# 3D non-rigid SLAM in minimally invasive surgery

by

Jingwei Song

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

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

March 2020

# Declaration of 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.

This research is supported by an Australian Government Research Training Program Scholarship.

Production Note:

Signed: Signature removed prior to publication.

Date: 22 March 2020

UNIVERSITY OF TECHNOLOGY SYDNEY

# *Abstract*

Faculty of Engineering and Information Technology
Centre for Autonomous Systems

Doctor of Philosophy

by Jingwei Song

Aiming at reducing trauma and morbidity associated with large incisions in open surgery, minimally invasive surgery (MIS) has been widely acquired in clinical practice as a powerful tool enabling patients with less pain, shorter hospital stay, and fewer complications. However, MIS narrows the surgeon's field of view which confines visual information when implementing MIS. Therefore, a stereoscope or monocular scope is an essential tool for capturing and transmitting 2D images during the procedure.

Although numbers of special sensors including laser, structured light, time-of-flight cameras have been applied or investigated in MIS, RGB scope is still widely applied in the intro-operative system because it is non-invasive and cheap to be installed. Thus it is an important topic to rebuild and visualize the latest deformed shape of soft-tissue surfaces to mitigate tissue damages from stereo or monocular scopes. This research aims at proposing innovative robocentric simultaneous localization and mapping (SLAM) algorithm for deformable dense reconstruction of soft-tissue surfaces using a sequence of images obtained from a stereoscope or monocular camera. In this paper, we try to solve the problem by introducing a warping field based on the embedded deformation (ED) nodes which makes full use of the 3D shapes recovered from consecutive pairs of stereo images by deforming the last updated model to the current live model. Our robocentric SLAM system (off-line and tested on stereo videos) can: (1) Incrementally build a live model by progressively fusing new observations with vivid accurate texture. (2) Estimate the deformed shape of

the unobserved region with the principle As-Rigid-As-Possible. (3) Perform the dynamic model shape deformation. (4) Estimate the current relative pose between the soft-tissue and the scope.

We further improve and optimize the proposed robocentric deformable SLAM algorithm to MIS-SLAM: a complete real-time large scale robocentric dense deformable SLAM system with stereoscope in MIS based on heterogeneous computing by making full use of CPU and GPU. Idled CPU is used to perform ORB-SLAM for providing robust global pose. Strategies are taken to integrate modules from CPU and GPU. We solve the key problem raised in previous work, that is, fast movement of scope and blurry images make the scope tracking fail. Benefiting from improved localization, MIS-SLAM can achieve large scale scope localizing and dense mapping in real-time. It transforms and deforms the current model and incrementally fuses new observation while keeping the vivid texture. In-vivo experiments conducted on publicly available datasets presented in the form of videos demonstrate the feasibility and practicality of MIS-SLAM for potential clinical purpose.

In MIS-SLAM, however, it remains challenging to keep constant speed in deformation nodes parameter estimation when the model grows larger. In practice, the processing time grows rapidly in accordance with the expansion of the maps. Therefore, we propose an approach to decouple nodes of deformation graph in large scale robocentric dense deformable SLAM and keep the estimation time to be constant. We discover that only partial deformable nodes in the graph are connected to visible points. Based on this principle, the sparsity of the original Hessian matrix is utilized to split parameter estimation into two independent steps. With this new formulation, we achieve faster parameter estimation with amortized computation complexity reduced from $O(n^2)$ to closing $O(1)$. As a result, the computation cost barely increases as the map keeps growing. By our strategy, the bottleneck of limited computation in estimating deformation field in large scale environment has been overcome. The effectiveness is validated by experiments, featuring large scale deformation scenarios.

In addition to robocentric SLAM, this thesis also aims at developing a general SLAM which estimates the scope poses correctly. An elaborate observability analysis is conducted on

the ED graph. We demonstrate and prove that the ED graph widely used in such scenarios is unobservable and leads to multiple solutions unless suitable priors are provided. Example, as well as theoretical prove, are provided to show the ambiguity of ED graph and scope pose. Different from robocentric SLAM, in modeling non-rigid scenario with ED graph, motion priors of the deforming environment is essential to separate robot pose and deforming environment. The conclusion can be extrapolated to any free form deformation formulation. In guaranteeing the observability, this research proposes a preliminary deformable SLAM approach to estimate robot pose in complex environments that exhibits regular motion. A strategy that approximates deformed shape using a linear combination of several previous shapes is proposed to avoid the ambiguity in robot movement and rigid and non-rigid motions of the environment. Fisher information matrix rank analysis is performed to prove the effectiveness. Moreover, the proposed algorithm is validated using Monte Carlo simulations and real experiments. It is demonstrated that the new algorithm significantly outperforms conventional SLAM and ED based SLAM especially in scenarios where there is large deformation.

# *Acknowledgements*

First and foremost I would like to express my sincere gratitude to my supervisor A/Prof. Shoudong Huang and Dr. Liang Zhao for the continuous support of my Ph.D study and related research, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I hope I can continue to collaborate with you in the future.

Besides my supervisor, I would like to thank Dr. Teresa Vidal Calleja, A/Prof. Jaime Valls Miro, Prof. Gamini Dissanayake, and A/Prof. Sarath Kodagoda for their guidance and friendship, but also for the hard question which incented me to widen my research from various perspectives.

My sincere thanks also go to Dr. Antonio Agudo and Prof. Francesc Moreno-Noguer, who provided me an opportunity to join their team as an exchange student at Institut de Robótica i Informática Industrial (IRI) of the Spanish Council for Scientific Research and the Technical University of Catalonia. And also for supporting me to win IRI-Maria de Maeztu intern scholarship. During my 4 months visit, I have terrific experience and memory in IRI as well as Barcelona. I am also grateful to Dr. Mitesh Patel for hosting me at FXPAL as an intern; the friendly environment and living in California touches me very much.

I acknowledge the scholarships International Research Training Program Scholarship (IRTP) funded by the Australian Government, Department of Education and Training to support 3.5 years for my research and make my study possible. I would also like to acknowledge Centre for Autonomous Systems and School of Mechanical and Mechatronic Engineering of the Faculty of Engineering and Information Technology at University of Technology Sydney for providing me funding to attend conferences held in London, Barcelona, Madrid, Brisbane, Vancouver and being an exchange student in Barcelona.

Many thanks go to my fellow friends and colleagues at the Centre for Autonomous Systems for the numerous stimulating and helpful discussions and their friendship. My special thanks go to Dr. Jun Wang for inspiring discussions on solving technical issues as well as reporting the state-to-art technologies. We work closely in achieving good publications. I also appreciate Dr. Raphael Falque for useful discussions on theories of deformation technologies. Raphael is senior to me and I benefit a lot from him. I would also like to thank my colleague Kristian Villavicencio for helping with the ElasticFusion experiments.

I thank my friends and labmates Karthick Thiyagarajan, Mahdi Hassan, Lakshitha Dantanarayana, Phillip Quin, Leo Shi, Nalika Ulapane, Buddhi Wijerathna, Asok Aravinda, Julien Collart, Alexander Virgona, Maani Ghaffari Jadidi, Kasra Khosoussi, Daobilige Su, Brendan Emery, Katherine Waldron, Cédric Le Gentil, Kanzhi Wu, Teng Zhang and many other colleagues. I thank my fellow labmates for the stimulating discussions and for all the fun we have had in the last four years, as well as their hands-on help on setting up the experimental environment to collect and process data; Many thanks to Herni Winarta and Katherine Waldron for their help on administrative work.

Additional thanks go to my other friends in Sydney, I enjoy the 3 years in this sunny city. The last but most important thanks give to my parents, who are always proud of me for my achievements, and supported and cared for me throughout these years.

# Contents

# List of Figures

# List of Tables

# Acronyms & Abbreviations

**CPU** Central Processing Unit

**CT** Computed Tomography

**CUDA** Compute Unified Device Architecture

**DFF** Distance Field Function

**ED** Embedded Deformation

**EKF** Extended Kalman Filter

**ELAS** Efficient Large-scale Stereo

**EM sensor** Electromagnetic sensor

**FEM** Finite Element Method

**FIM** Fisher Information Matrix

**GPGPU** General-Purpose computing on Graphics Processing Units

**GPU** Graphical Processing Unit

**ICP** Iterative Closest Point

**MIS** Minimally Invasive Surgery

**NRSfM** Non Rigid Structure from Motion

**PI** Points Irrelevant

**PR** Points Relevant

**RANSAC** RANdom SAmple consensus

**RMSE** Root-Mean-Square Error

**SfM** Structure from Motion

**SfS** Shape from Shading

**SfT** Structure from Template

**SIFT** Scale-invariant feature transform

**SLAM** Simultaneous Localization And Mapping

**SURF** Speeded Up Robust Feature

**TSDF** Truncated Signed Distance Function

**TSDW** Truncated Signed Distance Weight

# Nomenclature

| | **General Notations** |
|---|---|
| $O$ | Centers of the left image. |
| $O'$ | Centers of the right image. |
| $p$ | Pixel of the projected point on left image. |
| $p'$ | Pixel of the projected point on right image. |
| $u$ | First coordinate of 2D pixel. |
| $v$ | Second coordinate of 2D pixel. |
| $\mathbf{K}$ | Camera intrinsic matrix. |
| $\alpha(u, v)$ | Surface albedo. |
| $\hat{I}(u, v)$ | The measured pixel intensity at pixel $(u, v)$. |
| $O(\cdot)$ | Computational complexity. |
| $\Omega$ | Linear elastic solid parameter. |
| $\lambda$ and $\mathbf{G}$ | Lam'e parameters that define the material elastic properties. |
| $\mathcal{U}$ | Poisson's ratio. |
| $\mathcal{E}$ | Young's modulus. |
| $a_{i,jj}$ and $a_{j,ij}$ | Displacement vectors share the same edge. |
| $\mathbf{g}_j$ | Position of node $j$. |
| $\mathbf{A}_j$ | Affine matrix of node $j$. |
| $\mathbf{v}$ | Position of a point. |
| $\tilde{\mathbf{v}}$ | Target deformed vertex of $\mathbf{v}$. |
| $\mathbf{R}_c$ | Global rotation of the scope. |
| $\mathbf{T}_c$ | Global translation of the scope. |

| | |
|---|---|
| $w(\mathbf{v})$ | The quantified weight for transforming $\mathbf{v}$ exerted by each related ED node. |
| $d_{max}$ | The maximum distance of the vertex to $k+1$ nearest ED node. |
| $m$ | The number of ED nodes. |
| $\mathbf{c}_1$, $\mathbf{c}_2$ and $\mathbf{c}_3$ | The column vectors of $\mathbf{A}$. |
| $E_{rot}$ | The affine matrix close to $SO(3)$. |
| $E_{reg}$ | The sum the transformation errors from each ED node. |
| $\alpha_{jk}$ | The overlap influence of the two ED nodes. |
| $\mathbb{N}(j)$ | The set of neighboring node to node $\mathbf{g}_j$. |
| $\boldsymbol{\mathcal{F}}(\cdot)$ | A general function defining a point to target distance. |
| $\mathcal{D}(\cdot)$ | The corresponding voxel value recorded in DFF. |
| $E_{data}$ | The sum distance error. |
| $\mathbb{L}$ | The set for all the visible points. |
| $\mathbb{D}$ | Depth scan. |
| $\Gamma(\cdot)$ | Lift 2D depth pixel up to 3D point. |
| $\mathrm{H}(\cdot)$ | Lift 2D normal pixel up to 3D normal. |
| $P(\mathbf{v})$ | The projective function projecting 3D vertex to 2D pixel. |
| $\epsilon_d$ | Threshold for the distance. |
| $\epsilon_n$ | Threshold for the angle. |
| $\tilde{\mathbf{V}}_i$ | The 3D points of current frame from last frame of ORB features. |
| $\mathbf{V}_i$ | The 3D points of the deformed points from last frame of ORB features. |
| $\omega(\mathbf{v}_i)$ | Weight of model point $\mathbf{v}_i$. |
| $\mathbf{C}_i$ | Color of model point. |
| $d_{min}(\tilde{\mathbf{v}}_i)$ | The minimum distance of model point. |
| $\tilde{\mathbf{v}}_i$ | to its corresponding nodes. |
| $\epsilon$ | The average grid size of nodes. |
| $\tilde{\mathbf{v}}_i\|_z$ | The value of point $\tilde{\mathbf{v}}_i$ on the z direction. |
| $t_i$ | The time stamp of vertex $\mathbf{v}_i$. |
| $\mathbf{n}_i$ | The normal of vertex $\mathbf{v}_i$. |
| $\mathbf{s}_i$ | The boolean variable stability of vertex $\mathbf{v}_i$. |

| | |
|---|---|
| $E_{corr}$ | Sum errors of distances between deformed key points and target key points. |
| $E_r$ | The $SO(3)$ distance between estimated orientation and intilial orientation. |
| $E_p$ | The Euclidean distance between estimated position and intilial position. |
| $\epsilon_{dv}$ | The threshold for extracting visible points based on distance. |
| $\epsilon_{nv}$ | The threshold for extracting visible points based on angle. |
| $\mathbf{R}_c^n$ | Estimated rotation in step $n$. |
| $\mathbf{T}_c^n$ | Estimated translation in step $n$. |
| $\tilde{\mathbf{R}}_c^n$ | Rotation in step $n$ from ORB-SLAM. |
| $\tilde{\mathbf{T}}_c^n$ | Translation in step $n$ from ORB-SLAM. |
| $\mathbb{V}^n$ | Set of visible points in step $n$. |
| $\mathbb{D}^n$ | Observed depth in step $n$. |
| $\mathbb{P}^n$ | Fused point set. |
| $\epsilon_{df}$ | The threshold for fusing points based on distance. |
| $\epsilon_{nf}$ | The threshold for fusing visible points based on angle. |
| $\tilde{\mathbf{v}}_i^n\|_z$ | The value of deformed point $\tilde{\mathbf{v}}_i^n$ in the $z$ direction. |
| $\tilde{\mathbf{n}}_i^n$ | The deformed normal of $\mathbf{n}_i^n$. |
| $\omega_{max}$ | The maximum weight for each model point. |
| $\mathbf{P}$ | A group of predefined key source points. |
| $\tilde{\mathbf{P}}$ | Deformed key points set of $\mathbf{P}$. |
| $\otimes$ | The Kronecker product. |
| $\|\| \cdot \|\|_F^2$ | The Frobenius norm. |
| $X_i$ | The state vector is denoted as . |
| $skew(\cdot)$ | The skew symmetric operator. |
| $\mathbf{I}$ | A 3 by 3 identity matrix. |
| $\odot$ | The Hadamard product. |
| $\boldsymbol{\mathcal{H}}_{ed}$ | Hessian matrix of ED formulation. |
| $\mathbf{H}_1$ | One submatrix of $\boldsymbol{\mathcal{H}}_{ed}$. |
| $\mathbf{H}_2$ | One submatrix of $\boldsymbol{\mathcal{H}}_{ed}$. |

| | |
|---|---|
| $\mathbf{f}^n$ | A feature position in step $n$. |
| $N$ | The number of features. |
| $F$ | The number of steps. |
| $\mathbf{B}$ | The combination of all valid features. |
| $E_{obs}$ | The sum error of robot to feature observations. |
| $\mathbf{m}_i^j$ | The observation from robot to location of feature $i$ in step $j$. |
| $\mathcal{F}(\cdot)$ | The estimated observation from robot pose to feature position. |
| $E_f$ | The error between current feature and its estimation from historical locations. |
| $E_{ini}$ | The initial robot pose keeps static in the period size $t$. |
| $\ominus$ | Inverse retraction defining $SO(3)$ distance. |
| $\mathbf{c}$ | The coefficient matrix . |