Online Calibration of Stereo Rigs for Long-Term Autonomy

Michael Warren¹, David McKinnon² and Ben Upcroft¹

Abstract—Stereo-based visual odometry algorithms are heavily dependent on an accurate calibration of the rigidly fixed stereo pair. Even small shifts in the rigid transform between the cameras can impact on feature matching and 3D scene triangulation, adversely affecting pose estimates and applications dependent on long-term autonomy. In many field-based scenarios where vibration, knocks and pressure change affect a robotic vehicle, maintaining an accurate stereo calibration cannot be guaranteed over long periods.

This paper presents a novel method of recalibrating overlapping stereo camera rigs from online visual data while simultaneously providing an up-to-date and up-to-scale pose estimate. The proposed technique implements a novel form of partitioned bundle adjustment that explicitly includes the homogeneous transform between a stereo camera pair to generate an optimal calibration. Pose estimates are computed in parallel to the calibration, providing online recalibration which seamlessly integrates into a stereo visual odometry framework.

We present results demonstrating accurate performance of the algorithm on both simulated scenarios and real data gathered from a wide-baseline stereo pair on a ground vehicle traversing urban roads.

I. INTRODUCTION

Stereo visual odometry methods suffer from a key deficiency: sensitivity to inaccurate calibration between the rigidly linked stereo pair. For large baselines, even minute changes in translation or rotation can affect epipolar geometry and scene triangulation to such a degree that visual odometry is compromised or rendered impossible. For robotic experiments, this often means tedious re-calibrations of cameras using a calibration object (such as a checkerboard) to guarantee that accurate calibration is known for any one dataset, or deployment in limited scenarios where the geometry is guaranteed fixed.

With the progression of field robots into harsh environments and the push for long-term autonomy utilizing stereo-based visual odometry, the ability to maintain accurate calibration becomes a practical and fundamental issue. Furthermore, the desire to increase the effective operating range of stereo vision requires wide baseline stereo pairs, which in turn increases sensitivity to poor calibration. Vibration on flying vehicles, pressure changes underwater and impacts with vegetation and other obstacles can shift the rotation and translation between a stereo pair that must be accounted for in order to meet the increasing need to perform long-term in real world scenarios.

While engineering has a role to play in ensuring rigidity of a stereo camera rig, weight restrictions and cost can render such a rig impractical or ineffective. This consideration becomes even more pertinent for wide baseline rigs, where even the smallest change in the stereo transform will result in the need for a recalibration.

In this paper, we present a method capable of online estimation (and re-estimation at any time) of a stereo calibration for a rigidly-linked pair of cameras with overlapping views, while providing an up-to-scale pose estimate of a robotic vehicle. We achieve this by explicitly including the parameters of the stereo transform (i.e. the rotation and translation) between a rigid camera pair in a modified bundle adjustment routine and refine both the stereo transform, camera poses and scene geometry simultaneously.

To date, most stereo calibration methods have depended on using a calibration object visible to cameras with overlapping views to extract both the intrinsic properties and stereo transform of the cameras [1], [2], [3], and these are typically restricted to operating offline. Some methods exist for calibration with partially overlapping views [4], but many also specifically apply to cameras with non overlapping views. Carrera, et al. [5] use MonoSLAM as the pose prior to estimating the rigid transform of a non overlapping pair, and Lébraly, et al. [6] use an initial hand-estimated calibration optimized via a modified bundle adjustment algorithm. The first method is fundamentally offline, but both require either specific motions or known scene geometry to complete the calibration. Our method, however, is focused on the optimal solution to stereo calibration of cameras with a high degree of overlap, and specifically in the online case with no other requirement other than the known distance between cameras to provide scale. Alternative approaches exist to online calibration of overlapping stereo cameras, but these either depend on a known motion model [7] or constrain the problem by reducing the degrees of freedom [8]. Our method does not require or impose any of these constraints.

The presented theory is most similar to that of Lébraly, et al., where we extend the traditional bundle adjustment algorithm (best described by Triggs, et al.[9]) with the inclusion of the transform between the rigidly linked stereo camera pair. However, their work focussed on non-overlapping cameras, where complete monocular sequences were computed before any calibration was performed, reducing its potential as an online, in-the-field calibrator. As we are using cameras with overlapping views, we can simultaneously compute the calibration and the pose of both cameras in the stereo pair online. Furthermore, our method does not necessarily require any known scene geometry to recompute the stereo transform, and allows for recalibration at any time useful structure is available. This enables potential long term autonomy in field robotics applications where environmental conditions

¹M. Warren and B. Upcroft are with the Faculty of Science and Engineering, Queensland University of Technology, Brisbane, Qld, Australia. michael.warren, ben.upcroft at qut.edu.au
²D. McKinnon is with IPLEX, 11755 Wilshire Blvd, Suite 2000, Los Angeles, CA 90025, USA David at ipplex.com
can adversely affect calibration over long periods.

We present results of this method in simulation and also demonstrate the method on a ground-based stereo dataset, comparing the results to a known calibration and Inertial Navigation System (INS) ground truth pose.

For clarity, in this paper, ‘intrinsics’ refers to the the internal properties of a camera (such as focal length, principle point and distortions). In addition, we define the ‘stereo’ or ‘homogeneous’ transform as that between a base camera and other rigidly fixed cameras for the purposes of partitioning the parameters in our bundle adjustment algorithm. Extrinsics or pose denotes the position and orientation of cameras in a global co-ordinate frame. The rest of this paper is presented as follows: Section II presents an overview of the online calibration system, Section III presents the novel stereo camera bundle adjustment method. Section IV presents the online stereo calibration method, Section V presents the experiments demonstrating the online calibration method and finally, Section VI presents practical results of the method on gathered real-world datasets.

II. SYSTEM OVERVIEW

The purpose of our algorithm is to recover both the poses of a stereo pair at each time step $i$ similarly to traditional visual odometry [10], but also to recover the 6-DOF stereo transform between the rigidly linked camera pair from an initial approximation. We assume that the camera intrinsics have been estimated from a traditional monocular calibration for each camera and that the magnitude of the stereo transform $T_0^i = [R^k_t^k]$ between the 0th (base) camera and 1st camera has been estimated from an initial stereo calibration. We will show that this need only be done once when the cameras are first attached to the robot.

An overview of the system can be seen in Figure 1. Note that the pose of the cameras are computed at the same time as the calibration in the looped part of the stereo calibration routine. For clarity, we present the modified bundle adjustment separately (Sec III) from the online stereo calibration methodology (Sec IV).

III. STEREO BUNDLE ADJUSTMENT

The modified bundle adjustment algorithm closely follows the implementation given in [11]. However, we extend and specialize the algorithm through a modified partitioned parameter vector that separates parameters into the shared stereo transform, extrinsics and scene points, and include both physical cameras (not just the 0th or base camera) in addition to the homogenous transform of the stereo pair to form a truly optimal bundle adjustment. Additionally, we allow easy inclusion or exclusion of the stereo transform as parameters to optimize while maintaining a sparsified version of the modified Hessian for computational efficiency.

Given $m$ ($j \in [1, \ldots, m]$) scene points observed at $n$ unique timepoints/locations ($i \in [1, \ldots, n]$) by a single camera ($P$), the traditional model used for the projection of point $j$ in space ($X_j \in \mathbb{P}^3$) into its location in image $i$ ($x_{i,j} \in \mathbb{P}^2$) in a Euclidean coordinate frame is straightforward [11]:

$$x_{i,j} \simeq K[R_i t_i]X_j$$ (1)

where $K$ is the standard camera intrinsics matrix encoding the internal properties of the camera, and $R_i$ and $t_i$ denote the pose of the base camera $P$ at time $i$. Here we extend the single camera case to 2 unique rigidly linked cameras ($k \in [0,1]$) and express the second camera in terms of the base camera via the stereo transform $T_0^i = [R^k_t^k]$,

$$x_{i,j} \simeq K^k[R_i t_i]T_0^i X_j$$ (2)

additionally separating intrinsics $K^k$ to be unique to each camera. For generalized visual odometry the transform $T_0^i$ would remain fixed. However, in this paper we include these parameters as additional variables over which to optimize during online calibration, and then include them as fixed parameters in an optimal stereo camera bundle adjustment.

A. Modified Bundle Adjustment Algorithm

For clarity of our implementation, we re-derive bundle adjustment with three separated parts to optimize: scene points, shared and independent parameters; allowing the stereo transform to be shared amongst a group of camera pairs but the poses at each time step to remain independent. We present the modified bundle adjustment with the inclusion of the stereo parameters and implement a new, sparse, Hessian that is fundamental to maintain an efficient algorithm. This derivation closely follows that of [11], but results in a system with additional Jacobian components as a result of the additional parameters. The first deviation from
As we form expressions for the partial derivatives of (2) with Jacobian sparsity. In order to compute the Jacobian matrix the approximate Hessian (and ̂stereo parameters where the general bundle adjustment method is the presence of a partitioned parameter vector is expressed as a combination base camera at each time step) in a global frame: (, ). We also define the observation error:

\[ J_{x_{ij}} = \begin{bmatrix} \Delta \theta_S \\ \Delta \theta_I \\ \Delta \theta_P \end{bmatrix}, \quad N_{ij} = \begin{bmatrix} e_s \\ e_I \\ e_P \end{bmatrix} \]

From this modified partitioned parameter vector we set up the approximate Hessian \( N = J_{\theta}^T J_{\theta} \) to take advantage of Jacobian sparsity. In order to compute the Jacobian matrix we form expressions for the partial derivatives of (2) with respect to the parameters \( \theta \), defining the partial derivatives as \( A_{ij} = \frac{\partial x_{ij}}{\partial \theta_S}, B_{ij} = \frac{\partial x_{ij}}{\partial \theta_I} \) and \( C_{ij} = \frac{\partial x_{ij}}{\partial \theta_P} \) respectively. We also define the observation error: \( e_{ij} = x_{ij}^k - \hat{x}_{ij}^k \).

As a consequence of the sparsity of the Jacobian matrix \( J_{\theta} \) and the modified parameter vector the augmented normal equations have the following representation:

\[
\begin{bmatrix}
S^* & M & N \\
M^T & I^* & O \\
N^T & O^T & P^* \\
\end{bmatrix}
\begin{bmatrix}
\Delta \theta_S \\
\Delta \theta_I \\
\Delta \theta_P \\
\end{bmatrix} = 
\begin{bmatrix}
e_s \\
e_I \\
e_P \\
\end{bmatrix}
\]

This expression are block diagonal defined according to,

\[
S^k = \sum_{i,j} A^k_{ij} \Sigma^{-1} A^k_{ij} \\
I_i = \sum_{j,k} B^k_{ij} \Sigma^{-1} B^k_{ij} \\
O_j = \sum_{i,k} C^k_{ij} \Sigma^{-1} C^k_{ij}
\]

with the augmentation of the diagonals by multiplication of the Levenberg Marquadt damping parameter \( 1 + \lambda \) resulting in \( S \rightarrow S^*, I \rightarrow I^* \) and \( P \rightarrow P^* \). The other remaining terms in the augmented normal equations (3) are as follows.

\[
M^k = \sum_{j} A^k_{ij} \Sigma^{-1} B^k_{ij} \\
N^k = \sum_{i} A^k_{ij} \Sigma^{-1} C^k_{ij} \\
O_j = \sum_{i} B^k_{ij} \Sigma^{-1} C^k_{ij} \\
I_i = \sum_{j} B^k_{ij} \Sigma^{-1} B^k_{ij} \\
O_j = \sum_{i} B^k_{ij} \Sigma^{-1} C^k_{ij} \\
\]

To solve the system efficiently, we make the following substitutions,

\[
U \equiv \begin{bmatrix} S^* & M \\ M^T & I^* \\ \end{bmatrix}, \quad W \equiv \begin{bmatrix} N \\ O \end{bmatrix}, \\
\Delta \theta_U \equiv \begin{bmatrix} \Delta \theta_S \\ \Delta \theta_I \end{bmatrix} \quad \text{and} \quad \epsilon_U \equiv \begin{bmatrix} e_s \\ e_I \end{bmatrix}
\]

B. Analytical Jacobian

For an efficient implementation of the modified bundle adjustment routine, efforts should be taken to implement the Jacobian analytically. Special care can then be taken to reduce the number of floating point operations incurred.

Briefly, the modified Jacobian can be broken down into a chain of simpler components. Firstly assigning from (2),

\[
x_{ij} = \begin{bmatrix} x_{\beta} \\ y_{\beta} \\ w_{\beta} \end{bmatrix} = [R_i R^k[R_i t^k + t_i]]
\]

therefore,

\[
x_{ij}^k = \begin{bmatrix} x_{\alpha} \\ y_{\alpha} \\ 1 \end{bmatrix} = f(K^k x_{\beta})
\]

where \( f(y) \) results in the affine part of \( y \). Finally making the assignment:

\[
G = \begin{bmatrix} f_{sx} & 0 & \frac{1}{w_{\beta}} \\ 0 & f_{sy} & \frac{-x_{\beta}}{w_{\beta}} \\ \frac{1}{w_{\beta}} & \frac{-y_{\beta}}{w_{\beta}} & 0 \end{bmatrix}
\]

we can concisely state all the components of the Jacobian in simple terms (see Table I). Importantly, they are not considerably more complicated than those in the single camera case and thus can be implemented efficiently.
TABLE I
THE ANALYTICAL DERIVATIVES

<table>
<thead>
<tr>
<th>Stereo Transform</th>
<th>( \frac{\partial x_i}{\partial X} = GR_i )</th>
<th>( \frac{\partial x_i}{\partial \hat{X}} = GR_i \left[ \hat{R}^kX_j \right]_x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extrinsic</td>
<td>( \frac{\partial x_i}{\partial X} = G )</td>
<td>( \frac{\partial x_i}{\partial \hat{X}} = G \left[ R_i(R^kX_j + t^k) \right]_x )</td>
</tr>
<tr>
<td>Scene</td>
<td>( \frac{\partial x_i}{\partial X} = GR_iR^k )</td>
<td></td>
</tr>
</tbody>
</table>

IV. ONLINE STEREO CALIBRATION

The online stereo calibration routine contains four major components:

- Pose Initialisation
- Parallel Monocular Visual Odometry
- Online Stereo Calibration
- Stereo Visual Odometry

where each utilizes a different parameterization of the bundle adjustment algorithm. Typically, pose initialisation, monocular VO and the stereo calibration are completed within the first 20-30 frames. Once the calibration is completed stereo visual odometry can proceed as normal until a particular heuristic is met, such as a low number of inliers matched through epipolar geometry, that warrants recalibration.

It is assumed that before deployment a standard checkerboard or other calibration is used to recover the intrinsic properties of the camera, but also metric scale (Euclidean distance) between the camera pair to be used in re-initialisation. No other components of the stereo transform are kept.

A. Pose Initialisation

First, an initial estimate of pose and scene structure for the first stereo pair is required. Initial pose is setup by first computing an essential matrix \( E_0^{i,j} \) using 5 matched SURF [12] features using MLESAC [13] for robustness, from which the initial camera poses, \( P_0^0 \) and \( P_0^1 \), are extracted [14]. At this point it is possible to recover scale using the assumption that the Euclidean distance between the stereo pair will never vary significantly from the original or any subsequent calibration. The ratio of distances between the cameras from a previous calibration to the current scaled estimate can be applied to the extracted translation and newly generated scene, allowing the algorithm to generate an up to scale pose estimate from the very first frame.

B. Parallel Monocular Visual Odometry

Using SURF feature matches between the previous and current frames, as well as between the pair, each new camera pose is extracted independently \( (P_i) \) independent of \( P_0^i \) using calibrated 3-point pose estimation within a MLESAC robust estimator. The rigid link between the cameras is ignored here, effectively rendering it as a ‘parallel monocular’ visual odometry step, but common 3D scene is used to recover each camera. Following this, the entire set of camera poses and scene are refined inside a bundle adjustment routine, but parameterized such that the cameras are optimized independently, so that there are no rigid links between the left and right camera, in addition to the 3D scene. This means that the matrices \( S, M \) and \( N \) are eliminated in this configuration of bundle adjustment. We term this as a ‘parallel-monocular’ bundle adjustment. If warranted, the scene can be scaled at each time step using the Euclidean distance between the latest camera pair to achieve an up to date scale term. Any newly observed scene is triangulated from the independent cameras and the process is repeated. Any new scene is only reconstructed if it is seen in at least four independent images to ensure accurate triangulation and pose recovery.

C. Stereo Calibration

After an empirically derived number of frames, from which enough translation and scene structure has been accumulated, a stereo calibration is performed on what we term the initialisation set: the set of independent monocular poses (including both the left and right camera) computed in parallel from the monocular VO up until the current point. From this set the camera pair with the smallest average reprojection error and having at least a minimum number of inliers is chosen as the initialisation candidate. The transform \( T_0^1 \) is selected from this pair, and all right hand cameras \( P_0^i \) are re-initialized relative to the base camera \( P_0^i \top \) through this transform. From here, we use the full stereo bundle adjustment presented in Section III and allow optimisation of the stereo transform using this best candidate over the entire initialisation set. Once the bundle adjustment is complete and convergence is identified the rigid transform is fixed, removing the additional 6 shared stereo transform parameters from optimisation in the stereo bundle adjustment. Again, the matrices \( S, M \) and \( N \) are eliminated in this configuration of bundle adjustment.

D. Stereo Visual Odometry

Following the stereo calibration, the algorithm is then allowed to proceed in our standard stereo visual odometry pipeline, first presented in [10]. The camera pair is re-sectioned inside a MLESAC estimator, optimized, new scene structure added and the process repeated. Note that extrinsics for both cameras in the stereo pair are simultaneously computed while the stereo transform remains fixed.

If at any time a heuristic is met, such as the minimum inliers in the stereo pair drops below a threshold, the stereo calibration routine can be re-initialized and performed again to handle any long-term degeneration of the stereo transform.

V. EXPERIMENTS

To evaluate the performance of both the modified bundle adjustment algorithm and online calibration as a whole we present results from simulation and an experiment on real data.

- The first is a simulated 3D point cloud with snapshots of the scene taken by several rigidly linked stereo pairs.
From an artificially modified initial calibration we show the convergence of the solution to the known calibration parameters.

- In the second test, the calibration pipeline is demonstrated on a dataset gathered by driving a vehicle on urban roads with a stereo camera pair placed on the top of the vehicle. The pipeline is used to re-estimate the stereo calibration online from scratch and perform visual odometry in comparison to a known ground truth. We repeat this pipeline at three separate points in the dataset to show robustness to initialisation and consistency in the recovered calibration.

A. Simulation

In this simulation we examine the ability of the bundle adjustment algorithm to converge to an accurate estimate given poor calibration. With noise in both initialized camera positions and the stereo transform, the modified bundle adjustment algorithm should converge to a solution that accurately recomputes both of these parameters.

The test is setup as follows: a uniform point sphere with maximum radius $7m$ consisting of 100 points is generated. Ten random snapshots of the points are taken by a set of stereo cameras with a randomly selected baseline between $0.5 \leftrightarrow 1m$, from a position where the entire point cloud is observable (Fig. 3). We perform the simulation over 10 repeats with randomly distributed noise between $\pm 0.1m$ in base camera position, $\pm 0.1$ radians in stereo rotation and $\pm 0.05m$ in stereo translation. The bundle adjustment algorithm is allowed to converge until the parameter update size falls below a minimum threshold, or a maximum of 3000 steps. We perform the simulation with varying degrees of Gaussian pixel noise: 0, 0.1, 0.2 and 0.3 pixels. The output of the algorithm, including average reprojection error, the error in computed rotation and translation of the stereo calibration, and final error in base camera position is evaluated against the ground truth.

B. Ground Vehicle Dataset

To demonstrate the application of the algorithm to a real-world scenario, a stereo camera dataset was gathered using a passenger vehicle operating on local roads. We choose three sections of the dataset in this experiment (e.g. Fig. 4), covering about 1500 frames each, to show that recalibration could be performed at any point in the dataset.

The visual setup consists of a pair of Firewire 1394B colour Point Grey Flea 2 cameras. These are fixed via an aluminium bar with an approximately $0.75m$ baseline on top of the vehicle, facing in the direction of motion (Figs. 5 & 6). Both cameras use a 4.5mm lens with a field of view of approximately $60^\circ \times 45^\circ$, capturing synchronized stereo images at a resolution of $1024 \times 768$ pixels at $30Hz$.

The camera pair was calibrated before dataset capture using a checker-board and a modified version of the RADDODC toolbox [3] extended by the authors for automatic calibration of stereo pairs, the AMCC Toolbox\(^1\). The results from this calibration were used to generate the known intrinsic parameters for the cameras (Table III, ‘AMCC Toolbox’) and as the ground truth comparison for the recovered stereo calibration.

An Xsens MTi-g INS/GPS system is used as the ground truth for comparison of the stereo camera pose, providing a 6-DOF pose estimate at 120Hz. This output is sub-sampled according to the recorded timestamps to estimate camera pose at each frame capture. The INS unit and GPS antenna are rigidly attached to the stereo camera rig as seen in Fig. 6.

The vehicle covered a distance of approximately 9.5km, and acquired approximately 33,000 frames\(^2\).

---

\(^1\)http://code.google.com/p/amcctoolbox/

\(^2\)The dataset is available at: https://wiki.qut.edu.au/display/cyphy/datasets
The online pose estimator was applied to each of the three sections and used a total of 25 stereo images from which to perform the stereo calibration. Once a refined calibration was obtained from the parallel monocular bundle adjustment and stereo bundle adjustment, the computed stereo transform enabled stereo visual odometry to be performed on the next 1500 frames. As a comparison, stereo visual odometry was performed on the same sequences using the ground truth calibration. These were both registered to the INS/GPS and compared to show accuracy of the output solution.

VI. RESULTS

A. Simulation

In all cases for the simulation the algorithm converged towards the correct calibration and stereo pose (Table II). Of note, even in the presence of relatively large pixel noise, the algorithm is capable of converging to the correct estimates of rotation and translation of the stereo pair. In none of the simulations did the solution fail to converge. It can be seen, however, that the rotational components of the calibration are more consistently observable due to the setup of the simulation, and hence better approximated by the algorithm in this simulated case.

TABLE II
Final Errors for Stereo Bundle Adjustment Simulation with Varying Pixel Noise

| Error Type                    | Pixel Noise
|-------------------------------|-------------
|                              | 0           | 0.1         | 0.2          | 0.3          |
| Avg. Reprojection Error (pixel) | 2.4 × 10⁻³ | 1.9 × 10⁻² | 7.2 × 10⁻²  | 1.6 × 10⁻¹  |
| Avg. Pose Error (m)           | 9.5 × 10⁻³ | 1.0 × 10⁻² | 1.4 × 10⁻²  | 3.1 × 10⁻²  |
| Avg. Stereo Rotation Error (rad) | 4.4 × 10⁻⁴| 4.7 × 10⁻⁴| 9.1 × 10⁻⁴ | 8.7 × 10⁻⁴ |
| Avg. Stereo Translation Error (m) | 5.0 × 10⁻³ | 8.0 × 10⁻³ | 9.0 × 10⁻³  | 1.1 × 10⁻²   |

B. Online Pose Estimate

The results in Table III compare the results of the online stereo calibration algorithm for the three selected sections. Clearly, even without an initialisation to work from, the stereo transform is recoverable to a high degree of accuracy. While correct translations are recovered to within 37mm, rotations are recovered to within 1.3 × 10⁻³ radians or 8 × 10⁻⁵. The rotational and Tx components of these results are within the expected numerical error given by the RADDOC toolbox for the ground truth calibration (see column 2 of Table III). It can be seen, however, that the recovered Ty and Tx translations fall outside these values, partially due to the poor observability of these parameters in the forward facing stereo configuration.

TABLE III
Comparison of Ground Truth and Recovered Extrinsic Calibration on Ground Vehicle Dataset

<table>
<thead>
<tr>
<th>Parameter /Calibration Type</th>
<th>AMCC Toolbox + (Numerical error)</th>
<th>Section 1</th>
<th>Section 2</th>
<th>Section 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tz (mm)</td>
<td>750.7 (7.08 × 10⁻¹)</td>
<td>750.4</td>
<td>749.9</td>
<td>749.9</td>
</tr>
<tr>
<td>Ty (mm)</td>
<td>-3.8 (4.59 × 10⁻¹)</td>
<td>8.4</td>
<td>0.8</td>
<td>6.9</td>
</tr>
<tr>
<td>Tz (mm)</td>
<td>-3.0 (5.91)</td>
<td>-19.4</td>
<td>-33.6</td>
<td>34.6</td>
</tr>
<tr>
<td>Rx (rad)</td>
<td>-5.2 × 10⁻₃ (4.74 × 10⁻²)</td>
<td>-4.0 × 10⁻³</td>
<td>-5.2 × 10⁻³</td>
<td>-3.9 × 10⁻³</td>
</tr>
<tr>
<td>Ry (rad)</td>
<td>-5.3 × 10⁻₃ (7.92 × 10⁻²)</td>
<td>-5.3 × 10⁻³</td>
<td>-5.8 × 10⁻³</td>
<td>-5.9 × 10⁻³</td>
</tr>
<tr>
<td>Rz (rad)</td>
<td>-1.34 × 10⁻³ (5.10 × 10⁻⁴)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Fig. 7. The visual odometry result for (Top) Section 1, (Middle) Section 2 and (Bottom) Section 3 showing INS ground truth (red), VO based on the original RADDOC Toolbox calibration (blue) and VO based on the recovered calibration (green).

The results of the visual odometry based on both the ground truth calibration and recovered calibration are pre-
sent in Figure 7. For each section a total distance of approximately $400 - 500m$ was covered, related to the speed of the vehicle.

From these graphs, it can be seen that the visual odometry from the recovered calibration closely matches the result generated from the ground-truth calibration, and both closely match the ground-truth INS pose while remaining accurately scaled. This result validates the utility of the algorithm in recovering stereo calibration, up-to scale, from online data.

VII. CONCLUSIONS

This paper has presented a novel method of online stereo camera calibration applicable to field robots. Using a modified bundle adjustment algorithm that takes advantage of the rigidity between a stereo pair we have shown that a solution to both the stereo transform and pose of a stereo pair can be recovered from a poor initialisation. Additionally, we have shown that it is possible to recover an accurate stereo calibration online from real-world data. Using this calibration, it is possible to perform feature-based visual odometry that rivals that of using a calibration performed offline using standard algorithms. This result has implications for roboticists in removing the need to frequently perform offline calibrations on wide baseline rigs, but also for robots required to perform over long terms where robustness to environmental impacts is both warranted and necessary. Future work will include applying the technique to a number of additional field-robotics based datasets where repeated online calibrations are necessary. Finally, while we focus on the stereo case in this paper, the implementation of our modified bundle adjustment routine can be easily extended to $n$ rigidly linked cameras, provided that all views overlap to some degree. Additionally, the camera intrinsics can be easily included for each camera in the set for a complete solution.

REFERENCES