errors were: $0.781m$ using the original calibration, and $0.485m$ using online re-calibration.

[2]The GPS z-component was set to zero as the 3D solution was unreliable – the operating environment was approximately planar.



(a) KF estimates of the rotations angles.

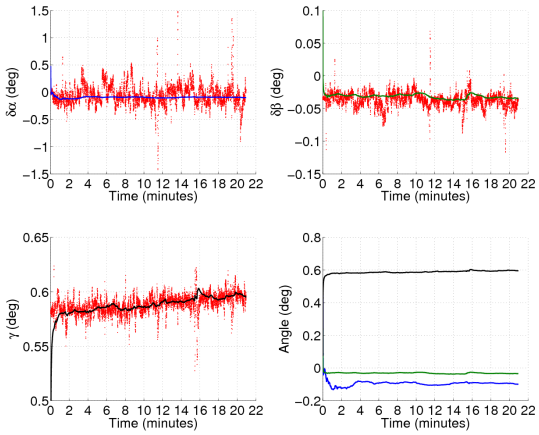


(b) VO (red), and the 5Hz GPS (blue). Left column is the original calibration, and right column the KF re-calibration.

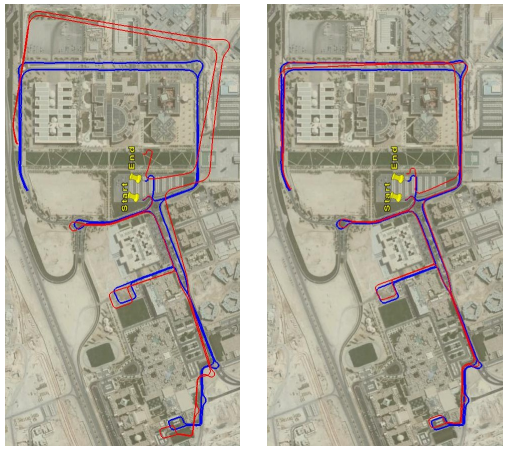
Figure 7: Results for 5.48km outdoor dataset 1 (commercial stereo camera). There are a total of 4 anti-clockwise loops.

Outdoor Dataset (Camera 2) The second outdoor dataset (see Fig.1c) uses a custom $342mm$ baseline stereo camera. Intrinsic and extrinsic parameters were calibrated offline and then we manually flexed the camera to alter the extrinsics. The KF estimates of the angles and visual odometry results are provided in Fig.8. GPS (3045 points at $5Hz$) formed the ground truth using the same techniques described previously. The absolute average distance errors were: $1.632m$ using the original calibration, and $0.700m$ using online re-calibration. As was the case with the first outdoor dataset, re-calibration reduced the rotational drift.

Indoor Scene Fig. 9a shows the stereo camera and a sample image from the left camera used for the indoor controlled test. The stereo head uses the same cameras as in the previous experiment, but with a baseline of $220mm$ and a configurable right camera pose. We collected three datasets



(a) KF estimates of the rotations angles.



(b) VO (red), and the 5Hz GPS (blue). Left column is the original calibration, and right column the KF re-calibration.

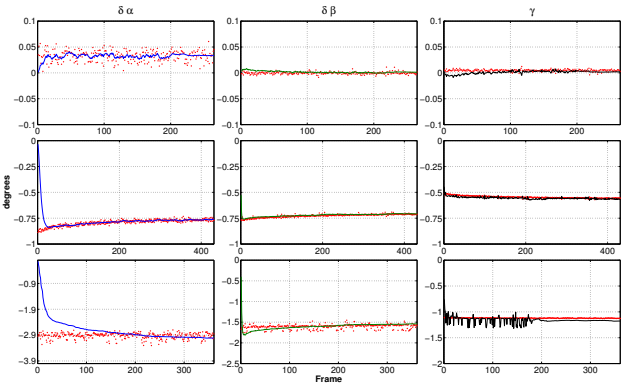
Figure 8: Results for 6.25km outdoor dataset 2.

(1, 2 and 3) observing the same indoor scene, each with a different right camera pose. Ground truth estimates of the extrinsic pose for each set were obtained using a checkerboard target. Dataset 1 was chosen as the *reference* calibration. The stereo correspondences for each set were found in rectified imagery using this reference calibration. The online KF re-calibration was used to estimate the changes from the reference calibration, as shown in Fig. 9b. The final KF results are compared to the ground truth in table 3.

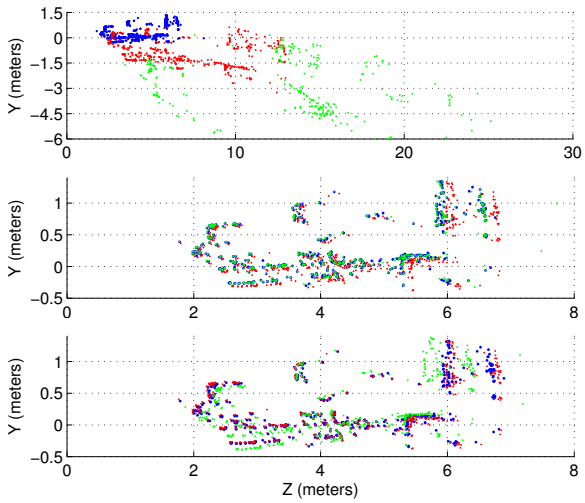
As expected, the performance degrades with large changes from the reference calibration. Although the errors for α_l and α_r appear large for set 1, the resulting change in the stereo disparity and scene reconstruction remained relatively small (see table 3). The standard deviation of the disparity (pix) is similar to the checkerboard calibration reprojection values of $(\sigma_x, \sigma_y) = (0.231, 0.212)pix$ which is itself only an estimate of the true extrinsic pose.

To better visualize the performance of the re-calibration, the overhead views of the scene reconstruction for each

(a) The stereo camera and sample image.



(b) The raw online calibration angle estimates (red points) and KF estimates (solid lines). Each row shows the differential angle estimates for each of the 3 datasets (changing right camera pose).



(c) The top view of the scene reconstructions for set 1 (blue), set 2 (red) and set 3 (green) using: original calibration (top row); checkerboard calibration (middle row); online KF re-calibration (bottom row).

Figure 9: Hardware and results for the indoor dataset.

set are shown in Fig. 9c: the first row uses the reference calibration for each set; the second row uses the checkerboard calibration; and the third row uses the online re-calibration. These reconstructions were produced using the exact same stereo correspondences detected in a single image pair from each set, and are all in the left camera coordinate frame. The results using online re-calibration are significantly more consistent than those using the reference

	calib_1	opt_1	calib_2	opt_2	calib_3	opt_3
α_l	0.00	-0.294	-0.362	-0.546	-1.137	0.829
α_r	0.00	-0.328	0.456	0.216	1.613	3.842
$\delta\alpha$	0.00	0.033	-0.818	-0.762	-2.750	-3.014
β_l	0.00	0.051	-0.127	-0.108	-0.367	-2.481
β_r	0.00	0.050	0.588	0.600	1.369	-0.940
$\delta\beta$	0.00	0.001	-0.716	-0.708	-1.736	-1.541
γ	0.00	0.002	-0.565	-0.566	-1.123	-1.179

Table 3: The changes in angles from the reference calibration using: offline checkerboard calibration (calib); online re-calibration (opt). All values have units of degrees. The subscripts calib_n and opt_n refer to the image set.

	mean	std. dev.
Euclidean Error (mm)	24.80	22.67
Euclidear Error (%)	0.436	0.329
Disparity Difference (pix)	1.076	0.212
Disparity Difference (%)	1.281	0.502

Table 4: Statistics for the Euclidean reconstruction and disparity differences between the checkerboard calibration and online re-calibration for set 1.

calibration for each set. Observe that there are some inconsistencies in the reconstructions for each set using the checkerboard calibration. Again, it too is only an estimate of the true extrinsic pose.

6. Conclusions

We presented an algorithm for online continuous stereo extrinsic re-calibration that estimates a separate extrinsic pose for each image pair using sparse stereo correspondences. An initial 5 DOF extrinsic pose estimate (relative camera orientations/fixed baseline) is found by minimizing stereo epipolar errors, and then refined using a Kalman Filter (KF). The KF measurement covariance is the lower bound of the per-frame solution uncertainty, which is dependent on the number and distribution of the scene point correspondences, as well as the camera focal length and stereo baseline. If only a small number of stereo correspondences can be found, they simply can be combined over multiple frames before estimating the extrinsic pose as no temporal constraints are used. Our results for visual odometry using a range of real datasets in different environments show that accuracy is improved using our technique compared to the original extrinsic calibration. Our future work will explore improved methods for estimating the change in baseline length.

7. Acknowledgements

This paper was made possible by the support of NPRP grants (# NPRP 08-589-2-245 and 09-980-2-380) from the Qatar National Research Fund. The statements made herein are solely the responsibility of the authors.

References

[1] M. Björkman and J. Eklundh. Real-time epipolar geometry estimation of binocular stereo heads. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(3):425–432, 2002. 1

[2] G. Carrera, A. Angeli, and A. Davison. SLAM-based automatic extrinsic calibration of a multi-camera rig. In *Int. Conference on Intelligent Robots and Systems*, 2011. 1

[3] T. Clarke and J. Fryer. The development of camera calibration methods and models. *Photogrammetric Record*, 16(91):51–66, April 1998. 1

[4] T. Dang, C. Hoffman, and C. Stiller. Continuous stereo self-calibration by camera parameter tracking. *IEEE Transactions on Image Processing*, 18(7):1536–1550, July 2009. 1

[5] T. Dang and C. Hoffmann. Tracking camera parameters of an active stereo rig. In *Pattern Recognition*, volume 4174, pages 627–636. Springer Berlin / Heidelberg, 2006. 1

[6] A. Davison, I. Reid, N. Molton, and O. Stasse. MonoSLAM: real-time single camera SLAM. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(6):1–16, June 2007. 1

[7] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comms. of the ACM*, pages 381–395, 1981. 5

[8] A. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *IEEE CVPR*, 2001. 1

[9] P. Hansen, H. Alismail, B. Browning, and P. Rander. Stereo visual odometry for pipe mapping. In *IROS*, 2011. 5

[10] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings Fourth Alvey Vision Conference*, pages 147–151, 1988. 5

[11] R. Hartley and S. B. Kang. Parameter free radial distortion correction with centre of distortion estimation. In *International Conference on Computer Vision*, 2005. 1

[12] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press, 2003. 2

[13] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960. 4

[14] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 5

[15] J. Shao. *Mathematical Statistics*. Springer Texts in Statistics. Springer-Verlag, second edition, 2003. 3

[16] R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, August 1987. 1

[17] Z. Zhang. On the epipolar geometry between two images with lens distortion. In *Proceedings International Conference on Pattern Recognition*, pages 407–411, 1996. 1

[18] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *ICCV*, 1999. 1