# Enabling Human-Robot Cooperation in Scientific Exploration of Bandwidth-Limited Environments

by

Stewart Christopher Jamieson

B.A.Sc., University of Toronto (2018)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics and Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY AND WOODS HOLE OCEANOGRAPHIC INSTITUTION

May 2020



Author .....
Joint Program in Oceanography/Applied Ocean Science and Engineering
Massachusetts Institute of Technology & Woods Hole Oceanographic Institution
May 19, 2020

Certified by .....
Yogesh Girdhar
Associate Scientist in Applied Ocean Physics and Engineering, WHOI
Thesis Supervisor

Certified by .....
Jonathan P. How
Richard Cockburn Maclaurin Professor of Aeronautics and Astronautics, MIT
Thesis Supervisor

Accepted by .....
Sertac Karaman
Associate Professor of Aeronautics and Astronautics, MIT
Chair, Graduate Program Committee

Accepted by .....
David Ralston
Associate Scientist with Tenure in Applied Ocean Physics and Engineering, WHOI
Chair, Joint Committee for Applied Ocean Science and Engineering

**Enabling Human-Robot Cooperation in Scientific Exploration of Bandwidth-Limited Environments**

by

Stewart Christopher Jamieson

Submitted to the Department of Aeronautics and Astronautics
on May 19, 2020, in partial fulfillment of the
requirements for the degree of
Master of Science in Aeronautics and Astronautics

## Abstract

Contemporary scientific exploration most often takes place in highly remote and dangerous environments, such as in the deep sea and on other planets. These environments are very hostile to humans, which makes robotic exploration the first and often the only option. However, they also impose restrictive limits on how much communication is possible, creating challenges in implementing remote command and control.

We propose an approach to enable more efficient *autonomous* robot-based scientific exploration of remote environments despite these limits on human-robot communication. We find this requires the robot to have a spatial observation model that can predict where to find various phenomena, a reward model which can measure how relevant these phenomena are to the scientific mission objectives, and an adaptive path planner which can use this information to plan high scientific value paths. We identified and addressed two main gaps: the lack of a general-purpose means for spatial observation modelling, and the challenge in learning a reward model based on images *online* given the limited bandwidth constraints.

Our first key contribution is enabling general-purpose spatial observation modelling through spatio-temporal topic models, which are well suited for unsupervised scientific exploration of novel environments. Our next key contribution is an active learning criterion which enables learning an image-based reward model during an exploration mission by communicating with the science team efficiently. We show that using these together can result in a robotic explorer collecting up to 230% more scientifically relevant observations in a single mission than when using lawnmower trajectories.

Thesis Supervisor: Yogesh Girdhar
Title: Associate Scientist in Applied Ocean Physics and Engineering, WHOI

Thesis Supervisor: Jonathan P. How
Title: Richard Cockburn Maclaurin Professor of Aeronautics and Astronautics, MIT

# Acknowledgments

None of this work would have been possible without the guidance and support of my co-supervisors, Dr. Yogesh Girdhar and Prof. Jonathan How. They have been exceedingly generous in giving their time, encouragement, and advice in matters both technical and related to professional development. I feel very grateful for the combined breadth and depth of their backgrounds, experiences, and technical knowledge, as this has been instrumental in allowing me to pursue such a cross-disciplinary and complex topic as co-robotic scientific exploration.

I would next like to thank all of my collaborators and peers involved in the research and discussion of the work and ideas presented in this thesis. My fellow members of the WARPLab and the Autonomous Controls Lab always created a positive, supportive, and productive atmosphere to work in, one I missed when working from home due to current world affairs. I'd like to particularly thank Vv, John, Genevieve, Victoria, Mike, Kaveh, Kasra, and Kevin for many long discussions and useful advice on the topics and ideas presented in this work. I'd also like to thank Nathan, Brian, and Stefano for their invaluable technical advice and support. Alongside them are all the others at MIT, WHOI, and in the MIT-WHOI Joint Program who have been excellent friends to me since coming to MA and made me feel at home in a new country.

Lastly, I would like to thank my family. In particular, my partner Victoria who provides me with constant support and inspires me with her incredible work ethic and dedication, as well my mom and dad and my brother Chris, whose love and encouragement have been ever-present and invaluable. I'd also like to thank Suzy and the rest of my extended family, who have likewise been major sources of support in both the best and worst of times. Finally, I dedicate this thesis in memory of my grandparents Chris & Molly, who were each deeply loved and are deeply missed; they always brought out the very best in me.

**Funding:** This work was partially supported by the National Science Foundation (NSF) Award #1734400, as well as by the Woods Hole Oceanographic Institution (WHOI). The author would like to thank both organizations for their support.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

> *"We shall not cease from exploration, and the end of all our exploring will be to arrive where we started and know the place for the first time."*
>
> – T.S. Eliot's *Little Gidding* [28]

The works presented in this thesis are all motivated by a single question: how can a team composed of humans and robots work together to explore new environments as effectively as possible even when it is difficult for them to communicate with each other? We are specifically interested in effective *scientific exploration*, where the success of an expedition is measured by the accumulated scientific value of the observations collected, rather than by the amount of area covered. Furthermore, we focus on teams which are structured such that the human scientists are located far away from where the actual exploration is being conducted by the robots. In fact, we are foremost interested in the case where exploration is happening in such a remote and inaccessible location that there are restrictive limits on how much the robots can communicate with the humans, and even among each other. This matches the structure and limitations of most modern scientific expeditions to the depths of Earth's oceans and to other planetoids in our solar system.

This chapter will provide some background in scientific exploration, leading into the structure of a modern expedition. It will also describe some of the capabilities of current robotic explorers as well as their relevant limitations.

## 1.1 A Brief Overview of Scientific Exploration

Scientific exploration is the pursuit of knowledge by travelling to observe the diversity of nature and the physical, biological, and geological activities of the natural world. In more concrete terms, it is about planning and making observations in places where we have significant *uncertainty* about what will be observed. Contemporary examples include probing Martian soil to look for hints of water [54], or searching for fish near the seafloor of the Gulf of California (an environment thought to be too hypoxic for fish to survive) [37]. These experiments were each accomplished using robots, and their results had major impacts in the scientific community. Scientific exploration is integral to the scientific method in many fields, such as (astro-)biology and (astro-)geology, because discovering the presence or absence of certain natural phenomena in specific environments can be key to supporting or discrediting scientific theories and models.

The prevalence of robots in modern scientific exploration is a consequence of humanity's boundless curiosity. Scientific exploration was born out of the transition from the Age of Discovery (an era of exploration) to the Age of Enlightenment (an era of science) which occurred in Europe at the turn of the 18th century. While most previous exploration had been driven by individuals in pursuit of fame and fortune, by the late 18th century many astronomers, physicists, biologists, ecologists and geologists would set out on global expeditions to collect the scientific observations necessary to prove new theories about everything from the shape of the Earth to the origin of species [21, 102]. Over the centuries since, the scientific community has collected detailed observations of most of the environments accessible to humans with modern technology. The only environments remaining to be explored are very hostile to humans, but our thirst for knowledge pushes us to reach beyond our grasp. Thus, the last few decades have seen the creation and rapid development of robotic explorers as our surrogates at the forefront of scientific exploration on Mars [31], in deep space [38], the Earth's oceans [4, 16, 36], and under Arctic ice sheets [105]. One commonality between these environments is how extremely difficult it is to communicate with robots operating within them; this presents some unique challenges we will address.

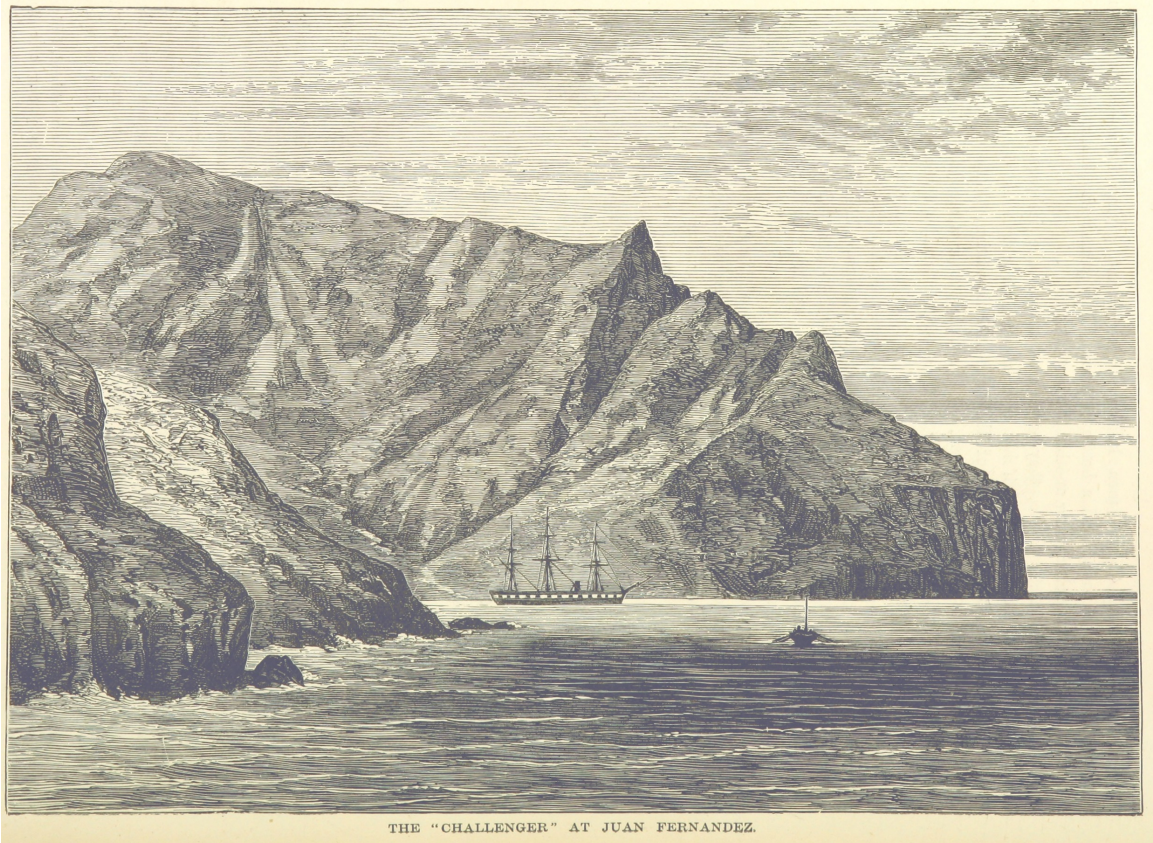


Figure 1-1: A depiction of the HMS Challenger (centre) at the Juan Fernández Islands, located in the South Pacific Ocean, in November of 1875, about 3 years into the eponymous Challenger Expedition [104].

## 1.2 Scientific Exploration in the Ocean

Oceanography is the study of the physical, chemical, and biological aspects of the Ocean, the largest and yet one of the least studied environments on Earth. As such, oceanographers are possibly the greatest modern examples of scientific explorers. While ships have been used in exploration for thousands of years, one of the first major attempts to study the Ocean itself was the Challenger Expedition of 1872-76. The expedition resulted in the discovery of, among many other things, approximately 4700 new species of plants and animals [46]. Almost 150 years later, nearly 2000 new species are discovered in the Ocean each year [51], even while humanity's capability to explore the oceans' depths is still very limited.

A great deal of my research is conducted at the Woods Hole Oceanographic


Figure 1-2: Photo by Mitch Elend showing the deployment of AUV Sentry from the R/V Atlantis. © Woods Hole Oceanographic Institution, 2019 [27].

Institute (WHOI), based in Woods Hole, MA. WHOI is one of the few organizations that owns vehicles capable of diving to depths more than 2000 metres. Most of these are unmanned underwater vehicles (UUVs) which can be further classified as either remotely operated vehicles (ROVs) or autonomous underwater vehicle (AUVs) [66, 107]. The paucity of manned deep sea submersibles is mainly due to the challenges and costs of their construction, as well as their operators requiring years of training and safety certification. Accordingly, UUVs (i.e, robots) are the main workforce in most deep sea exploration pursuits. As seen in Figure 1-2, UUVs are typically deployed into the Ocean from a research vessel by a team of scientists and engineers. The vehicle is later recovered, usually by the same ship, and the data collected is then downloaded.

### 1.2.1 Communication Capabilities & Limitations

An ROV must remain tethered to its parent ship by a cable that provides data transmission and power and enables remote control by a human pilot [66]. This tether limits the vertical and lateral distance the vehicle can move from the parent ship, and the complexity of tether management (i.e. preventing the tether from becoming

Table 1.1: The maximum theoretical communication bandwidth of the WHOI Micromodem for a variety of acoustic transmission modes [103].

| Rate | Frames Per Packet | Frame Size (Bytes) | Packet Size (Bytes) | Packet Rate (bps) | Time to Transmit 100 KB (s) |
|---|---|---|---|---|---|
| PSK-1 | 3 | 64 | 192 | 498 | 200.0 |
| PSK-2 | 3 | 64 | 192 | 520 | 192.3 |
| PSK-3 | 2 | 256 | 512 | 1223 | 81.8 |
| PSK-4 | 2 | 256 | 512 | 1301 | 76.9 |
| PSK-5 | 8 | 256 | 2048 | 5388 | 18.6 |
| PSK-6 | 6 | 32 | 192 | 490 | 204.1 |

tangled) means that typically only one ROV may be deployed at a time. These factors severely limit the productivity of ROV-based deep sea scientific exploration missions, which can cost up to $45,000 USD per day [89].

The only tetherless means of long-range underwater communication is acoustic communication, but acoustic data transmission rates are very low.[1] Details on the performance of the WHOI Micromodem are given in Table 1.1; other modern commercial acoustic modems perform similarly [96]. In practice, deep sea vehicles achieve a much lower effective data transmission rate than shown in the table due to the packet loss associated with transmitting an acoustic signal over long distances (kilometres). As the speed of sound in water is a little less than 1500 m/s [48], the round-trip communications latency is on the order of tens of seconds. The high latency and low rate of acoustic data transmission makes un-tethered remote control of deep sea vehicles infeasible for scientific exploration.

Unlike ROVs, AUVs are equipped with onboard power supplies and autonomous navigation systems, and can thereby explore without a tether or direct human guidance. AUV missions can last anywhere from hours, in the case of actively actuated vehicles such as Sentry (Figure 1-2), to weeks or months in the case of passively actuated vehicles like gliders [98]. In order to monitor the status and location of these vehicles,

[1] Optical modems are tetherless and capable of high-speed data transfer, but they are limited to ranges less than 200m [86].

they are typically equipped with some acoustic localization and communications device. These devices may also be used to occasionally send images back to ship, at a frequency dependent on the acoustic communication bandwidth and battery capacity (since acoustic communication requires significant energy).

### 1.2.2 Path Planning

While deep sea AUVs are very complex modern machines, their high-level autonomous behaviour is usually limited to moving through a sequence of pre-programmed waypoints. These waypoints are typically specified in coordinates of latitude, longitude, and elevation relative to sea level (i.e. depth), and usually form a familiar lawnmower pattern like the one in Figure A-1. This Boustrophedonic coverage is the most efficient way for an AUV to exhaustively search a pre-specified area [14]. However, exhaustive searches are highly inefficient when the goal is to collect observations of spatially sparse phenomena, or non-stationary phenomena such as animals. These are often the phenomena that are most scientifically valuable to observe.

### 1.2.3 State Estimation & Control

The ability of AUVs to track a trajectory is limited by the accuracy of their sensors used for state estimation, which may include acoustic rangefinders, multibeam sonar, Doppler Velocity Logs (DVLs), Inertial Measurement Units (IMUs), Fibre Optic Gyroscopes (FOGs), and cameras for visual odometry [5, 88]. SLAM systems have been developed using some of the sensors commonly found on deep sea AUVs [59]. The controllers used to execute the planned path, given the current state estimate, are typically simple but well tuned and effective. The control bandwidth does not need to be high since AUVs used in scientific exploration are usually slow-moving and can see static obstacles from afar using sonar, while most dynamic obstacles, such as fish, tend to avoid them.

### 1.2.4 Scientific Sensing

The sensor payload of an AUV is very dependant on the science goals of the mission. Commonly, an AUV is equipped with a standard complement including a pressure sensor (to measure depth), external thermometer, and conductivity sensor (to measure salinity). For many modern missions, the most important sensors are cameras because most geological and biological phenomena are best observed through vision. Furthermore, the deep sea is highly dynamic and poorly understood, so it is not unusual for a deep sea AUV to capture unexpected images with high scientific value (e.g., [79]).

The deep sea is almost completely devoid of light, so AUV camera systems are equipped with extremely bright lamps. Due to factors such as limited energy and processing power, images are usually captured at a rate of around $0.1 - 1$ Hz. This is still much greater than the rate at which the images can be streamed to the science team. As captured images are typically at least 2 megapixels (6 MB uncompressed), it can take a very long time to send one back to the ship via the acoustic modem. Thus, scientists usually need to wait until the AUV has docked with the ship in order to inspect its observations; by this point it may no longer be feasible to send the AUV back to an area if they find it glanced at something of scientific value.

## 1.3 Scientific Exploration in Outer Space

Robots have been essential to space exploration since humanity's first attempts to explore a celestial body that started a little more than 60 years ago. In fact, robots have visited far more planets and planetoids than humans have: they've landed on Venus, Mars, and Titan, the asteroids Eros, Itokawa, and Ryugu, and the comet 67P/Churyumov-Gerasimenko. Likely the most famous of these robots are the Mars exploration rovers: Pathfinder & Sojourner (1997), Spirit & Opportunity (2004), and Curiosity (2012). Each of these missions had a variety of unique objectives, but they all typically include scientific measurements pertaining to human habitability (e.g., the presence of liquid water), the detection of desirable resources, and searching for hints about the planet's history. Perhaps surprisingly, there is a great deal in common

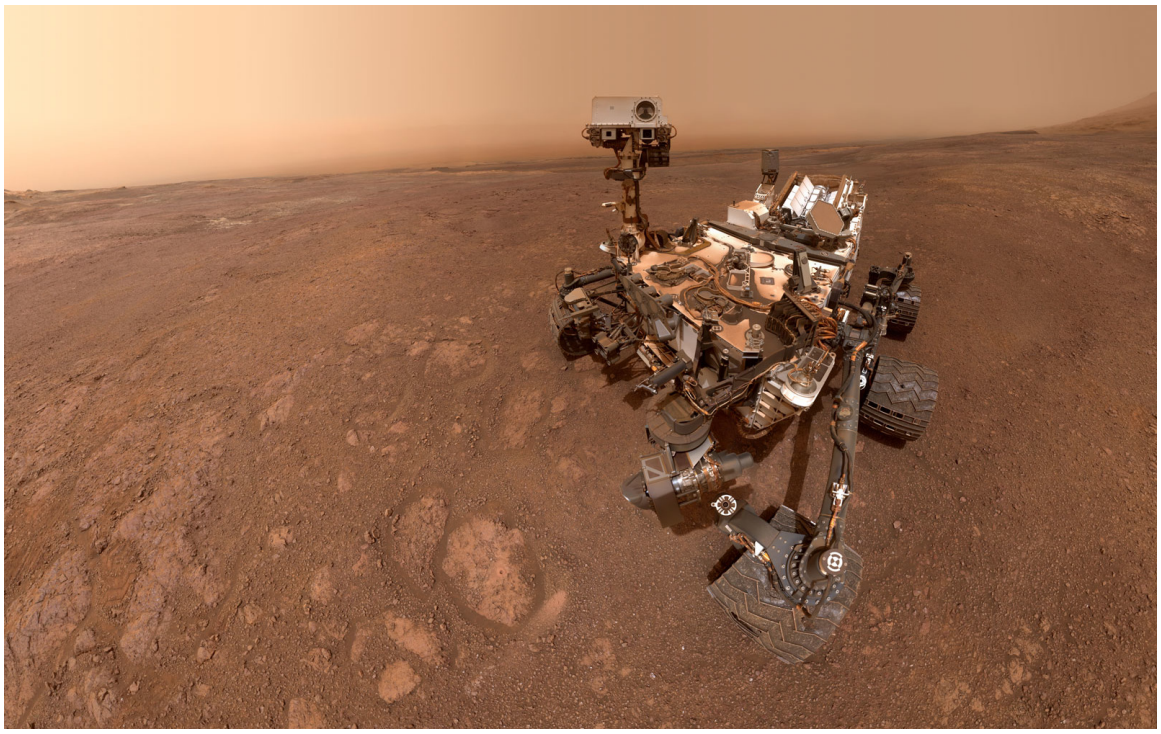

Figure 1-3: The Curiosity Rover at Rock Hall, Mars on January 29, 2019. The image was stitched together from selfies taken by a camera mounted at the end of rover's arm. Courtesy NASA/JPL-Caltech, retrieved from `https://mars.nasa.gov/resources/22273/curiositys-selfie-at-rock-hall/`.

between extraterrestrial and oceanic exploration.

As it is one of the most modern and technologically advanced robots in outer space, the following sections will focus on the capabilities and limitations of the Curiosity rover, a part of NASA's Mars Science Laboratory (MSL).

### 1.3.1 Communication Capabilities & Limitations

The Curiosity rover has two ways to send data back the scientists on Earth. The preferred method is to use Ultra-High Frequency (UHF) radio waves to communicate with one of the NASA's Mars orbiters, which will store the data and re-transmit it to Earth at the next available opportunity [63]. The orbiters are capable of high transmission rates due to their large antennas and power supplies [63]. The backup method is to transmit to Earth directly via its X-band radio, but due to power constraints the X-band radio is typically only used to receive commands from

Table 1.2: The return link (rover-to-Earth) data transmission rate of the Curiosity rover via either its UHF radio to the orbiters around Mars, or via its X-band radio to Earth directly [62, 63, 95].

| Band | Radio Frequency | Via | Min. Data Rate (bps) | Max. Data Rate (bps) | Min. Time to Transmit 100 KB (s) |
|---|---|---|---|---|---|
| UHF | 0.3 - 3 GHz | Odyssey Orbiter | 128,000 | 128,000 | 0.39 |
| UHF | 0.3 - 3 GHz | Mars Reconnaissance Orbiter | 500,000 | 2,000,000 | 0.07 - 0.2 |
| X-Band | 8 - 12 GHz | Direct to Earth | 500 | 32,000 | 3.12 - 200 |

Earth [63]. The data transmission rates of each method are presented in Table 1.2. Where the actual data rate falls between the minimum and maximum values in the table is mainly dependent on the current distance between Earth and Mars. However, the table does not tell the full story; the rover can only communicate with the orbiters over two 15 minutes periods per day, and the one-way communications latency between Earth and Mars varies from 4 to 24 minutes depending on their positions along their respective orbits [76]. Furthermore, no Mars-Earth communication is possible while the rover and orbiters are on the far side of Mars from Earth which is up to 8 hours per day [68].

In some ways, these communication capabilities mirrors AUVs; the X-band has similar data transmission rates to an acoustic modem (c.f. Table 1.1), while the UHF band offloads the latest batch of observations every several hours similarly to how data is offloaded from an AUV whenever it is docked. It is possible that future scientific exploration missions to Mars and other near planets could utilize more power-efficient X-band radios with similar bandwidth capabilities, enabling them to send back individual observations more frequently. Also, while communication between Earth and Mars can be (by some metrics) much better than between a ship and a deep sea AUV, exploration missions to the more distant planets and moons beyond the asteroid belt will likely have much lower communications bandwidth.

### 1.3.2 Path Planning

Curiosity does not choose its own destination; it only plans its own path to the next waypoint specified by an operator back on Earth [97]. Due to the limited communication windows, transmission bandwidth, and long communication latency, this severely limits the rover's productivity. Fortunately, the rover is capable of performing some basic *opportunistic science*: this refers to pausing the rover's present task execution to perform an autonomous action, such as targeting and using a particular sensor, after being triggered by a specific phenomena recognized by an onboard algorithm [12, 13]. For example, the OASIS system looks for Martian rocks with unusual characteristics, and whenever it finds them it preempts the robot's other activities to examine them in more detail [32]. Future missions will likely have more autonomous path planning, as recent research has found significant benefits in enabling *autonomous science*, wherein the vehicle is empowered to autonomously decide where to explore and how to use its various sensors in order to maximize the number of interesting scientific observations it collects [1].

Given a path, Curiosity is able to navigate with a configurable level of autonomy ranging from no autonomy ("blind driving", typically only used over short distances) to fully autonomous navigation including traversability analysis, slip detection and correction, and obstacle avoidance [60]. There has been a great deal of research into further developing the autonomous navigation capabilities of planetary rovers [106].

### 1.3.3 State Estimation & Controls

Curiosity primarily relies on wheel odometry from wheel encoders and visual odometry from its monochromatic stereo cameras for state estimation [60]. Curiosity also uses an IMU to correct its odometry measurements and to measure tilt [70]. As the rover's top speed is only 3.75 cm/s and there are no dynamic obstacles, the state estimation and controls problems can be adequately solved using only dead reckoning and PID control [6].

### 1.3.4 Scientific Sensing

Curiosity is equipped with a variety of scientific instruments, including a laser, several cameras, and a spectrometer [71]. Among these, the "Mastcam" is a general-purpose colour stereo camera capable of taking ~2 megapixel images or video at up to 10 frames per second [69]. This is similar in capability to a typical AUV camera, however it can achieve much higher frame rates because it does not need to supply its own illumination; all rover operations take place during Martian daytime, while the rover draws solar power from the sun. The Mastcam is one of the most useful sensors on the rover since it provides the first observation of every new phenomenon (usually an odd-looking rock) before it is decided whether that phenomenon should be studied in more detail using the other sensors.

## 1.4 A Taxonomy of Robotic Scientific Exploration

Technology roadmaps and taxonomies are useful for coordinating research efforts by listing the technological capabilities necessary to reach the long-term goals of a technology [67].[2] They focus researchers and industry by setting specific milestones, such as the SAE levels of driving automation [85]. In order to contextualize the current state of co-robotic scientific exploration, we present in Table 1.3 a six-level classification for the autonomous capabilities of robotic explorers. The table classifies robotic explorers according to their ability to model the world, understand the scientific mission objectives, and act to collect observations of scientific interest. For example, the AUV Sentry described in Section 1.2 meets the requirements for Level 1, while the Curiosity rover described in Section 1.3 would be classified as Level 2.

There have been designs for robotic explorers that can achieve Level 3 autonomy with certain limitations. For example, Arora, Fitch, and Sukkarieh [1] presented a robotic explorer that could autonomously explore a Martian-analogue environment to find rocks of scientific interest, however the rigid modelling requires substantial

[2]Examples include the ERTRAC Autonomous Driving Roadmap [30], the CCC and AAAI Roadmap for Artificial Intelligence Research [40], and the NASA Technology Taxonomy [72].

Table 1.3: Proposed Levels of Autonomy in Co-Robotic Scientific Exploration.

| Level | Name | Description |
|---|---|---|
| 0 | Remote Control | The robotic explorer:<br>• Has no spatial model to locate phenomena of interest<br>• Has no understanding of mission objectives<br>• Cannot move autonomously |
| 1 | Autonomous Navigation | • Has no spatial model to locate phenomena of interest<br>• Has no understanding of mission objectives<br>• Can navigate to a given waypoint without crashing |
| 2 | Opportunistic Science | • Has no spatial model to locate phenomena of interest<br>• Knows the scientific value of some specific phenomena<br>• Can navigate to a given waypoint without crashing |
| 3 | Autonomous Exploration | • Builds a spatial observation model of phenomena<br>• Knows the scientific value of some specific phenomena<br>• Can plan and execute high scientific value paths |
| 4 | Human-Robot Cooperative Exploration | • Builds a spatial observation model of phenomena<br>• Communicates with scientist to understand objectives<br>• Can plan and execute high scientific value paths |
| 5 | Multi-Robot Federated Exploration | In addition to the Level 4 description, multiple robots:<br>• Communicate among each other to disseminate observations and mission objective understanding |

scientist input into the development of the system and made it inflexible to unexpected observations or dynamic science mission objectives. Conversely, Hitz et al. [50] demonstrated a robotic platform that can also autonomously explore and efficiently achieve given science objectives, but their approach depends on the low-dimensional nature of their observation space (plankton concentrations), and does not scale to exploration based on images.

In other subfields of robotics, research is already exploring capabilities related to Levels 4 and 5. In particular, the intent recognition subfield investigates techniques to understand humans' objectives by asking them questions or observing their actions [2]. Recent research in multi-agent reinforcement learning has explored how multiple agents can teach each other task-relevant information and share observations with each other [75, 99]. Unfortunately, these approaches generally rely on having a large amount of training data and on frequent communication, neither of which are possible in the scientific exploration of remote environments.

### 1.4.1 The Need for Human-Robot Cooperation

At this point, it is worth asking why there is need for to go beyond Level 0 in Table 1.3: why should robotic explorers have any autonomy at all? The first reason is that lower levels of autonomy require far more work by the scientist to plan paths and guide the robot, when they could be working on other, more interesting tasks like analyzing data or writing papers. Scientists would rather have the robot operate mostly autonomously while still sharing relevant data with them in a timely manner, and ask them for their "expert advice" only as needed.

The next reason is that vision-guided remote control over a strictly limited communication channel is anywhere from highly challenging and inefficient to entirely infeasible. As discussed in Section 1.1, most modern scientific exploration is happening in places that are extremely remote and have very limited communication bandwidth. In fact, communication constraints are perhaps the biggest bottleneck to exploration in remote environments [10, 56]. Camera imagery is often the primary sensing modality, but transmitting images from the robot platform to the scientists running the mission

can be very expensive in terms of time and energy usage. As seen in Table 1.1 and Table 1.2, it can take several minutes to transmit even a very small image (~100 to 1000 KB) from a robotic explorer to its human operator, and this comes with a significant energy cost. Fortunately, robot state estimation, obstacle avoidance, and controls technologies can all be incorporated into the design of robotic explorers with relative ease, so most can navigate autonomously (i.e., are at least Level 1). However, the lack of communication creates a problem:

**Problem 1.** Since scientists do not have timely access to the robot's observations, then by the time they realize the robot passed by something of scientific value then it is often too late for the scientist to change the robot's path to better observe it.

Problem 1 highlights the need for at least Level 2 autonomy, where the robot can change its own path to better observe phenomena of interest. Level 3 is clear improvement from 2, since the robot is then able to autonomously plan high scientific value paths, making exploration much more efficient. But why is human-robot cooperation necessary, as opposed to stopping here? This is likewise related to the lack of communication making the following problem difficult to solve:

**Problem 2.** Planning a path with high scientific value requires having a "reward model" that defines the value of all possible observations according to the science objectives of the mission. However, *it is impossible to fully specify the reward model in advance of the mission.* Why? We can not define the scientific value of all possible visual observations *a priori* because there are an infinite variety of things the robot could find in a remote and novel environment, and many could have very high scientific value (e.g., a new species) or very low scientific value (e.g., plastic waste products).

As discussed in the previous sections, due to communication limits only the robotic explorer has access to most of the observations collected during the mission until the mission ends. On the other hand, only the scientist can make the final determination as to whether a class of observations are scientifically valuable. Therefore, as new phenomena are encountered during the mission the robot and the scientist must communicate with each other in order to maximize the scientific return of the mission.

It is also worth noting here that as more robotic agents are added into the system the amount of communication required would increase linearly. This would quickly become burdensome for both the scientist and the communication channel. This is part of what drives the need to achieve Level 5 co-robotic scientific exploration, where the robots help each other to take some of the burden off of the scientists. They can do this by efficiently disseminating the information received from the scientists among each other. It may also be possible for the robots to improve each other's performance by sharing other pieces of information that they collect.

## 1.5 Thesis Structure and Contributions

This thesis will primarily focus on developing the spatial observation models and human-robot communication strategies necessary to achieve robotic exploration autonomy Level 4: Human-Robot Cooperative Exploration. The next chapter will present the current state of the art in spatial observation modelling and in planning high scientific-value paths. In particular, it will explore what kind of reward model is necessary, and how a reward model can be learned using human-robot communication. The main contributions of this thesis are:

- **Spatio-temporal Topic Models as Spatial Observation Models:** In Chapter 3, we explore the usage of "spatio-temporal topic models" in co-robotic scientific exploration as a tool to mitigate Problem 1 by enabling online mission summarization, and next as a spatial observation model that can be used for autonomous exploration.

- **Online Active Reward Learning for Efficient Mission Objective Understanding:** In Chapter 4, we explore how a robotic explorer can successfully learn a reward model online in limited bandwidth environments by efficiently querying the scientist using a novel "regret"-based active reward learning approach to understand the mission objectives. We also provide some theoretical motivation behind this new approach and suggest future work in this area.

### 1.5.1 Statement of Originality

All of the work presented in this thesis was performed jointly with, and under the supervision and guidance of, my advisors, Yogesh Girdhar and Jonathan P. How. Chapter 3 mostly contains work performed jointly with Yogesh Girdhar, Levi Cai, Nathan McGuire, and the rest of the WARPLab, and is published in [42]. However, the discussion of spatio-temporal topic models as spatial observation models is based on my work published in [53], and my work on hyperparameter tuning is unpublished. Next, Chapter 4 mainly covers my own work performed jointly with my supervisors Yogesh Girdhar and Jonathan P. How, and most of the material discussed was published in [53]; the rest of the chapter is being prepared for inclusion in a new manuscript. Finally, Chapter 5 will provide a summary of conclusions as well as directions for future work. This future work includes taking the first steps towards Level 5: Multi-Robot Federated Exploration through an investigation into multi-robot semantic label association led by me but performed jointly with Kaveh Fathian and Kasra Khosoussi.

# Chapter 2

# Background

As discussed in Section 1.4, the three main capabilities needed to enable higher levels of autonomy in co-robotic scientific exploration are spatial observation modelling, human-robot communication for mission objective understanding, and the ability to plan high scientific value paths. Furthermore, the robot is required to apply these techniques in a severely bandwidth limited environment. This chapter will explore the state of the art in each of these areas and point out the gaps in the necessary capabilities that later chapters will seek to address.

## 2.1 Spatial Observation Modelling

Spatial observation modelling refers to being able to connect *what* the robot has observed to *where* the robot observed it. We can broadly describe systems for spatial observation modelling based on the following two capabilities:

1. The capability to *describe* previous observations, producing outputs like "I observed this species of fish near these coordinates..."

2. The capability to *predict* what may be observed in places not yet visited by extrapolating from previous observations, producing outputs like "I think I will find more corals near these coordinates..."

Planning high scientific value paths requires both of these capabilities: the first

capability is essential to connect observations to the mission objectives, while the second is essential to be able to compare the scientific value of different possible paths. For example, if the robot knows that observations of a particular coral species are scientifically valuable, it must know which observations contain that coral and where to find more observations of it.

The most common spatial observation models are Gaussian Processes (GPs) [81], which are very effective at modelling low-dimensional and continuous observations. As such, autonomous science techniques have been successful for planning adaptive robot trajectories when the observations are continuous scalar quantities, such as temperatures or concentrations. In fact, for these types of observations spatial models have been used in conjunction with informative path planning to enable efficient underwater scientific exploration [22, 34, 50], including using the AUV Sentry described in Section 1.2 [52]. Unfortunately, visual observations are very high dimensional, and GPs are highly ineffective and inefficient at predicting images. However, recent work has explored using modified GPs to spatially model the lower-dimensional *semantic representations of images* instead of the image observations themselves [87].

Most spatial observation models that work for visual observations are hand-designed by scientists for the specific phenomenon of interest. They often employ a machine learning based model trained to recognize such a phenomenon, such as a neural network based image classifier. While this is an effective technique, these models cannot be easily adapted to new phenomena and do not handle novelty well, in the sense that they will not recognize things they were not trained in advance to look for. This limits their applicability to general-purpose scientific exploration. As such, the remainder of this section will focus on ways to generate general-purpose semantic image representations that can be spatially modelled using (e.g.) a Gaussian Process.

### 2.1.1 Semantic Image Representations

In vision based scientific exploration, semantic image representations have been explored for years as a useful and communication-efficient tool for *mission summarization* (e.g., [35]). One class of representations are "semantic segmentations", which are

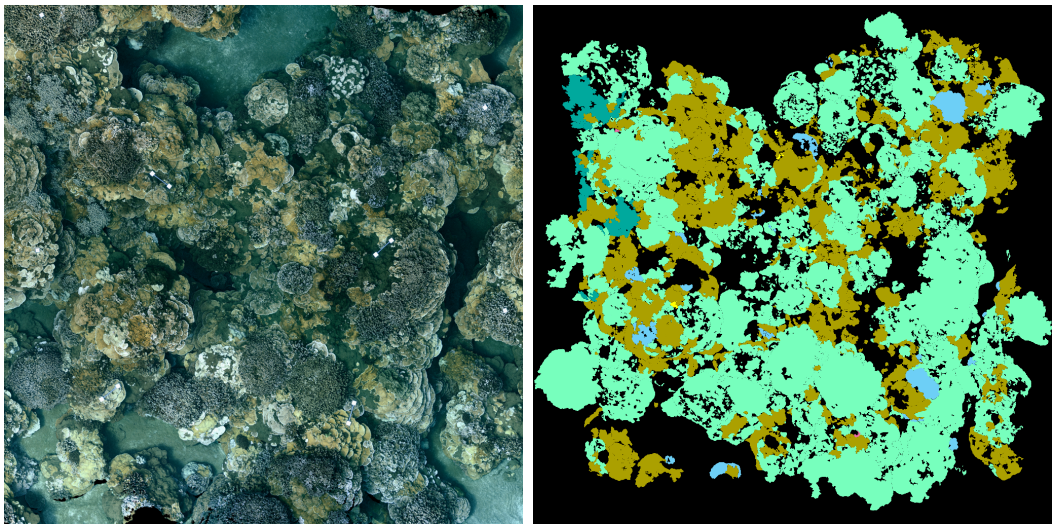

Figure 2-1: Left: A crop of the KAH_2016_3 photomosaic image from the 100 Islands Challenge [92], showing a coral reef near Kaho'olawe. Right: The photomosaic annotations where each color represents an expert label [92]. The image on the right is a semantic segmentation of the image on the left, where each species of coral is represented with a single label (colour). Like most semantic representations, the semantic image can be compressed much more efficiently than the original image; with lossless PNG compression, the semantic map can be transmitted using less than 12% of the bandwidth required to transmit the original.

images that use only a few distinct colours as labels for the semantically distinct parts of the original image. These semantic labels are consistent between images, and therefore provide a compact representation for describing one or more images. An example of a semantic segmentation is presented in Figure 2-1. There are many ways to generate semantic segmentations, including spatial topic models (e.g. [45, 101]) and deep feature extractors (e.g. [9, 24, 83]). When multiple images are collected over the course of the mission, their semantic representations can be combined to provide a mission summary.

Semantic image representations are much easier to spatially model than images because they tend to vary continuously. To be specific, while the pixels of two images taken near each other may be very different, their semantic content is usually very similar. For example, one may note that no two non-overlapping patches of the left image in Figure 2-1 are identical, but most nearby patches do contain similar species distributions (as seen in the right image). This is why it is much easier to

develop a spatial observation model for predicting what semantic representations will be observed in unvisited locations than one for predicting actual images.

Bayesian non-parametric (BNP) models are highly suitable for creating semantic image representations in scientific exploration. These models are typically unsupervised, meaning that they do not require training data, and they are capable of characterizing very complex inputs, such as natural images. Most of this thesis will use spatio-temporal topic models, a subclass of BNP models, to produce semantic image representations. Topic models were first developed almost 20 years ago for unsupervised understanding and organization of large text corpera.[1] Later works explored using spatial topic models for image understanding [11, 101]. However, these algorithms were relatively slow and computationally expensive, making them impractical to run on a robot. Over the last 6 years, efficient spatio-temporal topic models have been developed which can be run in realtime on a robotic platform [44, 45]. These will be discussed in much greater detail in Chapter 3.

**Aside: Online Mission Summarization**

As discussed in Problem 1, a major inefficiency in modern scientific exploration is that scientists do not have timely access to the robot's observations and are therefore unable to revise the robot's plan in response to an unexpected observation. This is an issue particularly when the primary sensing modality is vision because the communication bandwidth limitations of remote environments make sending images very difficult. One approach to this problem is to try to maximize the number of images which can be sent with the available bandwidth through efficient image compression [19], or through designing better communications technology. However, in many applications including scientific exploration, the rate at which data is collected already exceeds the capability of the owners of that data to analyze it. Thus, data compression or better communications are not by themselves a complete solution.

Online mission summarization is instead about summarizing the observations

[1] The original paper by Blei, Ng, and Jordan [8] showed topics models summarizing articles in the Associated Press news corpus as mixes of learned topics such as "Arts" and "Education".

collected during scientific exploration; *online* refers to creating and transmitting these summaries back to the scientists multiple times over the course of the mission. More abstract semantic representations require less bandwidth to transmit, and properly tuned semantic image representations are often easier for the scientist to understand and use. Figure A-2 presents a semantic timeline of a mission created using a spatio-temporal topic model, like the ones which will be discussed in Chapter 3.

## 2.2 Reward Model Learning for Understanding Mission Objectives

For a robot, understanding the mission objective means having an accurate "reward model". Given an observation (or the semantic representation of one), a reward model outputs the robot's prediction of the scientific value of that observation. Combined with a spatial observation model capable of predicting what will be observed in unvisited locations, this enables the robot to compare different paths and choose the one that it thinks will collect the most valuable observations. However, as discussed in Problem 2, it is not feasible for the scientist to provide the robot with a fully specified reward model in advance of the mission. Instead, the robot needs to learn parts of the model *online* (i.e., during the mission).

It is impossible to hand-design a reward model based on natural images because no meaningful scientific objective can be expressed as a simple function of the pixels in an image. Therefore, reward models for images are usually complex models such as deep neural networks, and trained on a dataset of examples using empirical risk minimization. Due to the high-dimensionality of natural images and the number of parameters in these models, it can take thousands or even millions of examples to learn even a simple classifier of natural images [49]. It is not feasible to collect this much training data for all of the phenomena that could be observed in scientific exploration.[2] It is especially not feasible to collect and label a large dataset online

[2]For example, a single coral reef explored by an AUV may be home to hundreds of unique species, and most of these species are rare and not well understood [25].

given the strict bandwidth limitations described in Chapter 1.

There are several techniques to train a machine learning model using a very small dataset of examples. We will focus on *active learning* and the advantages of finding *low-dimensional* semantic image representations, but also discuss potential techniques from *transfer learning*.

### 2.2.1 Active Learning

*Active learning* algorithms train a machine learning model by interactively querying an "oracle" to obtain the training labels most "helpful" for improving the model. This means that the model can be trained with far fewer labelled examples than would normally be required [3]. Active reward learning algorithms have efficiently learned reward models representing human ratings or preferences for robot behaviours by making on the order of 10-100 reward queries [20, 84]. However most of these and most other active learning objectives (e.g., those in [108]), assume that a large set of unlabelled training examples have already been collected, and the goal is to select a small subset to label and use to train the model for future usage in the same environment. This does not transfer well to scientific exploration because the unlabelled examples are collected sequentially online, can only be labelled one at a time, and the distribution of observations predicted to be collected along various potential robot paths may be very different than distribution of observations collected so far.

Doshi-Velez, Pineau, and Roy [26] explored using active learning *online* as a way for an autonomous robot to request help from its human operator if it was unsure about what to do next in order to maximize reward. They explored learning from policy queries, which consist of the robot asking the oracle (i.e., the human operator) to tell it the optimal action. This is a weaker oracle assumption than those made by previous works trying to solve the same problem [39]. However, it is still not practical for scientific exploration applications because while the scientist can accurately evaluate their own interests, they do not have enough information about the robot's previous observations or the robot's spatial model to be able to suggest the best next action.

The regret-based active learning approach we will present in Chapter 4 is similar in spirit to the approach used in [26], but focuses instead on identifying the information the scientist can provide that is most helpful in determining the optimal action (i.e., the path with highest scientific value).

### 2.2.2 Low-Dimensional Semantic Representations

Even when using active learning, it can take hundreds of queries to learn a reward model based on natural images [90]. This is partly due to the fact that the number of labelled examples that a model must be trained on in order to generalize well is proportional to the sample complexity of the model family [65], and for simple models the sample complexity is typically linear in the number of input dimensions [73]. For images, the number of input dimensions can be on the order of thousands to millions, but in bandwidth limited environments such as those discussed in Chapter 1 sending even just hundreds of examples for labelling during the span of a mission is typically not feasible. This motivates applying *dimensionality-reduction* techniques to the input images in order to reduce the number of input dimensions without discarding information key to learning the reward model. Fortunately, semantic image representations like those discussed in Subsection 2.1.1 are relatively low-dimensional, making them very helpful for learning new classification tasks with relatively few examples [7]. Topic models are particularly suitable for providing the lowest-dimensional semantic representation of the visual environment [45]. In Chapter 4, we will observe that using a combination of both active learning and topic model based low-dimensional image representations will enable efficient interactive visual exploration over very low bandwidth.

### 2.2.3 Transfer Learning

Transfer learning explores how to adapt a machine learning model trained for one task to a new task using fewer training examples than required for a new model to be trained from scratch [58]. It has been successfully applied to reduce the amount of training data required to train learned models for image classification [47] and other

robotics tasks [94]. The simplest form of transfer learning would be to initialize the robotic explorer with the reward model learned from the previous mission, assuming that the science objectives of the new mission are somewhat similar, which could be done in addition to using active learning and low-dimensional image representations. Furthermore, some low-dimensional semantic representations are based on using features from pretrained deep neural networks, which represents another form of transfer learning [35, 100].

## 2.3 Path Planning for Autonomous Science

Robotic path planning is typically about finding the safest and shortest route to a specified destination. In scientific exploration, however, the destination is often irrelevant, and the challenge is to find the path which would result in the robot collecting as many scientifically valuable observations as possible. For example, AUVs often encounter huge amounts of sand and gravel at the bottom of the ocean, observations of which are not usually of scientific value, whereas observations of a rare marine species or unusual geological phenomena may be highly valuable.

One robotics technique that is very applicable to scientific exploration is Informative Path Planning (IPP). Informative path planning studies how to design paths which maximize the "utility" of some path defined with respect to the environment, often subject to some constraints such as path length or avoiding obstacles [50, 64, 77]. In scientific exploration, the utility of a path would be computed based on the spatial observation's models predictions of what would be observed along it, and the reward models scores for how interesting those observations would be. Maximizing the sum of these scores would lead the robot to visit as many locations predicted to have scientifically interesting observations as possible. Modified utility functions could be used to, for example, have the robot try to find and observe the *most* interesting location, rather than all interesting locations [34].

Other, non-IPP based approaches to autonomous scientific exploration include AEGIS [31] and OASIS [12]; these systems enabled robots to opportunistically recognize

scientifically relevant image observations, given a predefined model, and schedule more detailed observations with other sensors. However, these algorithms required domain-specific feature engineering and lacked spatial observation models, so path planning was limited to moving the robot closer to a target that had already been detected. At the other extreme, "curious" robots use a generic unsupervised vision model and autonomously move towards anything in their environment that is surprising or novel to the model [43]; the lack of operator input makes it impossible to directly specify particular scientific objectives using this approach.

Arora, Fitch, and Sukkarieh [1] presented a novel approach to autonomous scientific exploration that had scientists model their domain knowledge with a pre-defined Bayesian Network (BN) that was used by the robot to estimate the reward of potential paths. They introduced a spatial observation model in their system, and enabled informative path planning using Monte-Carlo Tree Search (MCTS) [17, 57] to explore an action tree composed of movement and sensing actions [1]. Their approach requires the operator to specify the domain-specific BN *a priori*. In their example of a Martian exploration mission to find rocks which were once been part of a Martian riverbed, the solution required developing systems to detect rocks in an image and inspect them for certain visual features to feed into a specialized geological model [1]. Due to the high degree of complexity involved, their system was not a general purpose exploration tool that could be deployed in unfamiliar environments. However, it could be made into one by replacing the Bayesian Network with a general purpose spatial observation and reward model.

In Chapter 4 we will build upon the work of Arora, Fitch, and Sukkarieh [1] by using learned semantic image representations and online active reward learning to create a general-purpose approach to vision-based scientific exploration in bandwidth-limited environments.

# Chapter 3

# Spatial Observation Modelling with Topic Models

*"Joy in looking and comprehending is nature's most beautiful gift."*

– Albert Einstein

In order to decide where to go next, a robotic explorer must predict what it will observe along various candidate trajectories. Even in new and unfamiliar environments, it should be possible to predict some of the things that the robot will observe because the semantic contents of nearby natural images, such as terrain types and species present, have strong spatial correlation [35, 82]. However, these correlations are hard to model in the observation space (pixels), where even nearly identical images can be made distant by effects like sensor noise and slight changes in illumination [109]. Further, due to the high dimensionality of the image space, there are no spatial models with which it is computationally tractable to predict the image that would be observed in an unvisited location.

To overcome these challenges, current approaches to spatial observation modelling for images operate in a *semantic space* $\mathcal{Z}$. A natural image in the observation space $\mathbb{O}$ is mapped to a location in the semantic space called its *semantic representation*.[1]

[1]For example, a textual description of an image is a semantic representation of that image in a particular language. Another semantic representation is a vector of weights over classes describing what the image contains, saying something like 50% of an image is sand and 50% is gravel.

The robot trains a spatial observation model, denoted as $\boldsymbol{Z}(\boldsymbol{x}) : \mathbb{R}^3 \mapsto \mathcal{Z}$, using its previous observations and a semantic feature extractor $\boldsymbol{z}(I) : \mathbb{O} \mapsto \mathcal{Z}$. This approach requires that the semantic representations of two images $\boldsymbol{z}(I_1), \boldsymbol{z}(I_2)$ are close, as measured by Euclidean distance, if and only if the human-perceived similarity of $I_1$ and $I_2$ is high. Semantic representations derived from computer vision models developed for unsupervised natural image clustering, such as deep feature extractors [83, 109] and spatial topic models (STMs) [44, 101] have this property.

The Realtime Online Spatio-temporal Topic Modelling (ROST) [44] algorithm is particularly suitable as a spatial observation model because it

- Does not require any training data,
- Excels at recognizing and representing "anomalies" (novel observations) [45],
- Creates very low dimensional semantic representations which can be stored and communicated using very little bandwidth [53], and
- Defines a spatial prior function on these semantic representations which can be used as the spatial observation model $\boldsymbol{Z}(\boldsymbol{x})$ [53, 87].

ROST has been successfully used by a robotic explorer to enable online mission summarization [35] as well as unsupervised "curious" exploration [43].[2]

**Chapter Summary:** Section 3.1 will provide the mathematical background of ROST, summarizing material published in [41, 43–45]. Section 3.2 will present the novel extension of the ROST algorithm to fully three-dimensional semantic mapping; most of this work was performed jointly with Levi Cai and is published in [42]. To the best of our knowledge, no other unsupervised semantic mapping algorithm has been tested on a robotic platform for the purpose of scientific exploration. Finally, Section 3.3 will present strategies to ROST hyperparameter selection and tuning, which are unpublished.

[2] Curious meaning that the robot is attracted towards anything that looks sufficiently different to everything else that it has previously seen.

## 3.1 Spatio-Temporal Topic Modelling

This section will describe, mathematically, the process by which ROST performs semantic segmentation of the visual observations collected during scientific exploration. We denote the $m^{\text{th}}$ image collected as $I_m \in [0,1]^{w \times h \times d}$ where $w$ and $h$ are the width and height of the image in pixels, $d$ is the number of colour channels, and pixel intensities are in the range $[0, 1]$.

### 3.1.1 ROST Overview

ROST performs semantic segmentation by extracting "visual words" from each image, grouping them into "cells", and then assigning each word to a learned "topic". Then, an image segmentation is produced by treating labelling each cell with the distribution of topics assigned to the words within it.

Each visual word $w$ comes from finite a "vocabulary" of image features of size $V$, which is created in advance. For example, some of these words may correspond to various SIFT feature descriptors [61]; when a SIFT keypoint is extracted from the new image $I_m$, it is identified using the word in the vocabulary with the closest SIFT feature descriptor (by some distance metric in the descriptor space). Convolutional Neural Network (CNN) based feature extractors can also be produce suitable visual words [35], and further examples of visual words are presented in Section 2.3 of [41]. ROST differs from previous spatial topic models by assigning each visual word a three-dimensional location $(u, v, t) \in \mathbb{R}^3$ composed of the $(u, v)$ pixel coordinates of the feature centre concatenated with the time $t$ of the observation.

Next, ROST groups these visual words into identical non-overlapping three-dimensional "cells" $C_{ij\tau} \subset \mathbb{R}^3$, representing a spatio-temporal volume indexed by $(i, j, \tau) \in \mathbb{Z}^3$, such that

$$\bigcup_{(i,j,t) \in \mathbb{Z}^3} C_{ij\tau} = \mathbb{R}^3.$$

The dimensions of each cell are described by the tuple $(l_X, l_Y, l_T) \in \mathbb{R}^3$. This ensures

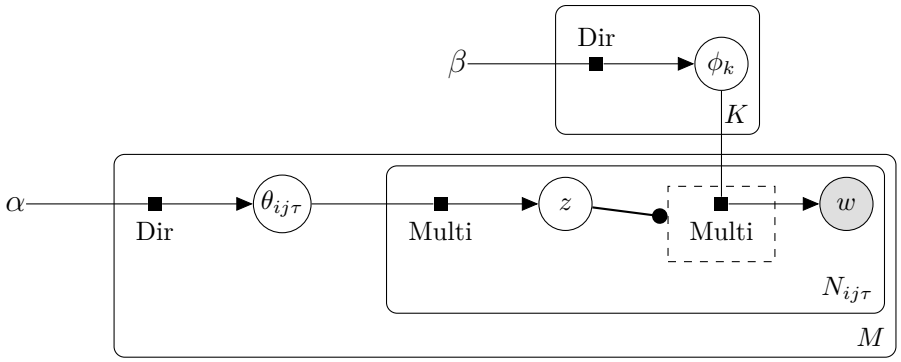


Figure 3-1: The graphical model for Latent Dirichlet Allocation, the simplest type of topic model. The terms "Dir" and "Multi" stand for Dirichlet Distribution and Multinomial Distribution, respectively, while the variables are defined in Section 3.1.

that each word $w$ can be assigned to a unique cell $C_{i,j,\tau}$ with indices given by

$$i = \left\lfloor \frac{u}{l_x} \right\rfloor,\ j = \left\lfloor \frac{v}{l_y} \right\rfloor,\ \tau = \left\lfloor \frac{t}{l_t} \right\rfloor.$$

We note that no more than $\left\lceil \frac{w}{l_x} \right\rceil \times \left\lceil \frac{h}{l_y} \right\rceil$ cells are necessary to contain all of the words extracted from a single image.

Finally, ROST assigns each word $w$ a topic label $z \in \{1, \ldots, K\}$ according to the probabilistic model described in the following subsection. A cell can thus be associated with a vector of topic weights $W_{ij\tau} \in \mathbb{N}^K$ such that $W_{ij\tau}[k]$ is the number of words in $C_{ij\tau}$ assigned to the $k^{\text{th}}$ topic. It is straightforward to create a semantic segmentation of an image by choosing $K$ colours and colouring the pixels contained in cell $C_{ij\tau}$ with the colour corresponding to $k^{\star}_{ij\tau} = \text{argmax}_k\, W_{ij\tau}[k]$.

### 3.1.2 The ROST Probabilistic Model

ROST builds upon the probabilistic model used in Latent Dirichlet Allocation (LDA), which is presented as a graphical model in Figure 3-1 and algorithmically in Algorithm 1. These models describe the process by which words are generated. The goal of LDA is to infer the latent variables $\{\theta_{ij\tau} \in \Delta^K\}$, the distributions of topics in each cell, as well as $\{\phi_k \in \Delta^V\}$, the distributions of words associated with each topic. This can be accomplished through any of a wide range of inference algorithms, such as variational

**Algorithm 1:** LDA Generative Model

1 **Given:** $\alpha \in \mathbb{R}^K_{++}, \beta \in \mathbb{R}^V_{++}, \zeta \in \mathbb{R}_{++}$
2 **Input:** Set of $M$ cells $\{C_{ij\tau}\}$
3 **for** $k = 1, \ldots, K$:
4 | Sample $\phi_k \sim \text{Dirichlet}(\beta \mathbf{1}^K)$
5 **foreach** $C_{ij\tau}$:
6 | Sample $\theta_{ij\tau} \sim \text{Dirichlet}(\alpha \mathbf{1}^K)$
7 | Sample $N_{ij\tau} \sim \text{Poisson}(\zeta)$
8 | **for** $n = 1, \ldots, N_{ij\tau}$:
9 | | Sample $z \sim \text{Multinomial}(\theta_{ij\tau})$
10 | | Sample $w \sim \text{Multinomial}(\phi_z)$
11 | | Add word $w$ to cell $C_{ij\tau}$

inference or Gibbs sampling.

LDA sets a symmetric Dirichlet distribution prior on the cell-topic distributions $\theta_{ij\tau}$ and the topic-word distributions $\phi_k$, parameterized by the concentration parameters $\alpha$ and $\beta$, respectively. The prior on $\theta_{ij\tau}$ is weakest when $\alpha = 1$; as $\alpha \to 0^+$ the prior strongly favours assigning all of the words in a cell to the same topic, whereas when $\alpha \to +\infty$ it strongly favours having each cell contain an equal mix of all $K$ topics. Likewise, compared to $\beta = 1$, moving $\beta \to 0^+$ encourages LDA to associate each topic with only a few unique words, whereas $\beta \to +\infty$ strongly favours associating each topic with many different words. These priors are discussed more thoroughly by Girdhar [41] and later in Section 3.3.

The main difference between ROST and LDA is that ROST introduces a prior on $\theta_{ij\tau}$ based on spatio-temporal correlation. In particular, line 6 of Algorithm 1 is replaced with:

$$\theta_{ij\tau} \sim \text{Dirichlet}\left(\alpha + \sum_{(i',j',\tau') \in G(i,j,\tau) \setminus \{(i,j,\tau)\}} \theta_{i'j'\tau'}\right) \tag{3.1}$$

where $G(i, j, \tau) \subset \mathbb{Z}^3$ defines the "neighbourhood" of cell $C_{i,j,\tau}$. For example, the L1 neighbourhood of size $d$ is defined as:

$$G_{\text{L1}}(i, j, \tau; d) = \{(i', j', \tau') \in \mathbb{Z}^3 : |i - i'| + |j - j'| + |\tau - \tau'| \leq d\}$$

This ensures that cells which are nearby in both space and time are correlated with each other, in the sense that they will should have similar topic distributions. The main advantage of introducing this prior is that, since nearby parts of an image and frames of a video really do tend to be correlated, it is easier for the inference algorithm to converge to a good $\theta_{ij\tau}$ in less time.

### 3.1.3 Bayesian Non-Parametric ROST (BNP-ROST)

One issue observed with the ROST algorithm is that it sometimes uses too many different topics to describe early observations, and then struggles to use these topics to represent different observations collected later on. This can happen because, if $K$ is set large enough to allow for enough topics to represent everything that could be observed in a mission, then the sampling performed in Equation 3.1 will be biased towards using several topics to describe each cell. This is more clear if we consider the posterior topic distribution of a word $w$. Let $n_k^v$ be the number of times we have have previously assigned the $v^{\text{th}}$ word to the $k^{\text{th}}$ topic, and let $W_{G(i,j,\tau)}[k]$ be the number of words assigned to topic $k$ in the neighbourhood of cell $C_{i,j,\tau}$. Then the likelihood that we assign a word $w = v$ to topic $z = k$ is:

$$\Pr(z = k \mid w = v, \{n_k^v\}, \{W_{ij\tau}\}) = \frac{n_k^v + \beta}{\sum_{v=1}^{V}(n_k^v + \beta)} \cdot \frac{W_{G(i,j,\tau)}[k] + \alpha}{\sum_{k=1}^{K}\left(W_{G(i,j,\tau)}[k] + \alpha\right)}. \quad (3.2)$$

When few observations have been collected, then each $n_k^v$ and $W_{G(i,j,\tau)}[k]$ are small. This makes the posterior distribution close to uniform over all $k$, and thus words are assigned to many different topics even if the observations are all similar. Thus, the topics are "used up" early in the mission, and it is difficult for ROST to label novel phenomena.

BNP-ROST addresses this issue by introducing a Chinese Restaurant Process (CRP) prior on the number of topics, which encourages using fewer topics at the beginning of the mission but allows for a potentially unbounded number of topics as

the mission progresses. It does so by once again redefining the prior for $\theta_{ij\tau}$:

$$\theta_{ij\tau} \sim \mathrm{CRP}_{i,j,\tau}(\gamma, W_{G(i,j,\tau)} + \boldsymbol{\alpha}) \tag{3.3}$$

As described in [42], in this model we assign a word $w$ to one of the $K_t$ occupied "tables" in a Chinese restaurant at time $t$ with probability proportional to $(W_{G(i,j,\tau)}[k] + \alpha), k \in \{1, \ldots, K_t\}$, where $W_{G(i,j,\tau)}[k]$ corresponds to the number of customers sitting at the $k^{\text{th}}$ table in that restaurant and at corresponding tables in the neighbouring restaurants. The word is assigned to a new table with probability proportional to $\gamma$, in which case we increment the number of topics in use for later timesteps: $K_{t+1} = K_t + 1$. The advantage of using a CRP is that we do not need to explicitly specify the number of topics *a priori*, as it grows automatically with the number and complexity of observations. By varying $\alpha$ and $\gamma$, it is possible to influence the number of topics used to describe each observation and the rate at which new topics are introduced. Ideally, these parameters should be set such that the topics and semantic segmentations created by ROST best match the way in which a scientist would segment the images.

## 3.2 Realtime Semantic Mapping with Sunshine

The main application we explored in [42] was semantic mapping, a form of mission summarization. This work sought to address Problem 1 described in Section 1.4.1 by giving scientists greater insight into the mission status and what the robot had found. In particular, the Sunshine system is to provide a near-realtime semantic summary of everything the robotic explorer has observed so that they can quickly recognize any anomalies, ask the robot for additional details (i.e., actual images of the anomaly), and decide whether to deviate from the initial exploration plan.

The main capability missing from BNP-ROST was handling 4-dimensional semantic mapping using real-world coordinates and time $(x, y, z, t)$, instead of the previous 3-dimensional mapping in image-based coordinates and time $(u, v, t)$. While this does not require changing the probabilistic model, it does require processing both RGB-D

(colour and depth) image data as well as the robot's pose estimates, whereas previous versions of ROST had only been applied to 2D colour images and video and did not model camera motion. To enable these new capabilities, we extended ROST and interfaced it with ROS, the Robot Operating System [80], which provides a number of software libraries for common robotics tasks. We call this new system "Sunshine".

After running ROST's visual word extraction, Sunshine projects each word into real world coordinates using the camera calibration matrix $K \in \mathbb{R}^{3\times3}$ and the camera extrinsic matrix $T = \begin{bmatrix} R & t \end{bmatrix}$ composed of the camera rotation $R \in \mathrm{SO}(3)$ and translation $t \in \mathbb{R}^3$ from the mission origin in the world frame. These were computed by ROS using the robot's current position estimate and the robot-camera transform. Letting $(u, v)$ be the word's location in image coordinates, and $d$ be the distance, in metres, of the pixel $(u, v)$ as reported by the RGB-D camera, then the real world coordinates $(x, y, z)$ of the word in the world frame are computed as:

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix}^\top = dRK^{-1} \begin{bmatrix} u & v & 1 \end{bmatrix}^\top + t \tag{3.4}$$

The word is then assigned to the unique cell $C_{ij\kappa\tau} \subset \mathbb{R}^4$ with dimensions $(l_X, l_Y, l_Z, l_T)$ that contains the point $(x, y, z, t)$.[3] Due to localization errors, there may be error in the camera transformation matrix $T$, however this should not have a significant effect on the semantic mapping process for the following reasons:

- For errors on a much smaller scale than $(l_X, l_Y, l_Z, l_T)$, the word will not be placed into the wrong cell unless it was already very close to the edge of the cell
- Even if the word is placed into an adjacent cell by mistake, the spatial neighbourhood prior means that it will similarly influence the semantic labels of nearby cells as though it had been correctly placed
- If the localization error is nearly constant between observations, then the semantic map will still have the correct local structure because locally it will be as though the true semantic map had undergone some transformation $T_{\mathrm{err}}$.

[3] Note that we have introduced $\kappa \in \mathbb{Z}$ to index cells along the $Z$-axis.

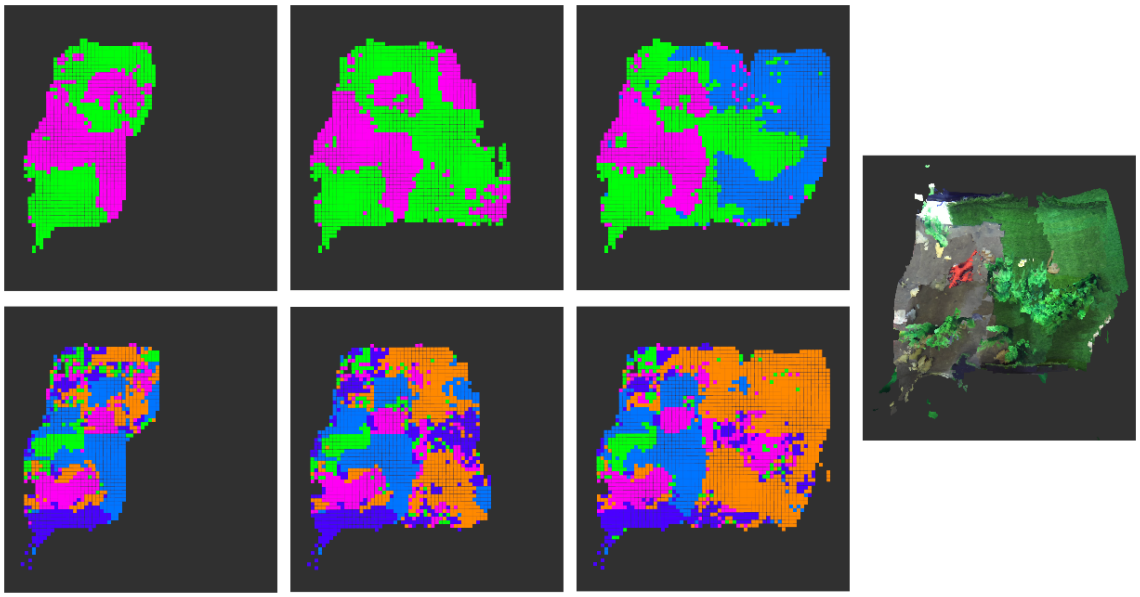

Figure 3-2: Progression of semantic maps generated as the robot explores the artificially created scene from Figure A-3. The top maps correspond to BNP-ROST hyperparameters $\alpha = 0.5, \beta = 0.1, \gamma = 10^{-5}$, while the bottom maps correspond to $\alpha = 5 \times 10^{-5}, \beta = 10^{-4}, \gamma = 10^{-4}$. The image on the right shows the stitched map of the visual observations. The robot localization system used was built by Levi Cai using AprilTags [74]. This figure appeared in [42].

Therefore, as long as the localization errors are small compared to the cell dimensions or the state estimates are filtered so that the errors between successive observations are similar, then the semantic map produced will be mostly unaffected.

### 3.2.1 Experimental Results

We performed several experiments to validate that Sunshine could be used for realtime semantic mapping. In one experiment we tested the effects of different values of the priors $\alpha$, $\beta$, and $\gamma$ when creating semantic maps of an artificial underwater environment. Two of these maps are presented in Figure 3-2. The maps show that different types of terrains are well characterized by the scene mode, and also that the system performs well even though the robot localization accuracy is not perfect. We demonstrated that by varying the priors, the semantic maps could be tuned according to a scientist's desired level of abstraction and the environment-specific bandwidth constraints. In particular, we described maps where adjacent cells are often labelled

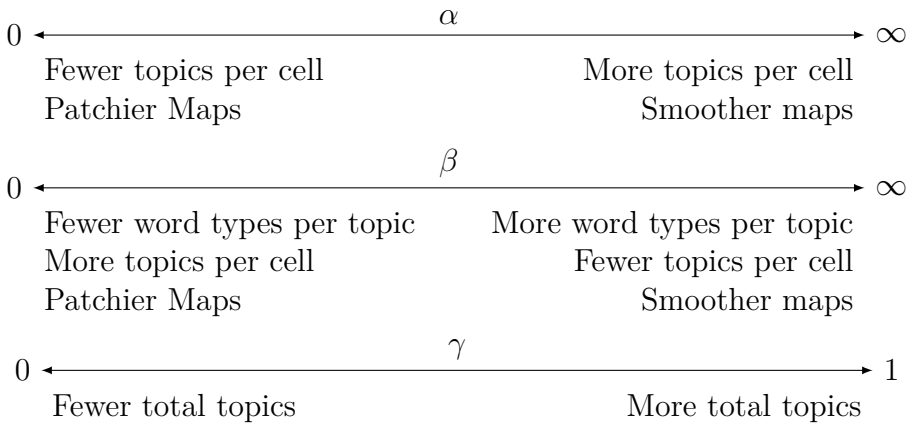


Figure 3-3: Qualitative effects of different values of $\alpha, \beta, \gamma$ based on the BNP-ROST probabilistic model and empirical observations of generated semantic maps.

differently as "patchier" than maps where they tend to be the same, and observed the results summarized in Figure 3-3.

The patchiness of maps and the number of topics used in each cell and in the entire map are meaningful metrics, making it easier to tune the values to produce meaningful semantic segmentations. For example, if the cell dimensions are small compared to the phenomena of interest, it is best to have only one or two topics per cell so $\alpha$ should be set very low. This is because for small cells it is likely that any two words in the same cell are associated with the same phenomenon, and we would generally prefer to use exactly one topic per unique phenomenon. The values of $\beta$ and $\gamma$ can likewise be tuned to create semantic maps which best match the scientist's expectations. Due to the general nature of these priors, this tuning can be performed using almost any dataset, even if it is not strictly representative of what the robot will observe on its next mission.

In addition to optimizing the hyperparameter values for semantic meaning, we also explored how they could affect the communication bandwidth required for mission summarization. The main effect is that *smoother maps are more compressible*, because most lossless image compression algorithms (e.g., the PNG algorithm) rely on compressing large contiguous patches of colour to save space. Thus, using larger values of $\alpha$ and $\beta$ and smaller values of $\gamma$ are associated with lower communication bandwidth usage when streaming the semantic maps.

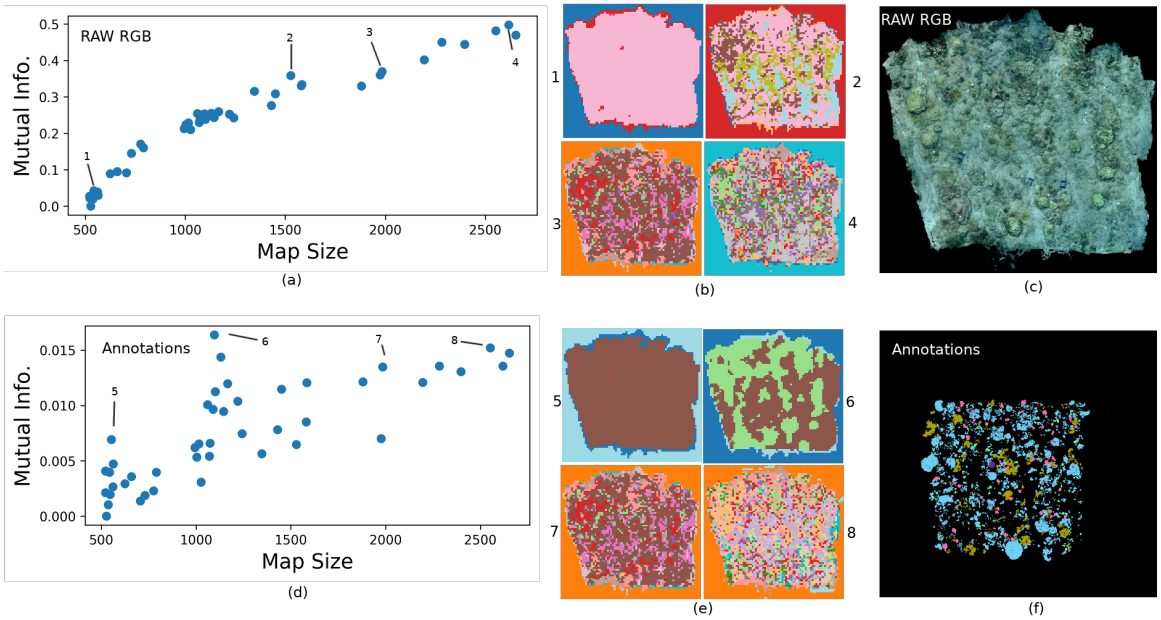


Figure 3-4: Topic Maps produced using the HAW-2016-48 Coral Reef Dataset [92]. (a) and (d) show scatter plots evaluating topic maps, like those in (b) and (e), generated by varying the hyperparameters $\alpha, \beta$ and $\gamma$. The x-axis shows the size, in bytes, of the topic map after PNG compression, while the y-axis shows the mutual information score of the map with, for (a) the RGB dataset (c), and for (d) the scientist's annotations (f). Note that mutual information was computed over the filled in (i.e., non-black) parts of each (c) and (f). The maps in (b) and (e) match the points in (a) and (d) labelled with the same numbers. This figure also appeared in [42].

Finally, we explored the relationship between semantic meaning in the maps and the size of the maps. The results are presented in Figure 3-4. The main takeaway is that patchier maps tend to be closer to the raw visual map, whereas they are not necessarily more semantically meaningful. This is deduced by noting that in subfigure (a), the mutual information[4] between the semantic map and visual map increases as the map size (patchiness) increases. Conversely, in subfigure (d), there is a "sweet-spot" that occurs for topic map #6, where the semantic map is quite well correlated with the scientist's segmentation of the map while also being very compressible.

In addition to demonstrating that these semantic maps could feasibly be streamed in realtime to a scientist during a mission, these results showed that spatio-temporal topic

[4]Mutual information is a measure of correlation between two images. A mutual information score of 1 means that one image could be entirely reconstructed from the other, whereas a score of 0 means that one image provides no information about the other.

models could be used as spatial observation models in low-bandwidth environments. In particular, since the semantic maps tend to be smooth and low-dimensional, it is fairly easy to predict the semantic labels which will be used in new locations near where the robot has previously visited. This can be done using a Gaussian Process based model [87], or by a simpler method of extrapolation based on nearby semantic labels. In fact, the Sunshine (BNP-ROST) spatial neighbourhood prior can be trivially adapted to be used as a spatial observation model. The last missing piece for a robotic explorer to be able to autonomously plan high scientific value paths is a learned reward model, which will be the topic of Chapter 4.

## 3.3 BNP-ROST Hyperparameter Tuning

While the BNP-ROST hyperparameters $\alpha, \beta, \gamma$ have clearly observable qualitative effects on the resulting semantic maps (as seen in Figures 3-2, 3-3, and A-4) they can still be challenging to tune because they can take on a wide range of values and it can take several minutes to generate a complete semantic map from scratch. In a broader sense, there are even more hyperparameters to consider and tune such as the cell dimensions $(l_X, l_Y, l_Z, l_T)$, the number of topics $K$, and the spatial neighbourhood prior $G$.

In [42], the hyperparameters $\alpha, \beta, \gamma$ were selected after running a gridsearch over 125 sets of values, which took several hours. This was slow and inefficient, and there were no guarantees that the 5 values tested per variable were close to the optimal values. To address this issue, we now use Bayesian optimization, an information-theoretic approach to function-maximization using a minimum number of function evaluations [93]. In particular, we use the Limbo library [18] to quickly find the hyperparameters that maximize the mutual information between the semantic maps and human annotations, which was the same objective of the grid search.

Another issue is that it can be difficult to find good initial values for the $\alpha$, $\beta$, and $\gamma$ hyperparameters. In the previous section we noted that, when the cell dimensions are set to be at or smaller than the dimensions of the phenomena of interest, $\alpha$ should

be set low to encourage the model to use only one or two topics in each cell. What range of values of $\alpha$ will accomplish this? One possible approach is to consider the expected *perplexity* of the cell topic distribution $\theta_{ij\kappa\tau}$ prior as a function of $\alpha$. Recall that, in the absence of a spatial neighbourhood prior, the prior $\theta_{ij\kappa\tau}$ is a symmetric $K$-dimensional Dirichlet distribution

$$\theta_{ij\kappa\tau} \sim \text{Dirichlet}(\alpha \mathbf{1}^K). \tag{3.5}$$

Girdhar [41] noted that the expected entropy of this prior is

$$\mathbb{E}\left[\mathbb{H}\left[\theta_{ij\kappa\tau}\right] \mid \alpha, K\right] = \Psi(K\alpha + 1) - \Psi(\alpha + 1), \tag{3.6}$$

where $\Psi(x) = \frac{d}{dx} \log \Gamma(x)$ is the digamma function. The perplexity of a random variable $X$ is defined as $PP(x) \coloneqq \exp \mathbb{H}[X]$. A relevant fact is that a perplexity value of $p$ implies that the random variable has as much uncertainty as a fair $p$-sided die. Thus, having a Dirichlet prior with expected perplexity $p$ encourages $\theta_{ij\kappa\tau}$ to distribute its probability mass split evenly over $p$ topics. Therefore, the perplexity of the prior is directly related to how many topics the prior will encourage per cell, so to encourage no more than two topics per cell, $\alpha$ should be set such that

$$\mathbb{E}[PP(\theta_{ij\kappa\tau})] = \mathbb{E}[\exp \mathbb{H}[\theta_{ij\kappa\tau}]] \approx \exp \mathbb{E}[\mathbb{H}[\theta_{ij\kappa\tau}]] = e^{\Psi(K\alpha+1)-\Psi(\alpha+1)} \leq 2. \tag{3.7}$$

Note that the approximate equality comes from the fact that the entropy of a discrete distribution is sharply peaked around its expected value [41].

It is relatively easy to solve for an $\alpha$ that satisfies Equation 3.7 despite it lacking a closed form solution because $\exp \mathbb{E}[\mathbb{H}[\theta_{ij\kappa\tau}]]$ is bounded below by 1 as $\alpha \to 0^+$

$$\lim_{\alpha \to 0^+} \exp(\Psi(K\alpha + 1) - \Psi(\alpha + 1)) = \exp(\Psi(1) - \Psi(1)) = 1. \tag{3.8}$$

Thus, by driving $\alpha$ sufficiently close to 0, any desired expected perplexity bound $PP_{\max} \geq 1$ can always be achieved. To expedite this process, one can numerically

approximate the derivatives of the digamma function by truncating their infinite series representations[5]

$$\frac{d^m}{dx^m}\Psi(x) = (-1)^{(m+1)}m!\sum_{k=0}^{\infty}\frac{1}{(x+k)^{m+1}},\ m > 0. \tag{3.9}$$

These can be used to find the optimal $\alpha$ using gradient descent or Newton's method.

This approach can also be used to optimize $\beta$ for a fixed vocabulary size $V$, if it is known approximately how many visual words in the vocabulary should be associated with each topic. With $\alpha$ and $\beta$ set, it is much easier to experiment and quickly find a good value for $\gamma$. This heuristic initialization has been successfully used to provide a good starting point for the Bayesian optimization procedure described previously.

[5]Note that the series representation for $\frac{d}{dx}\Psi(x) = \sum_{k=0}^{\infty}\frac{1}{(x+k)^2} > 0$ tells us that $\Psi(x)$ is continuous and monotonically increasing $\forall x > 0$. Empirically, we find that the expected perplexity is also monotonically increasing in $\alpha$, but have not derived the proof.

# Chapter 4

# Online Active Reward Learning

*"We are trying to prove ourselves wrong as quickly as possible, because only in that way can we find progress."*

– Richard P. Feynman

We proposed in Section 1.4 that the next level of autonomy in co-robotic scientific exploration would require a robot to possess:

- A spatial observation model to predict where it can find specific phenomena,
- A reward model that captures the value of observing a phenomenon with respect to the mission objectives, and
- A path planner that uses these to plan high scientific value trajectories.

We also discussed the need for human-robot communication during the mission to learn the reward model *online*, using scientist feedback to deduce which types of observations are most scientifically interesting. The particular approach that will be the focus of this chapter is using the limited communication bandwidth available to query the operator about the scientific value (i.e., reward) of past observations, and use their responses to learn the reward model.

We present our novel approach to achieving Level 4 autonomy in co-robotic scientific exploration, "Human-Robot Cooperative Exploration", in Figure 4-1. In the figure, the top map shows a visual map of the environment along with boxes representing

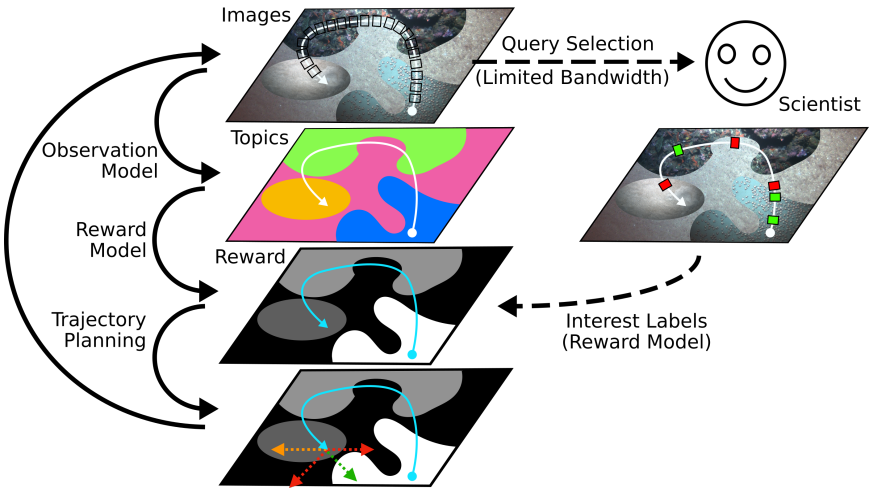


Figure 4-1: An approach to co-robotic exploration that models the interest of the operator over a low bandwidth communication channel and uses the learned reward model to plan the most scientifically rewarding robot paths.

the robot's past observations. The one below uses colour-coded topics to depict a semantic segmentation of the environment, including parts of the map not yet visited, created using a spatial observation model (e.g., Sunshine [Ch. 3]). The lower two maps use brightness to represent the predicted scientific value of observations in various locations, while the lowest map shows some candidate trajectories generated by a path planner (e.g., MCTS [17, 57]), colour-coded from red to green by predicted scientific value. The map to the right shows which observations were sent to and labelled by the scientist, with green indicating scientific value, and red indicating no scientific value. Importantly, only small percentage of observations are sent during the mission due to communication bandwidth constraints like those discussed in Chapter 1.

**Chapter Summary:** Most of this chapter will appear as it was published in [53]. Section 4.1 will present a partially observable Markov decision process (POMDP) formulation for vision-based scientific exploration and a solution that is generalizable to many environments. The approach is suitable for deployment in completely unknown environments, and can use (but does not require) prior knowledge about the environment and the phenomena being observed. Section 4.2 will present an analysis of active learning decision criteria that a robot could use for deciding which observations

to send to the operator. This includes a novel regret-based active learning algorithm designed for maximizing reward in the online setting. In Section 4.3, we compare these active learning criteria using simulations of a scientific exploration task, based on both real and artificial data. Finally, in Section 4.4, we present some new motivation for why regret-based methods should outperform information theoretic ones in certain online settings, especially under low communication bandwidth constraints.

## 4.1 The Co-Robotic Visual Exploration POMDP

We present the co-robotic visual exploration problem as a POMDP. The entire POMDP is characterized by the tuple $(\mathcal{S}, \mathcal{A}, \mathbb{O}, T, O, R, \gamma, b_0)$, defined in Table 4.1. We model the state of the robot at time $t$ as $S_t = (X_t, Y_t, L_t)$. $X_t = \{(\boldsymbol{x}_i, I_i)\}_{i=1}^{t}$ is the sequence of locations the robot has visited, with corresponding image observations, where the current location is $\boldsymbol{x}_t \in \mathbb{R}^3$ and the latest observation is the image $I_t \in \mathbb{O}$. Note we explicitly constrained our focus to dealing with high-dimensional observation spaces, such as images. $L_t$ is the set of indices of images sent to and labeled by the operator. $Y_t$ contains the reward labels for all images, including those that have not been sent; most of these are unknown, making the robot's state partially observable. For now we assume that these labels are binary, but in Section 4.4 we will explore relaxing this assumption.

The partial observability is a consequence of the robot's limited ability to query the operator during a mission; in bandwidth constrained environments the robot sends images at a much slower rate than it collects them, so it must decide which labels to observe. We assume that only the operator can evaluate the unknown, but deterministic, binary "interest" function $\mathcal{I}(I)$ such that $(Y_t)_i = \mathcal{I}(I_i)$. Further, it is assumed that the operator cannot express their interest function analytically (otherwise it would be computed onboard the robot), and would instead train an approximate model based on their labels for various example images. However, since exploration typically occurs in remote and unstudied environments like the ones discussed in Chapter 1, the operator does not have a representative dataset of what

Table 4.1: The Scientific Exploration POMDP Specification.

| Component | Definition | Our Assumptions |
|---|---|---|
| $\mathcal{S}$ | State space of the robot | $S = (X, Y, L)$ |
| $\mathcal{A}$ | Discrete set of robot actions | Motion primitives[1] |
| $\mathbb{O}$ | Observation space | Natural images |
| $T$ | Transition function | Deterministic, $\mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$ |
| $O$ | (Spatial) Observation model | $\mathcal{S} \times \mathcal{A} \mapsto \mathbb{O}$ |
| $R$ | Reward model | $\mathbb{O} \mapsto \mathbb{R}$ |
| $\gamma$ | Discount factor | $\gamma \in [0, 1]$ |
| $b_0$ | Initial belief state | $\mathcal{P}(Y)$ |

the robot will observe, and is unable to provide the robot with a complete model of $\mathcal{I}(\cdot)$ in advance of the mission.

At each timestep, the robot takes an action chosen from the set $\mathcal{A}$. These actions are often motion primitives, which represent a variety of the motions the robot is capable of executing over some short period of time. There may be a negative reward associated with actions depending on their energy usage. Using a sensor (e.g., a camera) or querying the operator may also be modelled (e.g., in [26]) as additional actions in $\mathcal{A}$ with costs, such as energy usage, included in the reward model. For simplicity, we assume these costs are negligible and that the robot takes an observation at every timestep and can send queries at some maximum rate but at no cost and concurrently with other actions.

The observation space $\mathbb{O}$ represents the space of images the robot could observe with its camera. The transition model $T$ is either a deterministic or probabilistic model of how the robot will behave when commanded to perform an action, and is used to model things like wheel slip on certain types of terrains.

The belief state $b_t$ is a probability distribution over the labels $Y_t$, based on all available information (i.e., the known labels) at time $t$. This represents how scientifically valuable the robot believes each unlabelled observation to be. This belief state is refined as the robot receives more labels and the reward model is updated. The belief state will typically also contain a probability distribution over the robot location $X$, since in a real-world environment there is always some localization uncertainty.

There are three key decisions to fully specify the remaining components of the co-robotic visual exploration POMDP. The first is defining a spatial observation model over the space of natural images, which was already discussed in Chapter 3. The second is choosing a reward model for estimating the reward of observations, which will be discussed in Subsection 4.1.1. The third is choosing an effective active learning strategy, which is the topic of Section 4.2.[2]

Given all of these specifications, the robot uses an online POMDP planner to approximate an optimal policy $\pi^\star : \mathcal{S} \mapsto \mathcal{A}$ in real-time. Algorithm 2 presents our approach to co-robotic exploration based on the assumptions listed above. The algorithm describes the process through which the robot moves along its planned trajectory while collecting observations and updating its spatial observation model accordingly, as well as concurrently requesting and receiving labels. Algorithm 2 makes use of the `PLAN_TRAJECTORY` subroutine (Algorithm 3) to run Monte Carlo Tree Search for some number of iterations in order to find the path with the highest scientific value.

### 4.1.1 Learning a Reward Model Online over Low Bandwidth

We define the robot's accumulated reward to be the total number of unique and interesting observations it has collected

$$R\left(X_t\right) = \sum_{i=1}^{t} \mathcal{I}\left(I_i\right) = \sum_{i=1}^{t} \left(Y_t\right)_i . \tag{4.1}$$

This can only be computed after the operator sees all images (i.e., after the mission), and thus it is not useful for planning. Instead, the planning algorithm (e.g., MCTS) requires a reward function $R(\boldsymbol{x})$ specifying the expected reward of visiting any location $\boldsymbol{x} \in \mathbb{R}^3$. The spatial observation model achieves part of this by predicting what will be observed at location $\boldsymbol{x}$ in the semantic representation space $\mathcal{Z}$. The robot then estimates the reward as a function of the semantic field $\boldsymbol{Z}_t\left(\boldsymbol{x}\right)$ by learning the model

[2]This could also be viewed as part of the reward model as an active learning strategy essentially assigns some "reward" to each question that could be asked, and chooses the highest scoring question.

**Algorithm 2:** Co-Robotic Exploration

**Given:** $(\mathcal{S}, \mathcal{A}, \mathbb{O}, T, \mathcal{O}, R, \gamma, \boldsymbol{x}_0)$, $t_{\max}$
$X_0 \leftarrow \varnothing$ // Stores the path and observations
$\boldsymbol{\tau} \leftarrow \{\boldsymbol{x}_0\}$ // The current trajectory plan
$q \leftarrow null$ // Index of next observation to label
$t \leftarrow 0$ // The current timestep
**while** $t < t_{\max}$:
  $\boldsymbol{x}_t \leftarrow$ Next_Step($\boldsymbol{\tau}$)
  $I_t \leftarrow$ Observe($\boldsymbol{x}_t$)
  $X_t \leftarrow X_{t-1} \cup \{\boldsymbol{x}_t, I_t\}$
  $\mathcal{O} \leftarrow$ Update_Observation_Model($\mathcal{O}, X_t$)
  **if** $q \neq null$ and Label_Ready($I_q$)
    $y_q \leftarrow$ Query_Result($I_q$)
    $Y_t \leftarrow Y_{t-1} \cup \{I_q, y_q\}$
    $R \leftarrow$ Update_Reward_Model($R, Y_t$)
    $q \leftarrow null$
  **endif**
  $\boldsymbol{\tau} \leftarrow$ Plan_Trajectory($X_t, \mathcal{S}, \mathcal{A}, T, \mathcal{O}, R, \gamma$) // See Algorithm 3
  **if** $q = null$
    $q \leftarrow$ Query_Selector($X_t, \mathcal{O}, R$)
    Request_Label($I_q$)
  **endif**
  $t \leftarrow t + 1$

**Algorithm 3:** Plan_Trajectory

**Input:** $X_t, \mathcal{S}, \mathcal{A}, T, \mathcal{O}, R, \gamma$
**Given:** $n$ // Number of trajectories to test
$\mathcal{T} \leftarrow$ Generate_Trajectories($X_t, \mathcal{S}, \mathcal{A}, T, n$)
**for** $i = 1, \ldots, n$:
  $\boldsymbol{s}_i \leftarrow$ Score_Trajectory($\mathcal{T}(i), O, R, \gamma$)
$\boldsymbol{\tau} \leftarrow \mathcal{T}(\text{argmax}_i\, \boldsymbol{s}_i)$
**return** $\boldsymbol{\tau}$

$g_\theta : \mathcal{Z} \mapsto [0, 1]$, where $\theta$ is a set of parameters for the model family. This means the expected reward for an observation at $\boldsymbol{x}$ is predicted as

$$R(\boldsymbol{x}) = g(\boldsymbol{Z}_t(\boldsymbol{x}); \theta), \tag{4.2}$$

Recall that $L_t$ is the set of labeled image indices at time $t$, and let $D_t = \{(I_i, (Y_t)_i)\}_{i \in L_t}$ be the corresponding training set. We choose $\theta$ to minimize the cross-entropy loss $\mathcal{L}$ on $D_t$, resulting in the final reward model

$$R(\boldsymbol{x}; D_t) \approx g\left(\boldsymbol{Z}(\boldsymbol{x}); \theta^\star_{D_t}\right) \tag{4.3}$$

$$\theta^\star_{D_t} = \operatorname*{argmin}_\theta \sum_{(I,y) \in D_t} \mathcal{L}(y, g(\boldsymbol{z}(I); \theta)). \tag{4.4}$$

As discussed in Subsection 2.2.2, for simple models the number of examples necessary to learn a simple model is typically linear in the number of input dimensions [73]. Thus, it is desirable to jointly pick a semantic representation and a model $g_\theta$ such that the total number of examples required to train $g_\theta$ is less than the number of examples that can be labelled during the mission. This further motivates the use of BNP-ROST [45] as the semantic feature extractor; due to its usage of the Chinese Restaurant Process, the dimensionality of its semantic representation grows as $\log t$, logarithmic in the number of images $t$. Conversely, the number of labelled images grows much faster (linearly) at $\frac{t}{n}$, where $n \geq 1$ is set by the bandwidth constraint. Thus, when using BNP-ROST in combination with a simple reward model, then the training process for $g_\theta$ is expected to quickly converge to good parameters $\theta$, even with few training examples. In our experiments in Section 4.3, we will use logistic regression as the reward model $g_\theta$, which has only $\dim(\theta) = \dim(\mathcal{Z}) + 1$ parameters.

## 4.2 Online Active Reward Learning for POMDPs

When the robot observes a novel phenomenon, it often needs to query the operator's interest in collecting more observations of that phenomenon. The only type of query

that the robot can perform while exploring a remote environment is sending an image to the scientist and receiving an interest label in return. However, the scientist cannot determine the scientific value of an image from the image's semantic representation alone and does not have access to enough information to advise the robot on the optimal policy. This presents a unique challenge for active learning.

Here we will consider active learning strategies to learn the parameters of a POMDP reward model online. We denote the set of unlabelled image indices at time $t$ as $\mathcal{U}_t$, and the active learning metric as $h(\boldsymbol{z})$, such that the next image to request a label for is chosen as

$$i^\star = \underset{i \in \mathcal{U}_t}{\operatorname{argmax}}\, h\left(\boldsymbol{z}\left(I_i\right)\right). \tag{4.5}$$

### 4.2.1 Non-Adaptive Query Selection

The simplest approaches to selecting images to be labelled do not depend on $\boldsymbol{z}$, and thus are good baselines to consider. *Random* selection chooses unlabelled observations uniformly at random. *Uniform* selection instead chooses every $n^{\text{th}}$ image, where $n$ is bounded below by the bandwidth constraint.

### 4.2.2 Informative Query Selection

Informative query selection involves defining some uncertainty metric on the model, and choosing to label the observation which results in the greatest reduction of uncertainty. There are many query selection strategies that fall into this category and are effective at learning a function in few examples [108]. A common uncertainty metric for classification problems is entropy, where the highest entropy values occur when an observation is on a decision boundary. A widely-used approach to informative query selection is "uncertainty sampling", which typically means picking the observation with the maximum entropy [108]

$$h_{\text{Entropy}}\left(\boldsymbol{z}; \theta^\star_{D_t}\right) = \mathbb{H}\left[g\left(\boldsymbol{z}; \theta^\star_{D_t}\right)\right]. \tag{4.6}$$

An issue with uncertainty sampling is that labeling the most uncertain observation might not have much effect on the model parameters $\theta$ – if the model parameters do not change, then the model performance does not increase. This suggests maximizing "error reduction" [108] instead

$$
\begin{aligned}
h_{\text{Info}}(\boldsymbol{z}) &= h_{\text{Entropy}}\left(\boldsymbol{z}; \theta^{\star}_{D_t}\right) - \mathbb{E}_{D'_t|D_t}\left[h_{\text{Entropy}}\left(\boldsymbol{z}; \theta^{\star}_{D'_t}\right)\right] \\
D'_t &= D_t \cup (\boldsymbol{z}, y) \\
p(D'_t \mid D_t) &\approx g(\boldsymbol{z}; \theta^{\star}_{D_t}).
\end{aligned} \tag{4.7}
$$

This *Information Gain* query selection method prioritizes labeling an observation by how much a new label $y$ is expected to reduce the entropy of similar future observations. This should maximize the rate at which entropy is reduced and thus the rate at which the reward function is learned.

### 4.2.3 Regret Minimizing Query Selection

We now introduce a novel *Regret* minimizing query selector that focuses on identifying the queries that help maximize the expected reward collected during the mission, rather than information gained about the reward function. Regret is typically defined for POMDPs as the difference in utility between the chosen action and the true optimal action based on complete information. To our knowledge, this was the first work to compare a regret-based heuristic against information-theoretic heuristics in online active learning [53].

Suppose that the robot is considering a finite set of trajectories $\mathcal{T} = \{\tau_i\}_{i=1}^{N_\tau}$. It uses it's spatial observation model to predict what it will observe along each trajectory $\tau$, predicts each trajectory's reward, and finally chooses the one with the highest reward (see Algorithm 3). However, given limited training data, the robot has significant uncertainty in the predicted rewards and thus is unlikely to have chosen the true optimal trajectory. This motivates a question for each unlabeled image: if this image were labeled, would the robot have chosen a different trajectory? If the answer is yes, then it must mean that, given this additional label, a different trajectory would be

predicted to have greater reward and thus the robot would "regret" not knowing the label. If it is no, then the robot would have no immediate regret for not knowing it, so it is a question less worth asking. We formalize this in the following objective:

$$h_{\text{Regret}}(\boldsymbol{z}) = \mathbb{E}_{D'_t|D_t}\left[R(\tau^\star_{D'_t}; D'_t) - R(\tau^\star_{D_t}; D'_t)\right] \tag{4.8}$$

$$R(\tau; D_t) = \sum_{\boldsymbol{x}\in\tau} g(\boldsymbol{Z}(\boldsymbol{x}); \theta^\star_{D_t}) \tag{4.9}$$

$$\tau^\star_D = \operatorname*{argmax}_{\tau\in\mathcal{T}} R(\tau; D). \tag{4.10}$$

An approach to computing $h_{\text{Regret}}$ is presented in Algorithms 4 and 5.

Equation 4.8 may be interpreted as the expected reward increase given the label for $\boldsymbol{z}$. Further interpretation and theoretical motivation for using this heuristic will be presented in Section 4.4.

**Algorithm 4:** Regret-Based Query Selection

**Given:** $X_t, \mathcal{S}, \mathcal{A}, T, \mathcal{O}, R, \gamma$
**Input:** $\mathcal{U}_t$ // Set of unlabeled image indices
$\tau_0 \leftarrow$ Plan_Trajectory$(X_t, \mathcal{S}, \mathcal{A}, T, \mathcal{O}, R, \gamma)$
**foreach** $i \in \mathcal{U}_t$:
  $\boldsymbol{z} \leftarrow$ semantic_representation$(I_i)$ // Apply semantic feature extractor
  $y_{\text{pred}} \leftarrow$ Predict_Reward$(\boldsymbol{z})$ // Apply reward model $g_\theta$
  $r_0 \leftarrow$ Compute_Regret$(\tau_0, \boldsymbol{z}, 0)$ // See Algorithm 5
  $r_1 \leftarrow$ Compute_Regret$(\tau_0, \boldsymbol{z}, 1)$
  $\text{regret}_i \leftarrow y_{\text{pred}} r_1 + (1 - y_{\text{pred}})\, r_0$
  // Expected regret for Bernoulli distributed $y$
**return** $\operatorname{argmax}_{i\in\mathcal{U}_t} \text{regret}_i$

## 4.3 Experiments

We performed two experiments, each based on simulations of the co-robotic exploration task with various bandwidth constraints to evaluate and compare the active learning heuristics described in Section 4.2. The first experiment used only artificial data, while the second one used data derived from a real coral reef dataset representative of what could be observed in scientific exploration of the Ocean. These experiments involved

**Algorithm 5:** Compute_Regret

**Given:** $X_t, \mathcal{S}, \mathcal{A}, T, \mathcal{O}, R, \gamma$
**Input**: $\tau_0, \boldsymbol{z}, y$ /* Reference trajectory, observation to label, temporary label */
Add_Temporary_Label$(\mathcal{O}, \boldsymbol{z}, y)$
$\tau^\star \leftarrow$ Plan_Trajectory$(X_t, \mathcal{S}, \mathcal{A}, T, \mathcal{O}, R, \gamma)$
$s^\star \leftarrow$ Score_Trajectory$(\tau^\star, \mathcal{O}, R, \gamma)$
$s_0 \leftarrow$ Score_Trajectory$(\tau_0, \mathcal{O}, R, \gamma)$
Remove_Temporary_Label$(\mathcal{O}, \boldsymbol{z})$
**return** $(s^\star - s_0)$ // Regret given the temp label

repeated rollouts of Algorithm 2, including simulated human-robot communication, and the robot was evaluated based on the amount of reward it was able to collect during each mission (rollout). To simplify the comparison, we gave the robot direct access to a pre-generated topic map in each rollout, rather than requiring it to use its own spatial observation model and learned semantic representation. This lets the comparison focus entirely on the active learning methods.

### 4.3.1 Experimental Methodology

The first experiment used 30 artificial "topic maps" (cf. [42]) created by randomly generating Voronoi partitions of a 100×100 image, assigning each cell a topic label, and then assigning each pixel's topic distribution as a distance-weighted mean over cell labels. This produced continuous topic maps with topics in varying concentrations, and each one was associated with a unique interest map (see Figure 4-2). In the second experiment, a single topic map was derived from the expert annotations of an actual coral reef image, and 30 interest maps were generated for it (see Figure 4-3).

The procedure for both experiments was:

1. Generate a map of topic distributions $\boldsymbol{z}(\boldsymbol{x}) \in \boldsymbol{\Delta}^d$ which represent the observations at each location $\boldsymbol{x}$

2. Generate an interest profile $\boldsymbol{p} \in [0, 1]^d$ so that $p = \boldsymbol{p}^T \boldsymbol{z}$ is the probability that the operator is interested in an observation with feature representation $\boldsymbol{z}$

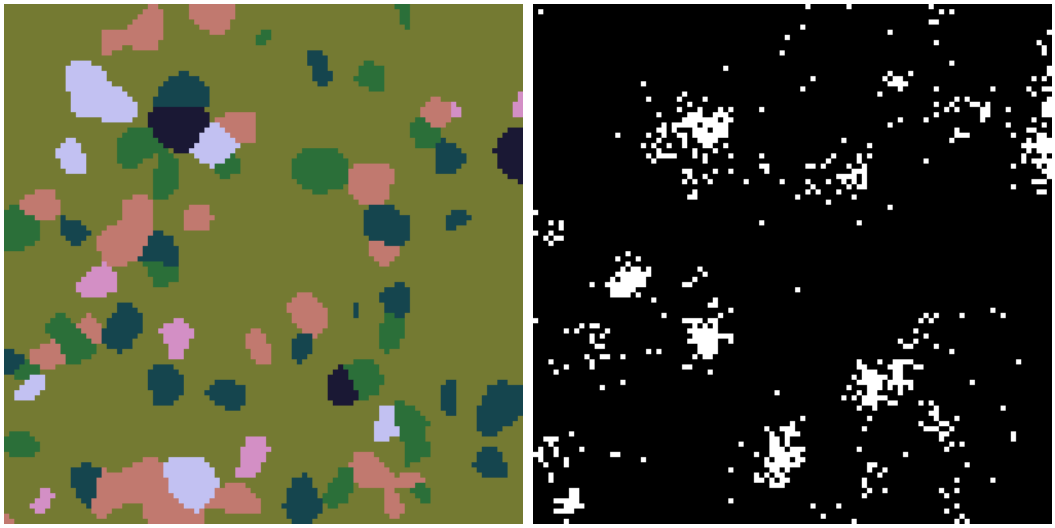

Figure 4-2: Left: A topic map where each location is described by the semantic representation $\boldsymbol{z}^{(i,j)} \in \boldsymbol{\Delta}^8$. The color of each pixel indicates the largest component of $\boldsymbol{z}^{(i,j)}$. Right: The reward at each location is randomly sampled as $R^{(i,j)} \sim \text{Bernoulli}(\boldsymbol{p}^T \boldsymbol{z}^{(i,j)})$, where $\boldsymbol{p} \in [0,1]^k$ represents how "interesting" each component of $\boldsymbol{z}$ is. Here, the pink and black topics are most interesting.

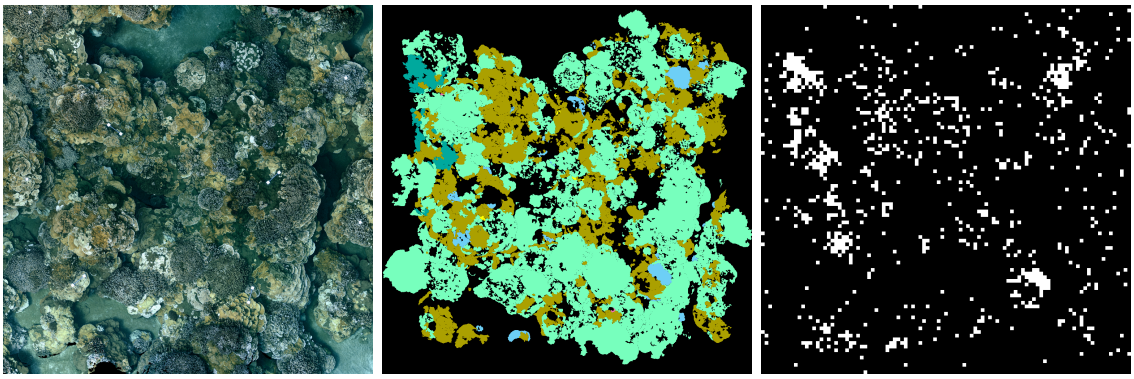

Figure 4-3: Left: A crop of the KAH_2016_3 photomosaic image from the 100 Islands Challenge [92], showing a coral reef near Kaho'olawe. Center: The photomosaic annotations where each color represents an expert label [92]. Right: One of 30 unique interest maps generated (cf. Figure 4-2).

3. Generate a binary "interest map" by sampling $R(\boldsymbol{x}) \sim \text{Bernoulli}(p(\boldsymbol{z}(\boldsymbol{x}))$ at each location $\boldsymbol{x}$ in the topic map

4. For each bandwidth limitation and each query selection algorithm: perform 36 rollouts of Algorithm 2 for a simulated robot making reward queries according to the bandwidth limitation and query selector

Each rollout in step (4) had a duration of 300 timesteps; robot movement was one pixel per timestep and bandwidth constraints were simulated by changing the number of timesteps for a label to be received after being requested. State transitions and observations were deterministic and noiseless. The robot started with no training data and used logistic regression (from [78]) as its reward model. Trajectories were generated by randomly sampling sequences of 5 motion primitives. The primitives were 13 straight lines, each 5 units long and at angles spaced uniformly between -135° to 135° from the robot's current direction. 50 trajectories were generated at each timestep and scored using the sum of the predicted rewards along the trajectory, less the scores of locations already visited. The highest scoring trajectory was followed.

### 4.3.2 Results and Discussion

We compared the Random, Uniform, Information Gain, and Regret query selectors described in Section 4.2 over a total of 69120 simulations. Some examples of the robot trajectories followed during these simulations are presented in Figure 4-4. The mean reward collection rates and interest map prediction losses for each experiment are presented in Figures 4-5 and 4-6. The Regret query selector matches, or outperforms, every other selection criterion at collecting reward, at any bandwidth availability, in these simulation configurations. The relative gains of non-random query selection are smaller when the time between queries is short (high-bandwidth) and thus almost every image is labeled, or when it is so long (low-bandwidth) that the robot barely learns anything before the mission ends. The results also demonstrate the vast improvement of autonomous exploration over preplanned trajectories: the adaptive planners collected up to 29.7% more reward at very low bandwidth, and up to 230%

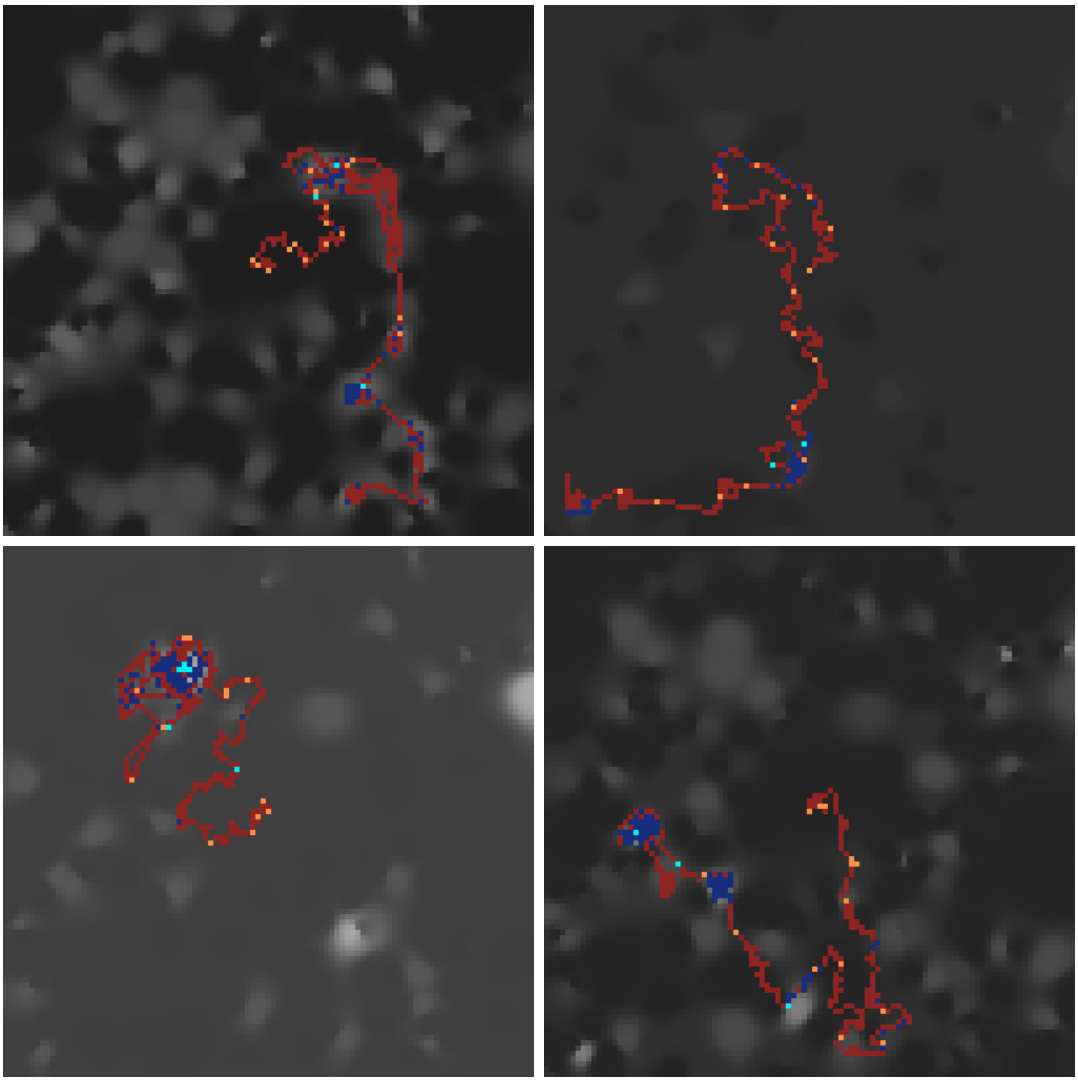

Figure 4-4: Best viewed on a screen. Sample trajectories followed by robots starting at the center of the map in Figure 4-2 with different query selectors. Along each trajectory, red-orange pixels correspond to no reward, and blue pixels to reward. Bright orange/blue pixels represent observations for which the query selector requested the label. The greyscale background intensities represent $g(Z(x); \theta^{\star}_{D})$: reward estimates of observations at each location, based on all labeled samples. Query Selectors: (top row) Random, Uniform; (bottom row) Info Gain, Regret. Appeared in [53].

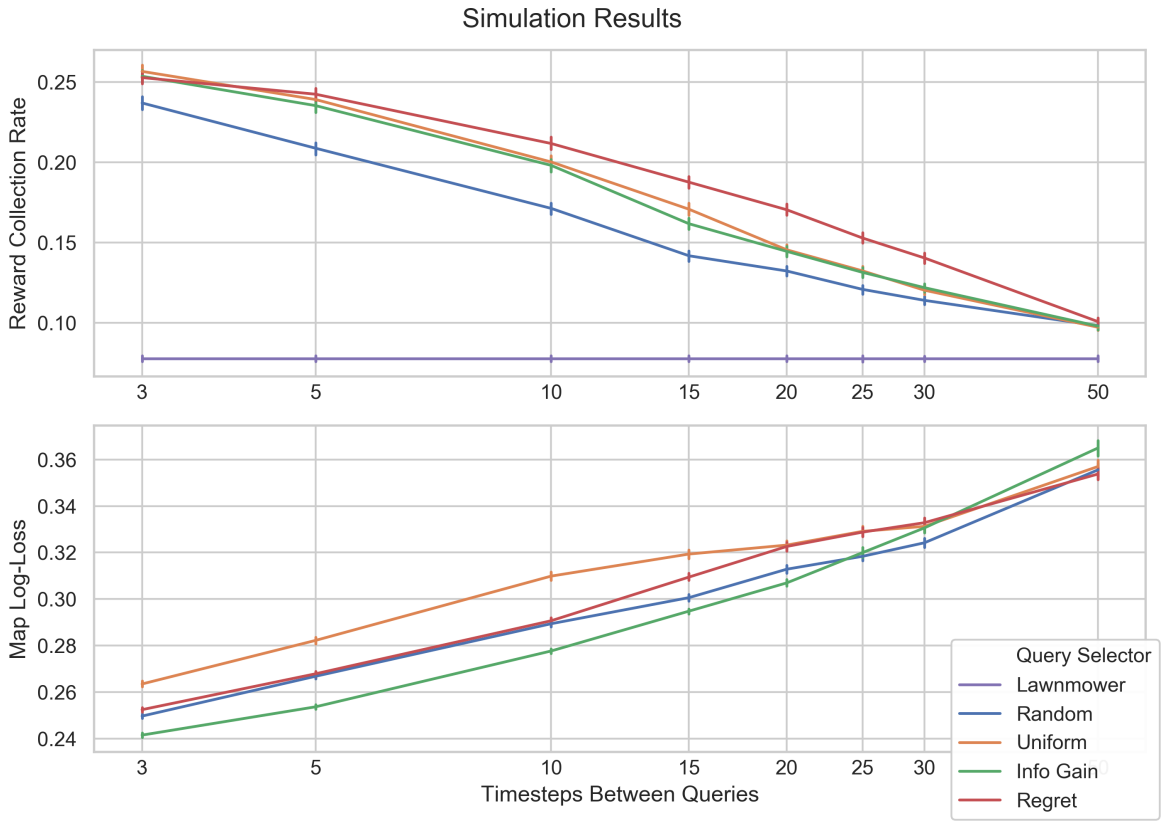


Figure 4-5: A comparison of the query selector performance for different bandwidth availability; the x-axis represents labeling period (time between making a call to Request_Label and Label_Ready returning true in Algorithm 2), which is inversely proportional to bandwidth. Each datapoint represents the mean of 1080 simulations (36 trials on 30 unique maps) and bars represent the 68% confidence bound of the mean. Top: The mean amount of reward collected by each robot per unit time (higher is better). Lawnmower is not a query selector, but rather represents the mean reward collected by 8 preplanned boustrophedonic trajectories [15] that each start at the center of the map and move towards a corner. Bottom: The mean cross-entropy loss between the ground truth interest maps, as seen in the right side of Figure 4-2, and the corresponding robots' predictions of the reward at each location, as seen as the backgrounds in Figure 4-4, at the end of each simulation (lower is better).

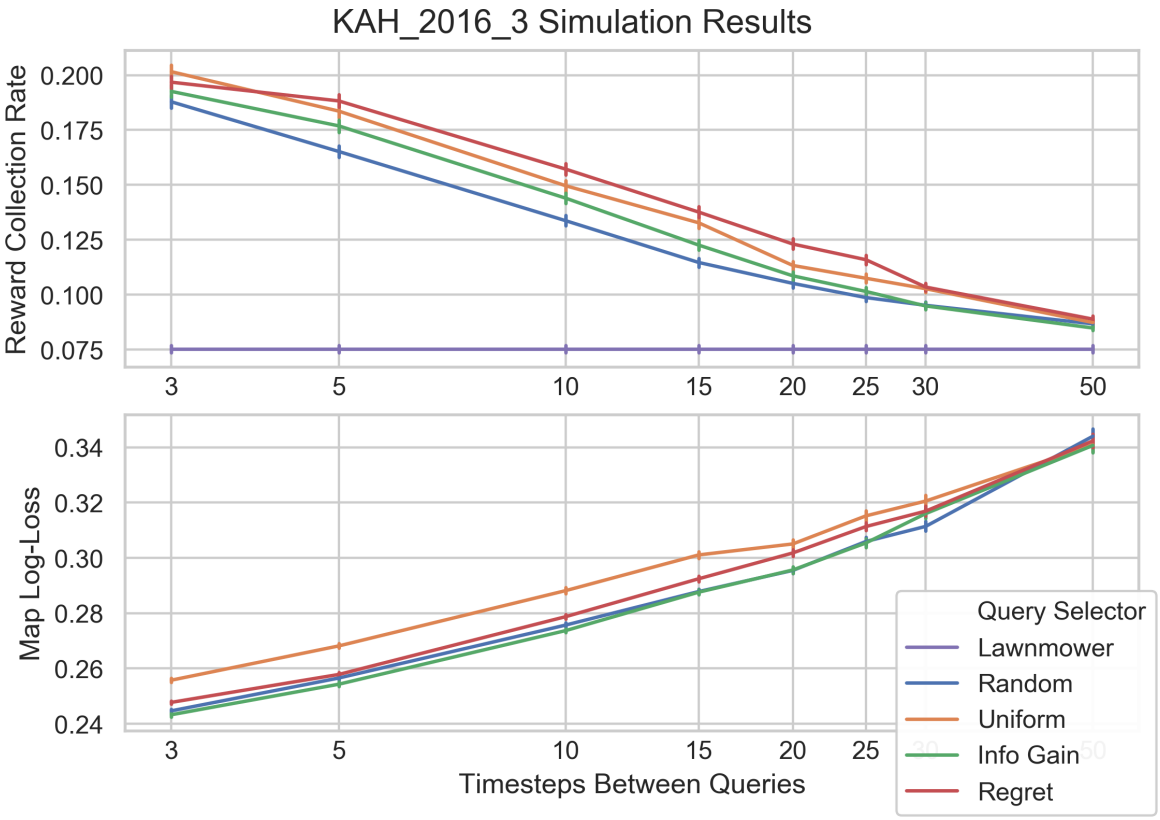


Figure 4-6: The Regret query selector continues to outperform the other active learning heuristics when the topic map is derived from a real image (see Figure 4-3).

more reward at high bandwidth.

The regret-based method did not learn the reward function as well as the information gain query selector, based on its higher map log-loss. This exemplifies the difference in the design criteria: the information theoretic criterion focuses on useful labels for learning a function, which is appropriate for active reward learning *offline*, during training. The regret criterion instead optimizes for the robot's reward, making it better suited for *online* active reward learning, which describes our usage of queries during a live mission.

## 4.4 Motivating Regret-Based Active Learning

As mentioned previously, *regret* is a term often used in robotics literature to denote the difference between how much reward would be collected by a real robot, with access to only limited information, versus how much would be collected by a robot with access to all (relevant) information. Intuitively, an agent experiences regret when it learns new information that causes it to realize its previous choice of action was not optimal. The magnitude of this regret corresponds to how much additional reward a different action would have resulted in. A regret-based active learning approach seeks to *anticipate regret* and ask questions to avoid it.

These intuitions should be captured in our definition of regret. Consider a robot which has at time $t$ a set of $M$ unlabelled examples $Q_t = \{\boldsymbol{z}_i\}_{i=1}^{M}$, a dataset of labelled examples $D_t$, and a reward model $R(\tau; D_t)$.[3] Let us define the augmented dataset $D_t^{Y_m} = D_t \cup \{(\boldsymbol{z}_i, y_i) : \boldsymbol{z}_i \in Q_t, y_i \in Y_t\}_{i=1}^{m}$ for any $m \leq M$. The regret of a trajectory choice $\tau$ given all information (the augmented dataset $D_t^{Y_M}$) is called the **true regret**, and is defined as

$$r(\tau; D_t^{Y_M}) \coloneqq \max_{\tau'} R(\tau'; D_t^{Y_M}) - R(\tau; D_t^{Y_M}). \tag{4.11}$$

Computing the true regret $r(\tau; D_t^{Y_M})$ requires knowing all the unknown labels in $Y_t$,

[3] We use the first $M$ indices of $\boldsymbol{z}$ and $y$ as the unlabelled examples for notational convenience; one could imagine that whenever an example is labelled we swap its index with the most recent example.

which is infeasible when communications bandwidth is limited. We instead consider the **expected true regret** of choosing $\tau$ given our current information $D_t$

$$\bar{r}(\tau; D_t) = \mathbb{E}_{Y_M|D_t}\left[r(\tau; D_t^{Y_M})\right]. \tag{4.12}$$

Referring back to Section 4.2, we can rewrite our original regret heuristic as:

$$h_{\text{Single Regret}}(\boldsymbol{z}_i) = \mathbb{E}_{y_i|D_t}\left[\max_{\tau'} R(\tau'; D_t^i) - R(\tau; D_t^i)\right] \tag{4.13}$$

$$= \mathbb{E}_{y_i|D_t}\left[r(\tau; D_t^i)\right]. \tag{4.14}$$

According to our new notation, $D_t^i = D_t \cup (\boldsymbol{z}_i, y_i) = D_t'$, so the new definition is equivalent to the original one. This definition is equivalent to expected true regret if there is only a single unlabelled question (i.e., if $M = 1$). In this section, we will argue that the success of the choosing queries to maximize the regret-based heuristic in the experiments of Section 4.3 is due to that being an effective approach to minimize $\bar{r}(\tau; D_t)$, which is much harder to compute or work with directly.

### 4.4.1 Minimizing Regret and Maximizing Reward

In some ways, regret and reward are very similar; for example, the maximum expected reward trajectory is the same as the minimum expected regret trajectory

$$\tau^\star_{\text{max reward}} \coloneqq \operatorname*{argmax}_\tau \mathbb{E}_{Y_t|D_t}\left[R(\tau; D_t^{Y_t})\right] \tag{4.15}$$

$$\tau^*_{\text{min regret}} \coloneqq \operatorname*{argmin}_\tau \mathbb{E}_{Y_t|D_t}\left[\max_{\tau'} R(\tau'; D_t^{Y_t}) - R(\tau; D_t^{Y_t})\right] \tag{4.16}$$

$$= \operatorname*{argmin}_\tau \mathbb{E}_{Y_t|D_t}\left[-R(\tau; D_t^{Y_t})\right] = \tau^\star_{\text{max reward}}.$$

This is why we can simply refer to either trajectory as $\tau^\star_{D_t}$, denoting that it is the optimal trajectory based on the available information $D_t$.

However, the intuitions behind the two ideas are very different. In particular, one must consider that *the true reward of each trajectory is fixed.* This means that no matter how many questions the robot asks, the reward associated with a particular

trajectory choice $\tau$ will not change; the only quantities that are updated are the robot's noisy estimates of the trajectories respective reward values. Hopefully, as more labels are received, these estimate converge towards the true reward of each trajectory. Uncertainty reduction techniques seek to reduce the uncertainty in these estimates as quickly as possible, but this is not always useful.

To see this, consider the fact that the robot does not need every label to know with high confidence that it has chosen the highest value path, as shown by the following propositions.

**Definition 1.** Let $\{\tau_i\}_{i=1}^{N}$ be candidate trajectories, and let $Y_M = \{y_j\}_{j=1}^{M} \subseteq Y_t$ be the unknown labels corresponding to the unlabeled queries in $Q_t$. As labels are received, the expected reward of each trajectory is assumed to converge to $R(\tau_i; D_t^{Y_M})$ regardless of the order in which those labels are selected. Furthermore, $\tau$ is called an *optimal trajectory* if and only if $R(\tau; D_t^{Y_M}) \geq R(\tau_i; D_t^{Y_M})\ \forall i$.

**Lemma 1.** *If $\bar{r}(\tau; D_t) = 0$, then $\tau$ is an optimal trajectory.*

*Proof.* If $\bar{r}(\tau; D_t) = 0$ then, for any non-zero probability realization of the unknown labels $\{y_j\}_{j=1}^{M}$, $r(\tau; D_t^{Y_M}) = 0$ and thus

$$\begin{aligned} &\max_i R(\tau_i; D_t^{Y_M}) - R(\tau; D_t^{Y_M}) = 0 \\ &\qquad\qquad \implies R(\tau; D_t^{Y_M}) \geq R(\tau_i; D_t^{Y_M})\ \forall i. \end{aligned}$$

□

**Proposition 1.** *If $\exists m \leq M$ such that, for some ordering of the first $m$ examples and labels, $\bar{r}(\tau; D_t^{Y_M}) = 0$, then $\tau$ is an optimal trajectory.*

*Proof.* Follows trivially from the proof of Lemma 1, where there is no regret for any non-zero probability realization of the remaining unknown labels. □

**Proposition 2.** *Assume the error in each reward estimate is bounded by some $f_i(m)$ that is monotonically decreasing and satisfies $f_i(M) = 0$, such that*

$$\left| R(\tau_i; D_t^{Y_m}) - R(\tau_i; D_t^{Y_M}) \right| \leq f_i(m). \tag{4.17}$$

*If* $\exists m \leq M$ *such that* $R(\tau_i; D_t^{Y_m}) - R(\tau_j; D_t^{Y_m}) \geq f_i(m) + f_j(m)$ *for all* $j \neq i$ *then* $\tau_i$ *is an optimal trajectory.*

*Proof.* Follows by simple substitution of the inequalities

$$R(\tau_i; D_t^{Y_m}) - R(\tau_j; D_t^{Y_m}) \geq f_i(m) + f_j(m) \tag{4.18}$$

$$\implies R(\tau_i; D_t^{Y_m}) - f_i(m) \geq R(\tau_j; D_t^{Y_m}) + f_j(m) \tag{4.19}$$

$$(4.17) \implies -f_i(m) \leq R(\tau_i; D_t^{Y_m}) - R(\tau_i; D_t^{Y_M}) \leq f_i(m) \tag{4.20}$$

$$R(\tau_i; D_t^{Y_m}) - f_i(m) \leq R(\tau_i; D_t^{Y_M}) \tag{4.21}$$

$$(4.17) \implies -f_j(m) \leq R(\tau_j; D_t^{Y_m}) - R(\tau_j; D_t^{Y_M}) \leq f_j(m) \tag{4.22}$$

$$R(\tau_j; D_t^{Y_m}) + f_j(m) \geq R(\tau_j; D_t^{Y_M}) \tag{4.23}$$

Thus, using (4.19), (4.21), and (4.23)

$$R(\tau_i; D_t^{Y_M}) \geq R(\tau_i; D_t^{Y_m}) - f_i(m) \geq R(\tau_j; D_t^{Y_m}) + f_j(m) \geq R(\tau_j; D_t^{Y_M})$$

$$\therefore R(\tau_i; D_t^{Y_M}) \geq R(\tau_j; D_t^{Y_M}) \; \forall j \neq i.$$

So $\tau_i$ is an optimal trajectory as defined in Definition 1. □

The importance of Propositions 1 and 2 is that, while the expected reward of each trajectory converges to some arbitrary value, the expected regret of any optimal trajectory converges to 0 once the robot has asked enough questions. Importantly, this is true as long as the uncertainty bounds decrease after each question. Given a probabilistic bound on $R(\tau_i; D_t^{Y_m})$ instead of the strict bound $f_i(m)$, we would have derived a probabilistic guarantees for Propositions 1 and 2.

The expected regret therefore represents the optimality gap: an estimate of how much additional reward could be collected if the robot had the opportunity to ask questions. Given this perspective, regret-based approaches are focused on asking questions which help to guarantee that the robot knows which actions are optimal or near-optimal.

### 4.4.2 Other Regret-Based Heuristics

Based on the previous results, a regret-based heuristic should prioritize labeling the queries that most quickly minimize the expected regret. A straightforward objective is to (greedily) label the query that results in the largest expected decrease in expected regret:

$$h_{\text{exp-regret}}(q) = \bar{r}(\tau^\star_{D_t}; D_t) - \mathbb{E}_{y|q,D_t}\left[\bar{r}(\tau^\star_{D_t^y}; D_t^y)\right] \tag{4.24}$$
$$D_t^y = D_t \cup \{(q, y)\}\,.$$

This computes the difference in expected regret between the optimal trajectory at time $t$ without knowing $(q, y)$, which is $\tau^\star_{D_t}$, compared to the optimal trajectory knowing $(q, y)$, which is $\tau^\star_{D_t^q}$. Unfortunately, it is generally intractable to compute expected regret $\bar{r}(\tau; D_t)$ for any $\tau$ since it is an expectation over the potentially large number of variables in $Y_t$ with unknown distributions. In fact, the number of unknown labels increases linearly over the course of the mission, so the complexity of the expectation grows exponentially. Thus, instead of computing this heuristic directly, we will begin with a very simple approximation and gradually build upon it. The heuristics considered are presented together in Table 4.2.

#### Single-Query Regret for Binary Reward

As discussed earlier, the single-query heuristic $h_{\text{Regret}}$ discussed in Section 4.2 is equivalent to expected regret if there is were a single unknown label. In the experiments in Section 4.3, we assumed that $y$ was binary valued $y \in \mathcal{R} = \{0, 1\}$ and Bernoulli distributed with estimated mean $\mathbb{E}[y \mid q, D_t] = \mu(y; q, D_t) \in [0, 1]$. The estimate of the mean was computed using the reward model as $\mu(y; q, D_t) = R(q; D_t)$. Thus, the heuristic was computed as:

$$h_{\text{binary-regret}}(q) = R(q; D_t) r\left(\tau^\star_{D_t}; D_t \cup (q, 1)\right) + (1 - R(q; D_t))\, r\left(\tau^\star_{D_t}; D_t \cup (q, 0)\right). \tag{4.25}$$

Table 4.2: Comparison of various regret-based single-query heuristics.

| Heuristic | Definition of $h(q)$ | Equivalent of $\max_q h(q)$ | Intuition |
|---|---|---|---|
| exp-regret | $\bar{r}(\tau^\star_{D_t}; D_t) - \mathbb{E}_{y\|q,D_t}\left[\bar{r}(\tau^\star_{D_t^y}; D_t^y)\right]$ | $\max_q \mathbb{E}_{Y\|D_t}\left[R\left(\tau^\star_{D_t^y}; D_t^Y\right) - R\left(\tau^\star_{D_t}; D_t^Y\right)\right]$ | Maximize increase in expected reward given $y \in Y$, the label for query $q$ |
| single-regret | $\mathbb{E}_{y\|q,D_t}\left[r\left(\tau^\star_{D_t}; D_t^y\right)\right]$ | $\max_q \mathbb{E}_{y\|q,D_t}\left[R\left(\tau^\star_{D_t^y}; D_t^y\right) - R\left(\tau^\star_{D_t}; D_t^y\right)\right]$ | Same as exp-regret, but compute regret as if $y \in Y$ was the only unknown label |
| max-regret | See Equation (4.30) | $\min_q \sup_{Y:\Pr(Y)\geq\delta}\left(R\left(\tau^\star_{D_t^y}; D_t^Y\right) - R\left(\tau^\star_{D_t}; D_t^Y\right)\right)$ | Minimize the regret in the worst case scenario of $Y$ with likelihood at least $\delta$ |

Unfortunately, this approach cannot be easily adapted for non-binary rewards, and especially not real-valued rewards.

### Real-Valued Regret

There are many ways in which the single-query regret heuristic could be extended to support real-valued rewards. However, in order to compute the expectation in Equation 4.13 we must have some knowledge of the probability distribution. For example, if we assume that the reward space $\mathcal{R} = [a, b] \subset \mathbb{R}$ is bounded, we can estimate the mean $\mu_y$ and variance $\sigma_y^2$ of the reward label $y$ for any particular query $q$. Nonetheless, we will *not* assume that the unknown labels are normally distributed; in fact, we will not assume any distribution.

Cantelli's inequalities provide probabilistic bounds on the value we will receive for any particular label, based on only its predicted mean and variance:

$$\Pr\left(y - \mu_y \leq \sigma_y \sqrt{\frac{1-\delta}{\delta}}\right) \geq 1 - \delta \tag{4.26}$$

$$\Pr\left(|y - \mu_y| \leq \sigma_y \sqrt{\frac{1-\delta}{\delta}}\right) \geq 1 - 2\delta. \tag{4.27}$$

One approach to handling real-valued regret is to consider the $(1-2\delta)$-probability case by computing the regret for each of $y = \mu_y \pm \sigma_y \sqrt{(1-\delta)/\delta}$, while ensuring that $y$ stays within the interval $[a, b]$. This can be used to upper bound regret even for real-valued reward and unknown distributions, as seen in the following subsection.

### Minimizing Worst-Case Regret

A limitation of the single-query regret heuristic is that it underestimates the true regret. In particular, the single-query regret may be zero even if the expected regret is relatively high. This occurs because the single-query heuristic does not take into account interactions between the answers of multiple queries, which are especially significant between correlated queries. To best understand these interactions and their impact on regret, we shift our analysis from focusing on queries and their unknown

labels to focus instead on how multiple queries interact with the trajectories.

While the robot does not know most of the labels, it can use the labels it does have to estimate the mean and variance of the rest. Then, with the following theorem it can compute a probabilistic upper bound of the regret of its current trajectory choice.

**Theorem 1.** *Let $\mathbf{y}_t \in \mathbb{R}^M$ be the random variable representing the unknown labels $Y_M$, with known $\boldsymbol{\mu}_t \coloneqq \mathbb{E}\left[\mathbf{y}_t \mid D_t\right]$ and $S_t \coloneqq \mathbb{E}\left[(\mathbf{y}_t - \boldsymbol{\mu}_t)(\mathbf{y}_t - \boldsymbol{\mu}_t)^\top \mid D_t\right]$. Suppose we are given $\boldsymbol{\rho}_{\tau_i} \in \mathbb{R}^M$ such that $\Delta R(\tau_i; D_t) \leq \boldsymbol{\rho}_{\tau_i}^\top(\mathbf{y}_t - \boldsymbol{\mu}_t), \forall i$. Then, the regret of a trajectory choice $\tau^\star_{D_t}$ has the following probabilistic upper bound:*

$$Pr\left(r(\tau^\star_{D_t}; D_t^{Y_t}) \geq \max_\tau \sqrt{\frac{1-\delta}{\delta}}\left\|\boldsymbol{m}_t(\tau)\right\| - r(\tau; D_t)\right) \leq \delta \tag{4.28}$$

$$\boldsymbol{m}_t(\tau) = S_t^{0.5}\left(\boldsymbol{\rho}_\tau - \boldsymbol{\rho}_{\tau^\star_{D_t}}\right) \tag{4.29}$$

*Proof.* Let $\Delta R(\tau; D_t) \coloneqq R(\tau; D_t^{Y_M}) - R(\tau; D_t)$ and rewrite (4.11) as

$$\begin{aligned} r(\tau^\star_{D_t}; D_t^{Y_M}) &= \max_\tau \left(R(\tau; D_t) + \Delta R(\tau; D_t)\right) - \left(R(\tau^\star_{D_t}; D_t) + \Delta R(\tau^\star_{D_t}; D_t)\right) \\ &= \max_\tau \left(\Delta R(\tau; D_t) - \Delta R(\tau^\star_{D_t}; D_t)\right) - r(\tau; D_t). \end{aligned}$$

Use Cantelli's inequality (4.26) to show that

$$\Pr\left(\Delta R(\tau; D_t) \geq \sqrt{\frac{1-\delta}{\delta}}\left\|S_t^{0.5}\boldsymbol{\rho}_\tau\right\|\right) \leq \delta$$

□

Thus, even when it is intractable to compute the expected regret (due to the unknown labels in $Y_t$ with unknown distributions), by using Theorem 1 the robot can upper bound the regret of any particular trajectory choice using only estimates of each unknown label's mean and variance. One heuristic to minimize this bound for some probability threshold $\delta$ is, for $\tau' = \text{argmax}_\tau \sqrt{\frac{1-\delta}{\delta}}\left\|\boldsymbol{m}_t(\tau)\right\| - r(\tau; D_t)$:

$$h(q) = \left|\left[\boldsymbol{m}_t(\tau')\right]_q\right|. \tag{4.30}$$

### Minimizing Maximum Expected Regret

Minimizing the worst-case regret requires setting the probability threshold $\delta$, which can have a significant impact on determining the query to label next. In particular, values of $\delta$ near 1 will give the trivial result that $\Pr(r(\tau^\star_{D_t}; D_t^{Y_t}) \geq 0) \leq 1$ whereas values near 0 will greatly overestimate the true regret. To set $\delta$, we use the following corollary to find the value that minimizes the maximum *expected* regret.

**Corollary 1.** *Suppose that the unknown reward labels are bounded as* $0 \leq y_i \leq 1$ *for* $i = 1, \ldots, |Y_t|$ *and that the maximum trajectory length is* $N$*. Then, it holds that:*

$$\bar{r}(\tau^\star_{D_t}; D_t) \leq \min_{\delta} \max_{\tau} (\gamma_\tau - \delta) \left( \sqrt{\frac{1-\delta}{\delta}} \|\boldsymbol{m}_t(\tau)\| - r(\tau; D_t) \right) + \delta N \tag{4.31}$$

$$\gamma_\tau = \frac{\|\boldsymbol{m}_t(\tau)\|^2}{\|\boldsymbol{m}_t(\tau)\|^2 + r(\tau; D_t)^2} \tag{4.32}$$

*Proof.* By Theorem 1, then $\delta \geq \gamma_\tau \implies \sqrt{\frac{1-\delta}{\delta}} \|\boldsymbol{m}_t(\tau)\| - r(\tau; D_t) \leq 0$. It also holds that $0 \leq r(\tau^\star_{D_t}; D_t^{Y_t}) \leq N$ due to the bounds on the reward labels. Since the bound in Theorem 1 must hold for any $\delta \in [0, 1]$, then $\max_\tau (\gamma_\tau - \delta) \left( \sqrt{\frac{1-\delta}{\delta}} \|\boldsymbol{m}_t(\tau)\| - r(\tau; D_t) \right) + \delta N$ is an upper bound of $\bar{r}(\tau^\star_{D_t}; D_t)$ for any $\delta$. By choosing $\delta$ to minimize this quantity we find the tightest upper bound on expected regret. □

The value of $\delta$ that minimizes (4.31) will represent a balance between regret bounds that are too optimistic or too conservative. Note that once $\delta$ is fixed to this value, choosing a query to minimize (4.31) is equivalent to choosing a query to minimize (4.28); thus Corollary 1 does not provide a new approach to query selection.

## 4.4.3 Multi-Query Regret

One more issue with the heuristics presented thus far is that they only consider greedy approaches to minimizing expected regret based on asking a single question. However, we have not found any theoretical results to suggest that the optimal next query can be identified without considering all possible combinations of queries. In fact, it is likely that the fastest way for the robot to reduce its regret will require consideration

of sets of questions that could be asked and the optimal order in which to ask the questions in those sets. This is because, for many reward models (including logistic regression) every unknown label is correlated to every other unknown label, and the effect of a new label on the model will depend on all previous labels.

Fortunately, some of the heuristics considered here might be easily adapted to consider combinations of queries. In particular, the vector $\boldsymbol{m}_t(\tau)$ defined in Equation 4.29 is interesting in that the $m^{\text{th}}$ element represents, in a sense, the expected impact of asking the $m^{\text{th}}$ label on trajectory $\tau$. It is strongly related to uncertainty in the reward of trajectory $\tau$ based on the relationship $\mathbb{E}[\Delta R(\tau; D_t)^2] = \|\boldsymbol{m}_t(\tau)\|^2$. By concatenating these vectors into a matrix $M_t = \Big[\boldsymbol{m}_t(\tau_1) \quad \dots \quad \boldsymbol{m}_t(\tau_N)\Big]$, it may be possible to use the rows of this vector to understand the cumulative impact of a query on the robot's reward model, turning the multi-query selection problem into a matrix row-selection problem. For certain objective functions, these problems are well studied in the field of spectral graph theory.

### 4.4.4 Conclusions

Regret-based online active learning criteria are effective at maximizing the reward collected because they explicitly model the impact of queries on the trajectory choices available. In particular, they focus on acquiring the information that makes the robot confident it has chosen the best trajectory, even if there is still uncertainty about the exact reward value of each trajectory. In bandwidth-limited environments, this is much more effective than simply reducing the largest uncertainties in the reward model which may not always be relevant to identifying the highest scientific value trajectories. This is why regret-based criteria can outperform information-theoretic criteria in these contexts. While the regret-based criterion presented in Section 4.2 was quite simple, more sophisticated regret based criteria are being explored that can better estimate the true change in expected regret associated with a query through modelling the interactions between queries.

# Chapter 5

# Conclusions and Future Work

This thesis began by asking how a team composed of humans and robots could work together to explore new environments as effectively as possible, even if it were difficult for them to communicate with each other. It turned out that there were few guiding principles in the field of human-robot cooperative exploration, and fewer still general-purpose approaches to scientific exploration. We presented three main axes along which a robotic explorer's capability for autonomous scientific exploration in bandwidth-limited environments could be measured: the ability to model the spatial distribution of relevant phenomena, the ability to learn and model the mission objectives, and the ability to plan trajectories with high scientific value according to these models. We labelled certain milestones along these axes in Table 1.3, and found that no current scientific explorers have yet reached the higher levels of autonomy. However, we have now successfully demonstrated in simulations the capability for autonomy at the second highest level, human-robot cooperative exploration, without requiring domain-specific modelling and with significant improvements in the effectiveness of scientific exploration regardless of the bandwidth limitation.

In our review of previous works, we found that existing path planning algorithms such as Monte Carlo Tree Search could be used to plan trajectories with high scientific value, if given a spatial model of relevant scientific phenomena and a reward model which estimated how relevant these phenomena were to the mission objectives. However, there was little prior research into the design of general-purpose spatial models and

online reward model learning over low bandwidth. This motivated our contributions into these two areas, wherein we extended a previous solution for unsupervised semantic mapping into a general-purpose approach to spatial observation modelling, and presented an approach for learning complex reward models online with as few examples as possible using a novel active learning strategy. These will be discussed further in the following section.

## 5.1 Thesis Contributions

Here we will present a brief summary of each of the contributions made in the previous chapters.

### 5.1.1 Topic-Model Based Spatial Observation Modelling

This thesis' first major contribution was identifying the utility of spatio-temporal topic models to aid in scientific exploration, and particularly their suitability as spatial observation models. In Chapter 3, we presented extensions to BNP-ROST [45] which made it more suitable for large-scale semantic mapping of 3D environments. This new "Sunshine" system enabled online mission summarization, a valuable tool for scientists. Sunshine was also revealed to be an effective spatial observation model, as it is highly effective at novelty detection and its Bayesian prior for the semantic labels of unexplored areas was exactly the low-dimensional semantic representation we needed for spatial observation modelling. We concluded this chapter by presenting techniques for tuning Sunshine in order for it to best model the environment to be explored; in particular, we demonstrated how Sunshine could be tuned to capture the spatial "patchiness" and complexity of natural environments.

### 5.1.2 Online Active Reward Learning for Efficient Mission Objective Understanding

This thesis' next major contribution was the novel "regret"-based active learning criterion presented in Chapter 4, which was much better suited than previous criteria for guiding human-robot communication in bandwidth-limited environments. In particular, Section 4.3 demonstrated how a regret-based criterion could outperform an information theoretic active learning criterion in enabling a robotic explorer to collect as much reward as possible during a mission. Perhaps counter-intuitively, this was accomplished despite the information-theoretic approach enabling the robot to learn the scientific objectives with less error by reducing the reward model uncertainty faster. Section 4.4 presented intuitive justification for why it is not necessary to eliminate *all* uncertainty in the reward model in order for the robot to plan high scientific value trajectories; it found that prioritizing the reduction of some uncertainties more than others could help the robot to more quickly reduce or eliminate the suboptimality of its trajectory choice.

### 5.1.3 Other Contributions

This thesis laid new groundwork for understanding how autonomy can play a role in efficient co-robotic scientific exploration. This began by presenting the taxonomy of robotic scientific exploration in Section 1.4, with which previous robotic exploration systems capable of varying degrees of autonomy could be organized. It continued with the Co-Robotic Visual Exploration POMDP presented in Section 4.1, which is a structured but flexible approach to managing human-robot collaboration and high-dimensional observation spaces. This thesis provided general principles for choosing the POMDP's observation model, reward model, and active learning criterion. Finally, the experiments in Section 4.3 revealed how communication bandwidth constraints could impact the effectiveness of an autonomous exploration system, and how even a small amount of communication can, with the right utilization, enable much more efficient scientific exploration than traditional approaches.

## 5.2 Future Work

This section will explore four avenues through which the effectiveness of co-robotic scientific exploration could be improved further. Subsection 5.2.1 will explore potential improvements to the spatial observation model presented in Chapter 3, while Subsection 5.2.2 will discuss one ongoing area of investigation into improving regret-based active learning by considering multiple queries at a time. Subsection 5.2.3 will revisit the idea of multi-robot federated exploration discussed in Section 1.4 and discuss approaches to address the challenge of associating the semantic representations learned by various robots' spatial observation models. Finally, Subsection 5.2.4 will discuss the importance of developing public datasets that could be used for comparing approaches to co-robotic scientific exploration.

### 5.2.1 Hierarchical and Long-Range Spatial Observation Models

One area for improvement in the Sunshine system presented in Chapter 3 is its vocabulary of visual words. Due to the vocabulary consisting of very low-level image features, many types of words are often required to describe a single phenomena, thus making the topic-word distributions quite complex and hard to learn. Previous work has explored replacing this vocabulary with higher-level image features derived from a Convolutional Neural Network [35]. This has the advantage of enabling the topic model to learn good topics much faster by requiring fewer word types to characterize each topic. Hierarchical topic models have shown remarkable flexibility for modelling highly complex visual hierarchies (e.g., [91]), but these are usually limited by requirement for heavy computational power not yet available on mobile robotic platforms. If the topic model is trained on a vocabulary that requires fewer words to characterize each topic, or if the available compute power on robotic platforms increases, learning these hierarchies may become more computationally feasible. In this case, future robotic explorers may be able to model the scientist's interest in more abstract phenomena like ecosystems, rather than simple visually distinct objects.

Another opportunity to improve Sunshine as a spatial observation model is to develop a more sophisticated spatial prior than the neighbourhood prior for making predictions about the semantic content of nearby locations. The current prior predicts that the phenomena at a new location will be a simple distance-weighted mixture of the phenomena observed at nearby locations, but natural phenomena can have much more complex spatial distributions and modelling these is an area of active research [29, 79]. Recent work has explored extending spatio-temporal topic models to capture more complex spatial distributions by combining them with Gaussian Processes (GPs) [87]. Incorporating such an extension, or domain-specific extensions for modelling what is known about specific phenomena of interest (e.g., marine fauna), may lead to much more accurate spatial observation modelling and therefore increased scientific return.

### 5.2.2 Multi-Query Regret-Based Active Learning Objectives

In Section 4.4, we demonstrated that the change in expected regret based on asking a single question is an effective active learning criterion. We also discussed how any active learning criterion which seeks to minimize the expected regret of the robots actions as quickly as possible (thus maximizing the amount of reward collected during the mission) must consider how the choice of query will affect future queries. This is a multi-query online active learning problem, and we have ongoing study into this area.

### 5.2.3 Multi-Robot Federated Exploration

The highest level of autonomy in co-robotic scientific exploration presented in Table 1.3 is Multi-Robot Federated Exploration, and describes a multi-robot system in which the robots communicate among each other to disseminate relevant information more efficiently. For example, sharing observations of known locations enables each robot to build a better spatial observation model, especially if the robots are far apart, while sharing observation-label pairs enable each robot to learn the reward model quickly without unnecessarily repeating queries to the scientist. If there is high communication

bandwidth availability between robots, it may be feasible to share the raw image observations between robots, making it easy to directly incorporate them in the new robot's models. However, in some environments such as the ocean (see Section 1.2), communication between robots is very limited if they are more than just a couple hundred metres apart from each other.

One approach to address this issue is for the robots to share their semantic representations of images among each other, rather than the raw images. However, in the case of topic-model based semantic image representations, one robot's representation is not easily interpretable to another robot because they learn these representations *online*. Thus, the robots are likely to use a variety of distinct semantic labels to represent the same phenomenon, especially if there are many different phonomena in the environment. Some previous work which started to address this issue explored sharing the topic-word distribution matrices between robots regularly, thus ensuring that all robots would eventually converge towards shared semantic representations [23]. However, this approach has some key limitations: first, it requires each robot to send and receive the entire topic model on each update, which may still be costly over a limited communications channel. Second, it required that after every merge each robot replaced its individual model with the new global model, which can cause problems if the round-trip communication latency involved in a merge is high (and thus the new global model becomes outdated). Finally, it fails to detect the case where two or more topics learned by one robot correspond to a single topic learned by another, and these failed merges cause the individual topic models to converge more slowly.

A new approach that we are exploring with Kaveh Fathian and Kasra Khosoussi is to solve the association problem between each robots unique set of semantic labels. This enables the robots to communicate labels and observations among each other, and potentially still merge topic models if appropriate, with fewer drawbacks and less communication required. Of particular interest is using the CLEAR algorithm [33], which is capable of efficiently solving the multi-agent data association problem to find all sets of equivalent semantic labels across all robots, to find much better correspondences than the 1-to-1 correspondences solved using the Hungarian algorithm in [23].

### 5.2.4 High Fidelity Robotic Exploration Datasets

It would be very beneficial to researchers in autonomous science and co-robotic exploration to have access to a consistent set of datasets. In the experiments in Section 4.3, the simulation environment was created from scratch, and while some of the data used was derived from the map of a real coral reef there are no community-recognized metrics by which to describe this dataset. For future works, it would be preferable if the research community had a set of datasets of natural environments representative of the spatial sparsity, concentration, and scale of various phenomena of interest. Given this, future work in areas like spatial observation modelling and online active reward learning could better support claims that they outperform baselines at collecting observations of more/less spatially sparse phenomena, more/less concentrated phenomena, larger/smaller scale phenomena, and so on.

## 5.3 Closing Remarks

Scientific exploration is an exciting endeavour that unifies humanity through our endless curiosity to understand the strange and the unknown. Given the extraordinary effort and expense that goes into robotic scientific exploration of environments like the deep sea and other planets, it is essential to maximize the productivity of our robotic explorers. Historically, however, this productivity has been strictly limited by our ability to command and control these explorers in highly remote and bandwidth-limited environments. In this thesis, we found that increasing the level of autonomy of these explorers to the point of human-robot cooperative exploration resulted in the mission producing several times as many scientifically interesting observations when compared to traditional techniques. In particular, we have shown that our novel approaches to spatial observation modelling and online active reward learning, which explicitly consider bandwidth limitations, enable autonomous planning of high scientific value trajectories with minimal communication. We hope that our contributions to these areas have a significant long-term impact in the future of co-robotic scientific exploration.

# Appendix A

# Figures

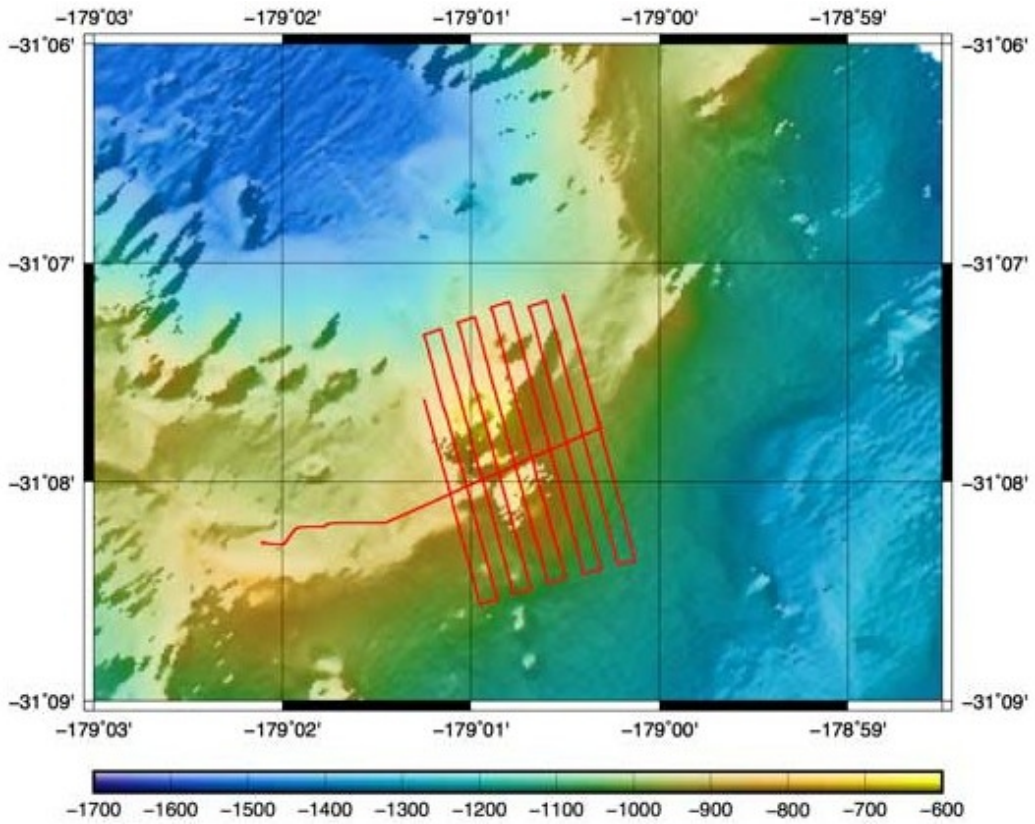


Figure A-1: An example of a path followed by the deep sea AUV Sentry. The red trackline resembles a lawnmower pattern, and is an example of Boustrophedonic coverage of the target area [14]. The target area is usually chosen carefully based on prior beliefs about something of interest existing in the general vicinity. Colour is used to indicate ocean depth in metres. © WHOI, 2015 [55].

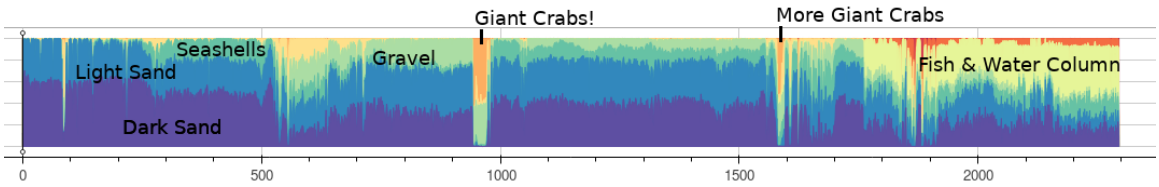


Figure A-2: This figure, which was adapted from Figure 4 in [35], shows the mission summary of an AUV deployment described in [79]. Each vertical slice of the timeline shows the proportion of semantic labels (e.g., colours in Figure 2-1 right image) observed at a particular moment during the mission. The timeline was produced by a spatio-temporal topic model, while the words used to label the colours were added manually. This timeline represents almost 2300 images (over 2.5 hours of mission data) using less than 100 KB, making it compact enough to stream over even very low communication bandwidth channels. Furthermore, since this representation uses only a handful of unique labels, it is feasible for a scientist to request representative examples of them; it only required inspecting around 5 images to be able to add words to annotate the meaning of each colour.

Figure A-3: Here we show the WARPLab AUV "Red", a modified BlueRobotics BlueROV, exploring an artificially created underwater environment at WHOI. Appeared in [42].

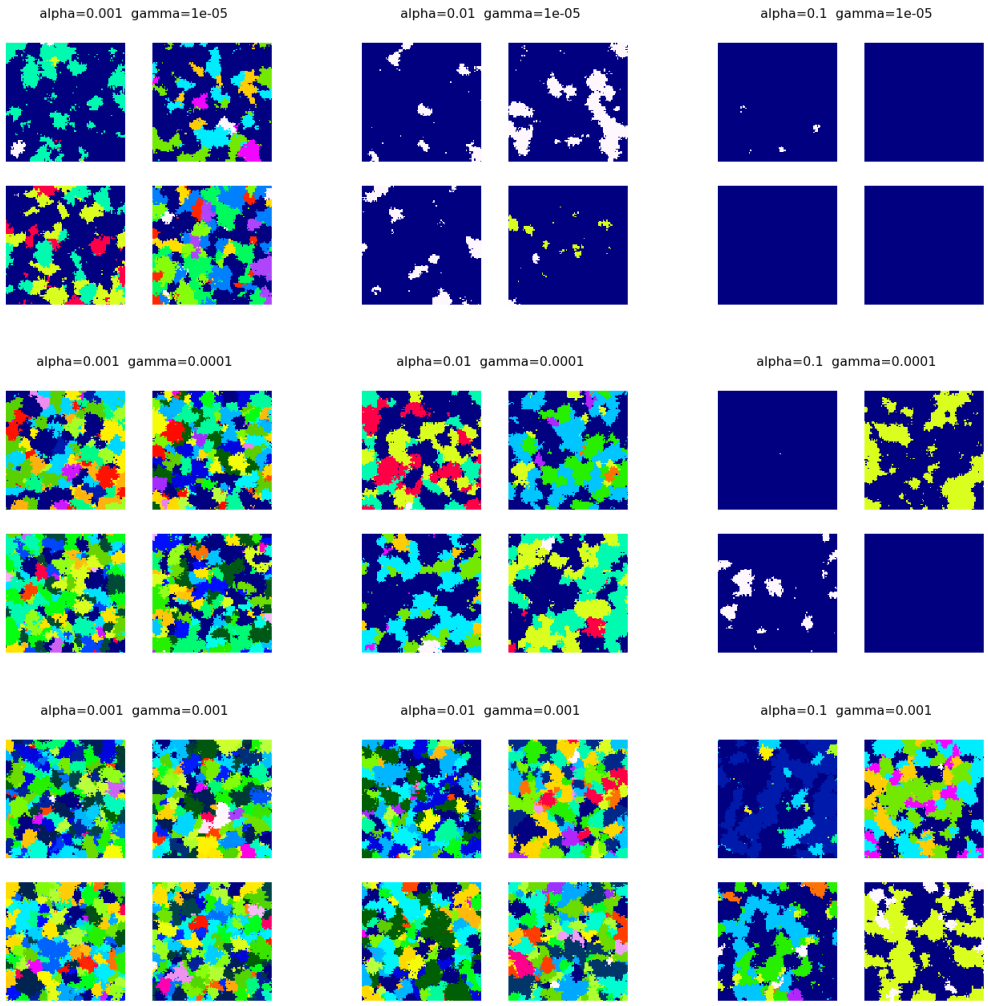


Figure A-4: These maps were randomly generated by sampling from the BNP-ROST prior described in Section 3.1 for different values of the hyperparameters $\alpha$ and $\gamma$. As discussed in Section 3.2, the prior can be tuned to create maps with varying degrees of "patchiness" and numbers of different semantic labels (topics). Appeared in [42].

# Bibliography


[1] Akash Arora, Robert Fitch, and Salah Sukkarieh. "An Approach to Autonomous Science by Modeling Geological Knowledge in a Bayesian Framework". In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Sept. 2017, pp. 3803–3810. ISBN: 978-1-5386-2682-5. DOI: 10/ggb45r. arXiv: 1703.03146. URL: http://dx.doi.org/10.1109/IROS.2017.8206230.

[2] Chris L. Baker and Joshua B. Tenenbaum. "Chapter 7 - Modeling Human Plan Recognition Using Bayesian Theory of Mind". In: *Plan, Activity, and Intent Recognition*. Ed. by Gita Sukthankar, Christopher Geib, Hung Hai Bui, David V. Pynadath, and Robert P. Goldman. Boston: Morgan Kaufmann, Jan. 2014, pp. 177–204. ISBN: 978-0-12-398532-3. DOI: 10.1016/B978-0-12-398532-3.00007-5. URL: http://www.sciencedirect.com/science/article/pii/B9780123985323000075.

[3] Maria Florina Balcan, Steve Hanneke, and Jennifer Wortman Vaughan. "The True Sample Complexity of Active Learning". In: *Machine Learning* 80.2-3 (2010), pp. 111–139. ISSN: 08856125. DOI: 10/btc9cq.

[4] Robert D Ballard. *WHOI-93-34: The JASON Remotely Operated Vehicle System*. Tech. rep. Woods Hole, Massachusetts: Woods Hole Oceanographic Institution, 1993.

[5] Jianhua Bao, Daoliang Li, Xi Qiao, and Thomas Rauschenbach. "Integrated Navigation for Autonomous Underwater Vehicles in Aquaculture: A Review". en. In: *Information Processing in Agriculture* 7.1 (Mar. 2020), pp. 139–151. ISSN: 2214-3173. DOI: 10/ggwnfk. URL: http://www.sciencedirect.com/science/article/pii/S221431731930071X (visited on 05/19/2020).

[6] J.J. Biesiadecki and M.W. Maimone. "The Mars Exploration Rover Surface Mobility Flight Software Driving Ambition". In: *2006 IEEE Aerospace Conference*. Big Sky, MT, USA: IEEE, Mar. 2006. ISBN: 0-7803-9545-X. DOI: 10/c646pv.

[7] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Ed. by M Jordan, J Kleinberg, and B Schölkopf. Springer-Verlag New York, 2006. ISBN: 978-0-387-31073-2.

[8] David M Blei, Andrew Y Ng, and Michael I Jordan. "Latent Dirichlet Allocation". In: *Journal of Machine Learning Research* 3 (2003), pp. 993–1022. ISSN: 15324435. DOI: 10/fc8s6g. arXiv: 1111.6189v1.

[9] Maxime Bucher, Tuan-Hung Vu, Matthieu Cord, and Patrick Pérez. "Zero-Shot Semantic Segmentation". In: (2019), pp. 1–15. arXiv: 1906.00817. URL: http://arxiv.org/abs/1906.00817.

[10] Guy Burroughes and Yang Gao. "Ontology-Based Self-Reconfiguring Guidance, Navigation, and Control for Planetary Rovers". In: *Journal of Aerospace Information Systems* 13.8 (2016), pp. 316–328. ISSN: 23273097. DOI: 10/ggb48k.

[11] Liangliang Cao and Li Fei-Fei. "Spatially Coherent Latent Topic Model for Concurrent Segmentation and Classification of Objects and Scenes". In: *Proceedings of the IEEE International Conference on Computer Vision* (2007). DOI: 10/cmp782.

[12] Rebecca Castano, Tara Estlin, Daniel Gaines, Andres Castano, Caroline Chouinard, Ben Bornstein, Robert C. Anderson, Steve Chien, Alex Fukunaga, and Michele Judd. "Opportunistic Rover Science: Finding and Reacting to Rocks, Clouds and Dust Devils". In: *IEEE Aerospace Conference Proceedings*. Vol. 2006. IEEE, 2006, pp. 1–16. ISBN: 0-7803-9546-8. DOI: 10/db33np. URL: http://ieeexplore.ieee.org/document/1656011/.

[13] Steve Chien, Rob Sherwood, Daniel Tran, Benjamin Cichy, Gregg Rabideau, Rebecca Castano, Ashley Davis, Dan Mandl, Stuart Frye, Bruce Trout, Seth Shulman, and Darrell Boyer. "Using Autonomy Flight Software to Improve Science Return on Earth Observing One". In: *Journal of Aerospace Computing, Information, and Communication* 2.April (2005), pp. 196–216. DOI: 10/cnpmnd.

[14] Howie Choset. "Coverage of Known Spaces: The Boustrophedon Cellular Decomposition". In: *Autonomous Robots* 9 (2000), pp. 247–253. DOI: 10/dkc9cw.

[15] Howie Choset and Philippe Pignon. "Coverage Path Planning: The Boustrophedon Cellular Decomposition". In: *Field and Service Robotics* (1998), pp. 203–209. ISSN: 09295593. DOI: 10/bkwz8h. arXiv: 1011.1669v3.

[16] M. Elizabeth Clarke, Nick Tolimieri, and Hanumant Singh. "Using the Seabed AUV to Assess Populations of Groundfish in Untrawlable Areas". In: *The Future of Fisheries Science in North America* (2009), pp. 357–372. DOI: 10/bg5cmv.

[17] Rémi Coulom. "Efficient Selectivity and Backup Operators in Monte-Carlo Tree Search". In: *5th International Conference on Computer and Games*. Vol. inria-00116992. Turin, Italy, May 2006, pp. 72–83. URL: https://hal.inria.fr/inria-00116992/document.

[18] Antoine Cully, Konstantinos Chatzilygeroudis, Federico Allocati, and Jean-Baptiste Mouret. "Limbo: A Flexible High-Performance Library for Gaussian Processes Modeling and Data-Efficient Optimization". In: *Journal of Open Source Software* 3.26 (June 2018), p. 545. ISSN: 2475-9066. DOI: 10/ggv6c4. URL: http://joss.theoj.org/papers/10.21105/joss.00545.

[19] Alice Danckaers and Mae L. Seto. "Transmission of Images by Unmanned Underwater Vehicles". en. In: *Autonomous Robots* 44.1 (Jan. 2020), pp. 3–24. ISSN: 0929-5593, 1573-7527. DOI: 10/ggjp5f. URL: http://link.springer.com/10.1007/s10514-019-09866-z (visited on 01/29/2020).

[20] Christian Daniel, Malte Viering, Jan Metz, Oliver Kroemer, and Jan Peters. "Active Reward Learning". In: *Proceedings of Robotics: Science and Systems (RSS)*. 2014. DOI: 10/ggb48v.

[21] Charles Darwin. *Journal and Remarks, 1832–1835*. Vol. 3. Voyages of the Adventure and Beagle. London: Henry Colburn, 1839.

[22] Jnaneshwar Das, Frédéric Py, Julio B.J. Harvey, John P. Ryan, Alyssa Gellene, Rishi Graham, David A. Caron, Kanna Rajan, and Gaurav S. Sukhatme. "Data-Driven Robotic Sampling for Marine Ecosystem Monitoring". en. In: *The International Journal of Robotics Research* 34.12 (Oct. 2015), pp. 1435–1452. ISSN: 0278-3649, 1741-3176. DOI: 10/f7s6q9. URL: http://journals.sagepub.com/doi/10.1177/0278364915587723.

[23] Kevin Doherty, Genevieve Flaspohler, Nicholas Roy, and Yogesh Girdhar. "Approximate Distributed Spatiotemporal Topic Models for Multi-Robot Terrain Characterization". In: *Intelligent Robots and Systems (IROS)*. 2018. DOI: 10/ggb47v.

[24] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition". In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by Eric P Xing and Tony Jebara. Vol. 32. Bejing, China: PMLR, 2014, pp. 647–655. URL: http://proceedings.mlr.press/v32/donahue14.html.

[25] Maria Dornelas and Sean R. Connolly. "Multiple Modes in a Coral Species Abundance Distribution". In: *Ecology Letters* 11.10 (2008), pp. 1008–1016. ISSN: 1461-0248. DOI: 10/cjwpkf. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2008.01208.x.

[26] Finale Doshi-Velez, Joelle Pineau, and Nicholas Roy. "Reinforcement Learning with Limited Reinforcement: Using Bayes Risk for Active Learning in POMDPs". In: *Artificial Intelligence* 187-188 (2012), pp. 115–132. ISSN: 00043702. DOI: 10/f997t2. URL: http://dx.doi.org/10.1016/j.artint.2012.04.006.

[27] Mitch Elend. *Lightning Deployment*. Oct. 2019. URL: https://www.whoi.edu/multimedia/lightning-deployment/ (visited on 05/06/2020).

[28] Thomas Stearns Eliot. "Little Gidding". In: *Four Quartets*. Harcourt, 1943.

[29] Jane Elith and John R. Leathwick. "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time". In: *Annual Review of Ecology, Evolution, and Systematics* 40.1 (2009), pp. 677–697. ISSN: 1543-592X. DOI: 10/ffggsk. URL: http://www.annualreviews.org/doi/10.1146/annurev.ecolsys.110308.120159.

[30] ERTRAC Task Force on "Connectivity and Automated Driving". *ERTRAC Automated Driving Roadmap*. Tech. rep. Version: 5.0. July 2015. URL: https://www.ertrac.org/uploads/documentsearch/id38/ERTRAC_Automated-Driving-2015.pdf.

[31] Tara A. Estlin, Benjamin J. Bornstein, Daniel M. Gaines, Robert C. Anderson, David R. Thompson, Michael Burl, Rebecca Castaño, and Michele Judd. "AEGIS Automated Science Targeting for the MER Opportunity Rover". In: *ACM Transactions on Intelligent Systems and Technology* 3.3 (2012), pp. 1–25. ISSN: 21576904. DOI: 10/gfvsn5.

[32] Tara Estlin, Daniel Gaines, Caroline Chouinard, Rebecca Castano, Benjamin Borastein, Michele Judd, Issa Nesnas, and Robert Anderson. "Increased Mars Rover Autonomy Using AI Planning, Scheduling and Execution". In: *Proceedings of the IEEE International Conference on Robotics and Automation*. 2007, pp. 4911–4918. ISBN: 1-4244-0602-1. DOI: 10/dmqc4f.

[33] Kaveh Fathian, Kasra Khosoussi, Yulun Tian, Parker Lusk, and Jonathan P. How. "CLEAR: A Consistent Lifting, Embedding, and Alignment Rectification Algorithm for Multi-View Data Association". en. In: *arXiv:1902.02256 [cs]* (July 2019). arXiv: 1902.02256 [cs]. URL: http://arxiv.org/abs/1902.02256 (visited on 11/06/2019).

[34] Genevieve Flaspohler, Victoria Preston, Anna P. M. Michel, Yogesh Girdhar, and Nicholas Roy. "Information-Guided Robotic Maximum Seek-and-Sample in Partially Observable Continuous Environments". In: *IEEE Robotics and Automation Letters* 4.4 (Oct. 2019), pp. 3782–3789. ISSN: 2377-3766. DOI: 10/ggb48n.

[35] Genevieve Flaspohler, Nicholas Roy, and Yogesh Girdhar. "Feature Discovery and Visualization of Robot Mission Data Using Convolutional Autoencoders and Bayesian Nonparametric Topic Models". In: *IEEE International Conference on Intelligent Robots and Systems*. 2017, pp. 1–8. ISBN: 978-1-5386-2682-5. DOI: 10/ggb47x. arXiv: 1712.00028.

[36] Brendan P. Foley, Ryan M. Eustice, Katerina Dellaporta, Dionysis Evagelistis, Dimitris Sakellariou, Vicki Lynn Ferrini, Brian S. Bingham, Kostas Katsaros, Richard Camilli, Dimitris Kourkoumelis, Aggelos Mallios, Hanumant Singh, Paraskevi Micha, David S. Switzer, David A. Mindell, Theotokis Theodoulou, and Christopher Roman. "The 2005 Chios Ancient Shipwreck Survey: New Methods for Underwater Archaeology". In: *Hesperia* 78.2 (2009), pp. 269–305. ISSN: 0018098X. DOI: 10/bcz76v.

[37] Kim Fulton-Bennett. *Biologists Discover Deep-Sea Fish Living Where There Is Virtually No Oxygen*. Jan. 2019. URL: https://www.mbari.org/low-oxygen-fish/.

[38] Yang Gao and Steve Chien. "Review on Space Robotics: Toward Top-Level Science through Space Exploration". In: *Science Robotics* 2 (June 2017). ISSN: 24709476. DOI: 10/gd5jt4. URL: http://robotics.sciencemag.org/lookup/doi/10.1126/scirobotics.aan5074 (visited on 03/18/2019).

[39] Mohammad Ghavamzadeh, Shie Mannor, Joelle Pineau, and Aviv Tamar. "Bayesian Reinforcement Learning: A Survey". In: *Foundations and Trends® in Machine Learning* 8.5-6 (2015), pp. 359–483. ISSN: 1935-8237, 1935-8245. DOI: 10/gfgwdh. arXiv: 1609.04436. URL: http://arxiv.org/abs/1609.04436 (visited on 05/19/2020).

[40] Yolanda Gil and Bart Selmen. *A 20-Year Community Roadmap for Artificial Intelligence Research in the US*. Tech. rep. Computing Community Consortium (CCC) and Association for the Advancement of Artificial Intelligence (AAAI), Aug. 2019. URL: https://arxiv.org/pdf/1908.02624.pdf (visited on 05/14/2020).

[41] Yogesh Girdhar. "Unsupervised Semantic Perception, Summarization, and Autonomous Exploration for Robots in Unstructured Environments". PhD thesis. 2014. URL: http://digitool.library.mcgill.ca:80/R/-?func=dbin-jump-full%5C&object_id=129641%5C&silo_library=GEN01 (visited on 04/01/2018).

[42] Yogesh Girdhar, Levi Cai, Stewart Jamieson, Nathan McGuire, Genevieve Flaspohler, Stefano Suman, and Brian Claus. "Streaming Scene Maps for Co-Robotic Exploration in Bandwidth Limited Environments". In: *2019 International Conference on Robotics and Automation (ICRA)*. Montreal, Canada: IEEE, May 2019, pp. 7940–7946. ISBN: 978-1-5386-6027-0. DOI: 10/ggb46q. URL: https://arxiv.org/abs/1903.03214.

[43] Yogesh Girdhar and Gregory Dudek. "Modeling Curiosity in a Mobile Robot for Long-Term Autonomous Exploration and Monitoring". In: *Autonomous Robots* 40.7 (2016), pp. 1267–1278. ISSN: 15737527. DOI: 10/744. arXiv: 1509.07975.

[44] Yogesh Girdhar, Philippe Giguère, and Gregory Dudek. "Autonomous Adaptive Exploration Using Realtime Online Spatiotemporal Topic Modeling". In: *The International Journal of Robotics Research* 33.4 (Apr. 2014), pp. 645–657. ISSN: 0278-3649. DOI: 10/f539cj. URL: http://journals.sagepub.com/doi/10.1177/0278364913507325.

[45] Yogesh Girdhar, Walter Cho, Matthew Campbell, Jesus Pineda, Elizabeth Clarke, and Hanumant Singh. "Anomaly Detection in Unstructured Environments Using Bayesian Nonparametric Scene Modeling". In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. May 2016, pp. 2651–2656. DOI: 10/ggb48m.

[46] Kate Golembiewski. *H.M.S. Challenger: Humanity's First Real Glimpse of the Deep Oceans*. Apr. 2019. URL: https://www.discovermagazine.com/planet-earth/hms-challenger-humanitys-first-real-glimpse-of-the-deep-oceans (visited on 05/06/2020).

[47] Kasthurirangan Gopalakrishnan, Siddhartha K. Khaitan, Alok Choudhary, and Ankit Agrawal. "Deep Convolutional Neural Networks with Transfer Learning for Computer Vision-Based Data-Driven Pavement Distress Detection". In: *Construction and Building Materials* 157.September (2017), pp. 322–330.

ISSN: 09500618. DOI: 10/gf3m3c. URL: https://doi.org/10.1016/j.conbuildmat.2017.09.110.

[48] Martin Greenspan and Carroll E. Tschiegg. "Speed of Sound in Water by a Direct Method". en. In: *Journal of Research of the National Bureau of Standards* 59.4 (Oct. 1957), p. 249. ISSN: 0091-0635. DOI: 10/ggwnfz. URL: https://nvlpubs.nist.gov/nistpubs/jres/59/jresv59n4p249_A1b.pdf (visited on 05/19/2020).

[49] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition". In: (2015). ISSN: 1664-1078. DOI: 10/gdcfkn. arXiv: 1512.03385. URL: http://arxiv.org/abs/1512.03385.

[50] Gregory Hitz, Enric Galceran, Marie-Ève Garneau, François Pomerleau, and Roland Siegwart. "Adaptive Continuous-Space Informative Path Planning for Online Environmental Monitoring". In: *Journal of Field Robotics* 34.8 (2017), pp. 1427–1449. DOI: 10/gcj7sp.

[51] Tammy Horton, Andreas Kroh, and Leen Vandepitte. *How Many Undiscovered Creatures Are There in the Ocean?* Nov. 2017. URL: http://theconversation.com/how-many-undiscovered-creatures-are-there-in-the-ocean-86705 (visited on 05/06/2020).

[52] Michael V. Jakuba, Daniel Steinberg, James C. Kinsey, Dana R. Yoerger, Richard Camilli, Oscar Pizarro, and Stefan B. Williams. "Toward Automatic Classification of Chemical Sensor Data from Autonomous Underwater Vehicles". In: *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Sept. 2011, pp. 4722–4727. DOI: 10/ckth5b.

[53] Stewart Jamieson, Jonathan P. How, and Yogesh Girdhar. "Active Reward Learning for Co-Robotic Vision Based Exploration in Bandwidth Limited Environments". In: *2020 International Conference on Robotics and Automation*. Paris, France: IEEE, May 2020. URL: https://arxiv.org/abs/2003.05016.

[54] Alok Jha. "Nasa's Curiosity Rover Finds Water in Martian Soil". In: *The Guardian* (Sept. 2013). ISSN: 0261-3077. URL: https://www.theguardian.com/science/2013/sep/26/nasa-curiosity-rover-mars-soil-water.

[55] Meghan Jones and Woods Hole Oceanographic Institute. *Sentry*. Apr. 2015. URL: https://web.whoi.edu/mesh/sentry/ (visited on 05/06/2020).

[56] Jeffrey W. Kaeli, John J. Leonard, and Hanumant Singh. "Visual Summaries for Low-Bandwidth Semantic Mapping with Autonomous Underwater Vehicles". In: *2014 IEEE/OES Autonomous Underwater Vehicles (AUV)*. IEEE, Oct. 2014, pp. 1–7. ISBN: 978-1-4799-4344-9. DOI: 10/ggb452. URL: http://ieeexplore.ieee.org/document/7054429/ (visited on 03/20/2019).

[57] Levente Kocsis and Csaba Szepesvári. "Bandit Based Monte-Carlo Planning". In: *Machine Learning: ECML 2006*. Ed. by Johannes Fürnkranz, Tobias Scheffer, and Myra Spiliopoulou. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2006, pp. 282–293. ISBN: 978-3-540-46056-5. DOI: 10/dp5gj5.

[58] Wouter M. Kouw and Marco Loog. "An Introduction to Domain Adaptation and Transfer Learning". In: (2018). arXiv: 1812.11806. URL: http://arxiv.org/abs/1812.11806.

[59] Clayton Kunz and Hanumant Singh. "Map Building Fusing Acoustic and Visual Information Using Autonomous Underwater Vehicles". In: *Journal of Field Robotics* 30.5 (Sept. 2013), pp. 763–783. ISSN: 15564959. DOI: 10/f45xt8. URL: http://doi.wiley.com/10.1002/rob.21473.

[60] Olivier Lamarre and Jonathan Kelly. "Overcoming the Challenges of Solar Rover Autonomy: Enabling Long-Duration Planetary Navigation". In: *International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*. Madrid, Spain: European Space Agency (ESA), June 2018.

[61] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2. IEEE, 2006, pp. 2169–2178. ISBN: 0-7695-2597-0. DOI: 10/chsk82. URL: http://ieeexplore.ieee.org/document/1641019/ (visited on 05/10/2019).

[62] Andre Makovsky, Andrea Barbieri, and Ramona Tung. *Odyssey Telecommunications*. Tech. rep. 6. Pasadena, California: Jet Propulsion Laboratory, Oct. 2002. URL: https://descanso.jpl.nasa.gov/DPSummary/odyssey_telecom.pdf.

[63] Andre Makovsky, Peter Ilott, and Jim Taylor. *Mars Science Laboratory Telecommunications System Design*. Tech. rep. 14. Pasadena, California: Jet Propulsion Laboratory, Nov. 2009. URL: https://descanso.jpl.nasa.gov/DPSummary/Descanso14_MSL_Telecom.pdf.

[64] Ajith Anil Meera, Marija Popovic, Alexander Millane, and Roland Siegwart. "Obstacle-Aware Adaptive Informative Path Planning for UAV-Based Target Search". In: *IEEE International Conference on Robotics and Automation*. 2019. ISBN: 978-1-5386-6026-3. arXiv: 1902.10182. URL: http://arxiv.org/abs/1902.10182.

[65] Michael Mitzenmacher and Eli Upfal. "Sample Complexity, VC Dimension, Rademacher Complexity". In: *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Second. Cambridge University Press, 2017, pp. 361–391. ISBN: 1-108-10799-0 978-1-108-10799-0.

[66] Monterey Bay Aquarium Research Institute (MBARI). *Vehicle Technology*. Dec. 2015. URL: https://www.mbari.org/technology/emerging-current-tools/vehicle-technology/ (visited on 05/19/2020).

[67] Hadi Moradi. "Road Maps for Robotics and Automation". In: *IEEE Robotics Automation Magazine*. Industrial Activities 16.3 (Sept. 2009), pp. 98–98. ISSN: 1558-223X. DOI: 10/dmr4mz.

[68] NASA/JPL. *Communications with Earth*. URL: https://mars.nasa.gov/msl/mission/communications (visited on 05/11/2020).

[69] NASA/JPL. *Mastcam*. URL: https://mars.nasa.gov/msl/spacecraft/instruments/mastcam (visited on 05/11/2020).

[70] NASA/JPL. *Rover Brains*. URL: https://mars.nasa.gov/msl/spacecraft/rover/brains (visited on 05/11/2020).

[71] NASA/JPL. *Summary | Instruments*. URL: https://mars.nasa.gov/msl/spacecraft/instruments/summary (visited on 05/11/2020).

[72] National Aeronautics and Space Administration (NASA). *2020 NASA Technology Taxonomy*. Tech. rep. 2020. URL: https://www.nasa.gov/sites/default/files/atoms/files/2020_nasa_technology_taxonomy.pdf.

[73] Andrew Y. Ng and Michael I. Jordan. "On Discriminative vs. Generative Classifiers: A Comparison of Logistic Regression and Naive Bayes". In: *Advances in Neural Information Processing Systems*. 2002. DOI: 10/dj3wcc.

[74] Edwin Olson. "AprilTag: A Robust and Flexible Visual Fiducial System". In: *2011 IEEE International Conference on Robotics and Automation*. Shanghai, China: IEEE, May 2011, pp. 3400–3407. ISBN: 978-1-61284-386-5. DOI: 10/cfp5qs. URL: http://ieeexplore.ieee.org/document/5979561/.

[75] Shayegan Omidshafiei, Dong-Ki Kim, Miao Liu, Gerald Tesauro, Matthew Riemer, Christopher Amato, Murray Campbell, and Jonathan P. How. "Learning to Teach in Cooperative Multiagent Reinforcement Learning". In: *Lifelong Learning: A Reinforcement Learning Approach Workshop at the Federated Artificial Intelligence Meeting (FAIM)*. Aug. 2018. arXiv: 1805.07830. URL: http://arxiv.org/abs/1805.07830.

[76] Thomas Ormston. *Time Delay between Mars and Earth*. Aug. 2012. URL: https://blogs.esa.int/mex/2012/08/05/time-delay-between-mars-and-earth/ (visited on 05/19/2020).

[77] Sooho Park. "Learning for Informative Path Planning". In: *Electrical Engineering* (2008). URL: https://dspace.mit.edu/handle/1721.1/45887 (visited on 12/14/2017).

[78] F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot, and E Duchesnay. "Scikit-Learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[79] Jesús Pineda, Walter Cho, Victoria Starczak, Annette F. Govindarajan, Héctor M. Guzman, Yogesh Girdhar, Rusty C. Holleman, James Churchill, Hanumant Singh, and David K. Ralston. "A Crab Swarm at an Ecological Hotspot: Patchiness and Population Density from AUV Observations at a Coastal, Tropical Seamount". In: *PeerJ* 4 (Apr. 2016). ISSN: 2167-8359. DOI: 10/ggvr7h. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4841253/.

[80] Morgan Quigley, Brian Gerkey, Ken Conley, Josh Faust, Tully Foote, Jeremy Leibs, Eric Berger, Rob Wheeler, and Andrew Ng. "ROS: An Open-Source Robot Operating System". en. In: *ICRA Workshop on Open Source Software*. Vol. 3. 2009, p. 6.

[81] Carl E Rasmussen and Christopher K I Williams. *Gaussian Processes for Machine Learning*. Ed. by Thomas Dietterich. Vol. 14. MIT Press, 2006. ISBN: 0-262-18253-X. DOI: 10.1142/S0129065704001899.

[82] Henning Reiss, Henning Cunze, Konstantin König, Konstantin Neumann, and Ingrid Kröncke. "Species Distribution Modelling of Marine Benthos: A North Sea Case Study". In: *Marine Ecology Progress Series* 442.December (2011), pp. 71–86. ISSN: 01718630. DOI: 10/bsgp62.

[83] Adriana Romero, Carlo Gatta, and Gustau Camps-Valls. "Unsupervised Deep Feature Extraction for Remote Sensing Image Classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.3 (2015), pp. 1349–1362. ISSN: 01962892. DOI: 10/f8fqzq.

[84] Dorsa Sadigh, Anca D. Dragan, Shankar Sastry, and Sanjit A. Seshia. "Active Preference-Based Learning of Reward Functions". In: *Proceedings of Robotics: Science and Systems (RSS)*. 2017. DOI: 10/ggb48w. arXiv: 1810.04303. URL: http://arxiv.org/abs/1810.04303.

[85] SAE International. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Tech. rep. J3016. June 2018.

[86] Nasir Saeed, Abdulkadir Celik, Tareq Y. Al-Naffouri, and Mohamed-Slim Alouini. "Underwater Optical Wireless Communications, Networking, and Localization: A Survey". In: *Ad Hoc Networks* 94.101935 (Nov. 2019). ISSN: 1570-8705. DOI: 10/ggwncp. URL: http://www.sciencedirect.com/science/article/pii/S1570870518309776 (visited on 05/19/2020).

[87] John E. San Soucie, Heidi M. Sosik, and Yogesh Girdhar. "Gaussian-Dirichlet Random Fields for Inference over High Dimensional Categorical Observations". In: *2020 International Conference on Robotics and Automation*. Paris, France: IEEE, May 2020. arXiv: 2003.12120. URL: http://arxiv.org/abs/2003.12120 (visited on 05/06/2020).

[88] Thomas R. Sayre-McCord, Chris Murphy, Jeffrey Kaeli, Clayton Kunz, Peter Kimball, and Hanumant Singh. "Advances in Platforms and Algorithms for High Resolution Mapping in the Marine Environment". In: *Lecture Notes in Control and Information Sciences*. Ed. by T.I. Fossen et al. Vol. 474. Springer, Cham, 2017, pp. 89–119. ISBN: 978-3-319-55371-9. DOI: 10.1007/978-3-319-55372-6_5. URL: http://link.springer.com/10.1007/978-3-319-55372-6_5 (visited on 03/20/2019).

[89] Tim Shank, Rick Chandler, and Woods Hole Oceanographic Institution. *Alvin FAQs*. URL: https://www.whoi.edu/what-we-do/explore/underwater-vehicles/hov-alvin/faqs/ (visited on 05/06/2020).

[90] Florian Shkurti. “Algorithms and Systems for Robot Videography from Human Specifications”. PhD thesis. McGill University, 2018.

[91] Josef Sivic, Bryan C. Russell, Andrew Zisserman, William T. Freeman, and Alexei A. Efros. “Unsupervised Discovery of Visual Object Class Hierarchies”. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition* (2008), pp. 1–8. ISSN: 1063-6919. DOI: 10/cfhcxt. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4587622.

[92] Jennifer E. Smith, Rusty Brainard, Amanda Carter, Saray Dugas, Clinton Edwards, Jill Harris, Levi Lewis, David Obura, Forest Rohwer, Enric Sala, Peter S. Vroom, Stuart Sandin, Saray Grillo, Clinton Edwards, Jill Harris, Levi Lewis, David Obura, Forest Rohwer, Enric Sala, Peter S. Vroom, and Stuart Sandin. “Re-Evaluating the Health of Coral Reef Communities: Baselines and Evidence for Human Impacts across the Central Pacific”. In: *Proceedings of the Royal Society B: Biological Sciences* 283.1822 (2016). ISSN: 0962-8452. DOI: 10/ggb48x. URL: http://rspb.royalsocietypublishing.org/lookup/doi/10.1098/rspb.2015.1985.

[93] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. “Practical Bayesian Optimization of Machine Learning Algorithms”. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Vol. 2. 2012, pp. 2951–2959. arXiv: 1206.2944. URL: http://arxiv.org/abs/1206.2944.

[94] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. “A Survey on Deep Transfer Learning”. In: *The 27th International Conference on Artificial Neural Networks*. 2018, pp. 1–10. arXiv: 1808.01974v1.

[95] Jim Taylor, Dennis K. Lee, and Shervin Shambayati. *Mars Reconnaissance Orbiter Telecommunications*. Tech. rep. 12. Pasadena, California: Jet Propulsion Laboratory, Sept. 2006. URL: https://descanso.jpl.nasa.gov/DPSummary/MRO_092106.pdf.

[96] Teledyne Marine. *Underwater Acoustic Modems for Wireless Communication*. URL: http://www.teledynemarine.com/acoustic-modems (visited on 05/06/2020).

[97] Paul Tompkins, Anthony Stentz, and David Wettergreen. “Global Path Planning for Mars Rover Exploration”. In: *2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720)*. Vol. 2. Big Sky, MT, USA: IEEE, Mar. 2004, 801–815 Vol.2. DOI: 10.1109/AERO.2004.1367681. URL: https://www.ri.cmu.edu/pub_files/pub4/tompkins_paul_2004_3/tompkins_paul_2004_3.pdf.

[98] Ursula K. Verfuss, Ana Sofia Aniceto, Danielle V. Harris, Douglas Gillespie, Sophie Fielding, Guillermo Jiménez, Phil Johnston, Rachael R. Sinclair, Agnar Sivertsen, Stian A. Solbø, Rune Storvold, Martin Biuw, and Roy Wyatt. “A Review of Unmanned Vehicles for the Detection and Monitoring of Marine

Fauna". en. In: *Marine Pollution Bulletin* 140 (Mar. 2019), pp. 17–29. ISSN: 0025326X. DOI: 10/ggt53t. URL: https://linkinghub.elsevier.com/retrieve/pii/S0025326X19300098.

[99] Samir Wadhwania, Dong-Ki Kim, Shayegan Omidshafiei, and Jonathan P. How. "Policy Distillation and Value Matching in Multiagent Reinforcement Learning". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Macau, China: IEEE, Nov. 2019. DOI: 10.1109/IROS40897.2019.8967849. arXiv: 1903.06592. URL: http://arxiv.org/abs/1903.06592.

[100] Wenlin Wang, Yunchen Pu, Vinay Kumar Verma, Kai Fan, Yizhe Zhang, Changyou Chen, Piyush Rai, and Lawrence Carin. "Zero-Shot Learning via Class-Conditioned Deep Generative Models". In: *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)* (2018), pp. 4211–4218. ISSN: 0976-9234. DOI: 10/ggb46j. arXiv: 1711.05820. URL: http://arxiv.org/abs/1711.05820.

[101] Xaiogang Wang and Eric Grimson. "Spatial Latent Dirichlet Allocation". In: *Neural Information Processing Systems*. 2007, pp. 1–8. ISBN: 1-60560-352-X.

[102] Stewart A. Weaver. "Exploration and the Enlightenment". In: *Exploration: A Very Short Introduction*. United States of America: Oxford University Press, 2015. ISBN: 978-0-19-994697-6.

[103] WHOI Acoustic Communications Group. *Micro-Modem Software Interface Guide*. Tech. rep. 401040. Version: 3.24. Woods Hole Oceanographic Institution, Nov. 2014. URL: https://acomms.whoi.edu/wp-content/uploads/sites/20/2014/09/401040-SIG-Micromodem-Software-Interface-Guide.pdf.

[104] Frederick Whymper. *The Sea: Its Stirring Story of Adventure, Peril & Heroism*. Vol. 1. Cassell, Petter & Galpin, 1887.

[105] Guy D. Williams, Ted Maksym, Jeremy Wilkinson, Clayton Kunz, Chris Murphy, Peter Kimball, and Hanumant Singh. "Thick and Deformed Antarctic Sea Ice Mapped with Autonomous Underwater Vehicles". In: *Nature Geoscience* 8.1 (2015), pp. 61–67. ISSN: 17520908. DOI: 10/xcg.

[106] Cuebong Wong, Erfu Yang, Xiu-Tian Yan, and Dongbing Gu. "Adaptive and Intelligent Navigation of Autonomous Planetary Rovers — A Survey". In: *2017 NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*. Pasadena, California: IEEE, July 2017, pp. 237–244. DOI: 10.1109/AHS.2017.8046384.

[107] Woods Hole Oceanographic Institution. *Underwater Vehicles*. URL: https://www.whoi.edu/what-we-do/explore/underwater-vehicles/ (visited on 05/06/2020).

[108] Yazhou Yang and Marco Loog. "A Benchmark and Comparison of Active Learning for Logistic Regression". In: *Pattern Recognition* 83 (2018), pp. 401–415. ISSN: 00313203. DOI: 10/ggb483. URL: https://doi.org/10.1016/j.patcog.2018.06.004.

[109] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018. DOI: 10/gfz33w.