Title of project placed here

Name of author placed here

School of Computing Science
Sir Alwyn Williams Building
University of Glasgow
G12 8QQ

A dissertation presented in part fulfilment of the requirements of the Degree of Master of Science at The University of Glasgow

Date of submission placed here

Abstract

abstract goes here

Education Use Consent

I hereby give my permission for this project to be shown to other University of Glasgow students and to be distributed in an electronic format. **Please note that you are under no obligation to sign this declaration, but doing so would help future students.**

Name: _____ Signature: _____

Acknowledgements

acknowledgements go here sdfsdf

Contents

Chapter 1

Introduction

1.1 Sentence

I completed a program which allowed a turtlebot to autonomously map an environment using SLAM as well as create an inventory of items found in the environment.

1.2 Importance/Context/Motivation

Mobile robots are becoming more and more affordable and accessible which has allowed developers to take advantage of their applications in many different ways.

SLAM has it's application in rescue robots which http://www.aaai.org/Pressroom/Releases/release-02-0910.php these robots required realtime control and utilised only video streams to identify and rescue people.

This compares to this which utilises rugged mobile robots to create SLAM maps of mine shafts with minimal supersvision. This can then be applied to areas which are too unsafe/ small for humans to access. https://miningrox.informatik.tu-freiberg.de/en/

More affordable Drones can also be used to increasingly accurate create maps of property as well

Object detection can be used to detect humans via heat sensors etc. Aswell as identifying bombs etc.

1.3 Objectives/Hypothesis Karl Popper/Problem statement

1.4 Description of Objectives

1.5 How I achieved it

1.5.1 System Diagram

1.6 Outline of the dissertation

Chapter 2

Background

This chapter will provide research into the hardware and software required to run the turtlebot and perform autonomous SLAM mapping. Currently there is a selection of applications which already perform individual tasks required to achieve this project and thus it is a case of discussing and evaluating each component. The area of object detection and recognition is firstly considered in context of natural visual systems and then in software applications where a discussion of two seminal methods of detection are discussed.

2.1 Robot Operating System (ROS)

2.1.1 What is ROS

ROS is a meta-operating system which provides a collection of tools and conventions to aid the writing of robot software. This includes an abstraction of hardware and communications as well as tools including message passing, package management and low level controllers.

2.1.2 Why do people use ROS

ROS features open-source licenses and a large growing collection of packages which contribute to a vibrant ecosystem of developers and researchers working on applications. Some of these packages includes powerful applications such as the Gazebo simulator and the visualisation tool RViz as well as complex algorithms such as SLAM and drivers such as OpenNI.

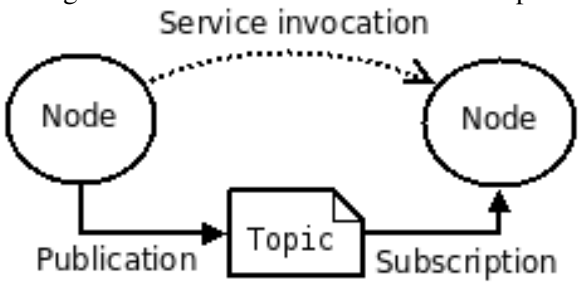
ROS encourages software engineering principles of loose coupling and abstraction which simplifies software integration and reuse as well as meaning that ROS is generally independent of hardware specifics. ROS features standard implementations in popular programming languages such as Python and C++ which increases adoption.

2.1.3 How does ROS works?

ROS consists of some key components within the communications infrastructure.

- Nodes are modular processes performing computation which can communicate with each other using topics or services. A system may consist of many nodes which perform many different tasks.
- Topics are named buses which provide a clear message passing interface between nodes. This is achieved through an anonymous publish/subscribe mechanism. This mechanism allows many-to-many transport.
- Messages are a data structure which are published by nodes to topics. They support strongly typed fields such as primitives or arrays and are either predefined or user-defined.
- The ROS core or Master provides a centralized node which locates and negotiates communications between nodes as well providing naming and registration services.
- Services forgo the many-to-many paradigm and provide a request/reply interaction via servers and clients.

Figure 2.1: Illustration of the ROS concepts



For example a simple robot vehicle system can consist of three nodes. The first node controls the robot's ultrasonic range finder and is publishing a stream of range messages taken from the sensor along the `sensor/sonar` topic.

The second node controls the robot's movement hardware and is publishing the robot's orientation on the `geometry_msgs/pose` topic. It also contains an actionlib server which responses to requests to change the robot's positioning.

The third node controls the robot's movement and is subscribed to the `sensor/sonar` topic and contains a client of the actionlib server. While processing the range messages if the range indicated is very small the node can use the actionlib service to request an adjustment in the orientation of the robot thus meaning the robot will avoid collisions.

Include diagram describing example

RViz

Rviz is a visualisation tool in ROS which provides 3d visualisations of sensor data and robot states, such as cameras, lasers and joint states. Add modular componets which can be customized.

Gazebo

Gazebo provides a simulated environment to build and test applications in ROS. A simulated robot will feature all the same sensors as a real one and a model world can be created. Physics, models, programmable features such as doors orlights.

2.2 Mobile Robots

Mobile robots are robots that are not fixed to one location and are capable of moving around their environments. Movement is achieved generally through legs, wheels or tracks but aerial and nautical robots can use propellors. The robot can be controlled through manual-tele operation involving a human driver, line-following which follow visual cues such as painted lines to navigate or autonomously. These robots have many applications in areas such as military, agricultural, rescue, transport and domestic usage and include Unmanned Aerial Vehicles (UAV or commonly drones), rescue robots and self-driving cars.

2.2.1 Turtlebot 2

The Turtlebot 2 is a part of a series of low cost, open-source mobile robots by Clearpath Robotics**. These robots are intended for educational and research purposes and provide an 'low-barrier-of-entry' platform for development. The robot consists of a Kobuki base, a computer running ROS and some form of RGB-D or Stereo Camera such as the Xbox Kinect or Asus Xtion.

Developed by ... for the intended purpose...

The Kobuki Base contains IR, bump, cliff and wheel drop sensors as well as different power connectors and ports.

The Kobuki Base requires a computer running ROS to communicate with, in most cases this can simply involve a laptop attached to the top of the base connected by a usb cable. However in some cases it may be more convienent to attach a small netbook or Raspberry Pi to the base and use a seperate computer to communicate with the host.

Why use the turtlebot?

- Lots of software standard packages support, gazebo models, teleoperation, slam as well as large and growing community of stuff... - Abstracted hardware, no need to understand how it works, just recieve the data,

How it works in ros

Odometry

2.3 Cameras

Why do we need them Why discuss both of these? - exemplar

2.3.1 RGB-D

- How it works
- Active camera - range finder
- Kinect specs

2.3.2 Stereo

- Passive camera, works well outdoors
- How depth is calculated
- How images are combined
- About the zed cam - specs

2.4 SLAM

- What is the problem, include references e.g. to correlation
- explain the proble
- solutions to the problem

SLAM is the problem of simultaneously localising (finding the pose and orientation) of a camera within it's surroundings at the same time as mapping the structure of the environment. This requires a robot or camera with the ability to produce odometry readings as well as camera with a range measurement device, either a range finder or found in a stereo camera.

SLAM forms the basis of navigation in most mobile robots, meaning unknown environments can be explored and mapped without the need for technology such as GPS which becomes innaccurate within two metres or in interiors. It has applications in a range of manned and autonomous robots. This includes UAV's, underwater robots and domestic robots such automatic lawnmowers. SLAM is a key component in the development of self-driving cars These cars which are driven along routes while performing SLAM capturing location, feature and obstacle data. Once the map is completed it is processed and the cars are driven autonomously along these routes updating the map as necessary.

This is considered to be a solved problem in Computer Science and their are many different approaches for finding a solution. [2]

During the development of the SLAM problem researchers established the correlation between estimates of a robots location and the landmarks which consituate the map

Building maps once can work, but it can be distrupted by changes in the environment, e.g lighting

We know the position of the landmarks Localisation requires landmarks, estimate the location of these and estimate it's own pose relative to it. This can be inaccurate due to odometry or inaccurate readings, rely on each. 95GEt estimate of where the robot probably is, get it's location relative to landmarks

Cannot fully decouple localisation and mapping as map is needed for localisation and a pose estimate is needed for mapping

Areas which are missing. Passive.

Given: Robot's controls, Observations(laser scans) Wanted: map of environment Path of the robot.

Probablistic techniques, uncertainity of robots motions and observations... represent uncertnity typically in a Gaussian distribution, error can accumalate over time. Take the uncertanity into account.

Slam is difficult because both path and map are unknown, but they are correlated and depend on each other

On a high level SLAM is solved by using the environment to update the pose of the robot. Using odometry as the sole measurement of localisation has an element of uncertainity due to extraneous factors such as wheels slipping on different environments meaning that a stated distance given by an odometry reading may be over or under estimated. Therefore laser scans or other forms of depth readings are used to correct the robot's position by extracting features from the surrounding environment. These are called landmarks and can be extracted by various methods such as Random Sampling Consensus (RANSAC) and provide a growing map of the enviroment.

The robot estimates the location of it's own pose relative to these different landmarks which are all correlated together and increase with successive observations. [1]

The combined localisation and mapping estimates as a single estimation problem are convergent

Extended Kalman Filter is used to update where the robot presumes it is based on these landmarks and keeps track of the uncertainity of the robots position and the landmarks seen.

When robot moves the uncertainity of the new position is updated. Landmarks are extracted and reobserved ones are used to update the robots position. New landmarks are added to the EKF.

***Diagram. of sequence of reaffirmation.

Uncertanity is kept track of.

2.4.1 RtabMap

- intro RTAB-Map is a RGB-D graph based SLAM approach. What it is

- Loop closure technique. How it works

Online slam, seeks to recover only the most recent pose, not the entire path algorithm

Localisation of a known map can be achieved with similar to orientation flight on animals insects -

2.4.2 Others

Compare rtabmap and other papers Cartographer

2.5 Frontier Exploration

2.6 Object Detection and Classification

2.6.1 Introduction

Object detection is the task of finding the different objects in an image and classifying them

Within Computer Science object detection refers to finding real-world objects within images or videos. Alongside robotics, such technology has applications in areas such as surveillance, medical image analysis and human computer interaction. It is useful to think of object detection software in terms of an intelligent agent which is an autonomous entity that gathers and observers information through sensors and acts upon an environment using actuators to direct its activity to achieve goals.[**?**] Why?

Understood in natural world with particular emphasis on lower organisms.

2.6.2 Detection in the natural world

It is easy to assume that an eye is comparable to a camera in that it streams data through the lense. Many organism's vision systems have much less resolution, sensitivity and field of vision than a modern camera yet are able to perform incredibly complex tasks successfully.

Modelling a intelligent agent that interprets objects in scenes requires an understanding of natural visual systems. For most systems while everything is seen for the first time our senses do not keep telling us things we already know, this is because between scenes most important environmental information stays constant. This process is a form of sensory adaption where perception is temporarily changed when exposed to new stimuli in an attempt to normalise visual experience.[3] This happens at a retinal and neural level, where information provided by past experience have a

greater say on how a scene has been interpreted than immediate information provided by external organs.

Contrast human vs lower

For example an ellipse projected into a human retina can be interpreted as a circle due to our experiences with perspective.

Many species exhibit a form of pattern or feature detection to trigger neural responses to visual changes. In contrast to humans which detect generic images features such as shapes many lower organisms utilise goal-based feature detection, this is useful to the field of computer science because it allows us to model intelligence agents on such processes and tend to engage in one activity at a time using only one way of doing it [] while achieving complex goals and tasks.pg 17 easier to model intelligence agents.

For example arthopods such as honey bees readily distinguish features such as flowers in the environment that pertain to the goal of gathering food. Similarly a frog's eye is stimulated when a black disc moves in an arc rapidly within the receptor field indicating that a flying insect is near and triggers the frogs feeding response [] thus satisfying the goal of feeding.

It is important to note that these examples all feature a form of active vision whereby the eye will "actively acquire the visual cues needed to cope with a particular task" pg 17 in the case of a human the fovea will align with the area of interest in the environment, comparable to something like a pan and tilt surveillance system which focus on areas of interest.

2.6.3 Extracting Features

What is a feature? Somewhere!

In contrast to higher organisms Lettvin et al. [] determined that a frog's retina has four classes of ganglion cells which specialize in extracting features from a visual signal.

- edge detectors for borders between light and dark areas
- moving edge detectors
- dimming detectors
- convex detectors that react when a black disc is in the field of vision

Image processing software has a similar definition of feature of extraction but in most cases they are generic. Explain

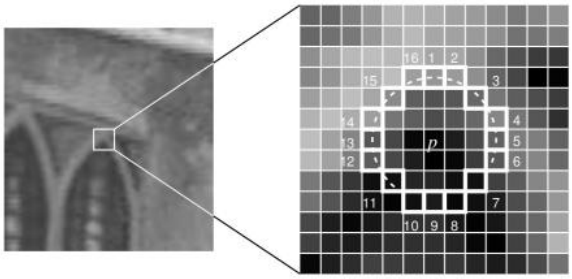
FAST Feature Detection

There are many different software applications of edge and corner detection, this includes SIFT (Scale-Invariant Feature Transform) which performs extremely well in most contexts. [**?**] However for applications in real-time image processing such as SLAM and Augmented Reality it is often not

fast enough. FAST (Features from Accelerated Segment Test) [**?**][**?**] avoids the costly difference of Gaussians (DoG) method found in SIFT and is subsequently considerably faster.

Using an appropriate threshold T which is usually 12, the algorithm selects a pixel P which is the centre of a circle with a circumference of 16 pixels n and the pixel's p intensity is Ip. The pixel P is determined to be a corner if n are all brighter that Ip + t or are all darker than Ip - t.

Figure 2.2: Illustration of the FAST



If a threshold T of 12 or greater is used a high speed test is performed to reject a large number of non-corner points which involves examining the pixels at 1, 9, 5, 13. If either pixel 1 or 9 are brighter or darker than T then 5 and 13 are checked. If three of these pixels are either all darker than the threshold or all brighter than p is a corner.

software appliations involving fast

2.6.4 Object Classification

However in contrast to simple organisms whose focus is on detection resulting in a sensitive and bespoke! and broad field of vision the programme must have the ability to classify which requires precise resolution when required.

Not feasible to search for objects at all locations in a visual field Partioning or perceptual organisation.

Law of proximity = Stimulus that are close together are percieved to be a group.

Geometric, Photometric modeling scene degmentation, naming+labeling,

Image matching: correlation approach

feature matching approach eges remain across two images of the same scene. relational matching approach

From page 281 –¿ talk about image labeling

Deep Leaning

Tensorflow

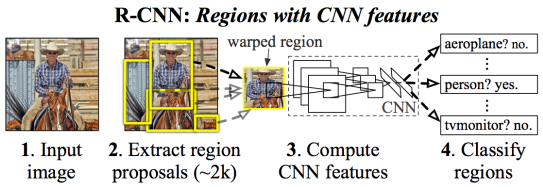
Haar Cascades

Google Vision API

2.6.5 Regions with Convolutional Neural Networks (R-CNN)

Detection systems utilise classifiers to evaluate potential objects. Yet there is know standard way of localising objects within an image. One such approach is called Regions with Convolutional Neural Networks (R-CNN) and is considerable more efficient than previous approaches and at the time of it's release, R-CNN had the best detection performance on the PASCAL VOC 2012 image dataset. In the paper "Rich feature hierarchies for accurate object detection and semantic segmentation"[**?**] the authors suggest a method which involves taking an image and identifying objects using bounding boxes and then performing a classification on these areas to create a label for each object.

Figure 2.3: Illustration of the process stages found in R-CNN



Roughly two thousand bounding boxes, or region proposals are created using the process of Selective Search which performs a segmentation algorithm that groups regions together by color, intensity or texture.[**?**] This is in contrast to using an exhaustive sliding window approach found in Deformable Parts Models.[**?**] Each selected region is warped to a 227 x 227 square RGB image and fed through a convulutional neural network (CNN) which computes features. The final stage involves running a Support Vector Machine (SVM) on the feature vector of each region classify and score the object within the region (if any). Greedy non-maximum suppression is applied to merge the regions which share the same object resulting in accurate bounding boxes for each object.

This method is slow as it requires the CNN to be run on every region created for the image.

2.6.6 Attention Systems - Memory Hierachy

Spatial temporal Attention Models

Neurophysiologic studies have shown that, in humans, these two factors are the main responsible ones to drive attention. Bottom-up factors emanate from the scene and focus attention on regions whose features are sufficiently discriminative with respect to the features of their surroundings. On the other hand, top-down factors are derived from cognitive issues, such as knowledge about the current task. http://www.prip.tuwien.ac.at/people/krw/more/papers/2012/Antunez2012a.pdf

How long do organism retain visual information.

How spatial temporal reasoning - don't overload memory - Have a Buffer - reaffirm spatial temporal Attention Memory Hierarchy - 9 seconds, How long? - What is long term? - What is short term? - What is important - How to decide - Intelligence agents

2.6.7 Summary and Discussion

What currently exists for completing the objectives

What are the Gaps, What do I have to do to fill the gaps

What methods are not appropriate to fill the gap

Chapter 3

Approach/Implementation/System Implementation

3.1 Requirements

3.2 Anaylsis the approaches

What problems I encounted etc.

3.3 Analysis of approaches

3.4 How did I go about it

3.5 Mapping

3.5.1 Transforming data

3.5.2 Calibration

3.6 Frontier Exploration

Hybrid approach taken from RCNN, currently only has an official matlab binding of RCNN.

Do not use bounding boxes, as not entirely useful in mapping as this requires new detections and classifiations when changing the object pose. A simple sspherical location detector is used to show the object is located within the this region of alpha

3.7 Object Detection and Recognition

3.7.1 Detecting ROI in FOV

Masking

Depth Mask

HSV Mask

3.7.2 Detecting Objects Clustering - different methods

3.7.3 Creating Boxes with DBScan

3.7.4 Tracking Boxes

3.7.5 Recognising Objects

Using google vision means that speed does not decreae with more objects detected, statys constant

3.7.6 Publishing to Rviz/Rtabmap

3.8 The whole package - how to utilise

I did this because of that which allows me to do this

Chapter 4

Evaluation

4.1 Testing

https://www.koen.me/research/pub/vandesande-iccv2011.pdf

Experiment 1 Question Experiment 2 Experiment 3

I think by doing this I will achieve... Experiments to answer questions

Chapter 5

Conclusion

5.0.1 Future work

incorporate active detection of features next time.

Appendix A

First appendix

A.1 Section of first appendix

Appendix B

Second appendix

Bibliography

[1] Hugh Durrant-Whyte. Uncertain geometry in robotics. *IEEE Trans. Robotics and Automation*, 1988.

[2] Hugh Durrant-Whyte, Fellow, IEEE, and Tim Bailey. Simultaneous localisation and mapping (slam): Part i the essential algorithms, 2006.

[3] Michael A. Webster. Visual adaptation. *Annual Review of Vision Science*, 1(1):547–567, 2015. PMID: 26858985.