

Accurate 3D Reconstruction of Underwater Infrastructure using Stereo Vision

Brenton Leighton

Master of Engineering (Research)

Faculty of Engineering and Information Technology

University of Technology Sydney

2019

Certificate of Original Authorship

I, Brenton Leighton, declare that this thesis is submitted in fulfilment of the requirements for the award of Master of Engineering (Research) in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

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

Signature:

Production Note:

Signature removed prior to publication.

Date: 05/08/2019

Acknowledgements

I would like to thank my supervisor Associate Professor Shoudong Huang for his guidance of my research work and assistance in the preparation of this thesis. I would also like to thank my co-supervisor Distinguished Professor Dikai Liu and alternate supervisor Dr. Liang Zhao for their support and assistance.

I would like to thank members of the SPIR project Dr. Andrew To, Dr. Khoa Le, and Craig Borrows, for their work in developing the ROV and assistance with data collection.

Finally I would like to thank the staff and students at the Centre for Autonomous Systems, in particular Li Yang Liu, for any advice that helped with my research.

Table of Contents

Certificate of Original Authorship	i
Acknowledgements	iii
Table of Contents	vi
List of Figures	viii
List of Tables	x
Acronyms	xi
Abstract	xiii
1 Introduction	1
1.1 Motivation	1
1.2 Contribution	2
1.3 Publications	3
2 Literature Review	5
2.1 Methods of Underwater Perception	5
2.2 Underwater Vision	7
2.3 Visual Odometry	13
2.4 Bundle Adjustment	17
2.5 Structure from Motion and Visual SLAM	18
2.6 3D Reconstruction	20
3 Underwater Visual Odometry	25
3.1 Introduction	25
3.2 Camera Calibration	26
3.3 Exposure Control	27
3.4 Image Enhancement	36
3.5 Visual Odometry	43
3.6 Summary	55

4	Extending Parallax Parameterised Bundle Adjustment to Stereo	57
4.1	Introduction	57
4.2	Algorithm	57
4.3	Evaluation	62
4.4	Summary	72
5	Underwater 3D Reconstruction	73
5.1	Introduction	73
5.2	Single Viewpoint Perception	73
5.3	Visual SLAM and Multi-View 3D Reconstruction	88
5.4	Summary	99
6	Conclusion	101
7	Bibliography	105

List of Figures

1.1	The third version of the SPIR ROV	2
3.1	The DUO M stereo camera	26
3.2	Examples of well exposed images at each position	29
3.3	Examples of subjective descriptions of exposure	30
3.4	The left images of each image pair used to evaluate image enhancement	37
3.5	Typical images from each data set	44
3.6	Number matches between sequential image pairs for each data set using SURF, KAZE, FREAK, and BRISK	48
3.7	Reduced camera viewpoints and landmarks from data set 1	49
3.8	All landmarks (shown in blue) generated from data set 1, viewed top down after aligning with a model (shown in red)	50
3.9	Reduced camera viewpoints and landmarks from data set 2	51
3.10	All landmarks (shown in blue) generated from data set 2 (SURF and KAZE only), viewed top down after aligning with a model (shown in red)	52
3.11	All landmarks generated from data set 2 (FREAK and BRISK only)	52
3.12	Reduced camera viewpoints and landmarks from data set 3	53
3.13	All landmarks (shown in blue) generated from data set 3 (SURF and KAZE only), viewed top down after aligning with a model (shown in red)	53
3.14	Reduced camera viewpoints and landmarks from data set 4	54
3.15	All landmarks generated from data set 4 (SURF and KAZE only)	54
4.1	Estimated landmarks (blue points) and viewpoints (red line) of a simulated data set with landmarks between 0.1 m and 2 m.	64
4.2	An example stereo image pair from the New College data set.	65
4.3	Estimated landmarks (blue points) and viewpoints (yellow line) of the New College data set. Distant landmarks are not shown.	68
4.4	An example stereo image pair from the cardboard box data set.	69
4.5	Estimated landmarks (blue points) and viewpoints (red line) of the card- board box data set. Distant landmarks are not shown.	70
4.6	An example stereo image pair from the underwater data set.	70
4.7	Estimated landmarks (blue points) and viewpoints (red line) of the under- water data set.	71

5.1	Image pair 1 before rectification	74
5.2	Image pair 1 after rectification	81
5.3	Point clouds generated from image pair 1	81
5.4	Image pair 2 after rectification	82
5.5	Point clouds generated from image pair 2	82
5.6	Image pair 3 after rectification	83
5.7	Point clouds generated from image pair 3	83
5.8	Image pair 4 after rectification	84
5.9	Point clouds generated from image pair 4	84
5.10	Image pair 5 after rectification	85
5.11	Point clouds generated from image pair 5	85
5.12	Image pair 6 after rectification	86
5.13	Point clouds generated from image pair 6	86
5.14	Camera trajectory and landmarks of underwater visual SLAM with data capture issues	91
5.15	Landmarks from ORB-SLAM2	93
5.16	3D reconstructions of data set 1	95
5.17	3D reconstructions of data set 2	96
5.18	3D reconstructions of data set 3	97
5.19	3D reconstructions of data set 4	98

List of Tables

3.1	Results of varying artificial lighting and exposure time with the pile and background weakly illuminated by the sun	31
3.2	Results of varying artificial lighting and exposure time with the pile moderately illuminated, and the background weakly illuminated, by the sun . .	32
3.3	Results of varying artificial lighting and exposure time with the pile weakly illuminated, and the background strongly illuminated, by the sun	33
3.4	Results of varying artificial lighting and exposure time with the pile strongly illuminated, and the background moderately illuminated, by the sun	34
3.5	Number of features detected with and without CLAHE	38
3.6	Number of SURF features (detected without contrast limited adaptive histogram equalisation (CLAHE)) matched with and without CLAHE before descriptor extraction	38
3.7	Number of SURF features (detected with CLAHE) matched with and without CLAHE before descriptor extraction	39
3.8	Number of KAZE features (detected without CLAHE) matched with and without CLAHE before descriptor extraction	39
3.9	Number of KAZE features (detected with CLAHE) matched with and without CLAHE before descriptor extraction	40
3.10	Number of FREAK features (detected without CLAHE) matched with and without CLAHE before descriptor extraction	40
3.11	Number of FREAK features (detected with CLAHE) matched with and without CLAHE before descriptor extraction	41
3.12	Number of BRISK features (detected without CLAHE) matched with and without CLAHE before descriptor extraction	41
3.13	Number of BRISK features (detected with CLAHE) matched with and without CLAHE before descriptor extraction	42
4.1	Results of the simulated data sets with 0.1 to 2 m landmarks	66
4.2	Results of the simulated data sets with 1 to 3 m landmarks	66
4.3	Results of the simulated data sets with 2 to 5 m landmarks	67
4.4	Results of the simulated data sets with 3 to 10 m landmarks	67
4.5	Results of the New College data set	68

4.6	Results of the cardboard box data set	69
4.7	Results of the underwater data set	71
5.1	Quantity and area (in megapixels) of landmarks produced by stereo feature detection and matching	79
5.2	Quantity (in millions) of points produced by dense stereo matching methods	79
5.3	Duration of each data set and number of key frames and landmarks gener- ated by ORB-SLAM2	92

Acronyms

AGAST adaptive and generic accelerated segment test.

AUV autonomous underwater vehicle.

BRIEF binary robust independent elementary features.

BRISK binary robust invariant scalable keypoints.

CLAHE contrast limited adaptive histogram equalisation.

CMVS clustering views for multi-view stereo.

DSO direct sparse odometry.

DTAM dense tracking and mapping.

ELAS efficient large-scale stereo matching.

FAST features from accelerated segment test.

FREAK fast retina keypoint.

HDR high dynamic range.

IMU inertial measurement unit.

KLT Kanade-Lucas-Tomasi.

LiDAR light detection and ranging.

LSD-SLAM large-scale direct SLAM.

MAPSAC maximum a posterior estimation sample consensus.

MLESAC maximum likelihood sample consensus.

MVE multi-view environment.

ORB oriented FAST and rotated BRIEF.

PMVS patch-based multi-view stereo.

PTAM parallel tracking and mapping.

RANSAC random sample consensus.

ROV remotely operated underwater vehicle.

SfM structure from motion.

SIFT scale-invariant feature transform.

SLAM simultaneous localisation and mapping.

SPiR Submersible Pile Inspection Robot.

SURF speeded-up robust features.

SVO semi-direct visual odometry.

USB universal serial bus.

VIO visual-inertial odometry.

VO visual odometry.

Abstract

Modern vehicle and pedestrian bridges over water are built on concrete piles; foundations that penetrate the soft soil of the bed of the body of water to sit on the solid rock below. Like all built structures these concrete piles require regular inspection to determine if any preventative maintenance is needed. Stereo vision and Structure from Motion techniques offer a cost effective method of creating a 3D reconstructions of a scene, but the underwater environment around a bridge pile has unique challenges. Poor visibility, strong and varying sunlight, and floating material in the water create difficulties for computer vision.

This thesis evaluates exposure control, image enhancement, and feature detection and description algorithms, for the purpose of localising images captured around a bridge pile. Stereo correspondence algorithms are evaluated and used to create a single viewpoint 3D reconstruction of a scene, then a visual SLAM system is used to localise the single viewpoint reconstructions, so that they can be merged together to create a 3D reconstruction of a bridge pile.

Visual odometry, using KAZE with CLAHE image enhancement for feature detection, was successfully performed in the underwater environment. ORB-SLAM2 can also perform well, and 3D reconstructions from a single viewpoint (created with block matching, semi-global matching, or ELAS) were merged to create 3D reconstructions of submerged bridge piles.