Proceedings of Australasian Conference on Robotics and Automation, 3-5 Dec 2012, Victoria University of Wellington, New Zealand.

Visual Sea-floor Mapping from Low Overlap Imagery using Bi-objective Bundle Adjustment and Constrained Motion

Michael Warren, Peter Corke & Ben Upcroft
Queensland University of Technology, Australia
{michael.warren, peter.corke, ben.upcroft}@qut.edu.au

Oscar Pizzaro & Stefan Williams
University of Sydney, Australia
o.pizarro@cas.edu.au
stefanw@acfr.usyd.edu.au

Abstract

In most visual mapping applications suited to Autonomous Underwater Vehicles (AUVs), stereo visual odometry (VO) is rarely utilised as a pose estimator as imagery is typically of very low framerate due to energy conservation and data storage requirements. This adversely affects the robustness of a vision-based pose estimator and its ability to generate a smooth trajectory. This paper presents a novel VO pipeline for low-overlap imagery from an AUV that utilises constrained motion and integrates magnetometer data in a bi-objective bundle adjustment stage to achieve low-drift pose estimates over large trajectories.

We analyse the performance of a standard stereo VO algorithm and compare the results to the modified vo algorithm. Results are demonstrated in a virtual environment in addition to low-overlap imagery gathered from an AUV. The modified VO algorithm shows significantly improved pose accuracy and performance over trajectories of more than 300m. In addition, dense 3D meshes generated from the visual odometry pipeline are presented as a qualitative output of the solution.

1 Introduction

Visual sea-floor mapping is a rapidly growing application for Autonomous Underwater Vehicles (AUVs)[17]. AUVs are well-suited to benthic mapping and monitoring as they remove humans from a potentially dangerous environment, can reach depths human divers cannot, and are capable of long-term operation in adverse conditions. The output of sea-floor maps generated by AUVs has a number of applications in scientific monitoring: from classifying coral in high biological value sites [16] to surveying sea sponges to evaluate marine environment health [8].

In order to generate self consistent visual maps with properly geo-referenced imagery over large swathes, accurate localisation of the AUV is a strict requirement. While localisation is relatively easy for surface vehicles due to GPS access, subsurface vehicles are either dependent on beacon based infrastructure (analogous to GPS localisation) or Simultaneous Localisation and Mapping (SLAM) using on-board sensors. In many subsurface environments of interest, beacon based infrastructure is unavailable or extremely sparse, meaning that SLAM is the only viable option for accurate localisation.

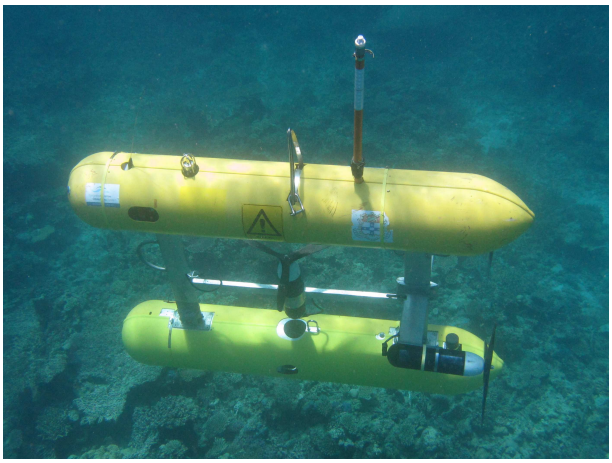

Figure 1: The Sirius AUV on deployment in Scott Reef, WA, Australia

In many AUV based sea-floor monitoring applications, an Information or Delayed state filtered SLAM solution [8, 2] is the standard method to integrate a large number of sensors and achieve an adequate pose solution. For visual mapping, using a set of downward facing cameras and active light strobes, imagery is taken at regular intervals and geo-referenced from the SLAM solution to generate 2D mosaics and 3D reconstructions of the environment [5]. In the current literature, visual information is rarely utilised in these filtered solutions for incre-

mental pose updates (typically termed Visual Odometry (VO)), mostly due to its high computational load and large storage requirements. In the presence of a number of specialised sensors for detecting pose underwater, visual odometry has remained outside most large-scale underwater applications. However, it gains significant benefit in loop-closure events, providing a method of constraining pose drift by detecting previously visited parts of the sea-floor and integrating this information into the pose filter [3].

Many terrestrial and airborne robots utilise VO to estimate vehicle pose from sequential monocular or stereo frames [14, 15, 1, 6, 11], covering distances of many tens of kilometres, with pose accuracy approaching 1% when loop closure is taken into account. In addition, the same has been performed in some underwater scenarios [12]. VO has been demonstrated to perform well as a single estimator for determining pose, but also has the potential to be used in combination with other sensors in a filtered framework [3]. By tracking visual features on the sea-floor it has distinct advantage as a passive pose estimator with a rich information output, and is capable of rivaling much more expensive inertial sensors in generating motion and orientation updates. With increasing speed and efficiency of computational resources, and demonstration over trajectories of tens of kilometres, VO has the potential to fully integrate into the real-time sensor suite in benthic monitoring vehicles, and even perform well as an independent pose estimator.

In contrast to other vision-based sensing scenarios, the imagery from the Sirius AUV [5] (Fig. 1, a model of the very popular SEABed AUV) presents some difficulties when performing 'traditional' VO. In order to conserve energy used for strobing, and access to limited storage and processing, imagery captured by Sirius is of very low frequency and low overlap ($\sim 30\%$), meaning that feature observations are fleeting and difficult to triangulate accurately. This adversely affects estimated pose using VO techniques typically suited to very high overlap imagery. Such limited visual information manifests itself in rapid pose estimate degeneration using standard 6DOF VO techniques. However, by taking advantage of the constrained motion of the AUV (see Sec. 2) and including some additional readings from a minimal set of other sensors, it is possible to constrain the error growth of a VO solution and produce accurate incremental pose estimates over large underwater trajectories, and ultimately combined with loop closure to generate a full SLAM solution. Applications of this research may assist future AUV research in two key ways: deployment of future vehicles at lower cost and increased operation time due to a reduced sensor suite, and capability improvement to existing vehicles by adding additional sensor information to the filtered solution.

This paper presents a method of performing high accuracy sea-floor mapping by integrating low-overlap stereo visual imagery and magnetometer data in a modified visual odometry algorithm. By taking advantage of the constrained motion of the AUV and integrating magnetometer data to correct yaw drift, accurate pose estimation is achieved using a minimal set of sensors. A brief introduction to the methodology, including a novel visual 2-point pose estimator and modified bundle adjustment are presented, and preliminary results on a 300m trajectory are shown. As a qualitative assessment of the trajectory estimation, 3D reconstructions of the observed scene are performed using the image data and pose estimates. The rest of this paper is outlined as follows: Section 2 outlines the experimental apparatus as a motivation for the problem, Section 3 describes the methodological approach to the problem, Section 4 details the experiments used to test the modified pipeline on both a simulated scenario and a selected dataset from the Sirius AUV and Section 5 shows the results of these experiments.

2 The Sirius AUV

The Sirius AUV (Fig. 1) is a modified version of the SEABed AUV, a mid-size underwater robotic vehicle primarily designed for large-scale sea-floor mapping for marine science and reef health monitoring. The AUV is equipped with a large set of oceanographic instruments (see Table 2) including a magnetometer and a high-resolution (1360×1024) downward facing stereo camera pair ($\sim 7.5cm$ baseline) with strobes for imagery. The vehicle typically captures imagery at $1Hz$ from a height of $2m$ above the sea-floor while maintaining a forward velocity of approximately $0.5m/s$. Key to the development of theory presented here, this AUV design is passively stable in pitch and roll, meaning its motion is effectively constrained to only four degrees of freedom. Typically, roll and pitch of the vehicle rarely exceeds 1°, particularly in the still water environments in which the AUV operates, actively avoiding impacts from strong currents and wave motion nearer the surface.

3 Methodological Approach

Here we present a modified visual odometry pipeline, divergent from the standard 6-Degree of Freedom methods typical of visual odometry in terrestrial applications.

Our algorithm is similar to others in consisting of four main repeating steps for each pair of images:

1. SURF based feature matching
2. Camera pose update
3. Structure triangulation
4. Pose and scene optimisation (bundle adjustment)

Sensor	Output	Accuracy
RDI Navigator WN-1200	Heading (Yaw) Roll/Pitch Velocity	$\pm 2^\circ$ $\pm 0.5^\circ$ $\pm 0.2\%$
Digiquartz Pressure Sensor	Depth	$\pm 0.1\%$
Tracklink 1,500 HA USBL	Relative Ship Position (m) Relative Ship Orientation (degrees)	$\pm 0.2m$ $\pm 0.25^\circ$
$2\times$ Prosilica 1360×1024 pixel CCD cameras	Imagery	-

Table 1: A summary of the pose estimation sensor suite on-board the Sirius AUV

The basic stereo visual odometry algorithm is described in detail in a previous paper [14]. In contrast to the basic algorithm, however, the novel component of this work modifies the visual odometry algorithm with two major differences:

- A novel 2-point camera pose estimator that assumes a zero or negligible roll and pitch in the solution (Sec. 3.1).
- The development of a bi-objective bundle adjustment that includes additional sensor inputs as objectives in the optimisation stage, assisting to minimise angular drift in the final pose estimate (Sec. 2).

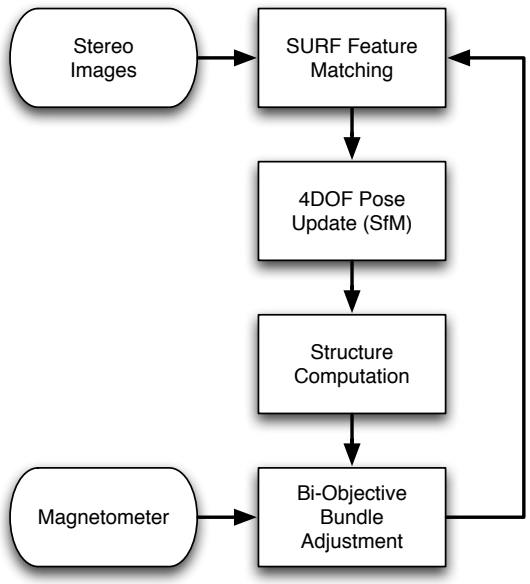

Figure 2: The modified visual odometry pipeline

An overview of this specialised pipeline is shown in Figure 2. In addition, we address 3D mesh generation and texturing from the final pose output as the useful output of such a system, the main application and use of visual imagery of the sea-floor.

We emphasise here that the only input to the proposed pipeline is stereo images and temporally registered magnetometer data, no additional sensors are included.

3.1 Constrained Camera Pose Update

Given scene structure generated from a previous pose update and a set of matched features in the current images, a new camera pose is usually generated in a full 6DOF solution for the orientation and position of the camera. This is achieved by solving a linear system of equations including the observed scene points $\mathbf{X} = \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T$ and their projections (matched features) $\mathbf{x} = \begin{bmatrix} u & v & 1 \end{bmatrix}^T$ into the image. These are used to find the elements of the matrix encoding the camera pose: $\mathbf{M} = [\mathbf{R}|\mathbf{t}]$ via the projection equation:

$$\mathbf{x} = \mathbf{PX}$$

where $\mathbf{P}$, termed the camera matrix, is composed of the camera intrinsics matrix ($\mathbf{K}$) and the camera pose matrix:

$$\mathbf{P} = \mathbf{KM}$$

In most cases, $\mathbf{K}$ is known and fixed, but the parameters of $\mathbf{M}$ are needed for a successful pose update. Expanding the projection equation:

$$\mathbf{K} \begin{bmatrix} r11 & r12 & r13 & t_x \\ r21 & r22 & r23 & t_y \\ r31 & r32 & r33 & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

and multiplying by the skew-symmetric form of $\mathbf{u}$, $[\mathbf{u}]_\times$ gives

$$\begin{bmatrix} 0 & -1 & v \\ 1 & 0 & -u \\ -v & u & 0 \end{bmatrix} \mathbf{K} \begin{bmatrix} r11 & r12 & r13 & t_x \\ r21 & r22 & r23 & t_y \\ r31 & r32 & r33 & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = 0$$

since $[\mathbf{u}]_\times \mathbf{u} = 0$. This matrix equation results in three polynomial equations from which only two are linearly independent, caused by $[\mathbf{u}]_\times$ having rank two. Hence, in the standard 6DOF case, a minimum of 3 points is required to extract the elements which define the pose: $x, y, z, \gamma, \phi, \theta$.

However, by taking advantage of passive stability of the Sirius AUV and assuming that the roll γ and pitch ϕ movement in sequential poses is negligible (*i.e.* zero) a new, constrained 4DOF pose estimate can be developed from the observation of only two points. This concept is similar to the absolute camera pose problem with known vertical direction given by an IMU [7]. Here, the rotation matrix $\mathbf{R}$ is simplified to the following case (we parameterise yaw, θ, in terms of variable q, where $\cos\theta = \frac{1-q^2}{1+q^2}$ and $\sin\theta = \frac{2q}{1+q^2}$):

$$\mathbf{R} = \begin{bmatrix} \frac{1-q^2}{1+q^2} & \frac{-2q}{1+q^2} & 0 \\ \frac{2q}{1+q^2} & \frac{1-q^2}{1+q^2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Hence, the required solution for $\mathbf{M}$ becomes:

$$\begin{bmatrix} 0 & -1 & v \\ 1 & 0 & -u \\ -v & u & 0 \end{bmatrix} \mathbf{K} \begin{bmatrix} \frac{1-q^2}{1+q^2} & \frac{-2q}{1+q^2} & 0 & t_x \\ \frac{2q}{1+q^2} & \frac{1-q^2}{1+q^2} & 0 & t_y \\ 0 & 0 & 0 & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = 0$$

Analytically solving this linear system of equations given two scene points $\mathbf{X}_1, \mathbf{X}_2$ and their projections $\mathbf{x}_1, \mathbf{x}_2$ gives two closed form solutions for q, from which can be extracted four potential values of theta: $\theta_1, -\theta_1, \theta_2, -\theta_2$. By checking the residual of the projections two values are immediately rejected, and the residual of a third point is used to find the correct θ. It is then possible to substitute the correct value for q and recover the other three degrees of freedom. This 2-point pose estimator is placed in a MLESAC[13]-based iterative estimator to achieve robustness in the presence of outliers.

It must be noted here that physical bias in roll and pitch due to poor balance do not adversely affect the solution, as the pose update is only concerned with incremental positions. Frame to frame roll and pitch motions will remain negligible, meaning the algorithm is capable of generating a pose update no matter the initialisation.

3.2 Bi-Objective Bundle Adjustment

Bundle adjustment is often performed after a camera pose update and the triangulation of new structure, in an attempt to minimise the error in both pose and scene structure estimation by posing the problem in a non-linear least-squares iterative optimiser. In visual odometry applications it is normally composed of a sliding window of the most recent camera positions $\hat{\mathbf{P}} = \left[\hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2 ... \hat{\mathbf{P}}_n\right]$ and observed structure $\hat{\mathbf{X}} = \left[\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2 ... \hat{\mathbf{X}}_m\right]$ and optimised by minimizing the residual error in the projection of each estimated 3D point $\hat{\mathbf{X}}_j$ into each camera $\hat{\mathbf{P}}_i$: $\epsilon_{ij(c)} = \mathbf{x}_{ij} - \hat{\mathbf{x}}_{ij}$, where $\mathbf{x}_{ij}$ is the projection of scene point $\mathbf{X}_j$ into camera $\mathbf{P}_i$, and $\hat{\mathbf{x}}_{ij}$ is the projection of the corresponding estimate.

The convergence of the algorithm is quantified by the reduction in the residual cost function over the estimated camera poses and scene structure:

$$\varepsilon_c^2 = \frac{1}{nm} \Sigma_i^n \Sigma_j^m \parallel \epsilon_{ij(c)} \parallel^2$$

that is, the minimisation of reprojection error between the detected projection of a point and its re-projection based on its estimation from multiple views. However, even with bundle adjustment to optimise the pose and scene structure, noise in the detector means that triangulation is non-perfect and drift is still present in the trajectory. In the low-overlap imagery scenario presented by Sirius this is most obvious in yaw estimation of the downward facing cameras, where global camera orientation can drift by up to $40°$ over $500m$.

Introducing Additional Objectives

Bundle adjustment is a special case of nonlinear least-squares solving, often implemented to take advantage of matrix sparsity for increased efficiency. By modifying the typical matrix setup and introducing a more general framework, without loss of this efficiency, additional objectives can then be introduced to the bundle adjustment scenerio. This means it is then possible to optimise not only on the image re-projection error, but additional constraints provided by other sensors [10] in an attempt to constrain or minimise drift.

Additional objectives can be provided by any sensor, provided it gives a measurement compatible with those terms optimised by bundle adjustment. A Doppler Velocity Log can provide incremental translational objectives, while an Inertial Measurement Unit can provide incremental orientation objectives. In addition, a magnetometer or compass can provide a global, rather than local, orientation objective.

By introducing a rotational cost term, ε_r, it is possible to optimize camera pose using both re-projection error and readings from an IMU or magnetometer by way of a rotational residual: $\epsilon_{i(r)} = \mathbf{r}_i - \hat{\mathbf{r}}_i$, where $\mathbf{r}_i$ is the orientation estimate provided by the additional sensor and $\hat{\mathbf{r}}_i$ is the corresponding estimate from visual odometry:

$$\varepsilon_r^2 = \frac{1}{n} \Sigma_i^n \parallel \epsilon_{i(r)} \parallel^2$$

Here, we parameterise the orientation in the form of a Rodriguez vector: $\mathbf{r} = \begin{bmatrix} \gamma & \phi & \theta \end{bmatrix}^{\mathbf{T}}$ and assume the difference $\epsilon_{i(r)}$ is small given a satisfactorily good estimate from a pose update.

In the case of our constrained motion estimate, and because of the parameterisation of the rotation, it is possible to introduce a cost dependent only on one dimension,

yaw, and use a magnetometer to provide the additional data. Since a magnetometer provides a global orientation it is possible to correct the orientation of the vehicle globally to maintain straight trajectories over large distances. In addition, again taking advantage of the passive stability of the vehicle, we can introduce additional objectives of 0° in both pitch and yaw, while still allowing the estimates of these parameters to drift slightly and account for the slight motion in these dimensions.

The error in both the re-projection and orientation can be considered independent and Gaussian, hence weighted by a covariance, and the costs can be added to give a bi-objective cost:

$$E(\mathbf{x},\mathbf{r}) = \frac{1}{(\sigma_x)^2 mn}\Sigma_i^n\Sigma_j^m \parallel \epsilon_{ij(c)} \parallel^2 + \frac{1}{(\sigma_r)^2 n_i}\Sigma_i^n \parallel \epsilon_{i(r)} \parallel^2$$
$$= \varepsilon_c^2 + \lambda^2\varepsilon_k^2$$

where $\lambda = \frac{\sigma_x}{\sigma_r}$, indicating the ratio of the two covariances. Implementing this bi-objective bundle adjustment using magnetometer data to constrain the yaw motion will reduce angular drift and give a better pose estimate.

A Modified Parameterisation

We implement a sparsified bundle adjustment algorithm that closely follows the implementation given in [4], but with some modifications to allow the inclusion of multiple rigidly fixed cameras and additional optimisation objectives.

Given m ($j \in [1,\ldots,m]$) scene points observed at n unique timepoints/locations ($i \in [1,\ldots,n]$) by 2 rigidly fixed physical cameras ($k \in [0,1]$) freely moving through space, the model used for the observation of point j in space ($\mathbf{X}_j \in \mathbb{P}^3$) into its location in image i ($\mathbf{x}_{i,j}^k \in \mathbb{P}^2$) in a Euclidean coordinate frame is,

$$\mathbf{x}_{i,j}^k \simeq \mathbf{K}^k[\mathbf{R}_i|\mathbf{t}_i]\mathbf{H}^k\mathbf{X}_j \tag{1}$$

where $\mathbf{K}^k$ is the camera intrinsics matrix encoding the internal properties of the camera, and $\mathbf{R}_i$ and $\mathbf{t}_i$ denote the pose of the base camera $\mathbf{P}^0$ at time i. $\mathbf{H}^k$ denotes the homogeneous transform between the base (or left) camera and camera k.

This model allows us to include camera intrinsics, the homogenous transform between a rigid set of cameras and the pose of the camera rig through time. While our implementation allows the optimisation of intrinsics and the rigid homogenous transform of a rigid rig, we will omit these from further discussion as they are not optimised in this scenario for efficiency and robustness reasons.

A second sensor, such as an IMU or magnetometer, also gives estimates of orientation of the base camera at timepoints/locations ($i \in [1,\ldots,n]$). The model used for the observation of this parameter is a direct linear mapping:

$$\mathbf{r}_i \simeq \mathbf{R}_i \tag{2}$$

where $\mathbf{R}_i$ is the orientation of the base camera at time i. The observation $\mathbf{r}_i^k$ is parameterised as a three element Rodriguez vector.

The first deviation from [4] is the presence of a partitioned parameter vector ($\hat{\theta}$) that encodes all the variables over which to optimise, which allows us to parameterise the seperate objectives easily. The partitioned parameter vector is expressed as a combination of all of the sets $\hat{\theta} = [\hat{\theta}_E, \hat{\theta}_P]^\top$ corresponding to the extrinsics ($\hat{\theta}_e^\top$) which encodes the six parameters (roll, pitch, yaw, x, y z) comprising the pose of the camera rig (i.e. the base camera), and the scene points ($\hat{\theta}_p^\top$), with three parameters (x,y,z).

The bundle adjustment routine seeks to perform an iterative refinement on the parameter vector $\hat{\theta}$ at timestep l, by linearising Eqns. 1 and 2 with a good initialisation $\hat{\theta}_l$ to achieve a parameter update Δ via the traditional normal equations:

$$\mathbf{J}^\top\mathbf{J}\boldsymbol{\Delta} = -\mathbf{J}^\top\epsilon_0 \tag{3}$$

where $\mathbf{J} = \frac{\partial\hat{\mathbf{z}}}{\partial\hat{\theta}}$.

The setup of the bundle adjustment algorithm is performed in a similar way to [4] by exploiting the sparsity of the Jacobian matrix used to form the normal equations. In order to compute the Jacobian matrix we form expressions for the partial derivatives of (1) with respect to the parameters $\hat{\theta}$ as $\mathbf{A}_{ij}^k = \frac{\partial\hat{\mathbf{x}}_{ij}^k}{\partial\hat{\theta}_E}$, $\mathbf{B}_{ij}^k = \frac{\partial\hat{\mathbf{x}}_{ij}^k}{\partial\hat{\theta}_P}$, $\mathbf{C}_i = \frac{\partial\hat{\mathbf{r}}_i}{\partial\hat{\theta}_E}$ and $\mathbf{D}_i = \frac{\partial\hat{\mathbf{r}}_i}{\partial\hat{\theta}_P}$

As a consequence of the sparsity of the Jacobian matrix the partitioning of the parameter vector and the augmented normal equations (including the Hessian $N = \mathbf{J}^\top\mathbf{J}$) have the following representation.

$$\begin{bmatrix} \mathbf{E}^* & \mathbf{O} \\ \mathbf{O}^\top & \mathbf{P}^* \end{bmatrix}\begin{bmatrix} \Delta\hat{\theta}_E \\ \Delta\hat{\theta}_P \end{bmatrix} = \begin{bmatrix} \mathbf{e}_E \\ \mathbf{e}_P \end{bmatrix} \tag{4}$$

The matrices $\mathbf{E}$ and $\mathbf{P}$ included in this expression are block diagonal defined according to,

$$\mathbf{E}_i = \textstyle\sum_{j,k}\mathbf{A}_{ij}^{k\top}\Sigma_{\mathbf{x}_{ij}^k}^{-1}\mathbf{A}_{ij}^k + \mathbf{C}_i^\top\Sigma_{\mathbf{r}_i}^{-1}\mathbf{C}_i$$
$$\mathbf{P}_j = \textstyle\sum_{i,k}\mathbf{B}_{ij}^{k\top}\Sigma_{\mathbf{x}_{ij}^k}^{-1}\mathbf{B}_{ij}^k + \mathbf{D}_i^\top\Sigma_{\mathbf{r}_i}^{-1}\mathbf{D}_i$$

with the augmentation of the diagonals using the damping parameter β in a Levenberg-Marquadt framework resulting in $\mathbf{E} \rightarrow \mathbf{E}^*$ and $\mathbf{P} \rightarrow \mathbf{P}^*$. The other remaining terms in the augmented normal equations (4) are as follows.

$$\mathbf{O}_{ij} = \textstyle\sum_k \mathbf{A}_{ij}^{k\top}\Sigma_{\mathbf{x}_{ij}^k}^{-1}\mathbf{B}_{ij}^k + \mathbf{C}_i^\top\Sigma_{\mathbf{r}_i}^{-1}\mathbf{D}_i$$
$$\mathbf{e}_{E_i} = \textstyle\sum_{j,k}\mathbf{A}_{ij}^{k\top}\Sigma_{\mathbf{x}_{ij}^k}^{-1}\epsilon_{ij(c)}^k + \mathbf{C}_i^\top\Sigma_{\mathbf{r}_i}^{-1}\epsilon_{i(r)}^k$$
$$\mathbf{e}_{P_j} = \textstyle\sum_{i,k}\mathbf{B}_{ij}^{k\top}\Sigma_{\mathbf{x}_{ij}^k}^{-1}\epsilon_{ij(c)}^k + \mathbf{D}_i^\top\Sigma_{\mathbf{r}_i}^{-1}\epsilon_{i(r)}^k$$

Analytical Jacobian

For a robust, fast implementation of bundle adjustment, the Jacobian must be implemented efficiently by computing the derivatives analytically from the observation function. This is straightforward for the projective observation model through Equation 1. For the simple case of the rotational model (Eq. 2), the derivative can be expressed directly in terms of the corresponding cameras orientation. Therefore, the derivative of the rotational estimate is

$$\frac{\partial \hat{\mathbf{r}}_i}{\partial \hat{\mathbf{P}}_i^0} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}^\top \tag{5}$$

with respect to the six parameters, $\hat{\theta}_i = \begin{bmatrix} x & y & z & \gamma & \phi & \theta \end{bmatrix}$,

$$\frac{\partial \hat{\mathbf{r}}_i}{\partial \hat{\mathbf{P}}_l^k} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{6}$$

with respect to every other camera, and

$$\frac{\partial \hat{\mathbf{r}}_i}{\partial \hat{\mathbf{X}}_j} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \tag{7}$$

with respect to the parameters of every scene point. This renders the partial derivative $\mathbf{D}_i$ as simply null, as there is no partial derivative for rotation with respect to 3D scene. This simplifies some of the arithmetic of the previous section, increasing computational efficiency.

Estimating the Cost Ratio λ

In determining the variance ratio, $\lambda = \frac{\sigma_x}{\sigma_r}$, it is often the case that the variance for the projective cost is unknown or unestimated, or is not in the same units as the variance from the other. In this case, the ratio must be determined via alternative means. By selecting a range of λ and evaluating convergence of the bundle adjustment algorithm over this range, an L-curve criterion can be applied to the relative cost of each sensor to find the minimum trade-off point [10]. However, experimental evaluation shows that the L-curve estimation method performs poorly for a rotational objective. In this case, we set a fixed value of $\lambda = 2000$, which roughly equates the influence of the rotational objective with the cumulative influence of the projective objective. This ensures that the two costs are roughly equally balanced. Given an alternative scenario or configuration, a new λ must be empirically derived.

3.3 3D Meshing and Texturing

From a pose solution generated by visual odometry, it is possible to generate dense textured maps of the environment based on dense feature matching between stereo pairs.

We implement a 3D meshing and texturing pipeline from the point cloud data and camera poses to qualitively evaluate the accuracy of the visual odometry solution.

For each stereo pair, dense feature matching with a number of consistency checks and smoothing operations [9] is performed on the imagery to gain dense depth maps for each base (or left) camera.

Following a consistent depth map from each pair, a dense set of 3D oriented points is generated and a Poisson mesh fitted to the points. Each stereo mesh is arranged into a common reference frame denoted by the stereo poses, and a second Poisson surface fitted to 10 consecutive pairs with overlapping windows of single pairs. This process preserves local mesh quality to a high degree while smoothing any poorly reconstructed sections. Texture is added by projecting each vertex in the mesh back into the estimated camera poses and extracting the color of the associated image pixel. These surfaces are then stitched together and visualised in 3D to assist further research such as estimating individual coral growth and reef complexity.

4 Experiments

We evaluate our modified visual odometry algorithm presented in Section 3 with two experiments: a simulated sea-floor mapping scenario, and an experiment on real data captured by the AUV during a feld trip to Scott Reef in August 2011.

These experiments are split into three tests:

1. A traditional 6DOF Visual Odometry solution with standard bundle adjustment (termed 6DOF VO)
2. A constrained 4DOF Visual Odometry solution with standard bundle adjustment (termed 4DOF VO)
3. A constrained 4DOF Visual Odometry solution with bi-objective bundle adjustment including input from an additional sensor (termed 4DOF BO-VO)

4.1 Simulated Experiment

The three differing VO pipelines were run on a simulated dataset reflective of the normal operation of the Sirius AUV. A downward facing stereo pair of $70mm$ baseline traversed a $100 \times 100m$ square pattern over a simulated scene of 3D points. These points were randomly but evenly distributed as a mock sea-floor with varying depth of 1 to 3m from the stereo pair. The simulated cameras were set to capture at the equivalent of $1Hz$ while the vehicle moved forward at a velocity of $0.5m/s$, tracking roughly $100 - 300$ features per frame. To accurately reflect the normal motion of the vehicle, the simulated trajectory includes a slight oscilatory motion in the vertical direction, $0.1m$ sigma Gaussian noise on position,

3° noise in yaw, and 0.5° noise in pitch and roll. Simulated magnetometer data from the ground truth is fed into the bi-objective bundle adjustment with Gaussian noise of 0.5°.

4.2 Real Data Experiment

To evaluate the modified visual odometry routines on real experimental data a dataset was gathered by Sirius at Scott Reef, North of Western Australia, during a field trip in 2011. Images were captured at $1Hz$ from the stereo pair while the vehicle follows a straight trajectory from shallow to deep water over an area of interest. Data from the cameras and magnetometer were recorded in parallel with a number of other sensors, and their offset geometry has been pre-calculated to account for any motion bias.

Over 600 images of the dataset, a pose estimate was generated using both the 4DOF estimator, and again with the 4DOF estimator with bi-objective bundle adjustment that includes yaw data from the on-board magnetometer. The traditional 6DOF estimator was also evaluated over a limited component of this trajectory. The results of the three tests are compared to the output of the Information filter based SLAM system normally used to generate pose estimates [8] as a pseudo 'ground truth', utilising the sensors shown in Table 2. This ground truth trajectory utilises a number of other sensors, including a Doppler Velocity Log, Inertial Measurement Unit and Ultra Short Baseline ship communications, and stereo vision for loop closure detection, but does not use visual odometry.

From the poses generated by the bi-objective bundle adjustment based VO, a dense 3D textured mesh was generated using the methodology described in Section 3.3.

5 Results

5.1 Simulated Experiment

The results of the simulated data experiment are graphed in Figures. 3 and 4.

Over the $400m$ trajectory the 6DOF VO pipeline shows rapid deviation from ground truth due to a rapidly accumulating error in pitch (Fig 3). At the end of the trajectory, the pitch of the camera exceeds 60°, demonstrating the poor observability of this parameter from low overlap imagery in a downward facing configuration, and has a final position error of over $80m$.

On the same data, the 4DOF VO successfully avoids the rapid accumulation of pitch error by constraining the incremental motion update to the 2D plane, before allowing bundle adjustment to recover the process noise in all degrees of freedom. This difference demonstrates the suitability of constrained motion to this specific problem, significantly improving the result. It can be seen, however, that yaw of the cameras (in the Z axis), drifts significantly over the trajectory even with the standard bundle adjustment to optimise the motion, with a final position error of $12.3m$.

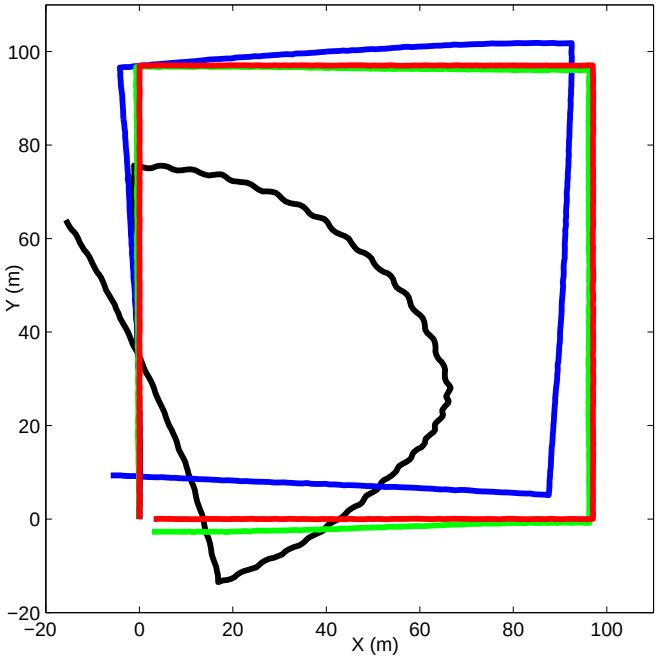

Figure 3: The simulation results of 6DOF VO with standard bundle adjustment (BA) (black), 4DOF VO with standard BA (blue) and 4DOF BO-VO (green) in comparison to ground truth (red)

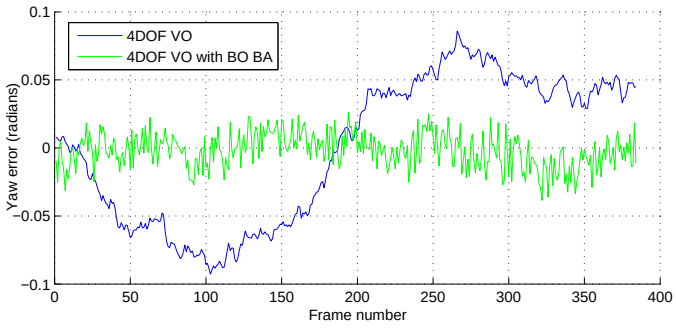

Figure 4: The error in yaw for the simulation between the pose estimate and ground truth for the 4DOF VO (blue) and 4DOF BO-VO (green)

By introducing the bi-objective bundle adjustment with input from a yaw sensor, the pose estimate is significantly improved, successfully constraining the motion such that the final position error is less than $2m$ over the total $400m$ trajectory.

The effect of the bi-objective bundle adjustment is evident in Figure 4, showing a plot of yaw error be-

tween ground truth and 4DOF VO (blue) and 4DOF BO BA (green). From the graph, it can be seen that bi-objective bundle adjustment successfully constrains yaw with a standard deviation $\sigma = 0.01$ radians, compared to standard deviation of standard bundle adjustment of $\sigma = 0.05$ radians over the length of this trajectory.

5.2 Real Data Experiment

The results of the real data experiment are graphed in Figs. 5, 6 and 7. In Figure 5, it can be seen that in comparison to the other VO methods, the 6DOF pose estimate quickly deteriorates and soon fails. Obvious from this result, pitch observability is poor, and the camera motion undergoes a rapidly looping motion over the trajectory.

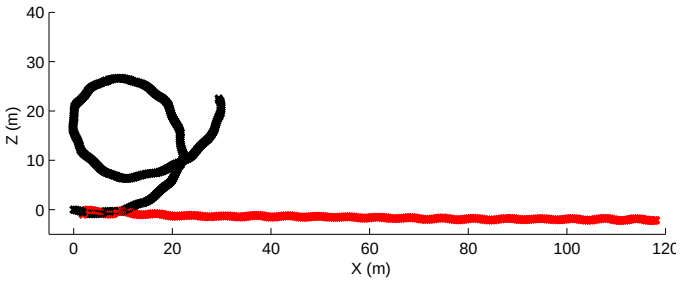

Figure 5: The pose estimate from Scott Reef imagery with 6DOF VO (black) over a 120 metre trajectory compared to the information filter SLAM solution (red)

In Figure 6 the two 4DOF pose estimators successfully approximate the 'ground truth' motion given by the SLAM solution. However, over the $300m$ trajectory, the 4DOF BO-VO algorithm (green) shows reduced drift over the length of the trajectory, with a final position error of $12.3m$ for the 4DOF VO and $6.4m$ for the 4DOF BO-VO. During pose estimation approximately $50-400$ features per tracked per frame, depending on the observed terrain, which ranged from near pure sandy bottom to high complexity coral outcrops.

This is further demonstrated in Figure 7, showing the ability of the bi-objective bundle adjustment to successfully constrain yaw drift over the trajectory. The 4DOF VO shows a standard deviation $\sigma = 0.06$ radians in yaw, while the 4DOF BO-VO shows a $\sigma = 0.01$ radians in yaw.

In Figures 8 and 9 examples of the 3D meshing and texturing pipeline are presented utilising a subsection of the poses generated in Figure 6.

6 Conclusion

A technique for performing accurate visual pose estimation using only low-overlap stereo images and yaw data has been presented. Quantitative results are shown from

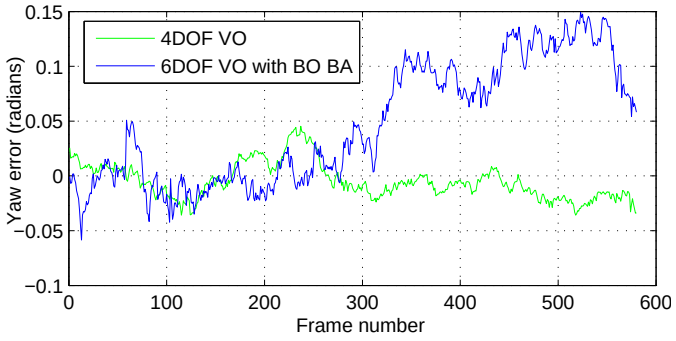

Figure 7: The error in yaw for the Sirius dataset between the pose estimate and ground truth for the 4DOF VO with BA (blue) and 4DOF VO with bi-objective BA (green)

the constrained visual odometry technique over a simulated $400m$ and real $300m$ trajectory and reconstructions generated from these pose estimates show qualitative accuracy. This research will enable future sub-sea mapping of high interest locations with increased accuracy on existing AUVs by integrating the odometry estimate into a filter, but also by enabling methods of producing seafloor maps with lower cost AUV hardware. Future work will involve demonstrating the technique on a full mission of the Sirius AUV, utilizing loop closure via open-FABMAP and graph relaxation to constrain VO drift over the entire mission, and the development of large scale environment reconstructions from the data.

Figure 9: A close up view of a sample of the reconstructed mesh indicating the quality of reconstruction (see Sec. 3.3)

References

[1] A I Comport, E Malis, and P Rives. Real-time Quadrifocal Visual Odometry. 1, 2008.

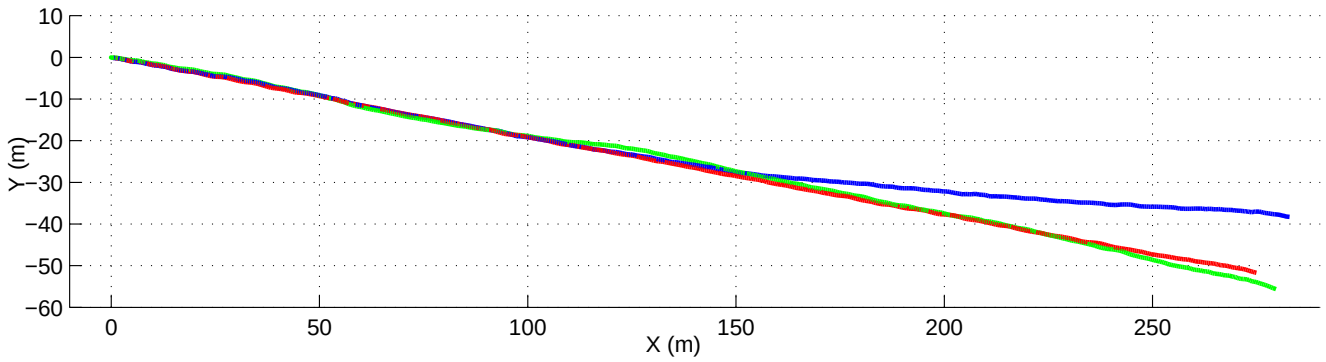

Figure 6: The pose estimate from Scott Reef imagery with 4DOF VO with standard BA (blue) and 4DOF VO with bi-objective BA (green) over a 300 metre trajectory compared to the information filter SLAM solution (red)

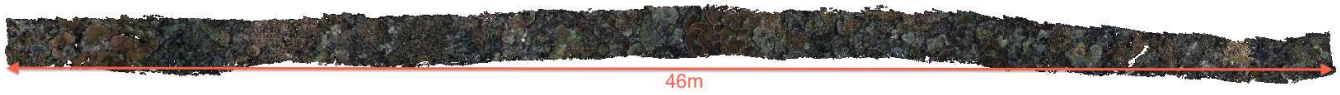

Figure 8: A high resolution mesh generated by the reconstruction pipeline from 100 camera poses covering a distance of 46m. (see Sec. 3.3)

[2] R M Eustice. Large-area visually augmented navigation for autonomous underwater vehicles. *Ph.D. dissertation, Massachusetts Inst. Technol. Woods Hole Oceanogr. Inst, Woods Hole, MA*, 2005.

[3] Ryan M Eustice, Oscar Pizarro, and Hanumant Singh. Visually Augmented Navigation for Autonomous Underwater Vehicles. pages 1–18, 1980.

[4] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. ISBN 0521540518.

[5] M Johnson-Roberson, Oscar Pizarro, S.B. Williams, and I. Mahon. Generation and Visualization of Large-Scale Three-Dimensional Reconstructions from Underwater Robotic Surveys. *Journal of Field Robotics*, 27(1):21–51, 2010.

[6] Kurt Konolige and Motilal Agrawal. Frameslam: From bundle adjustment to real-time visual mapping. *ieeexplore.ieee.org*, pages 1–11, 2008.

[7] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Closed-form solutions to minimal absolute pose problems with known vertical direction. In *Computer Vision - ACCV 2010*, pages 216–229, 2011.

[8] I. Mahon, S.B. Williams, O. Pizarro, and M. Johnson-Roberson. Efficient View-Based SLAM Using Visual Loop Closures. *IEEE Transactions on Robotics*, 24(5):1002–1014, October 2008. ISSN 1552-3098.

[9] David McKinnon, Ryan N Smith, and Ben Upcroft. A Semi-Local Method for Iterative Depth-Map Refinement. In *International Conference on Robotics and Automation (ICRA)*, 2012.

[10] J Michot and A Bartoli. Bi-objective bundle adjustment with application to multi-sensor slam. *3DPVT'10*, 2010.

[11] David Nistér, Oleg Naroditsky, and James Bergen. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23(1):3–20, January 2006. ISSN 1556-4959.

[12] O. Pizarro, R. Eustice, and H. Singh. Large area 3d reconstructions from underwater surveys. *Oceans '04 MTS/IEEE Techno-Ocean '04 (IEEE Cat. No.04CH37600)*, 2:678–687.

[13] P Torr. MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. *Computer Vision and Image Understanding*, 78(1):138–156, April 2000. ISSN 10773142.

[14] Michael Warren, D. McKinnon, H. He, and Ben Upcroft. Unaided stereo vision based pose estimation. In Gordon Wyeth and Ben Upcroft, editors, *Australasian Conference on Robotics and Automation*, Brisbane, 2010. Australian Robotics and Automation Association.

[15] Michael Warren, David Mckinnon, Hu He, Arren Glover, and Michael Shiel. Large Scale Monocular Vision-only Mapping from a Fixed-Wing sUAS. In *Field and Service Robotics*, pages 1–14, 2012.

[16] S Williams, Oscar Pizarro, and Michael Jakuba. AUV benthic habitat mapping in South Eastern Tasmania. *Field and Service Robotics*, pages 1–10, 2010.

[17] Stefan Williams, Oscar Pizarro, Michael Jakuba, Craig Johnson, Neville Barret, Russell Babcock, Gary Kendrick, Peter Steinberg, Andrew Heyward, Peter Doherty, Ian Mahon, Matthew Johnson-Roberson, Daniel Steinberg, and Ariell Friedman. Monitoring of Benthic Reference Sites: Using an Autonomous Underwater Vehicle. *IEEE Robotics and Automation Magazine*, 19(1):73–84, 2012.