
Autonomous Active Perception Framework for Large Object Recognition Tasks

A thesis submitted in partial fulfilment of the requirements

for the degree of

Master of Engineering (Research)

by

Thanh Long Vu

to

School of Mechanical and Mechatronic Engineering
Faculty of Engineering and Information Technology
University of Technology Sydney

NSW - 2007, Australia

July 2022

Certificate of Original Authorship

I, *Thanh Long Vu*, 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 IT* 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.

Production Note:

Signed: Signature removed prior to publication.

Date: 31/7/2022

Autonomous Active Perception Framework for Large Object Recognition Tasks

by

Thanh Long Vu

A thesis submitted in partial fulfilment of the requirements for the
degree of Master of Engineering (Research)

Abstract

In recent decades, the technology development in hardware and software has stimulated robotics systems' implementation in various fields. Robots are most commonly deployed in dangerous or mundane working environments, significantly reducing accidents, injuries and casualties in the workforce. Ideally, robots would be designed to perform the tasks autonomously, conducting calculations and making decisions based on sensory data. However, although robotics research has continuously advanced throughout the last half a century, there are still many complicated tasks where robots cannot achieve full autonomy yet. In these scenarios, interaction from a human supervisor may be required to make control decisions either locally or remotely. The quality of the decisions made by the supervisor or operator depends heavily on the available sensory feedback from the robotics system, which helps the human perceive the environment where the robot is operating. Therefore, perception capabilities are required for all autonomous and semi-autonomous robotics systems to process and make sense of the received data, so the system or the operator can perform necessary actions.

This thesis focuses on developing an active perception framework for robots working remotely, where the human operator cannot directly perceive the surrounding environment. The specific modality of sensor data received from the remote system is coloured Three-Dimensional (3D) point cloud data obtained from sensing devices such as Light Detection

and Ranging (LiDAR) or depth cameras. Additionally, this thesis investigates the practicality and benefits of utilising Virtual Reality (VR) as a tool to visualise the data obtained from a remote system.

This thesis firstly reviews multiple frontier detection algorithms for Two-Dimensional (2D) exploration, comparing these algorithms with other notable frontier detection algorithms. The algorithms are implemented to enable insight to be gained about the performance of the inspected algorithms compared with the other notable algorithms. A 3D interactive and active mapping framework for a mobile manipulator platform based on dynamic Gaussian Probabilistic Implicit Surfaces (GPIS) was investigated and implemented to validate its efficiency while simultaneously exploring and interacting with a large pile of objects. The framework is shown to perform near real-time map updates for a dynamically changing environment due to its probabilistic nature. Two perception systems are presented that employ the point cloud data in the framework to perform object detection, pose estimation and scene overlay annotation. Finally, a framework for sensor data visualisation in VR environments is presented, which acquires, transmits and renders real-time RGB-D sensor data with integrated automatic perception annotations in a VR environment. This enables various data modalities and limiting factors to be investigated and optimised to improve the subjective cognitive workload.

Experiments are conducted in both simulated and real-life scenarios. The two real robotic platforms used in the experiments are mobile manipulators. The first platform is composed of a 6 Degree of Freedom (DOF) manipulator and a commercial mobile base platform and was mainly used to validate exploration algorithms. The second mobile manipulator platform comprises a custom-designed mobile base and a 5 DOF manipulator, equipped with a camera system consisting of two RGB-D calibrated cameras and a real-time VR interface. This platform was used to conduct and investigate the optimal configuration for operator performance during collaborative autonomy tasks.

Contents

Declaration of Authorship	i
Abstract	ii
List of Figures	viii
List of Tables	xii
Glossary of Terms	xv
1 Introduction	1
1.1 Background	2
1.2 Motivation	3
1.3 Scope	5
1.3.1 Aims	6
1.3.2 Objectives	7
1.4 Contributions	7
1.5 Methodology	8
1.5.1 Evaluation of Frontier-Based Exploration and Dynamic Mapping Algorithms for Autonomous Robotic Systems	8
1.5.2 Enhancing Human-Robot Collaboration through Processing and Presentation of 3D Sensor Data in VR	9
1.6 Publications	10
1.7 Thesis Outline	10
2 Review of Related Work	13
2.1 Mobile Robot Exploration and Mapping	13
2.1.1 Frontier-based Exploration	14
2.1.2 Localisation and Mapping	18
2.1.3 Active and Interactive Perception	19
2.2 3D Point Cloud Processing	21
2.2.1 Filtering Algorithms	23
2.2.2 Segmentation Algorithms	24
2.2.3 Clustering Methods	28

2.3	Virtual and Augmented Reality in Robotics	30
2.3.1	Virtual and Augmented Applications in Training Programs	31
2.3.2	Virtual and Augmented Reality Applications in Control Interfaces	33
2.3.3	Dense 3D Data Transmission	35
2.4	Summary	36
3	Frontier Detection Algorithms	38
3.1	Introduction	38
3.2	Methodology	39
3.2.1	Experiment 1	41
3.2.2	Experiment 2	42
3.2.3	Experiment 3	44
3.3	Experimental Results	44
3.3.1	Discussion	48
3.4	Conclusion	51
4	Active and Interactive Mapping	52
4.1	Introduction	52
4.1.1	Active and Interactive Mapping Problem	53
4.2	Methodology	54
4.2.1	System Overview	54
4.2.2	Experimental Setup	56
4.2.3	Simulated Experiment	57
4.2.3.1	Ablation Study	57
4.2.3.2	Benchmark Test	58
4.2.4	Real-life Experiment	58
4.3	Experimental Result	58
4.3.1	Simulated Experiment	58
4.3.1.1	Ablation Study	58
4.3.1.2	Benchmark Test	59
4.3.2	Real-life Experiment	60
4.3.2.1	Dynamic GPIS Accuracy	60
4.3.2.2	Next Best View Test	60
4.4	Conclusion	62
5	Perception and Estimation	64
5.1	Introduction	64
5.1.1	Rectangular-shaped Object Pose Estimation	65
5.1.2	Overlay Bounding Box Annotation	67
5.2	Methodology	68
5.2.1	Core Background Fundamentals	68
5.2.1.1	Region Growing Segmentation	68
5.2.1.2	Principal Component Analysis	69
5.2.1.3	Rotation Averaging	72
5.2.2	System 1: Rectangular-shaped Object Pose Estimation	74

5.2.2.1	Filtering Strategies to Improve Object Detection in Noisy RGB-D Point Cloud Data	74
5.2.2.2	Region Growing Segmentation using Normal Vectors	75
5.2.2.3	Rectangular Object Recognition and Pose Estimation using PCA	75
5.2.2.4	Best Surface Selection for Rectangular Prism-Shaped Objects using Sigmoid Scores	76
5.2.2.5	Rotation Averaging for Surface Pose Estimation	77
5.2.3	System 2: Overlay Bounding Box Annotation	79
5.3	Experimental Result	80
5.3.1	System 1: Rectangular-shaped Object Pose Estimation	80
5.3.1.1	Comparative Study: System 1 and ICP-integrated System	81
5.3.1.2	Ratio Between Eigenvalue and Object Dimensions	83
5.3.1.3	Discussion	85
5.3.2	System 2: Overlay Bounding Box Annotation	86
5.3.2.1	Simulated Experiments	86
5.3.3	System 2: Background Removal and Clustering Process	87
5.3.3.1	Simulated Experiment	88
5.3.3.2	Real-world Experiments	90
5.4	Conclusion	92
6	Using Virtual Reality for Collaborative Autonomy	93
6.1	Introduction	93
6.1.1	User Study 1: The Influence of Sensor Data Visualisation Configuration on Users' Performance	94
6.1.2	User Study 2: An Investigation of User Performance in Virtual Reality-based Annotation-assisted Remote Robot Control	94
6.2	Methodology	95
6.2.1	The VR-ROS Framework For Collaboration Autonomy	95
6.2.1.1	Hardware and High-level Overview	95
6.2.1.2	Real-time Point Cloud Transmission for VR Rendering using Compressed Depth Images	95
6.2.1.3	Point Cloud Reconstruction	96
6.2.2	User Study 1: Experiment Setup	99
6.2.2.1	Experiment Procedure	101
6.2.2.2	Experimental Measurement 1: Objective Task Completion Time	102
6.2.2.3	Experimental Measurement 2: Subjective Cognitive Workload	102
6.2.3	User Study 2: Experimental Setup	103
6.2.3.1	Hardware and High-level Overview	103
6.2.3.2	Custom Robotic Manipulator	104
6.2.3.3	Data Transmission and Point Cloud Rendering	104
6.2.3.4	Experimental Procedure	105
6.3	Experimental Result	107

6.3.1	User Study 1 Results	107
6.3.1.1	Objective Task Completion Time	107
6.3.1.2	Subjective User Experience Results	108
6.3.1.3	Discussion	109
6.3.2	User Study 2 Results	110
6.3.2.1	Objective Task Completion Time	110
6.3.2.2	Objective Task Precision	111
6.3.2.3	Discussion	111
6.4	Conclusion	112
7	Conclusions	114
7.1	Summary Of Contributions	115
7.1.1	Efficiency Evaluation of Frontier Detection Algorithms for Map Exploration	115
7.1.2	Validation of an Active and Interactive 3D Mapping Framework for Mobile Manipulator Platforms	115
7.1.3	Development of Perception Systems for Object Detection and Pose Estimation using Point Cloud Data	116
7.1.4	Enhancing Perception for Collaborative Autonomy: A Framework and User Studies for Sensor Data Visualisation in VR Environments	116
7.2	Discussion of Limitations	117
7.3	Future Work	118
	Appendices	120
	A Human Ethics Application	120
	Bibliography	122

List of Figures

1.1	Abseiling workers performing manual rock-scaling operations.	4
2.1	The rays that traverse the occupancy grid (grey cells are unknown space, and white cells are freespace). The dashed lines show the cells that Bresenham's line algorithm traverses when determining the cells that are on the line between a ray's endpoints.	16
2.2	(a) Frontier-Tracing Frontier Detection (FTFD); (b) Expanding-Wavefront Frontier Detection (EWFD). Thick red borders indicate the cells that are determined to be potential frontiers; dashed borders are the cells that are removed from the queue but are discarded before full evaluation. The blue cells are frontier cells from the previous timestep, which are also evaluated. The thick black lines denote the sensor Field of View (FOV) [1].	18
2.3	Example of Point Cloud Library (PCL) applications [2]. (a) RangeImage display utilising PCL Visualisation (bottom) for a 3D data set (top). (b) PCL StaticalOutlierRemoval application. Left: Raw data. Middle: StaticalOutlierRemoval result. Right: The algorithm's rejected points.	22
2.4	Qualitative comparison of the filtering methods investigated by Moreno [3]. (a) - (d) The point cloud after the implementation of various filtering methods: (a) No filter, (b) Pass-through filter, (c) Voxel grid filter, (d) Approximate voxel grid filter.	24
2.5	GNG vs Voxel Grid Comparison. Top Left: Noisy model. Top Right: Original CAD model. Bottom left: filtered model using GNG method. Bottom right: filtered model using Voxel Grid [4].	25
2.6	Segmentation of natural colour image. (a_n) Challenging problem initialisation. (b_n) Final result with the proposed algorithm in [5].	25
2.7	Seeded region growing experimental results. (a) Artificial image with noise. (b) to (h) Results with different initial seeds sets [6].	27
2.8	Unseeded region growing experimental results. (a)(c) Artificial images with different noise levels. (b)(d) Results of (a) and (c), respectively [7].	27
2.9	Single-link vs Complete-Link on a data set containing two classes (1 And 2) connected by noisy patterns (*) [8].	29
2.10	Results of shared nearest neighbor algorithm [9].	30
2.11	Virtual Reality teleoperation. (a) Real-life implementation. (b) First-person view from inside VR [10].	32
2.12	(a) A system based on the GAVRe2 framework. (b) First-person view from inside VR [11].	33

2.13	The interfaces given to users. Left: 2D interface. Right: 3D VR interface [12].	34
2.14	Virtual hand operating in VR, controlled through motion-capture glove [13].	34
3.1	Freiburg lab environment with a preplanned trajectory for Experiment 1. .	41
3.2	The two map setups used in Experiment 2. (a) Small map top view. (b) Large map top view. (c) Small map perspective view. (d) Large map perspective view. (e) Scanned small map. (f) Scanned large map.	43
3.3	Experiment 2 setup.	44
3.4	Experiment 3 real-world platform: (a) The Neobotix MP-700 mobile robot [14]; (b) Robot and information diagram.	45
3.5	Experiment 3 setup. (a) Current real-world view of the robot. (b) Updating the map with the latest sensor observation (robot position indicated by frame annotation). (c) Newly detected frontiers (in red) in the constructed map.	45
3.6	Experiment 1 results. (a) The number of cells processed and evaluated by all algorithms. (b) The calculation time of the algorithms that consider the active area in relation to the maximum range of the simulated sensor (measured by the number of map cells).	46
3.7	Experiment 2 results for the small map case. (a) Average calculation time per iteration. (b) Calculation time as the small map is gradually explored for all algorithms.	46
3.8	Experiment 2 results for the large map case. (a) Average calculation time per iteration. (b) Calculation time as the large map is gradually explored for all algorithms.	47
3.9	Experiment 3 results. (a) Average calculation time per iteration with a horizontal line representing the updating rate of the map constructed by the Simultaneous Localisation and Mapping (SLAM) algorithm. (b) The calculation time as the map is explored for all algorithms.	47
3.10	Average calculation time in only the first iteration of each algorithm. . . .	48
3.11	Scanned map of Experiment 3.	50
4.1	The Active and Interactive Mapping cycle. The projection of GPIS mesh on the ground (red line) defines potential exploration segments, and the green bars indicate segments' information utilities. The Next Best View (NBV) is a red bar + arrow. Dark green dots are GPIS training points from the removed object [15].	54
4.2	Active and interactive mapping framework overview.	55
4.3	An illustration of full formulation and single factor utilities. First, identify pile segments (red contour) and compute the utilities (green bars). Max utility gives NBV (red bar+arrow). (b) Shows samples (blue dots) from the manipulability annulus. (c) Shows heights (orange bars). Grey bars on the ground indicate the uncertainty for imaginary segments.	56
4.4	Three Gazebo simulation scenarios.	57
4.5	Comparison of task and map coverage for simulation case 2.	60
4.6	Real-life experiment: comparing map accuracy between ours, [16] and [17]. .	61

4.7	Real-life experiment, active and interactive mapping cycle.	62
5.1	System 1 overview and its demonstration. (a) System 1 overview. (b) Simulation of an arm-equipped robot detecting the best brick for picking from a pile. Chosen brick is labelled with RGB-axes.	66
5.2	System 2 overview.	67
5.3	Principal Component Analysis (PCA) reveals the natural distribution of normally distributed data [18]. Here eigenvectors are labeled as \mathbf{u}_1 and \mathbf{u}_2 . The data mean is located at the centre of the surface. Standard deviations along each axis are the singular values $\lambda_1^{\frac{1}{2}}$ and $\lambda_2^{\frac{1}{2}}$. The ratio of length-to-height is $\lambda_1^{\frac{1}{2}} : \lambda_2^{\frac{1}{2}}$	71
5.4	Rotation of a vector, \mathbf{x} about the axis, \mathbf{u} by an angle, ϕ [19].	72
5.5	In unit 3-sphere manifold, quaternion \mathbf{q} defines an angle, $\theta = \frac{1}{2}\phi$ with a unit quaternion, \mathbf{q}_1 [19].	73
5.6	The five Experiment 1 scenarios.	82
5.7	Segmented point cloud results from Experiment 1.	83
5.8	Bounding boxes from the simulated experiments: (a), (b) individual objects experiment; (c), (d) multiple objects experiment with different cameras viewpoints.	86
5.9	System 2: Background removal and clustering process Gazebo simulated environment	88
5.10	System 2: Background removal and clustering process simulated environment view in RViz.	88
5.11	Results of the implemented ground removal and clustering processes. (a) The point cloud after ground plane removal through RANdom Sample And Consensus (RANSAC). (b) The point cloud after the removal of clusters below ground level and small clusters. (c) The position of clusters in (b) in the original point cloud.	89
5.12	System 2: Background removal and clustering process real-world experiments data set. (a) Dataset 1; (b) Dataset 2; (c) Dataset 3; (d) Dataset 4.	91
5.13	System 2: Background removal and clustering process real-world experiments result. (a)(b) Wall with minor roughness. (c)(d) Beach rock under sunlight.	91
6.1	Operator interacting with a remote robot in VR with the visual aid of processed point cloud data that is transmitted from RGB-D cameras. . . .	96
6.2	A flowchart of the lossless colour and depth image data compression and transmission process.	97
6.3	Point cloud reconstruction experimental results. (a) Raw point cloud data view in RViz. (b) Comparison of the reconstructed point cloud (white) and original point cloud (red).	98
6.4	Field of view comparison between depth and colour images.	99
6.5	Participant view in VR: (a) Participant view: colour point cloud (C-PC); (b) Participant view: mono-coloured point cloud (MC-PC); (c-d) Participant view: colour and depth image stream (C-D).	100

6.6	A participant assembling a puzzle while visualising it via RGB-D camera stream in a VR headset. Inset: The puzzle used in the experiment (left): Unassembled start configuration, (right) Successfully completed assembly.	101
6.7	Hardware block diagram showing data passed from a robotic arm with cameras to a remote operator controlling the system via a VR rig that displays the virtual robot and the transmitted perception data.	103
6.8	Participant's view in VR with differing settings: (a) Setting 1 - Point cloud only; (b) Setting 2 - Point cloud and two image streams; (c) Setting 3 - Point cloud and annotations; (d) Setting 4 - Point cloud, two image streams, and annotations.	105
6.9	A participant manipulating the end-effector to the green area of the target while in VR.	106
6.10	Experimental results: (a) Completion time; (b)-(d) for statements 1 to 3 the participants' responses on the Likert scale.	108
6.11	Experimental results for each setting: (a) Completion time; (b) Precision quantified score; (c) Histogram of repetitions where participants missed the green target area.	111

List of Tables

2.1	Example iteration times for Naïve frontier detection if cell evaluations take 1ms, 1us, or 1ns. The first two rows are the 2D areas of the “FR-079 corridor” and “Freiburg campus” data sets, respectively. The second two rows are the full 3D volumes of the same data sets, [20].	15
3.1	Properties and suggestions for using different frontier detection algorithms. Yes* is used to indicate where one algorithm performs markedly better than the others for a particular property.	49
4.1	Ablation analysis on Utility factors.	59
4.2	Benchmark Test	59
5.1	A comparison of the centroid estimation accuracy between the Iterative Closest Point (ICP)-integrated system and System 1.	84
5.2	A comparison of the orientation estimation accuracy between the ICP-integrated system and System 1.	84
5.3	A comparison of execution time (in seconds) between the ICP-integrated system and System 1.	84
5.4	A comparison of the brick’s PCA ratio with size ratio. $(\frac{L}{H})^2$ refers to the length-to-height ratio squared, and λ_i refers to the i ’th eigenvalue from PCA.	85
5.5	Bounding boxes centre and error values	87
6.1	Summary of experimental design information for the two user studies.	99
6.2	Experimental trial visualisation settings.	100
6.3	Experimental visualisation settings (i.e., images, coloured point cloud (PC) and bounding box annotations) indicating which information is available “A” or not available “N/A” to participants.	105

Acronyms & Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
AR	Augmented Reality
BFS	Breadth-First Search
DOF	Degree of Freedom
EWFD	Expanding-Wavefront Frontier Detection
FFD	Fast Frontier Detection
FOV	Field of View
FTFD	Frontier-Tracing Frontier Detection
GDP	Gross Domestic Product
GP	Gaussian Process
GPIS	Gaussian Probabilistic Implicit Surfaces
GPOM	Gaussian Processes Occupancy Maps
HALO	High Access Localised Operations
ICP	Iterative Closest Point
IG	Information Gain
IoU	Intersection Over Union

LiDAR	Light Detection and Ranging
MBZIRC	Mohamed Bin Zayed International Robotics Challenge
NBV	Next Best View
PCA	Principal Component Analysis
PCL	Point Cloud Library
RANSAC	RANdom Sample And Consensus
ROS	Robot Operating System
SLAM	Simultaneous Localisation and Mapping
UTS	University of Technology, Sydney
VR	Virtual Reality
WFD	Wavefront Frontier Detector
WFD-INC	Incremental Wavefront Frontier Detector

Glossary of Terms

Autonomous	Without human intervention.
Active and passive perception	Active perception involves actively seeking and processing sensory information through exploration and interaction with the environment, while passive perception involves receiving sensory information without any active involvement on the part of the system.
Collaborative Autonomy	Refers to a mode of operation in which humans and robots work together as a team to achieve a common goal, with each entity contributing their unique strengths and abilities. This approach leverages the strengths of both humans and robots to enhance performance and efficiency beyond what either entity could achieve individually.
Human operator	A person who manipulates the robot's action at a remote location.
Human supervisor	A person who oversees the robot's operations and provides the robot with instructions when needed.
Manipulator	Robotic arm
Mobile base	A robot that exclusively navigates on horizontal planes.
Occupancy map	A discretised grid representing the environment is made up of cells with values that indicate the likelihood that a given point in space is obstructed.
Voxel	Volumetric Pixel represents a 3D cube-like volume in Euclidean space.