Online Continuous Stereo Extrinsic Parameter Estimation

Peter Hansen  
*Carnegie Mellon University*

Hatem Alismail  
*Carnegie Mellon University*

Peter Rander  
*Carnegie Mellon University*

Brett Browning  
*Carnegie Mellon University*

Follow this and additional works at: [http://repository.cmu.edu/robotics](http://repository.cmu.edu/robotics)

Part of the [Robotics Commons](http://repository.cmu.edu/robotics)

Published In  
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 Crâmer-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
2. Stereo Geometry and Error Metric

2.1. Stereo Pose and Epipolar Constraints

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

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

$$\mathbf{u}_l \simeq K_l \mathbf{R}_l \mathbf{X}_l = K_l \tilde{\mathbf{X}}_l$$
$$\mathbf{u}_r \simeq K_r \mathbf{R}_r \mathbf{X}_r = K_r \tilde{\mathbf{X}}_r,$$

where $\mathbf{R}_l, \mathbf{R}_r$ 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

$$K_l = K_r = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}.$$  

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

The rectified coordinates are related by

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

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

2.2. Calibration Error Metric

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

$$\tilde{\mathbf{R}}_l = \mathbf{R}_l^T \mathbf{R}_l', \quad \tilde{\mathbf{R}}_r = \mathbf{R}_r^T \mathbf{R}_r'.$$

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

$$\mathbf{u}_l \simeq K_l \mathbf{R}_l \mathbf{R}_l'^{-1} \mathbf{u}_l', \quad \mathbf{u}_r \simeq K_r \mathbf{R}_r \mathbf{R}_r'^{-1} \mathbf{u}_r'.$$

For an estimate of $\tilde{\mathbf{R}}_l$ and $\tilde{\mathbf{R}}_r$, the epipolar error $\epsilon_i$ is

$$\epsilon_i = f \frac{\mathbf{R}_l'[3]}{\mathbf{R}_l'[2]} K_l^{-1} \mathbf{u}_l' - f \frac{\mathbf{R}_r'[3]}{\mathbf{R}_r'[2]} K_r^{-1} \mathbf{u}_r'.$$

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.
where $R^T_a b$ means row $b$ of matrix $R^T_a$. The re-calibration objective function is the sum of squared epipolar errors $\epsilon$:

$$\arg\min_{\hat{R}_l, \hat{R}_r} \sum_{i=1}^N \epsilon_i^2,$$  \hspace{1cm} (8)

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

$$\hat{S} = (\hat{Q}_r, \hat{Q}_l^T),$$  \hspace{1cm} (9)

$$\hat{Q}_l = \left[ (\hat{R}_l^T R_l^T)^T \mathbf{0} \right],$$  \hspace{1cm} (10)

$$\hat{Q}_r = \left[ (\hat{R}_r^T R_r^T)^T | (\hat{R}_l^T R_l^T)^T (b, 0, 0)^T \right].$$  \hspace{1cm} (11)

where $(\hat{Q}_r, \hat{Q}_l^T)$ is the projection $\hat{Q}_l^T$ followed by $\hat{Q}_r$.

As we can use only 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 $\hat{R}_l$ and $\hat{R}_r$ are

$$\hat{R}_l = R_X(\gamma/2) R_Z(\beta_l) R_Y(\alpha_l),$$  \hspace{1cm} (12)

$$\hat{R}_r = R_X(-\gamma/2) R_Z(\beta_r) R_Y(\alpha_r),$$  \hspace{1cm} (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 $u_i^l \leftrightarrow u_i^r$ will be corrupted with noise and the ability to accurately estimate $\Phi$ 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 $\Phi$ make over-fitting a concern.

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

#### 3.1. Solution Covariance

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

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

$$F = \mathcal{E} \left[ \frac{\partial \ln p(\epsilon|\Phi)}{\partial \Phi} \frac{\partial \ln p(\epsilon|\Phi)}{\partial \Phi}^T \right].$$  \hspace{1cm} (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

$$F = \frac{1}{\sigma^2} \sum_{i=1}^n \left( \frac{\partial \epsilon_i}{\partial \Phi} \right)^T \left( \frac{\partial \epsilon_i}{\partial \Phi} \right).$$  \hspace{1cm} (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 $\Phi$ at the solution:

$$J_i = \begin{bmatrix} \frac{\partial \epsilon_i}{\partial \alpha_l} & \frac{\partial \epsilon_i}{\partial \beta_l} & \frac{\partial \epsilon_i}{\partial \alpha_r} & \frac{\partial \epsilon_i}{\partial \beta_r} & \frac{\partial \epsilon_i}{\partial \gamma} \end{bmatrix},$$  \hspace{1cm} (17)

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

$$J_i|_{0^T} = \begin{bmatrix} -x_{i^l} y_{i^l} & x_{i^l} y_{i^l} & x_{i^l} & f^2 + y_{i^l}^2 + y_{i^r}^2 \end{bmatrix}. $$  \hspace{1cm} (18)
Table 1: Covariance matrices for the correspondences in (a) Fig. 1a and (b) Fig. 1b. The units are deg²/pix², and all values have been scaled by 1.0 × 10³ for display purposes.

<table>
<thead>
<tr>
<th>C</th>
<th>α₁</th>
<th>β₁</th>
<th>αᵣ</th>
<th>βᵣ</th>
<th>γ</th>
</tr>
</thead>
<tbody>
<tr>
<td>α₁</td>
<td>0.040</td>
<td>0.031</td>
<td>-0.010</td>
<td>0.030</td>
<td>0.019</td>
</tr>
<tr>
<td>β₁</td>
<td>0.031</td>
<td>3.142</td>
<td>0.070</td>
<td>3.127</td>
<td>1.969</td>
</tr>
<tr>
<td>αᵣ</td>
<td>-0.010</td>
<td>0.070</td>
<td>0.041</td>
<td>0.070</td>
<td>0.044</td>
</tr>
<tr>
<td>βᵣ</td>
<td>0.030</td>
<td>3.127</td>
<td>0.070</td>
<td>3.117</td>
<td>1.961</td>
</tr>
<tr>
<td>γ</td>
<td>0.019</td>
<td>1.969</td>
<td>0.044</td>
<td>1.961</td>
<td>1.235</td>
</tr>
</tbody>
</table>

(a) Pipe dataset (see Fig. 1a). All scene points are within 300mm of the camera. det(C) = 8.452 × 10⁻⁴².

<table>
<thead>
<tr>
<th>C</th>
<th>α₁</th>
<th>β₁</th>
<th>αᵣ</th>
<th>βᵣ</th>
<th>γ</th>
</tr>
</thead>
<tbody>
<tr>
<td>α₁</td>
<td>178.414</td>
<td>2.562</td>
<td>178.884</td>
<td>2.737</td>
<td>-0.013</td>
</tr>
<tr>
<td>β₁</td>
<td>2.562</td>
<td>0.967</td>
<td>2.710</td>
<td>0.979</td>
<td>0.007</td>
</tr>
<tr>
<td>αᵣ</td>
<td>178.884</td>
<td>2.710</td>
<td>180.958</td>
<td>2.979</td>
<td>-0.013</td>
</tr>
<tr>
<td>βᵣ</td>
<td>2.737</td>
<td>0.979</td>
<td>2.979</td>
<td>1.002</td>
<td>0.007</td>
</tr>
<tr>
<td>γ</td>
<td>-0.013</td>
<td>0.007</td>
<td>-0.013</td>
<td>0.007</td>
<td>0.001</td>
</tr>
</tbody>
</table>

(b) Outdoor dataset 1 (see Fig. 1b). Many scene points are > 10m from the camera. det(C) = 4.602 × 10⁻³⁷.

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 \(\frac{∂e}{∂\gamma}\) 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 example, and although the number of correspondences used was similar, the determinant of \(C\) 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 (\(α₁\) and \(αᵣ\), highlighted in blue), as well as the β angles (\(β₁\) and \(βᵣ\), highlighted in red)\(^1\).

This shows that it is primarily the relative angles \(δα = α₁ - αᵣ\) and \(δβ = β₁ - βᵣ\) being estimated (see Fig. 5). For example, if points at an infinite distance are observed in a perfectly rectified stereo pair, such that \(u_l' = u_r'\), the epipolar errors \(\sum ε_i^2\) will be zero for any rotations where \(β₁ = βᵣ\) (\(δβ = 0\)). In effect this is attempting to estimate a small translation using points at infinity (Fig. 4). It is only when \(β₁ ≠ βᵣ\) that \(\sum ε_i^2 > 0\).

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 \(\hat{Φ}\).

We use a stationary process model so that we have at time \(k\) \(\hat{Φ}_k = \hat{Φ}_{k-1}\), although more complex models could be used.

\[R^l(β_l) \quad R^r(β_r) \quad \beta_l = β_r\]

Figure 4: For a point at infinity, only relative angles can be estimated, for example \(δβ = β_l - βᵣ\). Rotating the cameras by the same angle \(β₁ = βᵣ\) (\(δβ = 0\)) is approximately equivalent to adding a small translation change \(δt\), and estimating small translations with distal points is problematic.

The lower bound \(C_k\) evaluated at time \(k\) is used as the measurement noise covariance. The process noise covariance \(Q\) is set to

\[Q = \left(\frac{π}{180}\right)^2 \left(\frac{τ}{60 \times fps}\right)^2 \text{Diag}(1, 1, 1, 1, 0.25),\]

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 \(\hat{Φ}_k\), error covariance \(P_k\), and Kalman gain \(K_k\) are

\[\hat{Φ}_k = \hat{Φ}_{k-1}\]
\[P_k = P_{k-1} + Q\]
\[K_k = P_k^{-1}(P_k^{-1} + C_k)^{-1}\]

from which the updated estimate of the camera state \(\hat{Φ}_k\) and error covariance \(P_k\) are evaluated as

\[\hat{Φ}_k = \hat{Φ}_{k-1} + K_k(Φ - \hat{Φ}_{k-1})\]
\[P_k = (I_{5 \times 5} - K_k)P_{k-1}\]

4.2. Initializing the State Covariance

We estimate the initial state covariance \(P_{k=0}\) 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 \(σ = 0.25\)pix. The reprojection errors are defined as a function of the Euler angles (6) — the \(y\) error component is (7). The initial estimate \(P_{k=0}\) is calculated from the lower bound of the solution uncertainty.

Figure 5 shows the KF results \(\hat{Φ}\) obtained from the original optimized estimates \(Φ\) in the example in Fig. 3 using the process noise rate \(τ = 1e^{-3}\). It is clear from Fig. 5 that the KF estimates of the individual angles \(α₁, αᵣ, β₁, βᵣ\) do not accurately estimate the simulated angles. However, the differential angles \(δα = α₁ - αᵣ\) and \(δβ = β₁ - βᵣ\) 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.

\(^1\)For any point at infinity, \(u_l' = u_r'\), so \(\frac{∂e}{∂α} = \frac{∂e}{∂α_l}\) and \(\frac{∂e}{∂β} = \frac{∂e}{∂β_l}\).
Figure 5: Ground truth angles $\Phi$ (black) and KF estimates $\hat{\Phi}$ (red) – original estimates $\Phi$ 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 used the following per-frame method to obtain the results in light measurement observable in both images. Here, we odometry, fixed reference fiduciary markers, or structured additional information. Examples include inertial or wheel

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 should be the same using both the original and recalibrated 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

$$
\hat{b} = \frac{b}{n} \sum_{i=1}^{n} l_i' \quad (25)
$$

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.

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 $P_{k=0}$ 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 $25 - 30^\circ C$ ambient at the start to $27 - 38^\circ C$ 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-

<table>
<thead>
<tr>
<th>Camera</th>
<th>Pipe</th>
<th>Outdoor 1</th>
<th>Outdoor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>Assembled</td>
<td>Commercial</td>
<td>Assembled</td>
</tr>
<tr>
<td>#images</td>
<td>971</td>
<td>8278</td>
<td>7567</td>
</tr>
<tr>
<td>fps</td>
<td>7.5</td>
<td>15</td>
<td>7.5</td>
</tr>
<tr>
<td>f (pix)</td>
<td>1203</td>
<td>811</td>
<td>1781</td>
</tr>
<tr>
<td>b (mm)</td>
<td>156</td>
<td>120</td>
<td>342</td>
</tr>
<tr>
<td># stereo</td>
<td>1236</td>
<td>947</td>
<td>885</td>
</tr>
<tr>
<td>Length (m)</td>
<td>7.1</td>
<td>5477</td>
<td>6247</td>
</tr>
</tbody>
</table>

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.
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: 100.1 mm for original calibration, 15.1 mm for re-calibrated.

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 342 mm 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 anglesler 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: 1.632 m using the original calibration, and 0.700 m 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 220 mm and a configurable right camera pose. We collected three datasets...
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 \( \alpha_L \) and \( \alpha_R \) 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 re-projection 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 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
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, and opt, refer to the image set.

<table>
<thead>
<tr>
<th></th>
<th>calib</th>
<th>calib</th>
<th>opt</th>
<th>opt</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>α_1</td>
<td>0.00</td>
<td>-0.294</td>
<td>-0.362</td>
<td>-0.546</td>
<td>-1.137</td>
</tr>
<tr>
<td>α_2</td>
<td>0.00</td>
<td>-0.328</td>
<td>0.456</td>
<td>0.216</td>
<td>1.613</td>
</tr>
<tr>
<td>δα_1</td>
<td>0.00</td>
<td>0.033</td>
<td>-0.818</td>
<td>-0.762</td>
<td>-2.750</td>
</tr>
<tr>
<td>β_1</td>
<td>0.00</td>
<td>0.051</td>
<td>-1.27</td>
<td>-1.018</td>
<td>-0.367</td>
</tr>
<tr>
<td>β_2</td>
<td>0.00</td>
<td>0.050</td>
<td>0.588</td>
<td>0.600</td>
<td>1.369</td>
</tr>
<tr>
<td>δβ_1</td>
<td>0.00</td>
<td>0.001</td>
<td>-0.716</td>
<td>-0.708</td>
<td>-1.736</td>
</tr>
<tr>
<td>γ_1</td>
<td>0.00</td>
<td>0.002</td>
<td>-0.565</td>
<td>-0.566</td>
<td>-1.123</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclidean Error (mm)</td>
<td>24.80</td>
<td>22.67</td>
</tr>
<tr>
<td>Euclidean Error (%)</td>
<td>0.436</td>
<td>0.329</td>
</tr>
<tr>
<td>Disparity Difference (pix)</td>
<td>1.076</td>
<td>0.212</td>
</tr>
<tr>
<td>Disparity Difference (%)</td>
<td>1.281</td>
<td>0.502</td>
</tr>
</tbody>
</table>

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