Towards Observable Urban Visual SLAM

by

Li Yang Liu

A thesis submitted in partial fulfilment of the
requirements for the degree of Doctor of Philosophy

Supervisor: Associate Professor Shoudong Huang
Co-supervisor: Professor Gamini Dissanayake

at the

Centre for Autonomous Systems
Faculty of Engineering and Information Technology
University of Technology Sydney

March 2020

Certificate of Original Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signed:

Production Note:
Signature removed prior to publication.

Date: 12/02/2020

Towards Observable Urban Visual SLAM

by

Li Yang Liu

Abstract

Visual Simultaneous Localisation and Mapping (V-SLAM) is the subject of robot state and environment map estimation by drawing inference on camera captured data. It has been a major branch of research and popular in application owing to the rich information and low cost in vision measurement acquisition. However, for applications in urban environments, where the camera-mounted vehicle moves along a straight line direction towards the road scene, a large number of features suffer difficulty in depth estimation due to their small parallax angles, as a result the classical V-SLAM algorithm encounters instability and the system state is often unobservable.

This thesis addresses the issue of Urban SLAM observability associated with monocular cameras. It proposes a novel Bundle Adjustment (BA) formulation that addresses the problem from a fundamental approach – by parameterising map points in an on-manifold ray parallax form the SLAM formulation has a stable configuration that guarantees local state observability despite of presence of low parallax features.

V-SLAM is known to be highly non-convex from its projective image formation principle. Slight off-optimal initial values easily lead to sub-optimal final state estimates. In Urban SLAM this is further exacerbated by collinear camera motion that causes ambiguity in initial state estimation. A robust initialisation method is proposed in this thesis to provide unique near-optimal initial estimates effectively addressing collinearity issues.

For practical use of our algorithm, we demonstrate how the urban scene friendly V-SLAM algorithms are integrated into a real-time Visual Inertial Navigation system (VINS).

A series of quantitative analyses are performed on a few benchmark datasets, demonstrating effectiveness of our algorithm in urban environments.

Acknowledgements

Throughout the last 4 years of research work that became the basis of this dissertation, I received a great deal of support and assistance. I would like to thank my supervisors A/Prof. Shoudong Huang, Prof. Gamini Dissanayake and Dr Liang Zhao for their guidance and support over the last four years.

It is worth noting that playing multiple roles of being a PhD student as well as a wife and mother of two young children has been an extremely challenging experience. I would like to thank my mother for looking after my children with total dedication, thank my deceased father for influencing the knowledge loving personality in me, thank my husband for providing support to fulfill my childhood dream, thank my children for putting up with a busy mum, thank my sister for listening to my anguish and thank my school time teachers for guiding me through the many wonders of mathematics. Also a big thank you to my church Corner Stone Concord which provided great sympathy and love during my difficult times.

Thanks to Dr Gavin Paul, Prof Dikai Liu for encouraging me to embark on the PhD course. I acknowledge Teng Zhang, my PhD classmate made contributions to some of the ideas presented in this thesis. Thanks to Ritesh Kumar for developing the VINS sensing platform with me for his capstone. Many thanks to fellow PhD classmates Jonathan Woolfrey and Liye Sun in CAS-UTS for your valuable discussions and open seminars. My sincere gratitude towards Dr Rory Niland, an ex-colleague and well-respected Physicist from whom I first learned SVD and Mark Jeffree the boss from previous work who taught me Signal Processing.

I acknowledge that this research is supported by an Australian Government Research Training Program Scholarship.

Contents

List of Figures

List of Tables

Acronyms

1D, 2D, 3D One-Dimensional, two-Dimensional, three-Dimensional

1DSfM One dimensional outlier removal and structure form motion

BA Bundle adjustment

CAS Centre for Autonomous Systems

CLS-NonLin A two stage position initialisation method: Constrained Least Squares with Non-Linear optimisation

CLT Central Limit Theorem

CPU Central Processing Unit

DL Powell's Dogleg optimisation method

EKF Extended Kalman Filter

EG Epipolar Geometry

g^2o General Graph Optimisation

GD Gradient descent

GI Global initialisation

GN Gauss Newton

GPS Global Positioning System

GPU Graphics Processing Unit

GT Ground Truth

LDL A variant of Cholesky decomposition $\mathbf{A} = \mathbf{LDL}^T$, $\mathbf{L}$ for lower unit triangular, $\mathbf{D}$ for diagonal

LM Levenberg-Marquardt algorithm for non-linear least squares regression problems

LS Least Squares

LUD Least Unsquared Deviation position initialisation method

IDP Inverse depth parameterisation of feature points

IMU Inertial Measurement Unit

IRLS Iterative Re-weighted Least Squares

iSAM Incremental Smoothing and Mapping

MAP Maximum a posteriori

MAV Micro aerial vehicle

ML Maximum-likelihood

MST Maximum spanning tree

MVG Multiple View Geometry

NEES Normalised Estimation Error Squared

NLS Non-linear Least Squares

PD Positive definite

PGILP PMBA global initialisation with Linear Programming

PBA Parallax Angle based Bundle Adjustment

PMBA Parallax on manifold bundle adjustment

PMBA+VINS Visual Inertial Navigation System using PMBA for feature paramtrisation and bundle adjustment

PSD Positive Semi-definite

QR QR decomposition, decomposition of matrix $\mathbf{A}$ into a product $\mathbf{A} = \mathbf{QR}$ of an orthogonal matrix $\mathbf{Q}$ and an upper triangular matrix $\mathbf{R}$

RANSAC Random Sample Consensus

RGB-D RGB image with corresponding and depth image

RMS Root mean square

RMSE Root mean square error

ROS Robotic Operating System

S^2 2-sphere: 2-dimensional manifold that can be embedded in Euclidean 3D space.

SD Steepest descent

SfM Structure form Motion

SE(3) Special Euclidean group 3D

SIFT Scale Invariant Feature Transform

SLAM Simultaneous localisation and mapping

SO(3) Special orthogonal group 3D

SURF Speeded-Up Robust Features

SVD Singular value decomposition

SWF Sliding Window Filter

UTS University of Technology, Sydney

VINS Visual Inertial Navigation System

VINS-Mono A realtime VINS opensource software package using monocular camera inputs

VIO	Visual Inertial Odometry
VO	Visual Odometry
WsNonLin	Non-Linear position initialisation method, by Wilson
V-SLAM	Visual SLAM
XYZ	3D Euclidean parametrisation of feature points

Nomenclature

	General Notations
$\mathcal{X}$	A generic symbol for robot state variable in the Manifold domain. Can refer to pose $\mathcal{T}$ and feature $\mathcal{F}$.
$\mathcal{T}_i$	the i'th camera pose, $\mathcal{T}_i = (\mathbf{R}_i, \mathbf{P}_i) \in \mathbb{SE}(3)$
$\mathcal{F}_j$	Feature j's PMBA parameters in manifold domain
$\mathbf{F}_j$	Feature j's coordinates in Euclidean space
$\mathbb{T}_j$	Indices of all camera poses observing featire j
	Distributed according to
$\approx$	approximately equal
$\xi(\cdot)$	the normalisation operator: $\xi(\mathbf{x}) = \frac{\mathbf{x}}{\lVert\mathbf{x}\rVert}$ gives vector $\mathbf{x}$'s direction.
$\pi(\cdot)$	the homogeneous normalisation operator: $\pi(\begin{bmatrix} x, y, z \end{bmatrix}^T) = \begin{bmatrix} x/z, y/z, 1 \end{bmatrix}^T$.
$\mathcal{V}$	Manifold variable displayed in calligraphic font
V	Euclidean variable displayed in roman font
Subscript $^{(l)}$	indicates the reference frame is local
$[...]^T$	Transpose
$\lvert ... \rvert$	Transpose
$\lVert \cdot \rVert$	Euclidean norm
$\mathbf{I}_n$	Identity matrix of size n
$\mathbf{0}_n$	Zero matrix of size n
$\mathcal{N}$	Normal distribution.
$O(x)$	Big O notation.
$\mathbb{R}^2$	The set of points in 2D Euclidean space

$\mathbb{R}^3$	The set of points in 3D Euclidean space
$\mathcal{M}$	The Manifold domain
$\mathbf{z}$	Measurement vector
$f(\cdot)$	Measurement function
$\mathbf{H}$	Hessian matrix and information matrix
$\boxplus$	Retraction operator
$\lfloor \mathbf{w} \rfloor_\times$	Anti-symmetric matrix based on $3x1$ vector $\mathbf{w}$, see Section 2.5.2 for definition
∂	Partial derivative
$\mathbf{J}_B^A$	Jacobian of variable A with respect to variable B
$\times$	cross product
$\tilde{\mathbf{x}}$	Noisy version of $\mathbf{x}$
$\bar{\mathbf{x}}$	Perfect estimate of $\mathbf{x}$