Extending Parallax Parameterised Bundle Adjustment to Stereo

Brenton Leighton, Liang Zhao, Shoudong Huang, Gamini Dissanayake
Centre for Autonomous Systems
University of Technology, Sydney, Australia
{brenton.leighton, liang.zhao, shoudong.huang, gamini.dissanayake}@uts.edu.au

Abstract

The main contribution of this paper is the extension of the ParllaxBA algorithm proposed by [Zhao *et al.*, 2015] into stereo. Simulated and experimental datasets are used to evaluate Cartesian and parallax angle parameterisation for stereo bundle adjustment. It is demonstrated that, like monocular ParallaxBA, under normal conditions the two algorithms perform similarly. However, when the parallax angle of landmarks is low, parallax parameterisation can converge to a lower cost and in less time than the traditional Cartesian parameterisation.

1 Introduction

Structure from motion (SfM) is the process of estimating camera motion and landmark locations from a sequence of images. A typical SfM system consists of:

- A front end to find features within images and feature correspondences between image pairs
- Motion estimation to determine camera movement from temporal feature correspondences
- Bundle adjustment to improve camera poses and landmark locations

In the context of robotics SfM and motion estimation are performed in real time, and referred to as visual simultaneous localisation and mapping (VSLAM) and visual odometry (VO) respectively.

Recent examples of VSLAM systems include PTAM [Klein and Murray, 2009], ORB-SLAM [Mur-Artal *et al.*, 2015] and S-PTAM [Pire *et al.*, 2017]. Bundle adjustment in these and other VSLAM systems commonly use Cartesian coordinates to define landmarks. [Zhao *et al.*, 2011] demonstrated that, when the parallax angle of landmarks is low (i.e. the lateral movement of the camera is small compared to the depth of the landmark, resulting in a small movement of the feature within the image), parallax parameterisation of landmarks can result in better and faster convergence.

Parallax parameterisation is similar to inverse depth parameterisation [Civera *et al.*, 2008], sharing the azimuth and elevation parameters. Inverse depth parameterisation encodes the distance to a landmark using an inverse depth parameter, whereas parallax parameterisation uses a parallax angle parameter.

Monocular VO and VSLAM determine camera poses and landmark locations from 2D bearing vectors, and therefore absolute scale is unknown. Monocular bundle adjustment is usually performed by constraining the distance between the first two poses to one [Scaramuzza and Fraundorfer, 2011].

Stereo VO and VSLAM can use correctly scaled landmarks (in the local frame) that have been triangulated from stereo feature correspondences to determine correctly scaled camera poses and landmark locations (in the global frame). This paper describes how to extend parallax parameterisation for bundle adjustment to stereo so that the resulting camera poses and landmark locations will be correctly scaled.

1.1 Nomenclature

The nomenclature used in this paper is similar to vocabulary used by the OpenGV project [Kneip and Furgale, 2014].

- Pose: A rotation and translation
- Absolute pose: A rotation and translation in the world reference frame
- Relative pose: A rotation and translation in the coordinate frame of another pose
- Viewpoint: The absolute pose of the left camera when a stereo image pair is captured
- Landmark: A point in the world defined by Cartesian coordinates $(x\ y\ z)$
- Feature: A salient point in an image defined by image coordinates $(u\ v)$

- Observation: The stereo features ($u_1\ v_1\ u_2\ v_2$) that describe a landmark observed from a viewpoint

2 Algorithm

2.1 Problem Formulation

Bundle adjustment is a non-linear optimisation problem. Viewpoint poses and landmark coordinates are produced by VO, which bundle adjustment improves by minimising reprojection error: the difference between the actual observations and the predicted observations determined from optimised poses and landmark coordinates.

Bundle adjustment can be described by

$$\min \sum_{i=1}^{n} \sum_{j=1}^{m} \left\| o_j^i - \tilde{o}_j^i \right\|^2 \tag{1}$$

where o_j^i is the observation of the jth landmark observed from the ith viewpoint

$$o_j^i = [u_1\ v_1\ u_2\ v_2]^\intercal \tag{2}$$

and $\tilde{o}_j^i$ is the predicted observation of a landmark j from pose i

$$\tilde{o}_j^i = [\tilde{u}_1\ \tilde{v}_1\ \tilde{u}_2\ \tilde{v}_2]^\intercal \tag{3}$$

Predicted observations are determined by:

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} = \begin{bmatrix} \hat{u}/\lambda \\ \hat{v}/\lambda \end{bmatrix} \tag{4}$$

where

$$\begin{bmatrix} \hat{u} \\ \hat{v} \\ \lambda \end{bmatrix} = K\ \mathbf{x}_j^i \tag{5}$$

for the first camera, and

$$\begin{bmatrix} \hat{u} \\ \hat{v} \\ \lambda \end{bmatrix} = K\ (R\ \mathbf{x}_j^i + \mathbf{t}) \tag{6}$$

for the second camera, where K is the intrinsic matrix of the camera, R and $\mathbf{t}$ are the rotation and translation of the first camera in relation to the second, and $\mathbf{x}_j^i$ is the predicted Cartesian coordinate of the jth landmark in the ith viewpoint, which is determined by (9) for Cartesian parameterisation, and (19) or (20) for parallax parameterisation.

When images are rectified the rotation between two cameras R is the identity matrix (i.e. no rotation) and the translation between two cameras $\mathbf{t}$ is only the translation in the x axis.

2.2 Cartesian Parameterisation

Initialisation

With Cartesian parameterisation, bundle adjustment uses the Cartesian coordinates of the landmark in the world reference frame as optimisation parameters. Landmarks in the world reference frame $\mathbf{x}_j$ are determined from the landmark in the coordinate frame of the first observation $\mathbf{x}_j^i$ with

$$\mathbf{x}_j = R_i\ \mathbf{x}_j^i + \mathbf{t}_i \tag{7}$$

where R_i and $\mathbf{t}_i$ are the absolute rotation and translation of the ith viewpoint respectively, and $\mathbf{x}_j^i$ is the Cartesian coordinates of the jth landmark in the coordinate frame of the ith viewpoint, which was determined by triangulation with

$$\mathbf{x}_j^i = K^{-1} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} \left(\frac{f\ b}{u_1 - u_2} \right) \tag{8}$$

where f is the focal length of the cameras and b is the baseline. After the images are rectified the vertical image coordinates of the features will be equal (i.e. $v_1 = v_2$).

Reprojection

To find reprojection error the Cartesian coordinates of a landmark in a viewpoint $\mathbf{x}_j^i$ is determined from the Cartesian coordinates of the landmark in the world reference frame $\mathbf{x}_j$ with

$$\mathbf{x}_j^i = R_i{}^\intercal (\mathbf{x}_j - \mathbf{t}_i) \tag{9}$$

which is the inverse of (7) used for initialisation.

The Cartesian coordinates of the landmark in the viewpoint are reprojected into the first camera with (4) and (5), and into the second camera with (4) and (6).

2.3 Parallax Parameterisation

Initialisation

Parallax parameterisation differs from Cartesian parameterisation in the way landmarks are defined. In parallax parameterisation the jth landmark is defined by the parameters ψ_j, θ_j, and ω_j. These parameters are determined in relation to a main and associate anchor. The main anchor is the pose of the first observation and the associate anchor is the pose of another observation. The associate anchor is chosen based on the resulting parallax: either the pose of the earliest observation that results in a parallax greater than a certain threshold or, if no pose results in a parallax greater than the threshold, the pose that results in the largest parallax. For the results in this paper the threshold used was 0.5 radians. [Zhao *et al.*, 2015] found that the threshold value was not critical to the performance of ParallaxBA.

For simplicity, only landmarks that were observed in both stereo images and from multiple viewpoints were used.

ψ_j and θ_j are determined with the equations

$$\psi_j = \arctan2\left(\vec{x}_j^m,\ \vec{z}_j^m\right) \tag{10}$$

$$\theta_j = \arctan2\left(\vec{y}_j^m,\ \sqrt{(\vec{x}_j^m)^2 + (\vec{z}_j^m)^2}\right) \tag{11}$$

and ω_j is determined with the equation

$$\omega_j = \arccos\left(\frac{\vec{\mathbf{x}}_j^m \cdot \vec{\mathbf{x}}_j^a}{\left\|\vec{\mathbf{x}}_j^m\right\| \left\|\vec{\mathbf{x}}_j^a\right\|}\right) \tag{12}$$

where $\vec{\mathbf{x}}_j^i$ is a vector from the absolute pose translation $\mathbf{t}_i$ to the jth landmark (where $i = m$ for the main anchor and $i = a$ for the associate anchor), and

$$\vec{\mathbf{x}}_j^i = \begin{bmatrix} \vec{x}_j^i \\ \vec{y}_j^i \\ \vec{z}_j^i \end{bmatrix} \tag{13}$$

To initialise bundle adjustment, the vector from the main anchor to landmark is determined by rotating the Cartesian coordinates of the landmark with

$$\vec{\mathbf{x}}_j^m = R_m\ \mathbf{x}_j^m \tag{14}$$

where $\mathbf{x}_j^m$ is triangulated from observations using (8).

The vector from the associate anchor is determined from the landmark in the main anchor with

$$\vec{\mathbf{x}}_j^a = R_m\ \mathbf{x}_j^m + \mathbf{t}_m - \mathbf{t}_a \tag{15}$$

Reprojection

To find reprojection error the predicted parallax parameters are converted back into Cartesian coordinates. From the parallax parameters ψ_j, θ_j, and ω_j, a unit vector from the main anchor $\mathbf{t}_m$ to the jth landmark is determined with

$$\vec{\mathbf{x}}_j^m = \begin{bmatrix} \sin\ \psi_j\ \cos\ \theta_j \\ \sin\ \theta_j \\ \cos\ \psi_j\ \cos\ \theta_j \end{bmatrix} \tag{16}$$

Then the angle φ_j between the vector $\vec{\mathbf{x}}_j^m$ and vector $\mathbf{t}_a - \mathbf{t}_m$ is determined with

$$\varphi_j = \arccos\left(\vec{\mathbf{x}}_j^m \cdot \frac{\mathbf{t}_a - \mathbf{t}_m}{\|\mathbf{t}_a - \mathbf{t}_m\|}\right) \tag{17}$$

To scale the unit vector correctly the Euclidean distance d_j^m between the main anchor and jth landmark is determined with

$$d_j^m = \frac{\sin\ (\omega_j + \varphi_j)}{\sin\ \omega_j} \|\mathbf{t}_a - \mathbf{t}_m\| \tag{18}$$

If the pose of the observation is the main anchor, the unit vector is scaled and rotated to find the predicted coordinates of the landmark in the main anchor with

$$\mathbf{x}_j^m = R_m{}^\intercal \left(\vec{\mathbf{x}}_j^m d_j^m\right) \tag{19}$$

If the observation pose is not the main anchor then a vector from the main anchor to observation pose ($\mathbf{t}_i - \mathbf{t}_m$) is subtracted from the scaled vector before it is rotated

$$\mathbf{x}_j^i = R_i{}^\intercal \left(\vec{\mathbf{x}}_j^m d_j^m - (\mathbf{t}_i - \mathbf{t}_m)\right) \tag{20}$$

The Cartesian coordinates of the landmark in the viewpoint are reprojected into the first camera with (4) and (5), and into the second camera with (4) and (6).

2.4 Implementation

The non-linear optimisation problem of bundle adjustment is solved using Ceres Solver [Agarwal *et al.*,]. The form of the bundle adjustment problem solved by Ceres is

$$\min \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \left\|o_j^i - \tilde{o}_j^i\right\|^2 \tag{21}$$

Both Cartesian and parallax parameterisation are solved using a trust region method. For parallax parameterisation the dogleg [Powell, 1970] strategy with sparse normal Cholesky solver was used for all datasets. The dogleg strategy can converge significantly faster than the Levenberg-Marquardt (LM) strategy [Madsen *et al.*, 2004]. For Cartesian parameterisation the dogleg strategy with sparse normal Cholesky solver was used where possible, but if the optimisation did not converge within 300 iterations the LM strategy and sparse Schur solver was used instead.

The stopping criteria used were:

- Parameter tolerance: 1e-9
- Function tolerance: 1e-9
- Gradient tolerance: 1e-9

3 Evaluation

3.1 Simulated Dataset

A simulated dataset was used to evaluate the bundle adjustment algorithms. The simulation generates a number of viewpoints and a number of landmarks. The landmarks are projected into the viewpoints as observations, sensor noise is added to the observations, and local landmarks are triangulated from the observations. To initialise bundle adjustment motion estimation error is added to the poses.

First the field of view is determined to generate landmarks that will be within the simulated image of a viewpoint. From focal length f, image width w and image height h, the field of view can be determined:

$$x_{\text{fov}} = 2 \arctan(\frac{w}{2}/f) \tag{22}$$

$$y_{\text{fov}} = 2 \arctan(\frac{h}{2}/f) \tag{23}$$

For the first viewpoint, a number of landmarks are randomly generated within the field of view and within a specified distance range. From each landmark $\mathbf{x}_j^i$ the image coordinates for each camera k are determined with

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \hat{u}/\lambda \\ \hat{v}/\lambda \end{bmatrix} \tag{24}$$

where $\hat{u}$, $\hat{v}$ and λ are determined by (5) for the first camera and (6) for the second camera.

Subsequent viewpoints are created by generating a relative rotation R_{i-1}^i and translation $\mathbf{t}_{i-1}^i$ which are applied to the previous viewpoint's global rotation R_{i-1} and global translation $\mathbf{t}_{i-1}$, to get the new viewpoint's global rotation R_i and translation $\mathbf{t}_i$:

$$R_i = R_{i-1}\ R_{i-1}^i \tag{25}$$

$$\mathbf{t}_i = \mathbf{t}_{i-1} + \left(R_{i-1}\ \mathbf{t}_{i-1}^i\right) \tag{26}$$

Using the relative rotation and translation, the Cartesian coordinates of each landmark in the previous viewpoint $\mathbf{x}_j^{i-1}$ are transformed into the new viewpoint with

$$\mathbf{x}_j^i = {R_{i-1}^i}^\intercal \left(\mathbf{x}_j^{i-1} - \mathbf{t}_{i-1}^i\right) \tag{27}$$

The image coordinates of the landmarks in the new viewpoint are determined with (24). Any landmarks that are outside of the field of view are removed from the viewpoint. Additionally a third of the landmarks are randomly removed from the viewpoint. New landmarks for the viewpoint are generated to replace the removed landmarks in the same way as for the first viewpoint. A random amount of error is added to the image coordinates of all landmarks to simulate sensor noise, and the Cartesian coordinates of the landmarks are triangulated from the observations with (8).

For the initialisation of bundle adjustment, a random amount of error is added to the rotation and translation of each viewpoint to simulate motion estimation error.

The camera parameters used to generate the simulation data were

- Camera focal length $f = 300$ px
- Stereo camera baseline $b = 30$ mm
- Relative rotation between the first and second camera $R = I$
- Relative translation between the first and second camera $\mathbf{t} = [b\ 0\ 0]^\intercal$
- Image width $w = 800$ px
- Image height $h = 600$ px

All simulated datasets have 100 viewpoints and 100 landmarks visible. There is no loop closure in the simulated datasets.

The rotation of each viewpoint relative to the previous viewpoint was $+\pi/64$ radians about the Y axis, with an additional uniformly distributed random value from $-\pi/32$ to $+\pi/32$ radians about each axis. The translation of each viewpoint relative to the previous viewpoint was +60 mm in the X axis, +2 mm in the Y axis, and +2 mm in the Z axis, with an additional uniformly distributed random value from −30 to +30 mm on each axis.

Uniformly distributed random values between −1 and +1 pixels is added to the image coordinates of every observation to simulate sensor noise.

To simulate motion estimation error the rotation of each viewpoint has uniformly distributed random values from $-0.3\pi/32$ to $+0.3\pi/32$ radians added to each axis, and the translation of each viewpoint has uniformly distributed random values from −18 to +18 mm added to each axis.

The distance range of generated landmarks was varied between datasets. For each range four datasets were generated and used to evaluate the bundle adjustment algorithms. The distance ranges used are:

- 0.1 to 2 m
- 1 to 3 m
- 2 to 5 m
- 3 to 10 m

Results

Table 1 shows the results of four datasets with landmark distance range of 0.1 to 2 m and Table 2 shows the results of four datasets with a landmark distance range of 1 to 3 m. The dogleg strategy was used for all datasets. In these datasets Cartesian and parallax parameterisation converged to the same final cost, in a similar number of iterations. Because of the complexity of parallax parameterisation it generally took a slightly longer time to run.

Table 3 shows the results of four datasets with a landmark distance range of 2 to 5 m. For these datasets Cartesian parameterisation with the dogleg strategy was unable to converge within 300 iterations. In two of the four datasets Cartesian and parallax converged to the same final cost, with parallax needing fewer iterations and slightly less time. In the other two datasets parallax converged to a lower final cost with fewer iterations (around 2 to 10 times fewer) and in less time.

Table 4 shows the results of four datasets with a landmark distance range of 3 to 10 m. For these datasets Cartesian parameterisation with the dogleg strategy was

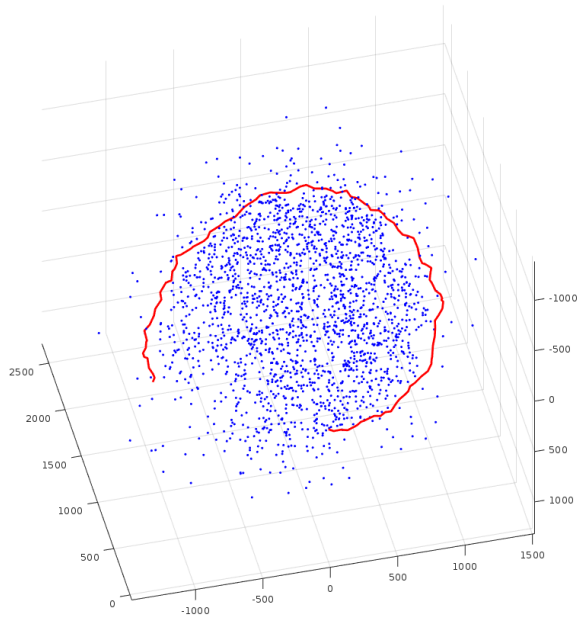

Figure 1: Estimated landmarks (blue points) and viewpoints (red line) of a simulated dataset with landmarks between 0.1 m and 2 m.

unable to converge within 300 iterations. In these datasets parallax converges to a final cost lower than Cartesian with significantly fewer iterations (around 10 to 20 times fewer) and in significantly less time (around 10 times less).

The simulation datasets show that, as the parallax angle of landmarks becomes low, Cartesian parameterisation has difficulty converging. In these cases parallax parameterisation converges to a lower final cost, in fewer iterations and less time.

3.2 New College Dataset

A section of the New College dataset [Smith *et al.*, 2009] that has been processed for use with g2o [Kümmerle *et al.*, 2011] was used to evaluate the bundle adjustment algorithms. The dataset contains 3,500 viewpoints, 491,640 landmarks, and 2,124,449 observations. An example of a stereo image pair used to generate the data for bundle adjustment is shown in Figure 2.

Results

The results of Cartesian and parallax bundle adjustment on the New College dataset are presented in Table 5. Parallax parameterisation converged to a lower final cost, in fewer iterations and less time.

3.3 Cardboard Box Dataset

A small dataset was captured using a stereo camera while walking around a cardboard box. Salient features were extracted from the images using the scale-invariant feature transform (SIFT) algorithm [Lowe, 1999]. The features are then rectified using the camera calibration parameters. Features are matched in the stereo image pairs and any epipolar outliers (where the vertical image coordinates of the matching features differ by more than two pixels) are removed, and the Cartesian coordinates of the landmarks are triangulated from the observations with (8).

Features from the temporal image pairs of the first camera are matched and, using the previous landmarks (Cartesian coordinates) and the current features (image coordinates), the relative motion of the camera is estimated using the perspective-three-point algorithm [Gao *et al.*, 2003] while eliminating erroneous matches with the M-estimator sample consensus algorithm [Torr and Zisserman, 2000].

Keyframing is used to reduce the amount of data for bundle adjustment. Frames are only added if the estimated motion is greater than 40 mm or 5° from the previous keyframe. The data used for both parameterisations of bundle adjustment is identical.

Feature matching is also performed on the first and last viewpoints. Since the camera has travelled back to its starting position this adds a loop closure to the data.

The data for bundle adjustment contains 60 viewpoints, 2,237 landmarks, and 12,933 observations. The stereo camera has a baseline of 30 mm, with the camera approximately 300 to 500 mm away from the cardboard box. The background provided more distant landmarks. An example of a stereo image pair used to generate the data for bundle adjustment is shown in Figure 4.

Results

The results of Cartesian and parallax bundle adjustment on the cardboard box dataset are presented in Table 6. Cartesian and Parallax parameterisation converged to the same final cost, however parallax converged in fewer iterations and less time.

4 Conclusions and Future Work

[Zhao *et al.*, 2015] proposed ParallaxBA: a monocular bundle adjustment algorithm using parallax angles for landmark parameterisation. In this paper we have described how to extend ParallaxBA into stereo, and demonstrated that parallax parameterisation in stereo bundle adjustment offers the same advantages as monocular while providing correct scale.

In all datasets evaluated parallax parameterisation converged to a cost equal to or less than Cartesian parameterisation. Generally parallax parameterisation can converge in fewer iterations however, because of the complexity of parallax parameterisation, each iteration takes longer to compute. For the datasets with landmarks near

Table 1: Results of the simulated datasets with 0.1 to 2 m landmarks

Simulation dataset	1		2		3		4	
Parameterisation	Cartesian	Parallax	Cartesian	Parallax	Cartesian	Parallax	Cartesian	Parallax
Strategy	Dogleg	Dogleg	Dogleg	Dogleg	Dogleg	Dogleg	Dogleg	Dogleg
Initial cost	7.16e+6	7.16e+6	7.71e+6	7.71e+6	1.46e+8	1.46e+8	3.24e+6	3.24e+6
Final cost	4.16e+3	4.16e+3	4.18e+3	4.18e+3	4.11e+3	4.11e+3	4.15e+3	4.15e+3
Iterations	8	8	7	7	13	12	5	5
Time	0.40	0.85	0.42	0.65	0.66	1.10	0.32	0.54

Table 2: Results of the simulated datasets with 1 to 3 m landmarks

Simulation dataset	5		6		7		8	
Parameterisation	Cartesian	Parallax	Cartesian	Parallax	Cartesian	Parallax	Cartesian	Parallax
Strategy	Dogleg	Dogleg	Dogleg	Dogleg	Dogleg	Dogleg	Dogleg	Dogleg
Initial cost	1.83e+6	1.83e+6	7.92e+6	7.92e+6	2.43e+6	2.43e+6	1.87e+6	1.87e+6
Final cost	3.92e+3	3.92e+3	4.04e+3	4.04e+3	3.96e+3	3.96e+3	4.01e+3	4.01e+3
Iterations	7	6	17	21	7	7	8	6
Time (sec)	0.40	0.62	0.67	2.01	0.40	0.72	0.42	0.62

Table 3: Results of the simulated datasets with 2 to 5 m landmarks

Simulation dataset	9		10		11		12	
Parameterisation	Cartesian	Parallax	Cartesian	Parallax	Cartesian	Parallax	Cartesian	Parallax
Strategy	LM	Dogleg	LM	Dogleg	LM	Dogleg	LM	Dogleg
Initial cost	1.94e+6	1.94e+6	1.64e+6	1.64e+6	2.17e+6	2.17e+6	1.51e+6	1.51e+6
Final cost	3.89e+3	3.89e+3	3.89e+3	3.89e+3	4.10e+3	3.85e+3	4.49e+3	3.89e+3
Iterations	21	9	34	9	83	8	104	8
Time (sec)	1.16	0.92	1.75	0.94	4.69	0.80	5.60	0.81

Table 4: Results of the simulated datasets with 3 to 10 m landmarks

Simulation dataset	13		14		15		16	
Parameterisation	Cartesian	Parallax	Cartesian	Parallax	Cartesian	Parallax	Cartesian	Parallax
Strategy	LM	Dogleg	LM	Dogleg	LM	Dogleg	LM	Dogleg
Initial cost	2.27e+6	2.27e+6	1.57e+6	1.57e+6	1.58e+6	1.58e+6	2.26e+6	2.26e+6
Final cost	5.85e+3	3.90e+3	5.75e+3	3.94e+3	6.30e+3	4.03e+3	5.36e+3	3.90e+3
Iterations	220	12	199	10	217	10	199	11
Time (sec)	12.03	1.22	10.94	1.05	12.26	1.03	11.05	1.10

Figure 2: An example stereo image pair from the New College dataset.

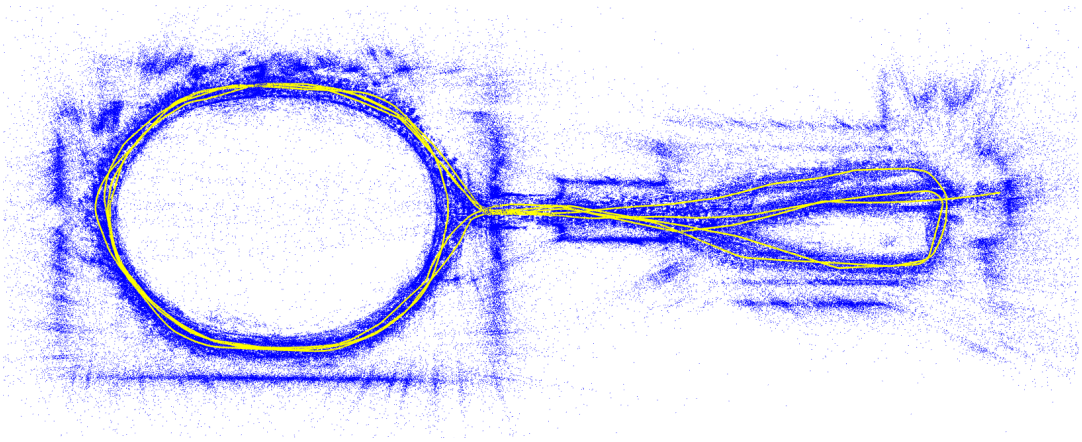

Figure 3: Estimated landmarks (blue points) and viewpoints (yellow line) of the New College dataset. Distant landmarks are not shown.

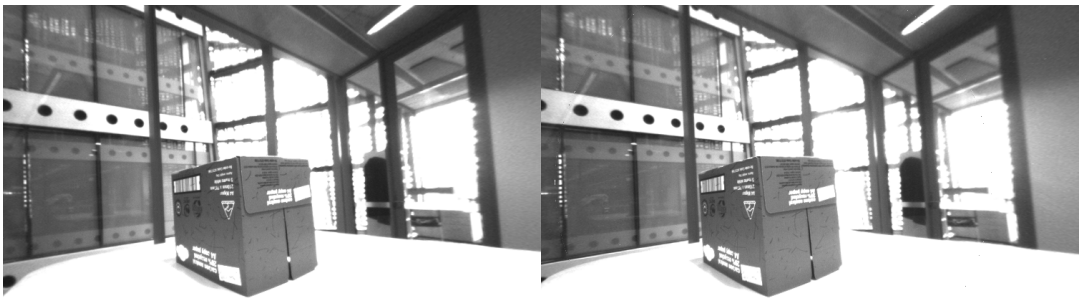

Figure 4: An example stereo image pair from the cardboard box dataset.

Table 5: Results of the New College dataset

	Cartesian	Parallax
Strategy	LM	Dogleg
Initial cost	5.120590e+7	5.120590e+7
Final cost	2.253884e+6	2.243549e+6
Iterations	74	10
Time (sec)	1990.65	290.53

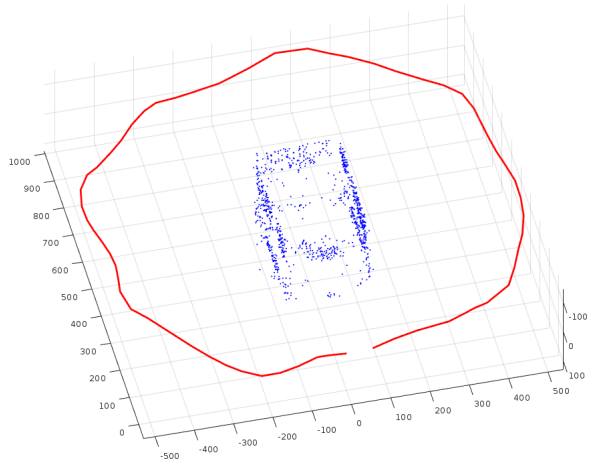

Figure 5: Estimated landmarks (blue points) and viewpoints (red line) of the cardboard box dataset. Distant landmarks are not shown.

Table 6: Results of the cardboard box dataset

	Cartesian	Parallax
Strategy	LM	Dogleg
Initial cost	9.763550e+4	9.763550e+4
Final cost	5.546365e+3	3.693787e+3
Iterations	99	6
Time (sec)	6.62	0.78

to the viewpoint (and thus well defined), parallax parameterisation converges to the same cost as Cartesian parameterisation, in slightly longer time. For the datasets with landmarks further from the viewpoint (and thus poorly defined), parallax parameterisation can converge to a lower cost than Cartesian parameterisation, in significantly shorter time.

The current implementation of stereo ParallaxBA ignores landmarks that have only been observed from a single viewpoint, and also ignores observations without a stereo correspondence. Future work could involve incorporating this information into bundle adjustment.

To evaluate the performance of parallax parameterisation with different front ends, our bundle adjustment algorithms could be integrated into existing VSLAM systems such as ORB-SLAM [Mur-Artal *et al.*, 2015] or S-PTAM [Pire *et al.*, 2017].

Acknowledgements

This research is supported by an Australian Government Research Training Program (RTP) Scholarship, and by the Australian Research Council and Roads and Maritime Services (Linkage Project LP150100935).

References

[Agarwal *et al.*,] Sameer Agarwal, Keir Mierle, and Others. Ceres solver.

[Civera *et al.*, 2008] Javier Civera, Andrew J Davison, and JM Martinez Montiel. Inverse depth parametrization for monocular slam. *IEEE transactions on robotics*, 24(5):932–945, 2008.

[Gao *et al.*, 2003] Xiao-Shan Gao, Xiao-Rong Hou, Jianliang Tang, and Hang-Fei Cheng. Complete solution classification for the perspective-three-point problem. *IEEE transactions on pattern analysis and machine intelligence*, 25(8):930–943, 2003.

[Klein and Murray, 2009] Georg Klein and David Murray. Parallel tracking and mapping on a camera phone. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 83–86. IEEE, 2009.

[Kneip and Furgale, 2014] Laurent Kneip and Paul Furgale. Opengv: A unified and generalized approach to real-time calibrated geometric vision. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 1–8. IEEE, 2014.

[Kümmerle *et al.*, 2011] Rainer Kümmerle, Giorgio Grisetti, Hauke Strasdat, Kurt Konolige, and Wolfram Burgard. g2o: A general framework for graph optimization. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 3607–3613. IEEE, 2011.

[Lowe, 1999] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[Madsen *et al.*, 2004] Kaj Madsen, Hans Bruun Nielsen, and Ole Tingleff. Methods for non-linear least squares problems. 2004.

[Mur-Artal *et al.*, 2015] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, 31(5):1147–1163, 2015.

[Pire *et al.*, 2017] Taihú Pire, Thomas Fischer, Gastón Castro, Pablo De Cristóforis, Javier Civera, and Julio Jacobo Berlles. S-ptam: Stereo parallel tracking and mapping. *Robotics and Autonomous Systems*, 2017.

[Powell, 1970] Michael JD Powell. A new algorithm for unconstrained optimization. *Nonlinear programming*, pages 31–65, 1970.

[Scaramuzza and Fraundorfer, 2011] Davide Scaramuzza and Friedrich Fraundorfer. Visual odometry [tutorial]. *IEEE robotics & automation magazine*, 18(4):80–92, 2011.

[Smith *et al.*, 2009] Mike Smith, Ian Baldwin, Winston Churchill, Rohan Paul, and Paul Newman. The new college vision and laser data set. *The International Journal of Robotics Research*, 28(5):595–599, 2009.

[Torr and Zisserman, 2000] Philip HS Torr and Andrew Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156, 2000.

[Zhao *et al.*, 2011] Liang Zhao, Shoudong Huang, Lei Yan, and Gamini Dissanayake. Parallax angle parametrization for monocular slam. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 3117–3124. IEEE, 2011.

[Zhao *et al.*, 2015] Liang Zhao, Shoudong Huang, Yanbiao Sun, Lei Yan, and Gamini Dissanayake. Parallaxba: bundle adjustment using parallax angle feature parametrization. *The International Journal of Robotics Research*, 34(4-5):493–516, 2015.