UNIVERSITY OF TECHNOLOGY SYDNEY

DOCTORAL THESIS

Action for Perception: Active Object Recognition and Pose Estimation in Cluttered Environments

Kanzhi WU

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

at the

Centre for Autonomous Systems

School of Electrical, Mechanical and Mechatronic Systems

August 8, 2017

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 part of the collaborative doctoral degree and/or 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.

Signature of Student:

_____

Date:

_____

Abstract

Object recognition and localisation are indispensable competency for service robots in everyday environments like offices and kitchens. Presence of similar objects that can only be differentiated from a small part of the surface together with clutter that leads to occlusions make it impossible to detect target objects accurately and reliably from a single observation. When the sensor observing the environment is mounted on a mobile platform, object detection and pose estimation can be facilitated by observing the environment from a series of different viewpoints. Computing *Active perception* strategies, with the aim of finding optimal actions to enhance object recognition and pose estimation performance is the focus of this thesis.

This thesis consists of two main parts:

In the **first** part, it focuses on object detection and pose estimation from a single frame of observation. Using an RGB-D sensor, we propose a modular 3D textured object detection and pose estimation framework which can recognise object under cluttered environment by taking advantage of the geometric information provided from the sensor. To handle less-textured objects and objects under severe illumination conditions, we propose a novel RGB-D feature which is robust to illumination, scale, rotation and viewpoint variations, and provides reliable feature matching results under challenging conditions. The proposed feature is validated for multiple applications including object detection and point cloud alignment. Parts of the above approaches are integrated with existing work to produce a practical and effective perception module for a warehouse automation task. The designed perception system can detect objects of different types and estimate their poses robustly thus guaranteeing a reliable object grasping and manipulation performances.

In the **second** part of the thesis, we investigate the problem of *active* object detection and pose estimation from two perspectives: with and without considering the uncertainties in the motion model and the observation model. First, we propose a model-driven active object recognition and pose estimation system via exploiting the feature association probability under scale and viewpoint variations. By explicitly modelling the feature association, the proposed system can predict future information more accurately thus laying the foundation of a successful active Next-Best-View planning system even with a naive greedy search technique. We also present a probabilistic framework which handles motion and observation uncertainties in the active object detection and pose estimation problem. We present an optimisation framework which computes

the optimal control at each step, using an objective function which incorporates uncertainties in state estimation, feature coverage for better recognition confidence and control consumption. The proposed framework can handle various issues such as object initialisation, collision avoidance, occlusion and changing the object hypothesis. Validations based on a simulation environment are also presented.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Gamini Dissanayake for his continuous support during my PhD study, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my PhD study. I would also like to thank my co-supervisor Dr. Ravindra Ranasinghe, the door of his office was always open whenever I ran into a trouble spot or had a question about my research or writing.

My sincere thanks also go to Assoc. Prof. Shoudong Huang and Mr Teng Zhang for the insightful discussions about SLAM and planning under uncertainty. During the past four years, I have experienced a welcoming and passionate research atmosphere in CAS and I would like to thank Prof. Dikai Liu for his management of CAS. I would like to thank Dr. Liang Zhao for patiently answering my questions about multiview geometry.

I would also like to thank my great friends and colleagues in CAS: Daobilige Su, Lakshitha Dantanarayana, Kasra Khosoussi, Raphael Falque and so many other colleagues who have helped me, too many to list here. The discussions with you not only have helped me in my academic research but have also enriched my life in the past four years.

Last but not the least, I would like to thank my parents and especially my wife Liye Sun for her understanding, support and companionship during my research career.

Contents

List of Figures

List of Tables

List of Abbreviations

ANN	**A**pproximate **N**earest **N**eighbour
APC	**A**mazon **P**icking **C**hallenge
BoW	**B**ag **o**f **W**ords
BRAND	**B**inary **R**obust **A**ppearance and **N**ormals **D**escriptor
BRIEF	**B**inary **R**obust **I**ndependent **E**lementary **F**eatures
BRISK	**B**inary **R**obust **I**nvariant **S**calable **K**eypoints
BRM	**B**elief **R**oad**M**ap
CNN	**C**onvolutional **N**eural **N**etworks
DLT	**D**irect **L**inear **T**ransformation
DoF	**D**egree of **F**reedom
FAST	**F**eatures from **A**ccelerated **S**egment **T**est
FIRM	**F**eedback-based **I**nformation **R**oad**M**ap
GLOH	**G**radient **L**ocation and **O**rientation **H**istogram
GBS	**G**eneral **B**elief **S**pace
ISS	**I**ntrinsic **S**hape **S**ignatures
KPQ	**K**ey**P**oint **Q**uality
LBSS	**L**aplace-**B**eltrami **S**cale-**S**pace
LDA	**L**inear **D**iscriminant **A**nalysis
LLC	**L**ocality-constrainted **L**inear **C**oding
LM	**L**evenburg **M**arquart
LQG	**L**inear **Q**uadratic **G**aussian
LQR	**L**inear **Q**uadratic **R**egulator
LOIND	**L**ocal **O**rdinal **I**ntensity and **N**ormal **D**escriptor
LSH	**L**ocal **S**ensitive **H**ashing
MDP	**M**arkov **D**ecision **P**rocess
MOPED	**M**ultiple **O**bject **P**ose **E**stimation and **D**etection

MPC	**M**odel **P**redictive **C**ontrol
NBV	**N**ext-**B**est-**V**iew
ORB	**O**riented FAST and **R**otated **B**RIEF
PCA	**P**rinciple **C**omponent **A**nalysis
PCL	**P**oint **C**loud **L**ibrary
PFH	**P**oint **F**eature **H**istogram
PnP	**P**erspective-**n**-**P**oint
POMDP	**P**artially **O**bservable **M**arkov **D**ecision **P**rocess
PRM	**P**robabilistic **R**oad**M**ap
RANSAC	**RAN**dom **SA**mple **C**onsensus
RISAS	A **R**otation, **I**llumination and **S**cale invariant **A**ppearance and **S**hape Feature
R-CNN	**R**egion-based **C**onvolutional **N**eural **N**etworks
RNN	**R**esidual **N**eural **N**etwork
ROI	**R**egion **O**f **I**nterest
RRG	**R**apid-exploring **R**andomised **G**raph
RRT	**R**apid-exploring **R**andomised **T**rees
RRBT	**R**apid-exploring **R**andomised **B**elief **T**rees
SfM	**S**tructure from **M**otion
SHOT	**S**ignature of **H**istogram **O**rien**T**ation
SIFT	**S**cale **I**nvariant **F**eature **T**ransform
SLAM	**S**imultaneous **L**ocalisation **A**nd **M**apping
SORSAC	**S**patially **OR**dered **SA**mple **C**onsensus
SQP	**S**equential **Q**uadratic **P**rogramming
SURF	**S**peed-**U**p **R**obust **F**eature
SVD	**S**ingular **V**alue **D**ecomposition
RANSAC	**RA**ndom **SA**mple **C**onsensus
USC	**U**nique **S**hape **C**ontext