University of Technology, Sydney

Doctoral Thesis

Motion Segmentation Based Robust RGB-D SLAM

Author:
Youbing Wang

Supervisor:
A/Prof. Shoudong Huang

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

in the

Centre for Autonomous Systems
School of Elec, Mech and Mechatronic Systems
Faculty of Engineering and Information Technology

May 2015

Declaration of Authorship

I, Youbing WANG, declare that this thesis titled, 'Motion Segmentation Based Robust RGB-D SLAM' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ____________________________________

Date: ____________________________________

UNIVERSITY OF TECHNOLOGY, SYDNEY

Abstract

Centre for Autonomous Systems
School of Elec, Mech and Mechatronic Systems
Faculty of Engineering and Information Technology

Doctor of Philosophy

Motion Segmentation Based Robust RGB-D SLAM

by Youbing Wang

While research on simultaneous localisation and mapping (SLAM) in static environments can be regarded as a significant success due to intensive work during the last several decades, conducting SLAM, especially vision-based SLAM, in dynamic scenarios is still at its early stage. Although it seems like just one step further, the dynamic elements have brought in many unanticipated challenges, including motion detection, segmentation, tracking and $3D$ reconstruction of both the static environments and the moving objects, in addition to the handling of motion blur.

Solely based on RGB-D data with no prior knowledge available, this work centres upon proposing new practical solution frameworks for conducting SLAM in dynamic environments with efficient and robust motion segmentation methods serving as the basis. After a detailed review of the related achievements for SLAM in static environments as well as dynamic ones, and an analysis of the unaddressed challenges, four different motion segmentation methods, which include two 2-view sparse feature based motion segmentation algorithms, a 2-view semi-dense motion segmentation algorithm and an extended n-view dense moving object segmentation algorithm, are firstly proposed and their advantages, disadvantages and feasibility for different practical SLAM application scenarios are evaluated.

Based on the proposed motion segmentation methods, two kinds of solution frameworks for performing SLAM in dynamic scenarios are then put forward: the first one is formulated by integrating our sparse feature based motion segmentation techniques with the available pose-graph SLAM framework; and the other one is built upon dense moving object segmentation and tailored for dense SLAM. Related simulation and experimental results have demonstrated the effectiveness of our approaches.

Acknowledgements

First of all, I would like to offer my sincere gratitude to my supervisor, A/Prof. Shoudong Huang, who has supported me during my PhD study continuously with his patience, motivation, enthusiasm and immense knowledge while allowing me the room to work in my own way. Without his encouragement, guidance, and efforts, this work would not have been possible.

Besides my supervisor, I would like to thank Prof. Dikai Liu and Prof. Gamini Dissanayake for their invaluable support for my study at the Centre for Autonomous System.

Sincere thanks also go to the other staff members of CAS, Dr. Jack Jianguo Wang, A/Prof. Sarath Kodagoda, A/Prof. Guang Hong, A/Prof. Jaime Valls Miro, Dr. Gabriel Aguirre-Ollinger, Dr. Liang Zhao and Dr. Lei Shi. They have given me support for my study and life in various ways.

In my daily work, I have been blessed with a friendly and cheerful group of fellow students: Gibson Hu, Kasra Khosoussi, Lakshitha Dantanarayana, Yue Wang, a visiting student from Zhejiang University, and many other people, for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last several years.

Last but not the least, I would like to thank my family and all my friends, for their unconditional support to me spiritually throughout my life.

Contents

Declaration of Authorship i

Abstract iii

Acknowledgements iv

Contents v

List of Figures ix

List of Tables xii

Abbreviations xiii

1 Introduction 1
- 1.1 Simultaneous Localisation and Mapping (SLAM) . . . 1
- 1.2 SLAM in Dynamic Scenarios . . . 3
  - 1.2.1 Various Dynamic Scenarios to Handle . . . 3
  - 1.2.2 Conducting SLAM in Dynamic Environments . . . 3
  - 1.2.3 Another Chicken-and-Egg Problem? . . . 4
  - 1.2.4 Motivation . . . 4
- 1.3 The Scope and Limitations of This Thesis . . . 5
  - 1.3.1 The Targeted Problem . . . 5
  - 1.3.2 Research Scope . . . 5
  - 1.3.3 Limitations . . . 6
- 1.4 Main Contributions . . . 7
- 1.5 Publication List . . . 7
- 1.6 The Structure of This Thesis . . . 7

2 SLAM: From Static to Dynamic 9
- 2.1 SLAM in Static Environments . . . 10
  - 2.1.1 SfM and SLAM Problem . . . 10
  - 2.1.2 Two Kinds of Formulations . . . 10
  - 2.1.3 Two Categories of Typical Representations . . . 12
  - 2.1.4 Visual Odometry . . . 15

2.2 Available Methodologies for SLAM in Targeted Dynamic Scenarios 17
2.2.1 SLAMMOT, SMSaM and Multibody SLAM/MSaM 18
2.2.2 Motion Segmentation based Multibody SLAM/MSaM 22
2.2.3 Summary of Different Methodologies for Conducting SLAM in Dynamic Scenarios 22
2.3 Motion Segmentation 23
2.3.1 Definition 23
2.3.2 Different Implementations 23
2.3.3 Related Work 24
2.3.4 Evaluation Methods 25
2.3.5 Difficulties 25
2.3.6 Usability Analysis of Different Methods 26
2.4 Usability of Available SLAM Solution Frameworks for Conducting SLAM in Dynamic Scenarios 27
2.4.1 Pose-graph SLAM Versus Pose-feature-graph SLAM 27
2.4.2 Dense SLAM Versus Sparse Feature-based SLAM 28
2.5 Benchmark Datasets for RGB-D SLAM 28
2.6 Towards Motion Segmentation based Robust SLAM 29
2.6.1 Missing Elements 29
2.6.2 Our Basic Objectives 29

3 Sparse Feature Distance-based Motion Segmentation for Multibody RGB-D SLAM 31
3.1 Motivation 31
3.2 An Efficient RGB-D Motion Segmentation Method 32
3.2.1 Problem Restated 32
3.2.2 Theory 33
3.2.3 Algorithm 35
3.2.4 Discussion 36
3.2.5 Segmentation Results Using Simulated Data 37
3.2.6 Efficiency 37
3.3 Evaluation with A Simulated Multibody RGB-D SLAM Problem 39
3.3.1 The Simulated Multibody RGB-D SLAM Scenario 39
3.3.2 Motion Segmentation Results Using Simulated Noisy Data 39
3.3.3 Multibody SLAM Results Using Simulated Noisy Data 41
3.4 Evaluation Based on Our Own Dataset 42
3.4.1 Collecting Our Own Dataset 42
3.4.2 Motion Segmentation Results From Experimental Data With Outliers 43
3.4.3 Initial Values Based on Motion Segmentation and Visual Odometry 44
3.5 Summary 44

4 Motion Model based Segmentation for Robust RGB-D SLAM: From Sparse to Semi-dense 46
4.1 Related Work 47
4.1.1 RANSAC and Its Application in SLAM 47
4.1.2 Motion Segmentation Methods for SLAM in Dynamic Scenarios . 48
4.1.3 Different Solutions for Different Scenarios 49

4.2 Our Sparse Feature Motion Model based Motion Segmentation Algorithm 50
4.2.1 Sparse Feature Motion Model Based Motion Segmentation Algorithm . . . 51
4.2.2 Solution for Robust SLAM and Multibody SLAM . . . 51
4.3 Simulation and Experimental Results . . . 52
4.3.1 Results Using Simulated Dynamic Scenarios . . . 53
4.3.2 Experimental Results Using Static and Dynamic Scenarios . . . 54
4.4 Motivations to Semi-dense Feature Motion Model Based Motion Segmentation . . . 60
4.4.1 The Challenging Sequences and Its Impacts on Available Methods 61
4.4.2 Sparse Feature-based SLAM Cannot Deal with Image Blur and Textureless Areas . . . 63
4.4.3 Dense SLAM is Vulnerable to Moving Objects . . . 63
4.4.4 RANSAC and Robust Kernels Cannot Endure Too Many Outliers 64
4.5 Our Semi-dense Feature Motion Model based Motion Segmentation Method 64
4.5.1 SIFT Flow For Dense-Feature Detection And Matching . . . 64
4.5.2 Semi-dense Feature-based Motion Segmentation . . . 65
4.5.3 Pose-graph SLAM . . . 65
4.6 Experimental Evaluation Using Challenging Sequences . . . 66
4.6.1 The fr3_walking_static Sequence . . . 67
4.6.2 The fr3_walking_xyz Sequence . . . 69
4.6.3 The fr3_walking_rpy Sequence . . . 69
4.6.4 The fr3_walking_halfsphere Sequence . . . 72
4.6.5 Discussion . . . 73
4.7 Summary . . . 74

5 Towards Dense Moving Object Segmentation based Robust Dense RGB-D SLAM in Dynamic Scenarios 75
5.1 Introduction . . . 76
5.2 Related Work . . . 77
5.2.1 Two-view/N-view Based Motion Segmentation Versus Extended N-view Based Moving Object Segmentation [50] . . . 77
5.2.2 RGB-D SLAM: Sparse Versus Dense . . . 78
5.3 Moving Object Segmentation Based Robust Dense SLAM . . . 79
5.3.1 Proposed Robust Moving Object Segmentation Method . . . 79
5.3.2 Dense Visual SLAM . . . 82
5.4 SLAM Results for The Benchmark Dataset . . . 84
5.4.1 RGB-D SLAM Benchmark and Methods Involved . . . 84
5.4.2 Challenging Sequences . . . 86
5.4.3 Comparison of the Experimental Results . . . 86
5.4.4 Discussion . . . 87
5.5 Summary . . . 87

6 Conclusions and Future Work 88
6.1 What We Have Done: Motion Segmentation based Robust RGB-D SLAM 88
6.2 Future Work: Beyond Motion Segmentation . . . 89

A Analysis of The Computational Cost of Algorithm 3 **91**

Bibliography **93**

List of Figures

1.1 SLAM in relation with the other research topics in mobile robotics. 1

1.2 The flowchart of modern SLAM. 2

2.1 A sketch of the relative positions of the major topics to be covered in this chapter, where items in red represent the work has been proposed for conducting SLAM in static environments, those in green belong to visual SLAM, and those in blue are closely related to performing SLAM in dynamic scenarios. 9

2.2 Representing SLAM problems with DBN, in which $\boldsymbol{x}_{1..T}$ are invisible system states, $\boldsymbol{m}$ the unknown static environment, $\boldsymbol{u}_{1..T}$ input control and $\boldsymbol{z}_{1..T}$ the observations. 13

2.3 A graph representing SLAM problems in which $\boldsymbol{z}_{ij}$ represents the observation/odometry of $\boldsymbol{x}_j$ at the position of $\boldsymbol{x}_i$ and $\boldsymbol{Q}_{ij}$ is the corresponding covariance matrix. 14

3.1 Constrain a new point P using three non-collinear points P1, P2 and P3 . 35

3.2 Motion segmentation result for the simulated 3 planes scenario. Here different colours are only used to differentiate different motion groups, so not necessarily the same as the ground-truth marked. Best viewed in colour. 38

3.3 Motion segmentation result for the simulated 1 plane and 2 balls scenario. Here different colours are only used to differentiate different motion groups, so not necessarily the same as the ground-truth marked. Best viewed in colour. 38

3.4 Motion segmentation result for the simulated 3 balls scenario. Here different colours are only used to differentiate different motion groups, so not necessarily the same as the ground-truth marked. Best viewed in colour. . 39

3.5 Simulated multibody RGB-D SLAM scenario: blue and red points represent static features, and red ones are currently seen by the camera; green ones stand for camera's positions; yellow ones correspond to the paths of the two moving objects 40

3.6 One segmentation result from the simulated data: blue, red and green points represent the three groups the algorithm has found, corresponding exactly to the ground-truth groups 40

3.7 Angular, translational and position errors of the multibody RGB-D SLAM results using the simulated data. 41

3.8 Experimental setup: one Kinect, two boxes (one in blue, one in black) are located at the further side of the image, and have been marked out . . 42

3.9 One sample feature matching result from the experimental data 43

3.10 One sample segmentation result from the experimental data 43

3.11 The angular and translational errors of camera poses' initial values as compared with the ground truth. Best viewed in color. 44

4.1 A sample motion segmentation result of two frames from the fr3_long_office sequence (best viewed in colour): blue points respresent detected inliers, while red points represent outliers caused by inaccurate depth data or feature matching. 56
4.2 Our SLAM results for the fr3_long_office sequence compared with the ground truth (best viewed in colour): gray lines represent ground truth, blue lines represent the SLAM results, and red lines represent the differences between them at various corresponding timestamps. 56
4.3 A sample motion segmentation result of two frames from the fr3_sitting_xyz sequence (best viewed in colour): blue points respresent detected inliers, while red points represent outliers caused by inaccurate depth data or feature matching. 57
4.4 Our SLAM results for the fr3_sitting_xyz sequence compared with the ground truth (best viewed in colour): gray lines represent ground truth, blue lines represent the SLAM results, and red lines represent the differences between them at various corresponding timestamps. 57
4.5 A sample motion segmentation result of two frames from the desk_with_person sequence (best viewed in colour): blue points respresent detected inliers, while points in other colours represent outliers caused by inaccurate depth data, feature matching, or moving objects. 58
4.6 Our SLAM results for the fr2_desk_with_person sequence compared with the ground truth (best viewed in colour): grey lines represent ground truth, blue lines represent the SLAM results, and red lines represent the differences between them at various corresponding timestamps. 59
4.7 Two snapshots of the fourth sequence: fr3_walking_static. 60
4.8 Samples of *fr3_walking_static* 61
4.9 Samples of *fr3_walking_xyz* 62
4.10 Samples of *fr3_walking_rpy* 62
4.11 Samples of *fr3_walking_halfsphere* 63
4.12 An sample SIFT flow and motion segmentation result (Best viewed in colour). The first column are the original images, and the second column is the SIFT flow output. In the third column, different colours in the upper image represent different motion groups we've found, and the blue points in the lower image represent the static group. 67
4.13 The RMSE of RPE for the Results of dense SLAM with the TUM *walking_static* sequence 68
4.14 The RMSE of RPE for the Results of Visual odometry of our method with the TUM *walking_static* sequence 68
4.15 The RMSE of RPE for the Pose-graph SLAM Results of our method with the TUM *walking_static* sequence 69
4.16 The RMSE of RPE for the Results of dense SLAM with the TUM *walking_xyz* sequence 70
4.17 The RMSE of RPE for the Visual odometry of our method with the TUM *walking_xyz* sequence 70
4.18 The RMSE of RPE for the Pose-graph SLAM results of our method with the TUM *walking_xyz* sequence 71

4.19 The RMSE of RPE for the Results of dense SLAM with the TUM *walking_rpy* sequence 71

4.20 The RMSE of RPE for the Visual odometry of our method with the TUM *walking_rpy* sequence 72

4.21 The RMSE of RPE for the Pose-graph SLAM results of our method with the TUM *walking_rpy* sequence 72

5.1 In this example, the biggest common part having depth information of the two non-consecutive images does not belong to the static group, making loop-closure detection harder to handle. 78

5.2 Flowchart of the whole process for robust SLAM 80

5.3 Flowchart of the original moving object segmentation method [50] with images showing the results at different stages. 80

5.4 The left column represents the result before the merging using fundamental matrix, while the right one represents that after merging. The lower images represent the segmented group masks in different color intensities, and the upper images are the combination of the original image and segmentation results. Best viewed in colour. 81

5.5 The left column represents the result before merging the connected neighbouring regions, while the right one represents that after merging. The lower images represent the segmented group masks in different color intensities, and the upper images are the combination of the original image and segmentation results. Best viewed in colour. 81

5.6 The original image and segmentation results (from the *walking_static* sequence). 83

5.7 Comparison of dense SLAM results before and after moving object segmentation using the four *walking* sequences. 85

List of Tables

3.1 Computation time in relation with the number of feature points and groups 39

4.1 Comparison of Average Computation Time: Scenarios Composed of Two Motion Groups . . . 54

4.2 Comparison of Average Computation Time: Scenarios Composed of Five Motion Groups . . . 54

4.3 Comparison of RMSE of RPE (m/s) and ATE (m) for SMS and DVSLAM 60

4.4 Comparison of Visual Odometry and SLAM Results in Terms of RMSE of RPE (m/s) . . . 73

4.5 Comparison of Visual Odometry and SLAM Results in Terms of RMSE of ATE (m) . . . 73

5.1 Comparison of the RPE (m/s) & ATE (m) Results of Dense SLAM Versus Those of Algorithm 6 (MS_DSLAM) Using the Four Sequences . . . 87

Abbreviations

SLAM	**S**imultaneous **L**ocalization **A**nd **M**apping
KF	**K**alman **F**ilter
EKF	**E**xtended **K**alman **F**ilter
EIF	**E**xtended **I**nformation **F**ilter
SfM	**S**tructure **f**rom **M**otion
MSaM	**M**ultiple **S**tructure **a**nd **M**otion
RANSAC	**RAN**dom **SA**mple **C**onsensus
GPS	**G**lobal **P**ositioning **S**ystem
IMU	**I**nertial **M**easurement **U**nit
RMSE	**R**oot-**M**ean-**S**quare **E**rror
RPE	**R**elative **P**ose **E**rror
ATE	**A**bsolute **T**ranslational **E**rror
SMSaM	**S**imultaneous **M**ultibody **S**tructure **a**nd **M**otion
MDL	**M**inimum **D**escription **L**anguage
KLT	**K**anade-**L**ucas-**T**omasi

Dedicated to my family. . .