Automated Observation of Spatial Distributions of Reef Fauna

Nader Boutros

A thesis submitted in fulfilment
of the requirements of the degree of
Doctor of Philosophy

Australian Centre for Robotics
School of Aerospace, Mechanical and Mechatronic Engineering
The University of Sydney

Submitted December 2023; revised September 2024

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the University or other institute of higher learning, except where due acknowledgement has been made in the text.

Nader Boutros
25 September 2024

Abstract

This thesis presents methods for observing high resolution spatial distributions of stationary and mobile fauna and habitat features from image-based, reef-scale, seafloor reconstructions. Advances in underwater imaging equipment and autonomous underwater survey platforms have made it possible to routinely collect large volumes of underwater images and to generate georeferenced 3D reconstructions of reef-scale seafloor extents. The scale of the datasets, and novel data products facilitated by these technologies present an opportunity to broaden the scope of our understanding of benthic habitats.

Two main barriers to the use of image-based habitat reconstructions for routine observation and monitoring of benthic ecosystems are addressed. Firstly, these technologies yield high volumes of image data which require time-consuming manual annotation to process. This has, to date, made them infeasible for routine or ongoing observation of benthic habitats, particularly where the objects or organisms of interest are densely aggregated. Secondly, mobile fauna are typically not observed from image based reconstruction type surveys since their movement between frames can cause them to be missed, duplicated, or distorted in the final reconstruction. This work addresses these challenges through the use of automated image analysis and by exploiting the vehicle navigation information and georeferencing contained in these data to detect and localise benthic targets.

Image based mapping approaches build a mosaic and 3D reconstruction of the surveyed habitat from a series of views. The mapped extent increases progressively as the imager (or imagers) move along the survey path. As a result, the technique used to observe features depends on the rate at which they move relative to the rate that the survey area is observed. The key contributions of this work lie in the development of a series of techniques for observation of the spatial distributions of variously mobile objects over large extents and using minimal manual annotation. In particular, the novel contributions of the work include:

- **A general approach to estimating the abundance and spatial distributions of benthic targets using observations from a moving imaging platform**. Previous studies observing benthic fauna in large-scale composite images have been limited to stationary or slow moving animals which can be treated as fixed features in the habitat over the course of the survey. The rate of movement of the targets relative to the rate at which the imaged seafloor area expands affects the methods needed to observe the distribution of targets. By considering the locomotory behaviour of the target species a probabilistic,

tracking based approach is developed by estimating the likelihood that a given detection is either a new individual, or a repeat detection of a previously observed individual. This results in predictions of the possible distributions of individuals based on the observed detections and the assumptions made about their behaviour.

- **Automated spatial mapping of stationary and slow moving benthic features**. Automated object detection and image segmentation techniques are used to estimate the spatial distributions of habitat features and slow moving benthic fauna directly from reef-scale mosaics. This facilitates observation of these key features at high resolution over scales that are impractical or impossible using traditional *in situ* observational techniques and allows for efficient surveying of even densely aggregated individuals which would otherwise require time-consuming manual annotation.

- **Automated detection of faster moving animals from unprocessed AUV images and reprojection into 3D**. Faster moving fauna may move between successive frames and therefore can appear distorted, appear multiple times or be absent in the final mosaic or reconstruction. An approach for observing the distributions of such animals is presented. Animals are detected in the unprocessed image frames where the objects can be clearly seen. These detections are projected into 3D space by intersecting a ray through the image plane to the surface of the 3D seafloor reconstruction. This results in a set of detections of individuals localised within the survey area. These localised, automated detections can be paired with the tracking based approach developed in the first contribution to estimate the abundance and distribution of mobile benthic targets at reef scales.

- **Segmentation of seafloor photomosaics using appearance and structure information**. An approach for incorporating the structural information measured using image based mapping techniques for segmenting objects of interest from reef-scale imagery is demonstrated. Benthic habitats, and reefs particularly, exhibit complex and diverse physical structure. The structure of reef features is often used to aid in the identification and delineation of reef features such as habitat forming organisms. Image based habitat modelling approaches generate aligned reconstructions of the appearance and 3D structure of the surveyed area. The work presents two methods for incorporating both visual and structural representations of seafloor habitats into an automated segmentation algorithm, and demonstrates their use for segmentation of kelp and hard corals from reef-scale reconstructions.

Together, the contributions of this thesis allow for efficient mapping of benthic habitat features and fauna over large extents. This allows for detailed analysis of the interactions between and among these inhabitants and the structure of their environment. The ability to efficiently observe the distributions of benthic fauna and habitat features at high resolution and over large seafloor extents facilitated by the methods presented in this thesis provide a unique way to observe benthic habitats.

Acknowledgements

I could not have completed this work without the contributions of many people. Firstly, my supervisor Professor Stefan Williams and co-supervisor Dr Oscar Pizarro. Thank you for your patience, invaluable support and guidance over the long process of my PhD, and for feeding my insatiable desire for fieldwork.

I have been fortunate to work with many inspiring researchers. In particular, Euan Harvey, who set me off on this journey and cemented my love for temperate reefs. Thank you for always giving excellent, practical advice and for being a great dive buddy. I also thank Jacquomo Monk who, despite having never met, has been so generous with his time. My colleagues at AIMS have been a great support and I've benefitted from their experience and advice in both my work and studies. In particular, I would like to acknowledge Mathew Wyatt who has had such empathy and been a great support while I balanced work and completing this thesis.

The ACFR is full of many kind, brilliant people. In particular I'd like to acknowledge Lachlan Toohey and Mitch Bryson, who are always happy to share their amazingly broad knowledge and skills. I would also like to thank the ACFR technical staff, especially Christian Lees, Jorja Martin and George Wakeham who gave me free reign in the workshop, but would graciously assist when I was out of my depth.

I have made many friends over the course of my PhD. To Rishi Ramakrishnan and Lloyd Windrim for (not so) gently pushing me into machine learning, and for our sometimes unhinged thought bubbles on the Rose Street whiteboards. Also, Phil, John, Jackson, Eric, Stuart, Nathan and the rest of the students at the ACFR for countless coffees, lunches and afternoons at the pub. To my friends and housemates in Ann Arbor, your warmth and fun made my time there so memorable. Lizard Island will always be a special place to me, I have such fond memories of the friendships I made there. To my friends in Perth, Sydney and Darwin and abroad. Thank you for being the best cheer-squad throughout my PhD. I'm looking forward to reconnecting with you all once I resurface.

My greatest thanks go to my family. To my parents, you are the epitome of selflessness and I am here today thanks to your sacrifices. To Samy, for always checking in and making sure I look after myself. Finally, Lucy, I cannot express how much your love and support mean to me. I could not have done it without you.

Contents

List of Figures

List of Tables

Nomenclature

List of Acronyms

ACFR	Australian Centre for Robotics
API	application programming interface
AUV	autonomous underwater vehicle
BRUV	baited remote underwater video
CNN	convolutional neural network
DBSCAN	density based spatial clustering of applications with noise
DEM	digital elevation model
DOV	diver operated video
DVL	Doppler velocity log
FPR	false positive rate
GUI	graphical user interface
lidar	light detection and ranging
MIS	maximal independent set
MBES	multibeam echosounder
MHT	multiple hypothesis tracking
MPA	marine protected area
ROC	receiver operating characteristic
SfM	structure from motion
SLAM	simultaneous localisation and mapping

SDM	species distribution modelling
SfM	structure from motion
TPR	true positive rate
USBL	ultra short baseline
UVC	underwater visual census

Chapter 1

Introduction

Earth's oceans are home to a rich diversity of organisms. This diversity is highly concentrated in coastal reef systems, which only make up a fraction of the total area of the world's oceans but are home to a disproportionately large amount of its biota. These environments rival terrestrial ecosystems in density, abundance, and diversity of lifeforms [21]. This concentration of biodiversity makes them a critical contributor to the oceans' health and functioning. Additionally, fish and other organisms targeted by commercial, recreational and subsistence fishers often rely on reef habitats for some or all of their life stages. As a result, reefs support an essential food source for millions of humans [168], particularly in developing nations [6]. Robust, efficient sampling procedures allow us to better understand and monitor change in these critical ecosystems, which is essential for their ongoing management and sustainability.

This thesis presents methods for observing the spatial distributions of physical structures, habitat forming organisms, and benthic fauna from high resolution, large scale, image based seafloor reconstructions. Automated image analysis is employed to allow for efficient processing of the large volumes of data generated by these techniques. A range of novel approaches are presented which allow for mapping the distributions of both stationary and mobile benthic fauna and habitat features. This is achieved by using automated detections of objects or features and reprojecting them into 3D space using the vehicle's navigation data, and a tracking based approach for estimat-

ing the abundance of mobile fauna using detections made from a moving platform by considering the targets' locomotory behaviour.

The approaches in this work combine to allow observations of how benthic fauna interact with one another and with their habitat at sub metre resolution over large spatial extents. The ability to implement these methods without the need for large amounts of manual processing of imagery means they can be efficiently applied, for both exploratory type surveys, and for routine monitoring of benthic ecosystems.

1.1 Motivation

Reef structure can be influenced by the underlying geology, local or regional oceanographic conditions, and by habitat forming organisms, such as calcifying corals, encrusting algae, or kelp. Reefs are therefore complex, spatially heterogeneous environments, comprised of microhabitats with varying physical, biological, and environmental conditions allowing them to support their characteristic diverse and abundant biological assemblages [16, 34, 7]. This variation in conditions within a reef means that the fauna that inhabit it are not uniformly distributed, but show affinities for particular habitat types or environmental conditions [102, 87, 84].

Seafloor associated fauna are often responsible for key ecosystem drivers including grazing and predation, and are often directly targeted by fisheries, or form a part of the food webs on which they rely. An understanding of how these organisms are spatially distributed and how they interact with their habitat facilitates more accurate estimates of population density. Additionally, observations of fine scale spatial distributions of reef features can be used to explore the mechanisms which drive the habitat preferences of benthic fauna and the ecological structure of marine ecosystems. This can guide management decisions such as the planning and evaluation of marine protected areas (MPAs).

Image-based seafloor reconstructions, which utilise structure from motion (SfM) techniques [194] to estimate the 3D structure and appearance of habitats, have become

an increasingly common tool for observing the seafloor, and are widely employed for surveying benthic habitats [61, 55, 155]. These approaches can be used to capture a "snapshot" of the reef at a point in time, and provide an estimate of the spatial distributions of observed features and 3D structure of the habitat. However, as the area of interest becomes larger it becomes increasingly difficult to maintain full coverage of the reef, and the time and computational resources required to process the imagery increases significantly [158]. Benthic imaging AUVs can be programmed to collect overlapping imagery over large extents of the seafloor. They often employ algorithms such as simultaneous localisation and mapping (SLAM) to efficiently find corresponding features among large numbers of images and build high resolution ($\mathcal{O}(10^{-3}m)$) habitat reconstructions over large spatial extents (routinely up to $\mathcal{O}(10^{4}m^{2})$) [180, 205, 24]. This level of detail allows observation of fine-scale processes over large regions, *e.g.* observing the branching structure of coral colonies over an entire reef patch [193].

Their ability to generate reconstructions of large seafloor extents has driven increasing use of these data for observation and monitoring of benthic habitats (*e.g.* [207, 31, 140, 182]). However, while reef fauna can be observed in these habitat reconstructions, and their positions determined (*e.g.* [126, 174, 190, 204]), there are a number of challenges which have limited their use for routine surveys of benthic fauna. This thesis aims to address these challenges and provide a cost and time effective process for spatial surveys of benthic targets.

1.2 Problem Statement

Using image-based habitat reconstructions for observing of benthic fauna allows their spatial distribution, and that of the underlying habitat, to be recorded at large spatial extents. There are two primary hurdles to performing such surveys:

1. Movement of the target organism while the survey imagery is being collected makes it difficult or impossible to detect them in the processed photomosaic.

Detecting them in the underlying images is possible but requires further steps to localise the detections and associate them to individuals.

2. These surveys generate massive volumes of image data. Relying on manual annotation to estimate the abundance and distribution of benthic targets is time and cost intensive and limits the ability to scale up observation or monitoring programs.

Therefore, there is a need for processes to detect and localise individuals in AUV benthic images, and to automate this analysis to facilitate routine application of these methods for observing the abundance and spatial distributions of benthic targets. The contributions of this work aim to address these gaps to pave the way for widespread adoption of these surveys. This would allow observation of fine-scale interactions between benthic fauna and their habitat at extents that are impossible to efficiently perform using existing tools.

1.3 Contributions

This thesis is concerned with estimating the abundance and spatial distributions of benthic targets, including mobile reef-associated fauna. The primary contributions of this work are:

- **A general approach to estimating the abundance and spatial distributions of benthic targets using observations from a moving imaging platform.**

  Existing studies using benthic imaging AUVs for census of benthic species largely involve observation of fixed habitat features or stationary and slow moving fauna (relative to the survey image collection timescales). They typically assume that each new detection of the target object represents a new individual,

not a repeat detection of a previously seen one. This is possible since the observed objects they are slow moving compared to the vehicle so their movement over the course of the survey is not a consideration (*e.g.* [174, 191]).

In contrast, this work develops methods which account for the relationship between the movement of the imaging platform and the likely movement of the targets for detection, localisation, and association to individuals. For more mobile targets, the algorithm uses the vehicle's navigation information, and knowledge about the target species' behaviours (e.g. movement speed, home range preference) to generate an estimate of the abundance and distribution of individuals within the survey area.

- **Automated spatial mapping of stationary and slow moving benthic features**

  Observing the spatial distributions of benthic fauna within a reef using traditional methods is either restricted to small extents (square metres) using quadrats or belt transects, or, requires large amounts of *in situ* observation (*e.g.* [129]), manual image analysis [43, 174], or a combination of both [193]. Even when large scale reef reconstructions are available, researchers often choose to observe target features within them, *e.g.* using virtual quadrats or belt transects (*e.g.* [126, 153]) since analysing the entire reconstruction would be infeasible.

  The trained algorithm developed in this work detects objects in the georeferenced photomosaic resulting in a map of individuals over the survey area. This approach was applied to estimate densities of sea urchins and detected similar inter-annual trends in abundance of benthic fauna compared with manual analysis of the same survey imagery, and *in situ* counts from the same plots. This shows that automated image analysis approaches can complement or replace manual methods to increase the efficiency of survey and monitoring regimes.

- **Automated detection of faster moving animals from unprocessed AUV images and reprojection into 3D**

Mobile benthic targets cannot be reliably observed in the georeferenced photomosaic since their movement during the survey image collection since distortions introduced during the reconstruction process can cause them to appear distorted, duplicated or be completely absent the composite images. Instead, this thesis presents an approach that allows individual organisms to be automatically detected, and exploits the vehicle's navigation solution and the 3D reconstruction to localise these detection within the surveyed reef.

- **Segmentation of seafloor photomosaics using appearance and structure information**

  The structure of reefs is shaped in part by the habitat forming organisms from which they are composed. For instance, on a coral reef, individual colonies of live coral show distinct growth patterns between one another and the surrounding reef substrate. This structural information is therefore useful for identifying reef elements. SfM habitat survey methods provide us with an approximation of the 3D structure of the seafloor and this can be used to segment regions of interest within reef-scale photomosaics.

  A novel segmentation model architecture has been developed which incorporates the appearance and structure of the seafloor. The network is applied to segment kelp from imagery of a temperate rocky reef. Models trained using both structural and image information are shown to improve segmentation performance in certain cases.

1.4 Thesis Structure

Chapter 2, which follows, describes existing techniques for observing reef habitats and fauna, and provides a background on the image processing and species distribution modelling techniques employed in the subsequent chapters.

Chapter 3 outlines how the relative movement between the image capture platform and observed benthic targets affects the necessary methods for detection, localisa-

tion, and aggregation of detections in order to estimate their abundance and spatial distribution using image-based seafloor reconstructions.

Chapter 4 demonstrates how this process can be used to automatically census stationary or slow moving benthic targets, using the long-spined sea urchin (*Centrostephanus rodgersii* as an example. Interannual trends in *C. rodgersii* abundance calculated using the automated estimates are compared to those calculated using manual counts from the survey imagery, as well as *in situ* estimates.

Chapter 5 applies the pipeline to mobile fauna, using the southern rock lobster (*Jasus edwardsii*) as a case study. Automatic detections of *J. edwardsii* individuals are made in the unprocessed imagery, then projected into 3D using the vehicle's navigation solution. These are then aggregated using an understanding of the locomotory behaviour of the species.

Chapter 6 demonstrates how estimates of the spatial distribution of benthic targets can be combined with fine scale measures of structural comlexity derived from 3D reconstructions to model the relationship between mobile benthic targets and their habitat. It also introduces a novel model architecture which combines appearance and structure information to segment diffuse habitat forming organisms (*e.g.* kelp).

Chapter 7 summarises the conclusions of this thesis, and proposes directions for future research and applications of this work.

Chapter 2

Background

Monitoring of marine ecosystems has traditionally relied on either direct, *in situ* observations, or catch based methods. Advances in the quality and availability of underwater imaging systems has lead to widespread uptake of image based systems to supplement and, in some cases, replace traditional observational techniques.

In this chapter, we will explore established and emerging methods for observing marine habitats and the animals that use them. Existing techniques for these surveys tend to experience a trade off between survey resolution and extents, so we will then discuss how emerging 3D habitat reconstruction processes and image processing techniques can facilitate observing fine-scale spatial distributions of stationary and mobile benthic fauna and habitat features allowing for routine, repeatable mapping of animals and habitats over large spatial extents.

2.1 Established Methods for Observing Marine Ecosystems

The wide range of environments, species, and research problems related to marine environments means there are a multitude of methods for observing these habitats. Typically, marine observational techniques are used to study a combination of the

physical structure of the habitat, the composition and condition of the water column (*e.g.* temperature, salinity, flow), and the composition of biological assemblages present in a given system. The choice of observational approach is strongly tied to the spatial scale at which the system is to be observed, with very different approaches being employed for surveying at, say, oceanic scales compared to reef scales (*i.e.* a descriptor of the spatial scale of individual reefs *e.g.* [127, 143, 77]).

2.1.1 Habitat and Habitat Forming Organisms

Physical seafloor structure

Variation in depth of the seafloor (called bathymetry) can be observed at broad scales using a range of remote sensing techniques (see review in [17]). The presence of seafloor structures can be predicted at oceanic scales from observations of depressions in the sea surface which can be detected using satellite altimetry [115]. Bathymetry can be measured at higher resolution and precision using acoustic methods [29], most commonly using ship-borne multibeam echosounders (MBESs). These surveys can observe structure with resolutions of tens of centimetres up to tens of metres depending on water depth and the frequency of soundings [109]. Much finer resolution bathymetry can be captured using aircraft mounted light detection and ranging (lidar) systems. However, due to the attenuation of light in the water column, lidar methods are restricted to relatively shallow coastal areas (typically $\mathcal{O}(10m)$ [111]), such as shallow coral reefs (*e.g.* [28]). In addition to estimating its physical structure, acoustic measures can also be used to provide information about the composition of seafloor substrate by measuring the properties of the return from a generated pulse of sound, known as backscatter [44].

The physical structure of the habitat is ecologically important. Habitat complexity has been shown to be a strong driver of biodiversity in both terrestrial and aquatic ecosystems (See review [114]). The ratio between the surface and planar areas of a habitat, called the surface rugosity, has become a widespread analogue for structural complexity. Studies have used habitat rugosity to predict the presence of hard

seafloor substrates from bathymetry [56], to measure the response of habitats to disturbance [62], and many have shown rugosity to be correlated with fish diversity and abundance [129, 67, 84, 45, 82]

The traditional technique for measuring surface complexity at reef scales involves comparing the length of a chain draped along the surface of the habitat between two points (see Figure 2.1). However, this approach is difficult and time-consuming to implement. Small changes in the positioning of the chain can result in large variation in the rugosity estimate, and the link size of the chain must be chosen to match the scale at which rugosity will be estimated. Due to the difficulties associated with the chain-tape method, researchers often visually score habitat complexity as low or high complexity (*e.g.* [119]), or into a number of classes based on complexity (*e.g.* [159, 85]). This allows a rapid, if coarse, estimate of habitat complexity [211].

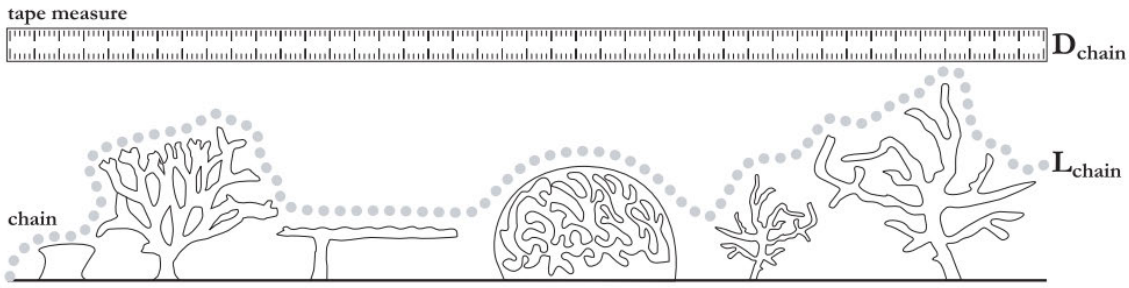


Figure 2.1 – From [68]. An illustration of the chain-tape method for estimating sea floor structural complexity. The ratio between the length of the chain and that of the tape is called the habitat rugosity and is a commonly reported analogue for habitat complexity [114].

Ecological composition of the seafloor

While acoustic and lidar generated bathymetric data is useful for observing habitat structure, and can hint at the geological (*i.e.* the physical structure) and biological composition of the seafloor (*i.e.* the abundance of individuals of different taxa in the habitat), different observational methods are generally necessary if more detailed information about the ecology of a site is required. Direct observations of the com-

position of seafloor sediments and biota can be made using techniques such as grab sampling [145]. However, many samples are required to create a complete picture of the variation within an area [98], so these approaches cannot be feasibly used to observe the distribution of habitat features at reef scales.

In situ observation of seafloor communities can also be used to estimate the community compositions. Typically the distribution of species in a sample area is observed and used to extrapolate the cover of each component over the survey sites. Sampling units can include quadrats draped over the reef (*e.g.* [15]), or belt-transects swum across a section of seafloor (*e.g.* [18, 209]).

Recently, image based surveys have become an increasingly common tool for estimating seafloor composition. These allow for rapid data collection in the field and facilitate improved consensus on identification of benthic species. However, the ease of data collection often means the bottleneck is the manual annotation of imagery required.

2.1.2 Mobile Fauna

Observational techniques for mobile marine fauna can be separated into two classes: catch-based methods, and fishery-independent techniques. Catch-based methods, as their name implies, involve capturing individuals of the target species using *e.g.* traps [50], nets [200] or lines [178]. Conversely, fishery-independent techniques do not rely on capturing individuals, rather they use direct observations to estimate abundance and/or community composition of the target species.

Rapid estimates of biomass (*i.e.* the mass of observed organisms) can be made using acoustic methods [25], however, more precise community composition estimates rely on *in situ* observation. Underwater visual census (UVC) involves a diver recording observed species in a survey area. Some studies employ a stationary observer scanning a fixed area (*e.g.* [82]), while others will make observations while swimming a transect(*e.g.* [121]).

UVC allows for accurate estimates of community composition of mobile species. However, there are several challenges associated with implementing UVC based studies. Firstly, observers require extensive training in order to correctly identify individuals' species [141]. Secondly, surveys are restricted in depth and time by the limitations of SCUBA diving. A further complication is that the presence of the observer can influence the behaviour of the subjects. However, techniques for minimising such impacts are widespread, *e.g.* waiting to begin point counts to allow aclimation or swimming transects from a greater distance to the seafloor [152]. Image based methods for observing mobile species address these challenges and have become more widespread in recent years. They result in a permanent record of the status of a site, allowing for confirmation by expert taxonomists of species ID, and reuse of the survey imagery in future studies. Remote image based surveys can also be performed at depths which divers cannot access and over longer periods than SCUBA allows. Remote cameras (if unbaited) may also have a smaller impact on target behaviour than human observers.

Stereo imaging allows for more accurate and precise length estimates of targets than UVC [91], facilitating repeatable estimates of size distributions and biomass of mobile species. A range of implementations of stereo imaging can be found in the literature including ROV mounted systems [172], drifters [202] and towed systems. However the most common stereo surveys involve either stereo diver operated video (DOV), or baited remote underwater video (BRUV). DOVs surveys are typically treated like belt transect UVC where each new target is assumed to be an individual so can be used to directly estimate abundance and community composition [79]. For BRUVs, the same cannot be assumed, since the attraction of the bait may cause the same individual to repeatedly enter the frame, and since it is typically impossible to distinguish between individuals of the same species. For this reason, BRUVs are typically used to estimate a relative abundance measure. Often the maximum number of individuals in a frame, called maxN, is used as the relative abundance metric [38].

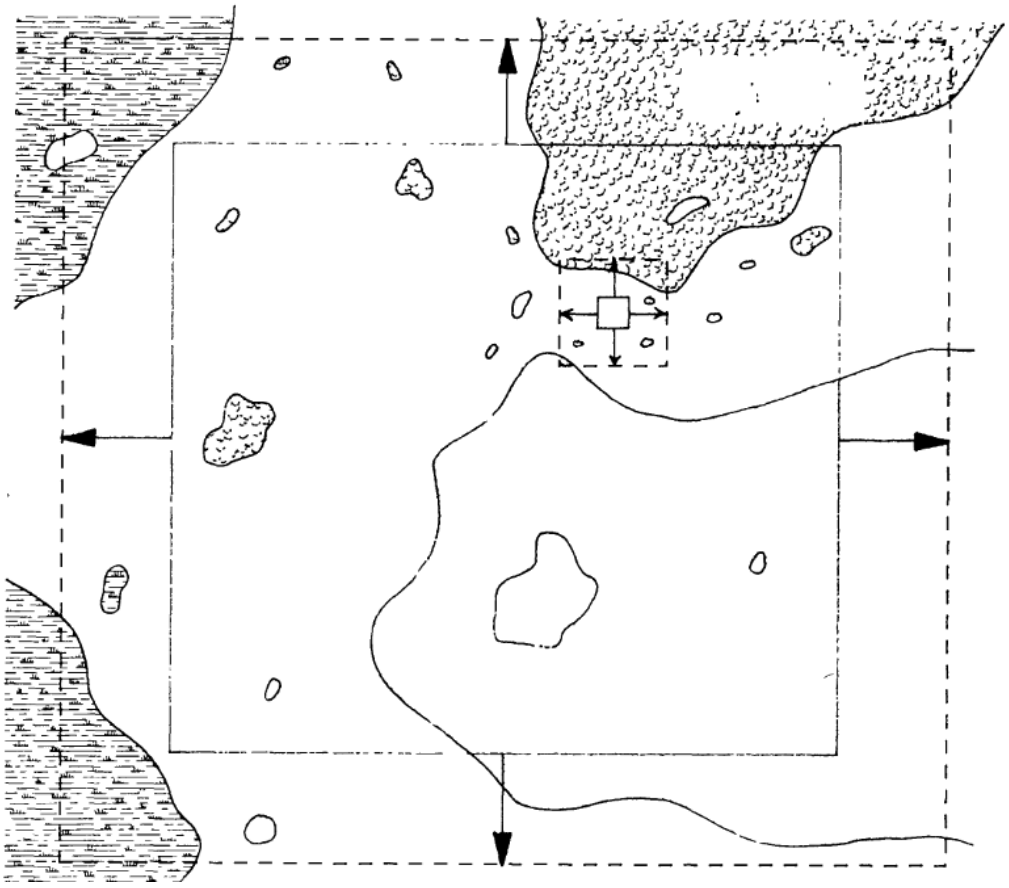


Figure 2.2 – From [203]. An illustration of the impact of observation scale in spatially heterogeneous habitats (such as reefs). Increasing the extent of the study (large squares) increases the likelihood that distinct features in the area are captured. Increasing the grain (small squares) means that smaller patches are combined into single samples and are therefore averaged out.

2.2 Effects of Survey Scale

It is impossible to observe an entire ecological community in all but the simplest of systems [195]. Therefore, an understanding of the ecology of a system, that is the structure, composition and interactions between the species of which it is comprised, must be developed based on observed samples. The characteristics of the sampling technique therefore dictate the inferences which can be made, so careful design of the method of observation is critical. In particular, the ability to make inferences from a sample based study depends on its extent, the area covered by samples, and its grain,

the resolution of each individual sample (Figure 2.2 [203]).

In practice, to observe features or patterns over a larger area, you generally need to increase the extent of your individual samples (*i.e.* increase the survey grain). However, by increasing the grain, you lose the ability to resolve patterns at scales smaller than this unit. Conversely, if you were interested in smaller patterns, you would reduce the extent your individual sampling units (*i.e.* reduce the survey grain), more of which would be required to observe the variation within your survey extent, essentially putting an upper limit on the area you can observe. Therefore, using traditional methods, there is a trade-off between the grain and extent of surveys.

Using traditional observation methods, the spatial information within a sampling unit is typically lost once the observation is made. Take the example of a diver counting organisms in a quadrat to estimate their abundance. The count of individuals within the quadrat is typically compressed to a single value for that sample, so their distribution within the quadrat is not recorded [76, 100, 118].

Reef-scale image reconstruction techniques allow for large areas of reef to be observed, while preserving the spatial distribution of the features within the mapped region. The use of AUVs to produce these reconstructions allows us to generate a continuous, georeferenced image of the seafloor over large extents [108]. The ability to capture large swathes of habitat in a single survey mean that the spatial distribution of elements within the survey area are preserved, allowing fine-grain observations over large extents and addressing the trade-off inherent to typical observational techniques.

2.3 Observing Species-Habitat Associations

Understanding how species interact with their habitat, and the mechanisms which drive variations in the diversity and distributions of organisms is a central problem in ecology. Countless studies exist that explore the relationships between fauna and their habitat.

In the marine environment, particular interest is shown on the effect of structural

complexity on populations of corals (*e.g.* [62, 84]) and mobile fauna (*e.g.* reef fish [129, 45, 136, 82, 73, 146], demersal fish [192, 9], and mobile invertebrates [5, 101]).

Another common field of research is exploring the associations between mobile fauna and habitat forming organisms. Examples of studies exploring such associations include relationships between fish distributions and coral cover [15, 136, 113], fish distributions and macroalgal community composition [154, 210], invertebrate distributions and macroalgal community [215], seagrass shoot density and epifaunal community composition [95], and reef fish abundance and coral settlement and recruitment [42].

Species distribution models (SDMs) are numerical models used to quantify relationships such as those observed in the studies above. Building an SDM involves associating the occurrence or abundance of a species of interest with habitat characteristics, and can be used to predict their influence on the spatial distribution of the target species [59]. SDMs have been widely applied to marine species, particularly fish (*e.g.* [72, 73, 139]), but also invertebrates such as mussels [96], barnacles [48] and rock lobster [101].

To observe species-habitat interactions (and to build robust SDMs) the studies presented in this section require observing where individuals of the target species are present, and the characteristics of the habitat in these locations. Implementation of these two steps are typically decoupled and these observations are constrained by the tradeoff between grain and extent described in Section 2.2. This means that observing such interactions at fine spatial scales is difficult to scale to large spatial extents. The tools described in the next section allow spatially conserved, simultaneous observation of organisms and habitat over entire sections of reefs. Utilising such data to observe species-habitat interactions, and build SDMs, will allow for a better understanding of how marine species interact with their environment.

2.4 3D Reef Reconstructions

Three dimensional scenes can be reconstructed by imaging them from different viewpoints in a process called structure from motion (SfM)[194]. These tools have become increasingly common for observing reef ecosystems [62, 55, 155]. SfM can be used to capture a "snapshot" of the reef at a point in time, and facilitates non-destructive, fishery-independent observations of the spatial distributions of observed features and 3D structure of the reef habitat.

At smaller scales, photogrammetric reconstructions have been used to observe the 3D structure of individual coral colonies to quantify their growth rates and patterns [218, 61] and to predict their responses to disturbances [217].

Moving up to the reef-scale, SfM methods are able to generate reconstructions of whole sections of reef by stitching together images captured over large swathes of seafloor (*e.g.* [197, 160, 155, 193]). These approaches have been applied to observe a range of benthic habitats (*e.g.* coral reefs [35, 62, 164], temperate, rocky reefs [126, 186] and volcanic fields [116])

There are three primary benefits to using these reconstructions to observe benthic habitats, firstly they allow cost and time effective observation of traditional metrics such as organism density or percentage cover of habitat types [126, 186]. This is achieved by transferring the analysis time from in the field (as would be required for typical *in situ* analyses, to in the lab after the data is collected, reducing the expense of performing the surveys.[1] Secondly, these methods preserve the distribution of observed targets within the surveyed area. This means they can be used to analyse spatial interactions among them. Example applications include quantifying the patterns of clustering among coral colonies [58], or observing size-frequency distributions of corals [97]. Finally, since these approaches allow for precise georeferencing of the

[1] A potential tradeoff of this shifting of analysis away from the field is that if targets are not observed in the imagery (*e.g.* due to occlusion or poor imagery), these observations are lost. This is often reflected in lower estimates of abundance or density in studies using imagery compared to *in situ* surveys (*e.g.* [126, 186]). Similarly, while imagery can be used to confirm identifications by consultation with experts, if the distinguishing characteristics are not visible in the collected images, a coarser classification must be used, or *in situ* confirmations of identifications are required [193]

generated reconstructions, they can be used to observe changes to a site over time. This has been exploited to observe changes in structural complexity of a coral reef in response to disturbance [62], observerving patterns of growth in individual coral colonies [117], and for generating baseline data and observing the response of coral communities to restoration activities such as in the Ecological Intelligence for Reef Restoration initiative (EcoRRAP).

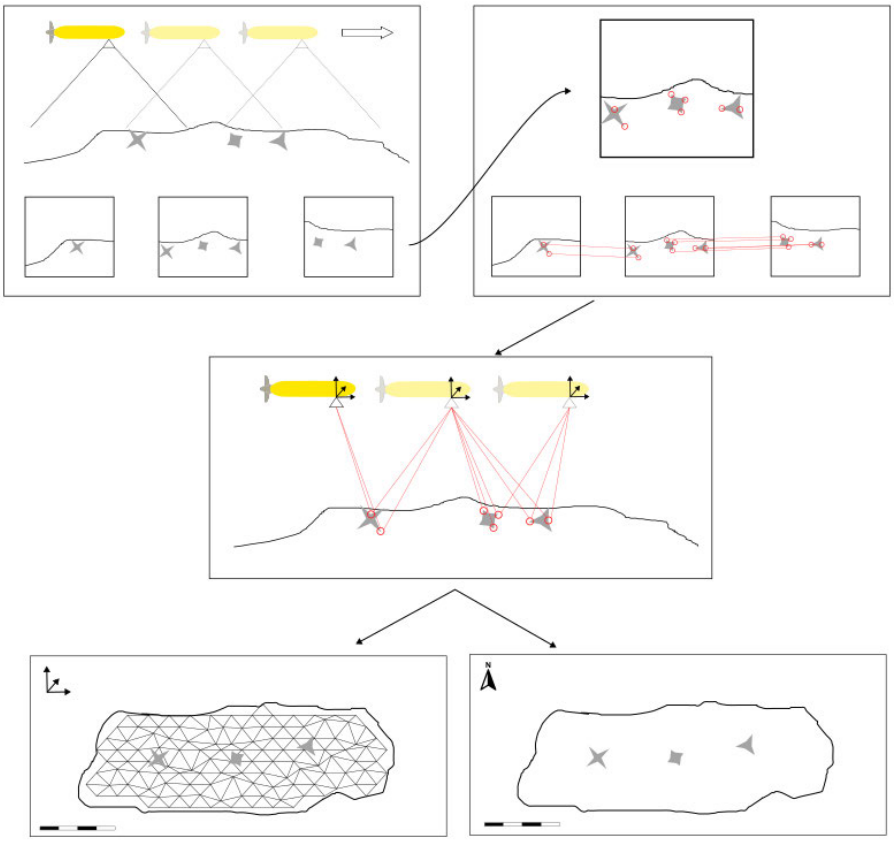


Figure 2.3 – Process for generating 3D seafloor reconstruction using terrain features. Overlapping seafloor images containing terrain features (grey shapes) are collected from a moving platform (**top left**). Keypoints are detected in each image and matching features between images are identified (**top right**). The keypoints are used to triangulate the position and orientation of the camera at each image (**middle**). The 3D positions of the keypoints can then be estimated and used to generate a 3D triangular mesh (**bottom left**) and a 2D georeferenced projection of the seafloor (**bottom right**).

Figure 2.3 demonstrates the process for collecting imagery and generating a recon-

struction at this scale. First, overlapping images of the seafloor are captured. A keypoint detector (see examples in Section 2.5.1) is used to characterise keypoints in the images. By matching keypoints between images, the relative trajectories of the keypoints can be estimated. Since these are terrain features and therefore are stationary, this relative motion can be solely attributed to the movement of the camera between frames, and therefore is used to estimate the 3D position and orientation of the camera as it captures images. This information can then be used to estimate the 3D position of the identified keypoints, and therefore build a model of the reef structure.

The reconstruction can be scaled using reference features of known dimension (referred to as ground control points). This allows that accurate size and distance measurements of observed features in the reconstruction and projections [83].

The 3D reconstructions generated using SfM can be used to create a more nuanced understanding of the geometry of the reef compared to *in situ* methods like chain-tape, or visual scoring. Friedmann *et al.* [69] describe an approach for generating multi-scale measures of habitat complexity from a 3D reconstruction.

Similarly, preserving the spatial distribution of habitat forming organisms (*e.g.* corals) can allow for unique ways of quantifying their communities. For example, it facilitates simple analysis of the distribution of different sized individuals, or trends in the adjacency of different species.

While purely appearance based SfM applications are convenient since they only require a camera in the field, the computational requirements increase significantly as the reconstruction extent increases. This is because the keypoint matching process needs to consider many candidates over all the collected images. This often makes the processing step the bottleneck in generating reconstructions.

2.4.1 Benthic Imaging AUVs for Seafloor Observation

The use of autonomous platforms to collect the imagery for these processes reduces the limitations imposed on scientific divers. Benthic imaging AUVs can be programmed

to collect overlapping imagery over large spatial extents, and can be fitted with navigation sensors. Precomputing the vehicle's path using a navigation algorithm with information the vehicles sensors allows for efficient detection of correspondences among large numbers of images facilitating the construction of sub-millimetre resolution habitat reconstructions over large spatial extents (up to $\mathcal{O}(10^4 m^2)$) [180, 205, 24, 81, 193].

A common solution employed for the navigation of mobile autonomous platforms involves concurrently estimating the robot's position and mapping landmarks of the environment as the robot moves. This approach is known as simultaneous localisation and mapping(SLAM)[183], and fuses information derived from all the robot's sensors (including the benthic images) to determine the optimal solution for the vehicle's path. By using the suite of sensors rather than purely matching image features (as in image only SfM), much larger reconstructions can be created with fewer computational resources.

The AUV *Sirius*, operated by the Australian Centre for Robotics (ACFR), is based on the Seabed platform built by Woods Hole Oceanographic Institution [180]. *Sirius* collects images of the seafloor at 2 m altitude using its onboard stereo camera pair[2]. These images are used to generate a three dimensional, georeferenced reconstruction with the aid of its onboard navigation sensors; multibeam sonar, a depth sensor, Doppler velocity log (DVL) including a compass with integrated roll and pitch sensors, and an ultra short baseline (USBL) acoustic positioning system. The imagery and navigation data are combined using a SLAM algorithm [206, 108]. *Sirius* has been in operation since 2007 and has collected over 5 million benthic images, predominantly in reef environments around Australia.

The survey data collected by *Sirius* used throughout this thesis was accessed through Australia's Integrated Marine Observing System (IMOS). IMOS is enabled by the National Collaborative Research Infrastructure Strategy (NCRIS). It is operated by a consortium of institutions as an unincorporated joint venture, with the University of Tasmania as Lead Agent.

[2]The use of a calibrated stereo camera pair allows for direct scaling of the captured images and therefore the generated 3D reconstructions and projections. This eliminates the need to use ground control points to scale the imagery.

2.4.2 Observing Benthic Fauna in Seafloor Reconstructions

Reef fauna are often seen in AUV generated habitat reconstructions, and their positions can be accurately determined using the vehicle's precise navigation information. This makes them an ideal platform for performing surveys of these fauna and their use for this purpose is becoming increasingly widespread. Ling *et al.* [126] utilised *Sirius* to observe urchin habitats and found that while their population density estimates (*i.e.* the number of individuals per unit area) manually derived from the AUV imagery were lower than those made using *in situ* diver surveys, underlying temporal changes in the populations could still be observed. Tolimieri *et al.* [191], Clarke *et al.* [43], and Seiler *et al.* [174] used AUV platforms to survey benthic and demersal fish in rocky and highly rugose habitats, which are typically inaccessible to traditional trawl surveys. The use of these platforms allowed the observation of spatial patterns within the surveyed areas such as age and size specific habitat preferences. Such patterns are not possible to capture using extractive methods, and *in situ* diver based techniques are difficult to scale to the survey extents that AUVs are capable of covering.

These studies demonstrate that the data collected by these vehicles can be used to support scientific exploration of marine habitats and can provide unique insights compared to existing survey techniques. However, they rely on manual annotation of the imagery or terrain models to extract the species of interest. AUV surveys can be composed of tens of thousands of individual images, so it can be prohibitively time consuming to manually find and identify all individual animals or habitat features captured in a survey. Additionally, it has been shown that multiple observers are often necessary in order to limit the biases introduced by annotators, further increasing the required annotation effort [57]. This time requirement can result in a bottleneck in data collection and processing pipelines therefore increasing the cost of performing surveys. To combat this, researchers often choose to collect smaller datasets, or to subsample their image sets, often only analysing a fraction of the collected data [157].

2.5 Automated Image Analysis

Image based ecological survey techniques can generate large volumes of image data, which typically requires a significant time investment to process and extract the desired information. This can result in a bottleneck in the data collection and processing pipeline [86, 131].

Machine learning based image analysis techniques can allow for objects of interest to be automatically detected, identified, and measured, with minimal human intervention. Recent advances in algorithm design, and increased availability of high performance computing has led to rapid development of artificial neural networks for solving computer vision problems [198, 1]. Deep neural networks learn patterns from data they are provided and reduce the need for manually crafting image features. In particular, convolutional neural networks (CNNs) which can learn abstract representations from image data [8], are well suited to the challenges of automating the analysis of underwater images and often outperform traditional feature based techniques in machine vision challenges [181, 120]. Automated image processing facilitates rapid processing of the high volumes of data generated using AUV collected image-based seafloor reconstructions and facilitate their use as large scale samples for measuring reef community compositions and distributions.

2.5.1 Feature Based Learning

Traditional machine learning techniques for image analysis involve identifying specific features in images (*e.g.* edges [37] or corners [88]) then classifying the images based on the presence and distribution of these features.

Such approaches have been applied for classifying underwater images, including for classifying kelps [20], and other benthic categories (*e.g.* [70]), as well as for census and measurement of fish [176, 185].

However, the need to handcraft image features mean these methods are fairly rigid,

and require much fine tuning to adapt to the complex underwater imaging environment.

2.5.2 Deep Learning

Deep learning models do away with the need for handcrafted image features, instead they learn the relevant representations from the training data [149, 142]. Their ability to learn abstract concepts from large datasets, mean they are useful for extracting information from complex data sources, such as audio and imagery. This means that, provided they are trained on sufficiently large and diverse datasets, they can become invariant to non-informative fluctuations in the data, and do not need to be manually tuned to new datasets. These characteristics make deep approaches extremely attractive for processing underwater images, since large amounts of noise and poor image quality are often encountered due to the light attenuating characteristics of water, and the ubiquity of suspended sediments in marine environments.

In particular, CNNs, which include layers that learn filters which are convolved over the data, are particularly suited to such tasks. This includes facial detection [188] and recognition [173], semantic segmentation of medical imagery [166] or to aid autonomous driving [220],

2.5.3 Automated Analysis of Underwater Imagery

Deep learning has been taken up rapidly for processing underwater imagery for observing marine habitats (see review by Rubbens *et al.* [167]). This includes classification of corals in benthic images [14, 133], detection of fauna on the seafloor [171, 144, 189] and in the water column [201], classification [177], detection, and measurement of fish [105, 169, 162], for planning AUV surveys [175], detection of harmful species to aid with their control [52], and automating analysis of video surveys of fish from BRUVs [134].

Additionally, deep learning based approaches have been used to process large-scale seafloor reconstructions. Mahmood *et al.* [132] utilised a CNN based network, VGGnet [179], to annotate coral reef mosaics generated by the AUV *Sirius* in North-Western Australia.

Mizuno *et al.* [138] employ the U-Net image segmentation algorithm to perform pixel-wise classification of large scale image mosaics of coral reefs. The U-Net algorithm was originally developed for processing medical images but has been commonly applied for many segmentation tasks such as for analysis of remote sensing images [94] *e.g.* for street scenes [11].

2.6 Summary

In this chapter, we've discussed traditional techniques for observing benthic ecosystems. These tools are limited in their ability to estimate fine-scale spatial distributions over large extents (*e.g.* for observing and modelling associations between reef associated fauna and habitat characteristics). Emerging sampling techniques such as AUVs and SfM pipelines allow large areas of seafloor to be mapped, facilitating direct observation of individuals and habitat features.

However, these technologies produce large volumes of data, which are time-consuming to manually annotate. Also, since they survey large areas by stitching together images taken from many viewpoints, estimating the abundance of mobile targets requires a way of associating detections to individuals.

This thesis will address these challenges by developing:

1. An aggregation approach which generates scored hypotheses for the association of detections to individuals. These can be used to generate a probabilistic estimate of the distribution of a population of mobile, reef-associated fauna (in Chapter 3).

2. A pipeline for automatically detecting and localising targets from AUV data (in Chapters 4, and 5). This addresses the bottleneck where human annotation is required to analyse AUV imagery, allowing for rapid, cost-effective surveys of benthic targets over large seafloor extents.

By addressing these gaps in current capabilities, the tools presented in this thesis will allow for mapping benthic features and fauna with minimal manual annotation and using existing field sampling techniques. Examples of the novel analysis techniques for these data made possible by the developments in this thesis are presented in Chapter 6.

Chapter 3

Spatial Census of Seafloor Targets from a Moving Benthic-Imaging Platform

3.1 Introduction

As discussed in Section 2.4, to create an image-based reconstruction, overlapping images from multiple viewpoints must be captured. A composite of these overlapping images can be produced by aligning matching features among images allowing for a 3D reconstruction and 2D projections of the entire surveyed area. Since the image collection process is not instantaneous, the observed region of seafloor grows as the imager traverses the reef, be it propelled by an autonomous vehicle, or swum by a human operator. The rate of this growth is largely a function of the seafloor area captured in each image. At the extreme, if the entire area is captured in a single image, the rate of expansion is instant. In a more realistic case, where multiple images are used, the rate of map expansion also depends on the vehicle's linear speed and the survey path used to capture the imagery (see Figure. 3.1).

Estimating the abundance and distribution of benthic targets using a 3D, image-based

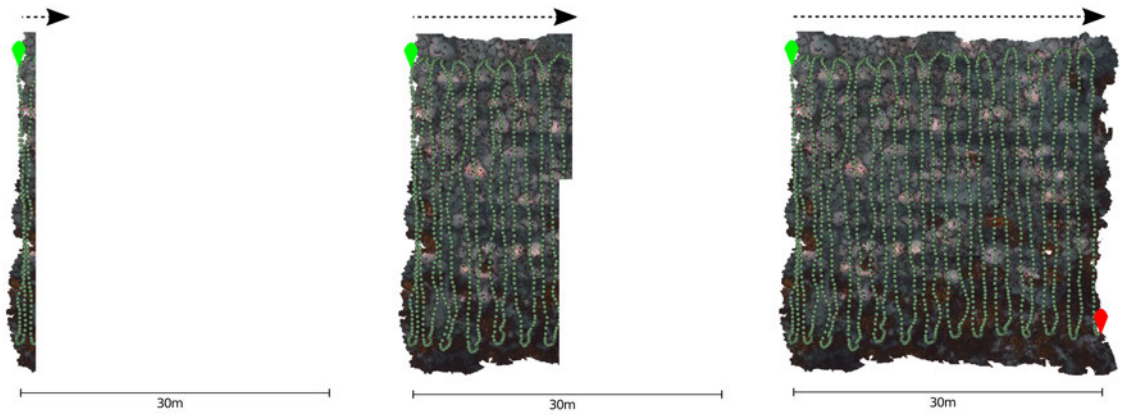


Figure 3.1 – A demonstration of the expansion of a mosaic generated from imagery collected by the AUV *Sirius* at St Helens Island, north eastern Tasmania. The vehicle began the survey at the green pin and ended at the red pin, travelling along the green dotted path. Each dot shows the location where an image was captured by the vehicle. The mapped area grows in the direction indicated by the dotted arrow as the AUV progresses along the survey path.

habitat reconstruction (*e.g.* as collected by a benthic imaging AUV) involves three steps:

1. Detection - identifying targets in the survey imagery
2. Localisation - estimating the position of target detections within the survey region
3. Aggregation - associating each detection to an individual

This chapter develops the approach needed to observe benthic targets from imagery collected from a moving platform (*e.g.* an AUV). The design of these surveys typically assumes a stationary scene, therefore the methods presented here adapt them for the observation of mobile targets.

If we wish to observe benthic targets using a moving imaging platform, the rate at which the targets are likely to move will impact their appearance in the generated

imagery and therefore the required approach for each of these steps. The effect of target motion on their appearance in the composite images is explored in section 3.2.

Where the rate of map expansion is much greater than the targets' rate of movement (*i.e.* stationary or slow moving targets), individuals can be detected in the generated photomosaic. This simplifies the localisation and aggregation process since the data is already georeferenced, and each detection can be assumed to be of a distinct individual.

Faster moving targets (*i.e.* those whose rate of motion approaches the rate of map expansion) cannot be reliably observed in the photomosaic, and need to be observed in the individual, unprocessed images. The steps required to detect benthic targets are outlined in section 3.3.

If such targets are detected in the source imagery, they cannot simply be localised using the georeferencing of the photomosaic. Section 3.4 outlines an approach for localising detections made this way. Additionally, since they may move while the survey is underway, each detection cannot be assumed to be a unique individual. Section 3.5 describes a proposed approach for associating localised detections of mobile benthic targets to individuals in order to estimate their abundance and spatial distributions.

3.2 Impact of Target Motion on Appearance and Observability in Benthic Composite Images

The process by which images can be used to estimate the 3D structure and spatial distribution of benthic terrain features is described in Figure 2.3. Similarly, benthic habitat features and fauna can be used as keypoints in generating the reconstruction and therefore their positions can be estimated within the survey area. Movement of these targets - whether due to the force of the water column acting on a non-rigid target, or as a result of the target's locomotion - will affect how they appear in the generated composite image and 3D reconstruction. This, in turn, will affect the

process required to detect, localise, and aggregate detections in order to estimate the spatial distribution of targets within the surveyed region.

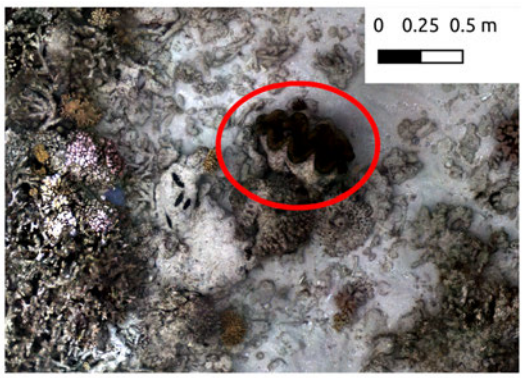
3.2.1 Stationary Benthic Targets

Physical habitat structure, sessile invertebrates, and habitat forming organisms (*e.g.* corals, sponges, and kelp) will typically be clearly visible in reef-scale photomosaics since they are attached to the substrate and cannot move while the survey images are collected. Many slow moving benthic fauna, including some echinoderms, arthropods, molluscs, and even species of bottom dwelling fish may also be observed in these reconstructions provided they do not move while the images are collected (Figure 3.2).

Rigid benthic features such as reef substrate, or hard corals will not move between subsequent images, regardless of the time between them within the survey. Therefore, parts of these features will produce image keypoints which can be readily matched between images and therefore will be present and undistorted in the final composite image products.

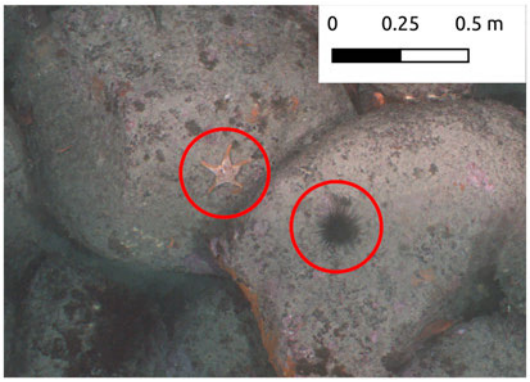
Non-rigid, attached habitat features and sessile organisms (*e.g.* kelps, soft corals) while not moving along the seafloor, can sway and distort between successive images (Figure 3.3). This makes it less likely that matching kepoints will be found for the moving parts of these organisms, and if they are, the solution that incorporates these will be incompatible with features taken from rigid points in the scene.

The appearance of the reconstruction is built using the images which provide the best views of the selected keypoints. So, targets which are present in the images but whose keypoints are not used due to their movement may still appear in the reconstructions since they are likely to be adjacent to other valid keypoints in the scene around them (or on rigid sections of the target in questions *e.g.* a kelp's holdfast). However, their relative position to these keypoints will be different in each image due to their movement, which can result in shadowing (*i.e.* multiple views of the target being superimposed), or distorted in the composite image.



(a) A giant clam (genus *Tridacna*)

(b) A patch of seaweeds



(c) A sea star and a long-spined sea urchin (*Centrostephanus rodgersii*)

(d) A bottom dwelling reef perch (*Helicolenus percoides*)

Figure 3.2 – Examples of benthic targets, with increasing levels of mobility, observed in photomosaics generated using a benthic imaging AUV. Rigid, stationary targets such as clams (a) appear undistorted in photmosaics. While sessile organisms such as seaweeds (b) are attached to the seafloor and cannot move from their positions, their non-rigid structure mean they often appear distorted in the composite image due to deformation between views by the imager (see Figure 3.3). Mobile organisms such as in (c) are capable of movement along the seafloor, but are typically slow moving so are often observed in the same position in the survey image sequence where they typically appear undistorted in the photomosaic. Observing fish in composite benthic images is typically difficult due to their capacity for rapid movement throughout the water column. However, some species such as the perch shown in (d) spend extended periods of time resting on the bottom where they can be observed in composite images .

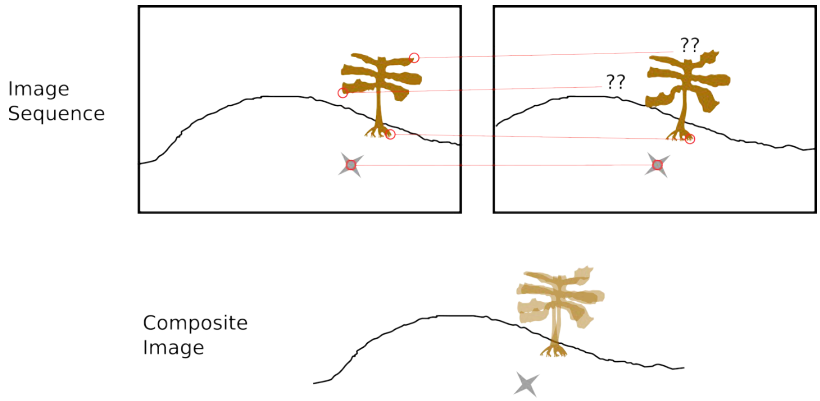


Figure 3.3 – Non-rigid benthic organisms such as kelp are challenging to observe using composite imaging techniques such as structure from motion (SfM). This figure shows two views of a section of seafloor containing a kelp. Matching keypoints can be found in rigid elements of the scene such as the terrain feature (shown as a grey diamond), and the kelp's attachment point to the reef (called a holdfast). However, movement of the blades of the kelp result will mean it is difficult to locate matching features on those parts. This results in a blurred appearance when the images are combined to create the final data products.

3.2.2 Mobile Benthic Targets

Observation of mobile benthic targets in composite images will result in similar effects due to the deformation of their body parts as they move. More significantly, their locomotion along the seafloor will again result in keypoints on the organism being incompatible with the larger solution and these keypoints are therefore less likely to be used to create the composite image products. This will again result in distortions (Figure 3.4) or shadowing of the target along its path (Figure 3.5).

Since overlapping images of the seafloor are required in order to build the reconstruction and composite images, it is possible that if a target moves out of a particular patch of reef during the survey, the algorithm could use only views where it is not present to generate that region of the reconstruction. This would result in the target not appearing at all in the final reconstruction and composite image.

These effects mean that the target's abundance in the survey area cannot be reliably estimated from detections in the composite images, since individual targets can appear in multiple instances or be missing entirely.

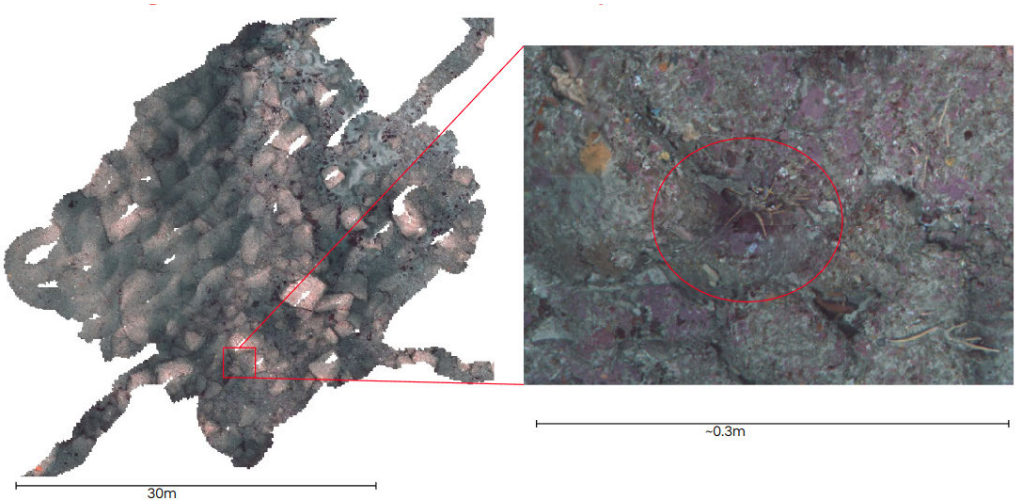


Figure 3.4 – An example of distortion of a mobile benthic target (a southern rock lobster) in a composite photomosaic. Movement of the target while the survey image is being collected can result in such distortion, but can also result in the target appearing multiple times (called ghosting) or being completely absent in the depending on the keypoints and images used to generate the final projection.

While their movement may mean that keypoints on the target are unlikely to be matched between frames, their proximity to the seafloor mean that instances can usually still be localised using their adjacency to features in the stationary scene.

3.2.3 Free Swimming Organisms

Organisms that move through the scene but are not necessarily associated with the seafloor (*e.g.* fish) can be observed in benthic composite images, however there are significant challenges associated with using these data to survey free swimming organisms.

Typically, when such organisms are observed in composite images, they are on or near the seafloor, so can be observed similarly to reef-associated fauna (see Figure 3.2). If they are swimming in the water column as the imager passes several factors make them unlikely to be observed in either the raw images or processed composite.

Since these targets are able to move around the water column, they are able to move out of the volume being observed (*i.e.* swim above the imager) where they cannot

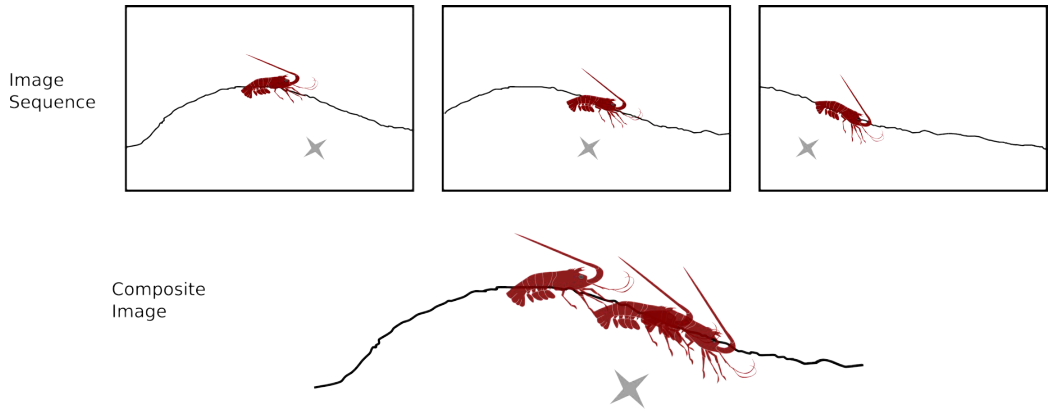


Figure 3.5 – Illustration of the effect of a target's locomotion on their appearance in a composite image. A sequence of three images captures a lobster moving along the seafloor (**top**). If these images are used to create a reconstruction of the seafloor, the individual would appear in three positions corresponding to their location in the individual frames. This duplication complicates estimating the spatial distribution of mobile benthic targets directly from composite images, since a single individual can result in multiple detections.

be detected. Additionally, benthic imaging systems, as the name implies tend to be suited to observing the seafloor, so targets above the seafloor are likely to be out of focus and undetected even if they are within the observed volume. It has been observed that some fish will, in fact, follow the AUV around, rendering them invisible to a vehicle imaging the seafloor [187]. Furthermore, this suggests that the imager may in fact have some impact on fish behaviour of target organisms. Considerations of the potential impact on survey operations is beyond the scope of this particular work.

In the case of mobile benthic fauna as described above, the target can be localised using the keypoints generated from stationary parts of the scene adjacent to them. For free swimming targets, there are unlikely to be any rigid, stationary objects nearby. This means they cannot be localised using the 3D reconstruction. For vehicles equipped with stereo cameras, the position of observed targets can be triangulated (similar to the approach used in stereo BRUVs and DOVs [90]) independently of the habitat reconstruction.

For these reasons, spatial census of free swimming mobile organisms is outside of the

scope of this thesis. Instead the focus will be on a pipeline for observing the spatial distributions of stationary and mobile reef associated targets. The following sections will describe how the appearance of benthic targets in composite images affects their detection, localisation, and aggregation required in order to estimate their spatial distributions.

3.3 Detection of Benthic Targets in Composite Benthic Images

The first step in estimating the spatial distribution of targets from composite images is to detect individuals. As discussed, the movement of a target affects its appearance in the imagery, and will therefore affect the way in which individuals can be detected for estimating their distribution and abundance.

3.3.1 Detection of Stationary Benthic Targets

Stationary benthic features which readily appear in 2D projections of reef-scale reconstructions can simply be detected in the composite image. Dense aggregations of targets in the survey area can make this process slow to complete manually, and therefore automated approaches can help in facilitating scalable, low cost analysis of these imagery.

Objects which occur as discrete instances - like slow moving benthic fauna such as sea-stars - can be detected as such either by manually marking the point or region they occupy in the image or mosaic, or using automated object detection algorithms (see Chapter 4). Other features, in particular habitat forming organisms like kelp or coral are more difficult to assign as an instance. Instead these tend to appear as regions of the reef which would need to be delineated either manually, or using an automated segmentation approach (see Section 6.3).

3.3.2 Detection of Mobile Benthic Targets

The distortion, shadowing, or absence of mobile benthic targets in composite images described in section 3.2.2 mean that they cannot be reliably detected in the generated composite images in order to estimate their abundance and distribution.

Instead, they must be detected in the unprocessed images. Here the limiting time scale becomes the shutter speed of the camera, rather than the rate of map expansion, facilitating detection of much faster moving targets.

SfM surveys can contain tens of thousands of images, so manually detecting targets in these images is likely to be prohibitive in terms of the cost and time required for comprehensive, manual annotation. Therefore, it becomes almost essential to employ automation in order to facilitate performing such surveys at scale (see Chapter 5).

Since the targets are not being detected in the georeferenced data products, localisation will involve reprojecting their position in the image onto the 3D reconstruction.

3.4 Localising Detected Benthic Targets

Composite images generated from 3D reconstructions on the seafloor can be goereferenced using either navigation sensors and calibrated stereo-cameras on the imaging platform, or using ground-control points with known positions in the scene. This allows estimation of the geographic position of points in the scene. For stationary elements of the imaged area - including attached benthic habitat features and fauna - this allows for straightforward estimation of their spatial distributions. Detections of individuals made in the 3D reconstruction or composite image are therefore also georeferenced. Since targets in the composite image represent unique individuals, these georeferenced detections can be used to estimate the abundance and distribution of targets in the survey area.

However, this approach must be modified for mobile targets which need to be observed in the unprocessed images. A process for reprojecting the location of the target detected in the images onto the 3D reconstruction is necessary in order to estimate the

position of observed targets. During the seafloor reconstruction process, the vehicle pose is estimated for each AUV image. The rigid transformation between the vehicle origin and the stereo-camera system is known, so the camera origin position and orientation is known for each image. Additionally, the stereo cameras are pre-calibrated so the intrinsic camera parameters and relative stereo orientation are known. This information can be used to transform an image point onto the 3D surface generated using the AUV data processing pipeline.

Consider a target detected in an unprocessed benthic image collected as part of an SfM survey. Assuming this target lies on the seafloor - as would be expected for a benthic species - a ray passing through the point of detection in the image intersects the seafloor at the position of that target in space (see Figure 3.6). Therefore, an estimate of the position of a detected target can be made by projecting a ray through the detection in the image and locating the intersection point on the generated 3D mesh. Section 3.4.1 describes the geometric solution for this transformation. Additionally, many SfM software suites (*e.g.* AgiSoft Metashape [3]) provide functions for performing this transformation in their graphical user interface (GUI) or application programming interface (API).

3.4.1 Reprojection Geometry

Since the camera model used to calibrate the system assumes linear projection of points, non-linear deviations introduced by imperfect optical equipment need to be corrected for. A four parameter distortion model [89] was used to perform this correction. The direction of a ray passing through a (distortion corrected) image point (u,v) in the reference frame of the camera is given by:

$$\boldsymbol{r_c} = C^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{3.1}$$

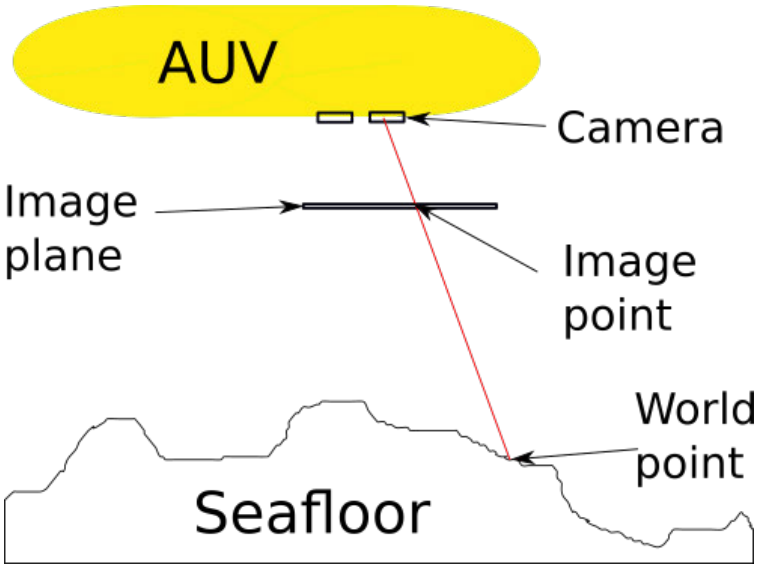


Figure 3.6 – Diagram showing the projection of an image point into world coordinates. A camera ray **(red)** passing through an image point (*e.g.* where an animal detection has been made) intersects the seafloor at the approximate 3D position of that point. This geometry can be used to localise a target detected in an unprocessed image by finding the intersection point between the camera ray and the 3D mesh.

where C is the camera matrix, which can be estimated using a camera calibration routine (*e.g.* [26]).

The vehicle's SLAM navigation solution provides an estimate of the vehicle's 3D position with respect to some local origin (v_1, v_2, v_3) and its orientation, in the form of Euler angles about the X, Y and Z axes (α, β, γ). The ray's direction, in the vehicle reference frame is given by:

$$\boldsymbol{r_v} = R\boldsymbol{r_c} \tag{3.2}$$

where

$$R = R_3 R_2 R_1 \tag{3.3}$$

and,

$$R_1 = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.4}$$

$$R_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & \sin\beta \\ 0 & -\sin\beta & \cos\beta \end{bmatrix} \tag{3.5}$$

$$R_3 = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.6}$$

So a point along the ray, in the vehicle's coordinate frame is given by:

$$\boldsymbol{P} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + t\boldsymbol{r_v} \tag{3.7}$$

The 3D seafloor reconstruction was output as a triangulated mesh surface comprising a collection of triangles joined by their shared edges. The normal vector of a plane on which the three vertices of a triangle $(\boldsymbol{p_1}, \boldsymbol{p_2}, \boldsymbol{p_3})$ lie can be given by:

$$\boldsymbol{N} = (A, B, C) = (\boldsymbol{p_2} - \boldsymbol{p_1}) \times (\boldsymbol{p_3} - \boldsymbol{p_1}) \tag{3.8}$$

And the distance from $\boldsymbol{p_1}$ to the origin is:

$$D = \sqrt{p_{11}^2 + p_{12}^2 + p_{13}^2} \tag{3.9}$$

The plane's equation can be given as:

$$A(p_{11}) + B(p_{12}) + C(p_{13}) = D \tag{3.10}$$

To find the location where a given ray intersects the mesh, the ray intersection between the planes containing each of the mesh triangles was calculated. To determine whether the intersection point lies within a given mesh triangle, the point is converted into barycentric coordinates [64] with respect to the mesh triangle being considered (Figure

3.7). A point (**P**) lying on the same plane as a triangle with vertices **A**, **B**, and **C** an be described using two scalars u and v, where:

$$P = u\vec{AC} + v\vec{AB} \tag{3.11}$$

If $u > 0$, $v > 0$, and $u + v < 1$, then **P** lies inside **ABC**.

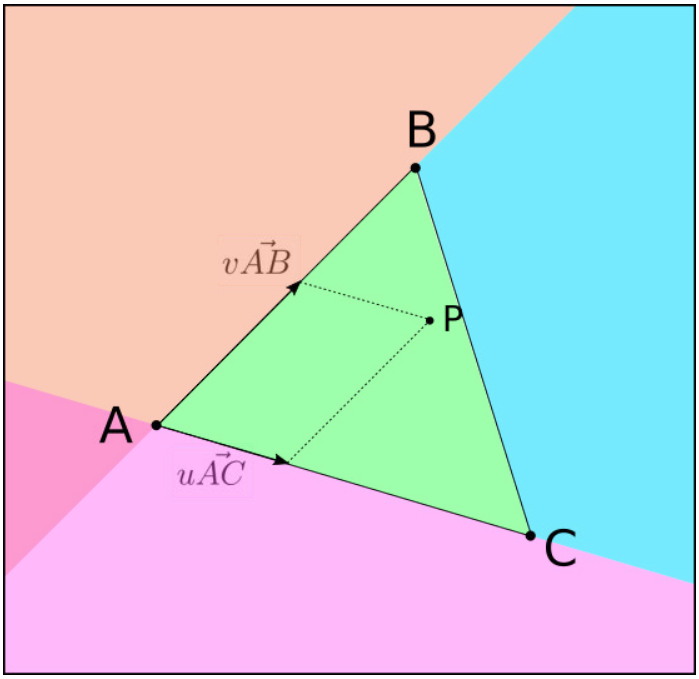


Figure 3.7 – A point **P** lying on the same plane as triangle **ABC** can be described in relation to the two vectors $\vec{AC}$ and $\vec{AB}$ using two scalars u and v. If u or v are less than zero (regions shown in orange and magenta respectively) or $u + v > 1$ (cyan region) then **P** lies outside **ABC**. This approach is used to determine whether the ray projected through a detected animal intersects with a given triangle in the 3D reconstruction mesh.

The barycentric method allows for fast determination of which mesh triangle the ray passes through. To further improve the computation time for calculating the ray-mesh intersection point, only the portion of the triangular mesh near to the location of the vehicle at the time of the detection are considered. It is possible that a ray could intersect the mesh at more than one point, particularly if the mesh is especially complex and folded upon itself. In this case, the first point where the ray intersects

the mesh is likely to be where the detected individual is located, since that surface would be visible to the camera. This scenario would also be largely avoided when using images from an AUV or other survey where a fixed altitude is maintained by the imaging platform, resulting in a predominantly overhead view of the seafloor.

We've discussed how localised detections of mobile benthic fauna can be generated from reef-scale composite images. These detections can be used to interrogate *e.g.* how habitat characteristics drive the distribution of mobile benthic fauna (see Chapter 5). However, individuals are likely to be observed in multiple images as the vehicle traverses the survey area, so are likely to result in more than one localised detection. Therefore we need an approach to aggregate these detections and associate them to individuals in order to make reliable estimates of the abundance of mobile benthic targets.

3.5 Aggregation of Detections of Benthic Targets

After localised detections of individuals are made, the final step in generating an estimate of their abundance and distribution is assigning each detection to a unique individual. This ensures that repeat observations of the same individual do not result in an overestimate of target abundance.

3.5.1 Aggregation of Stationary Benthic Targets

Stationary targets can typically be detected and localised directly from the photomosaic or 3D reconstruction. Since they would not have moved in between image captures, it can be assumed that each detection in the composite image is a unique individual. This means that these detections can directly be used to estimate the abundance and distributions of individuals in the survey area.

However, there are some cases where observing benthic targets in the underlying imagery is beneficial. Photomosaics are a 2D projection of a 3D surface. Consider

the case of a highly rugose habitat where the seafloor substrate forms overhangs. A top-down projection of this habitat will not necessarily capture all the places where benthic targets may inhabit.These regions are also often poorly resolved in 3D reconstructions, making observations of targets difficult. Therefore the underlying imagery may provide views of individuals not seen in the composite image or 3D reconstruction.

These points need to be reprojected onto the surface in order to estimate their positions. Errors in this process may result in observations of the same individual being projected to different locations. Assuming a well solved reconstruction, these reprojections should be tightly grouped so can therefore be aggregated to prevent multiple counting of detections originating from a single individual.

3.5.2 Aggregation of Mobile Benthic Targets

As discussed the movement of mobile benthic targets impacts their appearance in composite images. Observing these targets in the underlying images is necessary for performing a reliable spatial census of individuals.

Again, to turn these detections into an estimate of the abundance and distribution of targets over the survey area, it is necessary to determine which detections are repeat observations of previously seen individuals, and which are new. Otherwise, repeated detections of a potentially moving individual may appear multiple times in the reconstruction due to the image coverage and vehicle movement needed to reconstruct the habitat. This may result in an overestimate of the population. This problem is compounded by the targets' movement, as repeat observations of the same individual will not necessarily occur at the same point on the seafloor as it does for stationary targets.

The primary challenge here arises due to the difficulties in distinguishing between individuals of many marine species commonly observed in imagery. Even for experts, it is often impossible to distinguish between individuals, and many image based observational techniques for quantifying populations of mobile fauna consider this in their

design. Typically surveys will employ a measure of relative abundance proxy such as MaxN [161, 208, 39] rather than purporting to measure the actual abundance of targets. As a result, we need a way to associate detections with a given individual in order to prevent overestimation of abundance and to generate more consistent models of the distributions of individuals within the surveyed region.

The problem of aggregating detections and associating them to target individuals can be performed using naive methods which simply cluster adjacent observations, through to more complicated approaches which consider the timing of detections and the behaviour of the target species. Examples of these approaches will be applied in Chapter 5 for estimating the abundance of mobile benthic targets.

3.6 Summary

In this chapter we have explored the steps involved in estimating the abundance and spatial distribution of seafloor targets using data generated from a moving benthic-imaging platform. The difference between the rate of expansion of the observed seafloor area, and movement of the target organisms was shown to affect the steps required to spatially census them.

In cases where the rate of map expansion is much greater than the possible movement of the targets, detections can be made in the processed composite image, providing a straightforward way to localise each target individual.

For targets with a greater likely rate of movement, detections are made the underlying source imagery since their movement can cause artefacts in the final composite image which make detections difficult. The detections are localised by projecting the image points onto the 3D reconstruction using the vehicle's navigation solution. A tracking approach for associating these localised detections to individuals was also prevented.

Regardless of the movement of the observed targets, a requirement for performing census of benthic targets requires localised detections of individuals in the survey area. Generating these detections from SfM data manually can be extremely time

consuming since the volume of images is so large. The next two chapters describe how this process can be automated, and applied to stationary and slow moving fauna (Chapter 4), and more mobile benthic fauna (Chapter 5)

Chapter 4

Estimating Abundance and Spatial Distribution of Stationary and Slow Moving Benthic Fauna

4.1 Introduction

The previous chapter discussed how the movement of benthic targets affects their appearance in reconstructed composite images and therefore the methods required to observe them. In this chapter these principles will be applied to spatially census slow moving targets, using the long-spined sea urchin (*Centrostephanus rodgersii*) as a case study.

To observe these targets using SfM survey data we first need to determine how much they are likely to move over the course of a typical survey. Section 4.3 examines studies of urchin locomotion, directly observes their appearance in AUV survey imagery, and considers methods that researchers have used to estimate their abundance from reef-scale composite images.

Urchins often appear in dense aggregations, so image based census methods may require time consuming manual annotation. Section 4.4 presents an automated ap-

proach to analysing reef-scale photomosaics for estimating urchin abundance allowing low cost surveys. Abundance estimates made using this automated approach are compared to those from existing studies using both *in situ* estimates and manual analysis of imagery in Section 4.5.

4.2 Invasive Long-Spined Sea Urchin in Tasmania

Urchin barrens are areas of reef originally inhabited by kelp or other macroalgae, where population growth of herbivorous sea urchins has lead to high levels of grazing, leaving an exposed rocky region with high densities of urchins. These regions often occur due to the expansion of the ranges of urchin populations to higher latitudes facilitated by the global trend of warming ocean temperatures, and are likely to become more widespread as this phenomenon accelerates [124, 63]. An increase in the southward extent of the warm water, poleward flowing East Australian Current [165] has resulted in increased larval transport of the long-spined sea urchin *Centrostephanus rodgersii*, native to waters further north, into Tasmanian. Elevated water temperatures in their newly expanded range have allowed them to establish growing populations, leading to an explosion in the abundances of urchins. There is now an estimated population of 20 million individuals, and a rapid rise in barren areas in eastern Tasmania, increasing from 3.4% to 15.2% from 2001/02 to 2016/17 [123].

Interventions have been established, including harvesting and culling, to control populations of *C. rodgersii*, promote kelp recovery, and combat further expansion of barrens in Tasmania [107, 47]. A recent Australian Goverment Senate inquiry [147] recommended significant investment in urchin control measures, in particular establishing a fishery for extracting urchins in Tasmania.

Rapid, low cost monitoring of these environments would therefore aid in tracking the spread of these invasive organisms and assessing the effectiveness of any implemented management efforts. If automated census methods can detect the same trends as *in situ* or manual analysis of image based surveys, their use as a replacement or supplement to existing surveys can be justified, allowing for increased temporal and

spatial resolution in monitoring without the cost of expanded manual surveys or image analysis.

4.2.1 Using Benthic Imaging AUV Surveys of Urchin Barrens

Observing fine scale spatial distributions of benthic targets *in situ* is very time consuming. A diver would need to map out each individual with some reference to the habitat. Estimating habitat characteristics at the same time would add more time. This would not be scalable to the extents being observed here so AUVs (or other SfM methods) make these analyses feasible. AUVs allow for efficient, scalable observations of large seafloor extents, and can operate beyond the depth and time restrictions of SCUBA diving. This makes them ideal tools for performing scalable, routine monitoring of urchin barrens.

Their ability to simultaneously estimate the spatial distribution and abundance of urchins, and model the seafloor structure and habitat composition would allow for detailed studies of the relationship between these animals and their habitat. This information could be used to better control their populations and identify vulnerable habitat types for protections.

As discussed in Chapter 3, the rate of movement of a target during an AUV survey impacts it's appearance in the generated data, and therefore the required observation technique. Therefore the typical rate of movement over the timescales of an AUV survey must be investigated before estimates of their abundance and spatial distribution can be made.

4.3 Effects of Urchin Movement During AUV Surveys

Sea urchins are equipped with adhesive appendages, called tube feet, which they use in conjunction with their spines to move across the seafloor [214]. They are relatively slow moving, with field observations showing average rates of motion $\mathcal{O}(10)$cm per day [135]. Flukes *et al.* [65] specifically observed the locomotory behaviour of *C. rodgersii* and found that their rate of movement was highly variable and depended on the time of day and the substrate. Mean displacement of an individual over 12 hours during the night (when they tend to be most active) was around 40cm.

So, while capable of locomotion, over the course of a typical AUV survey ($\mathcal{O}(1)$hour), an individual urchin is unlikely to move very far. However, examining the movement of individuals in survey imagery can support this.

Figure 4.1 shows the positions of urchin in benthic images taken approximately 20 minutes apart during the course of a survey by the AUV *Sirius*. Interrogating the positions of the urchins present in each pair of images shows that none of the observed individuals appear to have moved from their starting positions. This lack of movement of urchins means that urchins appearing in the photomosaic are likely to be unique individuals and abundance estimates can be made by counting them directly.

A study by Ling *et al.* [126] compared abundance estimates made by counting *C. rodgersii* individuals in urchin barren photomosaics generated using images collected by the AUV *Sirius* with those made using *in situ* counts by divers. They were able to detect similar temporal trends in abundance using both methods also supporting the assumption that counts made directly from the composite images are not significantly impacted by the locomotion of individuals during the survey. In fact, abundance estimates from the photomosaic tended to be lower than the *in situ* counts (Figure 4.5) which suggests that obstruction of individuals by seafloor structures presented a greater impact on abundance estimates than repeat appearance in the mosaic due to movement of the urchins.

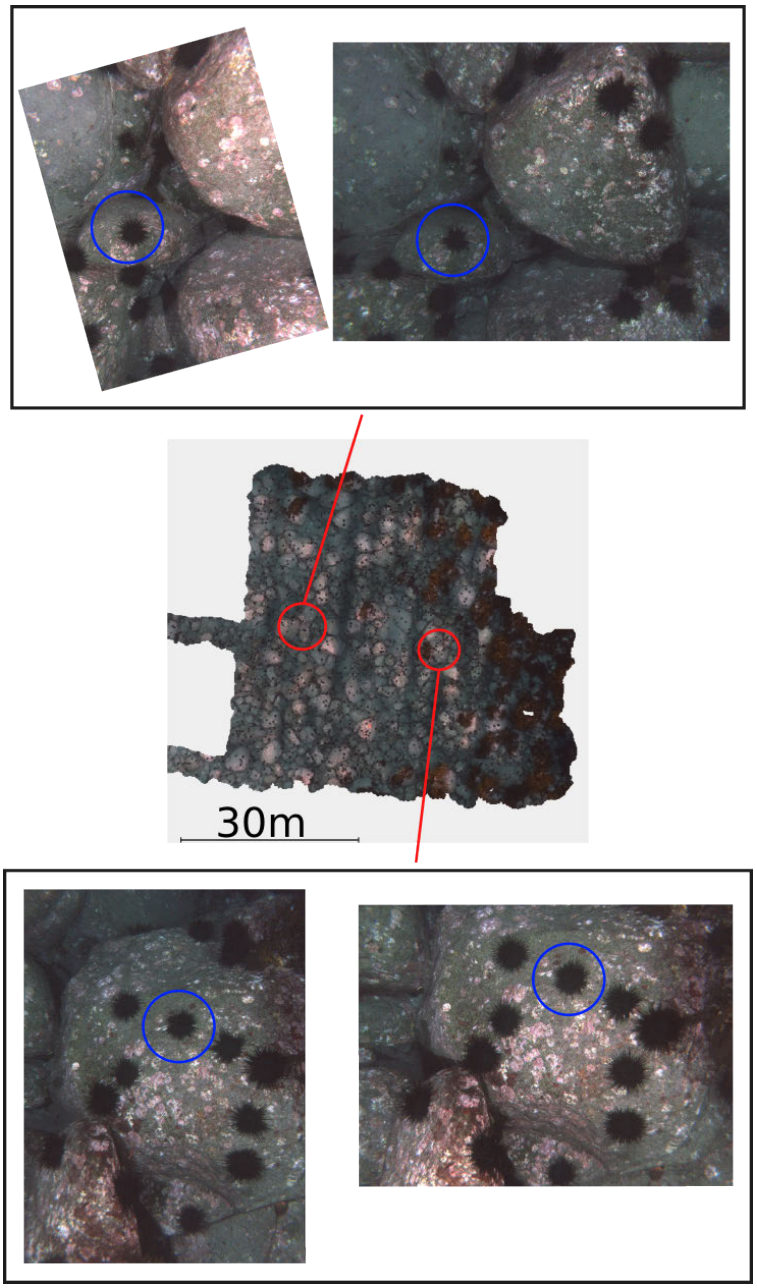


Figure 4.1 – Two pairs of benthic images (**top and bottom**) collected by the AUV *Sirius* at St Helens Island, Tasmania. Each pair of images captures the same region of seafloor (shown in red on the inset composite image generated using the images from the survey) but separated by approximately 20 minutes. Individual urchins captured in the frames have not moved along the seafloor over this time period (a matching individual is marked with a blue circle for each pair), suggesting that estimates of urchin abundance can be made by counting individuals appearing in the composite image. Note, the image pairs have been rotated to align the seafloor features captured.

4.4 Automated Detection of Urchins in Reef-Scale Photomosaics

Object detection algorithms delineate targets from the background, often using a bounding box. Applications include human face detection [219], vehicle [221] and pedestrian [30] detection for autonomous driving, and automated measurement of fish [134].

Here, we will use the RetinaNet object detection algorithm [122] to identify and localise urchins in reef-scale photomosaics collected by *Sirius*. RetinaNet uses a single-stage CNN-based approach to propose regions of images which match the target class or classes. Single-stage object detectors directly estimate regions within an input image which match the target class (or classes). This is in contrast with two-stage detectors, which firstly identify regions of interest in the input image which are then assigned to the appropriate class. Typically one-stage detectors process images faster, but achieve lower performance on benchmark datasets [184]. RetinaNet was developed to address this trade-off, and achieved state of the art scores for detection performance and speed at the time of its release. It does this by addressing the class imbalance issue which often arises in object detection problems due to the relative abundance of the "background" class compared to the target classes. RetinaNet applies a greater weighting to difficult (*i.e.* easily misclassified) classes compared to areas of obvious background. This makes the RetinaNet algorithm particularly suitable for object detection from benthic imagery, since typically the target organisms or features we are interested in are less abundant than the background substrate.

4.4.1 Training the Urchin Detection Algorithm

The data used to train and test the urchin detector were collected and described by Ling *et al.* [126]. Photomosaics were generated from imagery collected by *Sirius* at two sites, Elephant Rocks and St Helen's Island, in north-eastern Tasmania over three years (2009, 2011, 2013) using the method described in [206, 108]. All urchin

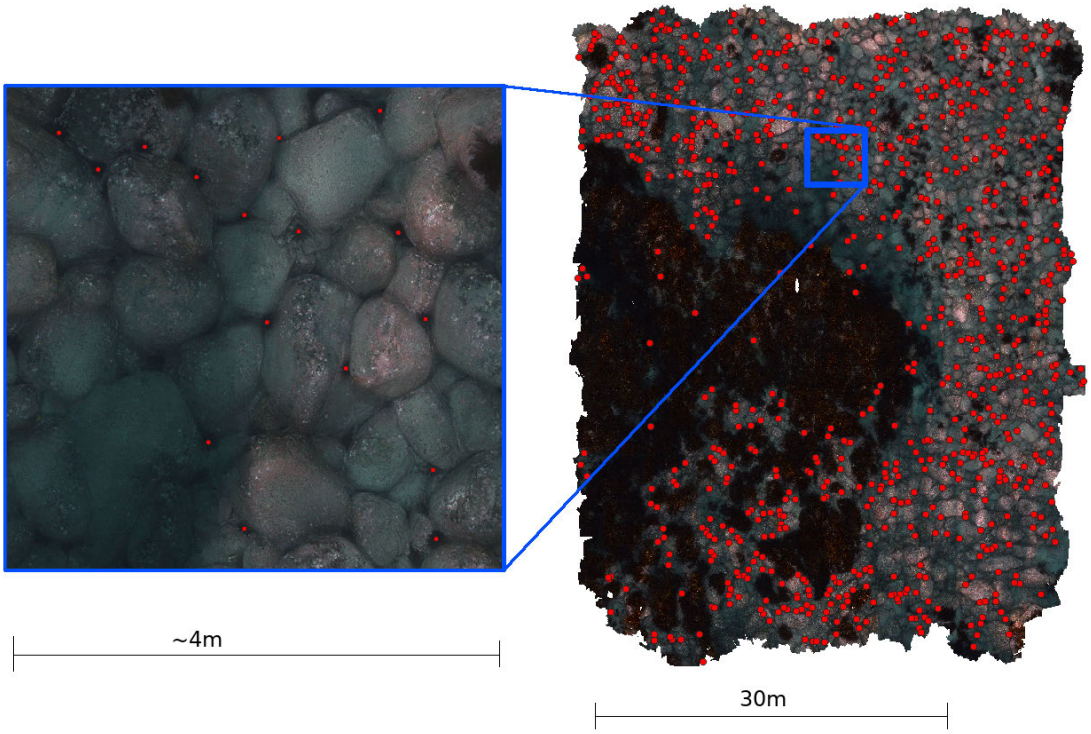


Figure 4.2 – Automated detections of long-spined sea urchins (*C. rodgersii*) by the trained RetinaNet algorithm in a large scale AUV photomosaic showing detail **(left)**, and the spatial distribution of detected urchins across the entire surveyed area **(right)**. Red dots show the predicted urchin positions.

individuals visible in one complete mosaic were manually annotated using bounding boxes. *Sirius* collects downward facing imagery and maintains a fixed height from the seafloor while performing a survey. From this perspective, urchins generally present in the imagery as circular cross sections of a relatively uniform size (see Figure 3.2). Due to this consistent appearance, and to allow for rapid annotation of urchins, square bounding boxes of fixed size were used to annotate the urchins. This meant that each individual could be labelled using only a single click at the centre of each urchin, rather than resizing a bounding box for each observed urchin in the mosaic. A custom Python script was used to perform this one-click annotation process on the photomosaics.

A 30m × 30m survey mosaic was used for training the object detection algorithm. All

visible urchins were manually annotated within this mosaic and formed the training set for the detection algorithm. 387 urchins were labelled in total on this reef. The trained algorithm was then presented with a new mosaic unseen at training time to propose bounding boxes around areas where it predicted the presence of an urchin. The centrepoint of each proposed bounding box detection was taken as an urchin's location. Since the photomosaic is georeferenced, the location of the detected object within the image can be converted into a geographic coordinate. This allows the position of each detected urchin to be projected onto the AUV generated habitat map resulting in a representation of the spatial distributions of all detected individuals in the surveyed area (Figure 4.2).

4.4.2 Evaluating Detection Performance

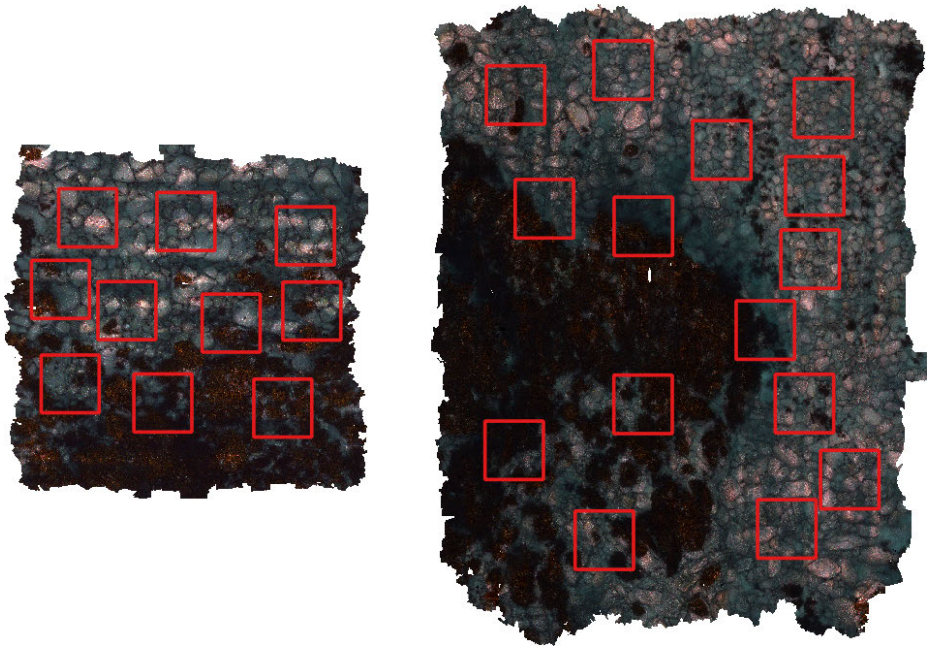


Figure 4.3 – Virtual 5 x 5 metre quadrats used to compare density estimates using manual annotation and automated detection

To validate the performance of the detection and projection process against a manual annotation process, the trained detection algorithm was presented with 30m $\times$ 30m photomosaics from two surveys unseen at training time, one at St Helens Island, and

one at Elephant Rock. Virtual $5m \times 5m$ quadrats were laid over these photomosaics (fig 4.3) and estimates of the density of the urchins were made using the automated detections, and by manually counting visible urchins in the quadrats (Figure 4.4). Automated density estimates approximated the manual estimates, with a mean error of $0.26m^{-2}$ over the 25 quadrats. This is well below the range of variation in abundance shown within each of the two sites (approximately $2.3m^{-2}$ and $0.8m^{-2}$ at St Helens Island and Elephant Rock respectively) and suggests that this automated approach can be used to complement or replace manual image annotation to estimate urchin abundance.

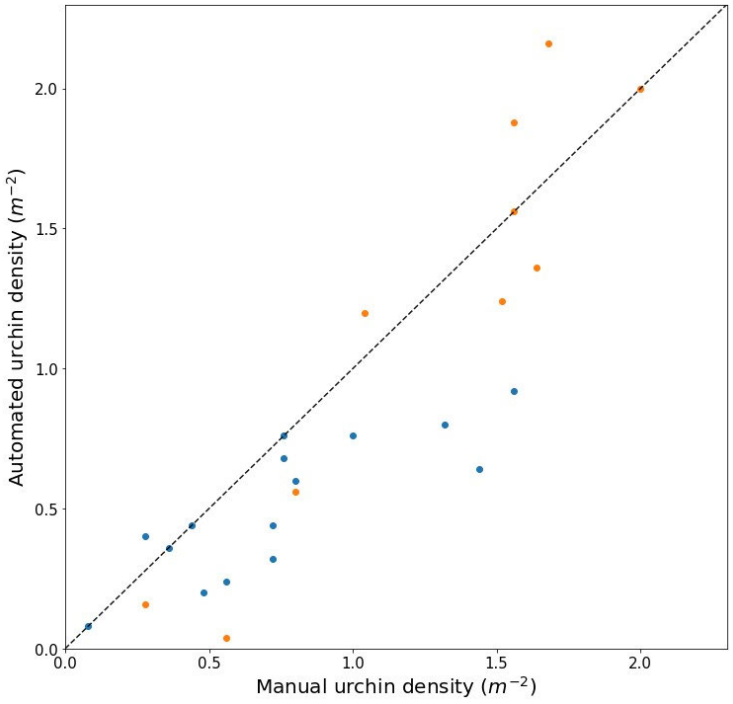


Figure 4.4 – Validation of automated estimates of urchin density by manual counts and automated detections in the 25 $5m \times 5m$ virtual quadrats at two sites in north-eastern Tasmania, St Helens Island (**orange**) and Elephant Rock (**blue**) shown in Figure 4.3. Overall the estimates made using the automated detections were lower than those made from manual counts from the imagery (paired T-Test p=0.009).

4.5 Observing Temporal Variation in Urchin Abundances

In this section, the automated spatial census approach for urchins developed above is used to estimate their abundance in AUV survey imagery of urchin barrens in eastern Tasmania. Estimates of urchin density made using these automated detections are compared with manual density observations from the imagery made by Ling *et al.* [126] from the same imagery to determine whether the automated pipeline developed in this thesis can be used to detect the same trends in urchin density as existing methods.

4.5.1 Dataset

Ling *et al.* [126] estimated urchin density using photomosaics generated using imagery collected by the AUV *Sirius*. Surveys were conducted over three years (2009, 2011 and 2013) at two sites on Tasmania's east coast, St Helens Island and Elephant Rock. At each site, the AUV surveyed 30x30m plots of reef. The authors estimated the density of *C. rodgersii* by manually counting urchins lying within virtual 2x25m quadrats on each photomosaic (Figure 4.5). Additionally, long-term *in situ* census of urchins by divers within the surveyed area were completed. While the overall density of urchin estimated by manual analysis of the imagery was typically lower than the *in situ* estimates by divers, the authors concluded that both methods were able to detect the same trends in densities. These showed that urchin densities remained relatively stable at the two sites over the survey time period.

4.5.2 Automated Detection of Urchins

The trained RetinaNet based urchin detection model described in section 4.4.1 was used to estimate the positions of urchins in the photomosaics from these surveys. Variations in lighting conditions and image quality affects the reef mosaic generated

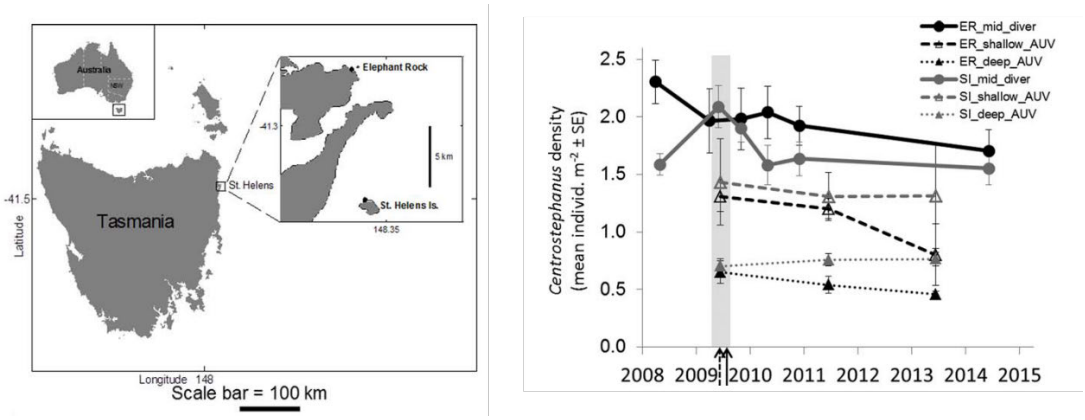


Figure 4.5 – From [126] Urchin barren sites St Helens Island and Elephant Rock surveyed by the AUV *Sirius* (**left**) and estimates of urchin density from counts by divers (**solid lines**), and from the AUV generated photomosaic (**dashed/dotted lines**) (**right**). Counts made from the AUV tended to estimate lower urchin density, however both methods observe the same stable trend over time. Automated estimates of urchin density are made on the shallow sites (**dashed lines**) and compared to these estimates in Section 4.5.2.

during different dives. This appears to affect the detection of urchins when using the model trained on only the one training photomosaic. To address this, a small number (between 30 and 50) individual urchins was randomly selected in each survey. These individuals were annotated as described in Section 4.4.1 and re-trained the initial model using these site specific annotations. The manual annotation and retraining the network of this small for each dive takes little time (typically less than 10 minutes in total), but the fine-tuned models outperform the pre-trained one for all the surveys using their mean average precision (mAP[1]) (see Table 4.1).

The automated detector performed particularly poorly on the survey imagery from St Helens Island in 2011, even after retraining. The imagery from that survey did appear degraded compared to the other surveys, with several holes in the photomosaic and

[1]mAP is a commonly used metric for assessing the performance of automated object detection and instance segmentation models [151]. It assesses the performance on a scale from 0-1. The area under the precision-recall curve is calculated for each class and the average across classes is the mAP. The prcision-recall curve requires identifying whether each prediction is a true positive, false positive, true negative, or false negative. To do thise, the intersection over union (IoU) between the predictions and ground truch are calculated, and a threshold value is used to determine whether or not a prediction is correct (here I have used a threshold of 0.5 (*i.e.* if a predictions overlaps a true instance by over 50% it is determined to be a true positive).

artefacts from the stitching process present (Figure 4.6). This suggests that image and reconstruction quality has an impact on the performance of the automated detection process.

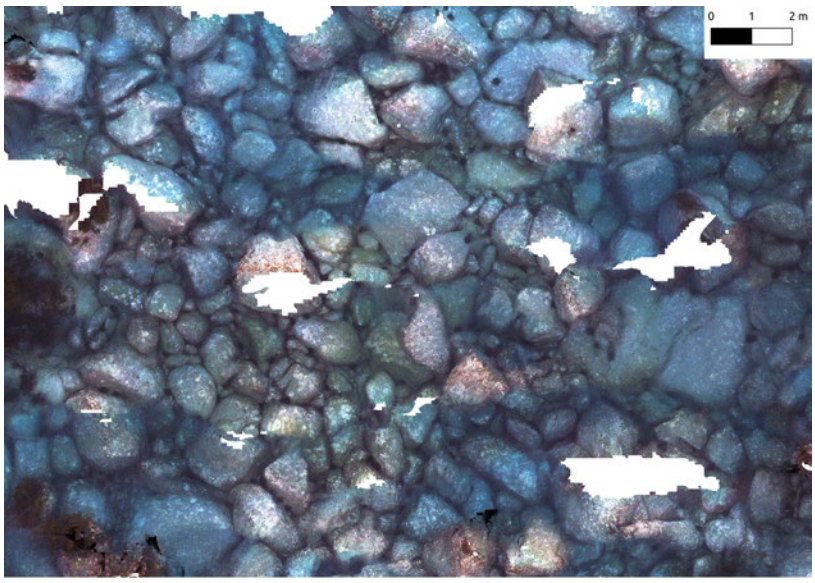


Figure 4.6 – Portion of a photomosiac of an urchin barren at St Helens Island in 2011. This mosaic shows poor quality compared to others collected using the same process (*e.g.* Figure 4.3). This may explain the poor performance of the automated urchin detector (see Table 4.1), which resulted in significantly lower estimates of urchin density compared to manual image analysis at this site (Figure 4.8).

4.5.3 Estimates of Urchin Density

To validate our automated pipeline against the manually generated estimates, the densities of urchins were calculated using counts of detected urchins within six 2x25m quadrats randomly placed within the photomosaic from each site (Fig 4.7). The average density per quadrat and the variance were calculated for each site.

The automated pipeline tended to underestimate density compared to the manual counts (Figure 4.8). This suggests that there are individuals missed using this method, and mirrors the trend between density estimates made *in situ* and from the imagery manually. The poor performance of the detector on the survey imagery from St Helens

Table 4.1 – Performance of a RetinaNet urchin detection network on reef-scale photomosaics generated at two sites in southeastern Tasmania over three years. At each site the mean average precision (mAP, see footnote 1) is shown for a generic model trained on a separate mosaic containing 387 individuals, and for the same model fine-tuned using a small number of manually annotated individuals (between 30 and 50) from that survey. Fine tuning the original model improved performance on the new data in all surveys. The 2011 survey at St Helens Island performed poorly with both models (shown in red), and this is reflected in a low estimate in urchin abundance (see Figure 4.8)

Site	**Year**	**Original Model mAP**	**Fine-tuned Model mAP**
Elephant Rock	2009	0.7003	0.8057
Elephant Rock	2011	0.6981	0.7632
Elephant Rock	2013	0.3875	0.6127
St Helens Island	2009	0.3721	0.5515
St Helens Island	2011	0.0000	0.0767
St Helens Island	2013	0.4004	0.5401

in 2011 appears to be reflected in a low estimate of urchin density for that survey. However, apart from this outlier, the use of automated detection did appear to capture a stable trend in urchin densities over the three sampling years as reported by Ling *et al.*. To test this hypothesis that there was no observed trend in the urchin density over the three years of surveys a two-way fixed effects analysis of variance (ANOVA; n=5 replicate quadrats per plot) was used with factors of *Site* (Elephant Rocks and St Helens Island) and *Year* (2009, 2011, 2013), as an analogue to the analysis performed in the study. The ANOVA showed no significant correlation between density and year confirming the automated approach observed the same stable observations of urchin density as the *in situ* and manual image based observations.

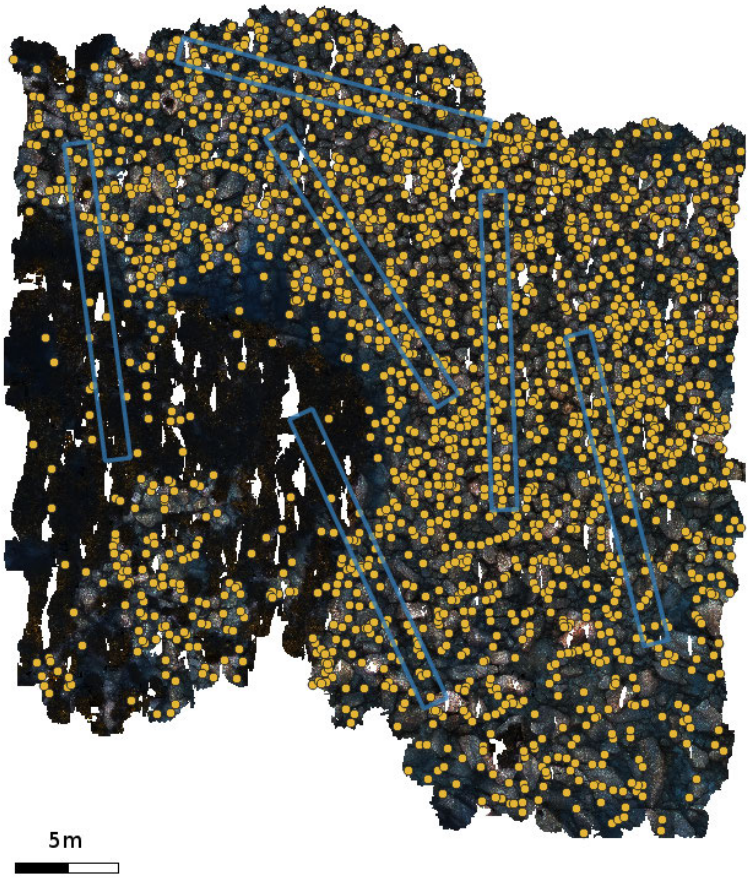


Figure 4.7 – AUV generated photomosaic at Elephant Rock, Tasmania in 2013 showing automated detections of *C. rodgersii* (**yellow dots**) and random virtual 2x25m quadrats (**blue**) used to estimate the urchins' density at this survey site.

4.6 Observing Urchin Density from Unprocessed AUV Images

So far in this chapter, detections of urchins made in photomosaics generated using AUV collected images have been used to estimate their density. As discussed, this is possible since urchins are slow moving so individuals are unlikely to move in between repeat views by the imager. This means that urchins detected in the photomosaic likely represent unique individuals.

Section 3.2.2 discussed how the movement of targets during survey image collection

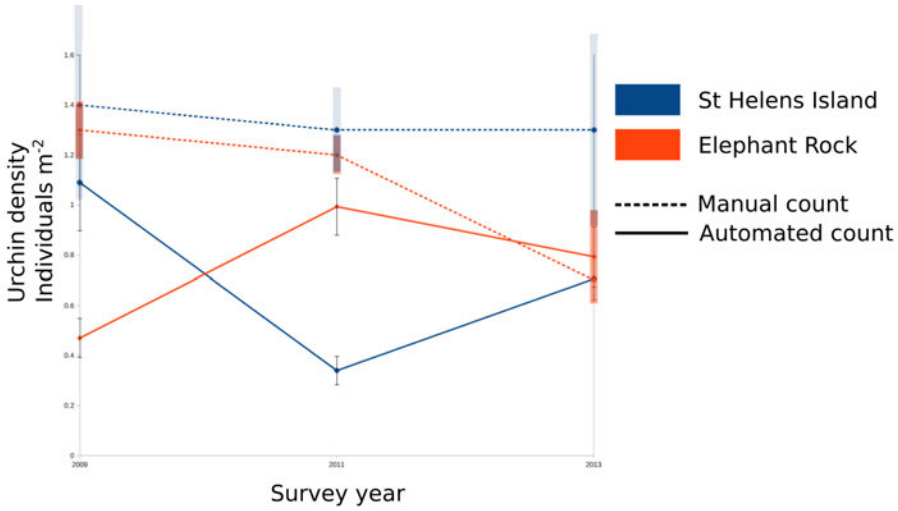


Figure 4.8 – Mean densities of long-spined sea urchins (*C. rodgersii*) at two study sites over three sampling efforts (2009, 2011, 2013) calculated using our automated process **(solid lines)** and manual counts from the AUV photomosaic **(dashed lines)** taken from [126]. Error bars for the automated estimates are ±1 standard error. Note, values and error bars for the manual analyses have been transcribed from the paper, see Fig 4.5 for the original plot.

affects their appearance in the generated composite images. Therefore, the same approach is unlikely to reliably estimate target densities if we are instead observing more mobile benthic targets. For these targets, detections need to be made in the unprocessed images (as opposed to in the composite photomosaic) then reprojected onto the composite images.

Before this can be done, we need to verify whether individuals detected in images can be accurately reprojected onto the mosaic, and therefore whether these detections can be used to estimate the abundance of targets. We can estimate the reprojection error by projecting repeat observations of an individual back onto the composite image and measuring the variation in the points. Large reprojection errors would affect the precision of estimates of the position of the detected targets. If these are systematic (*e.g.* in response to a particular habitat structure) they could be incorporated into the model of animal locomotion used in the following chapter. Random reprojection errors would add to the uncertainty in estimating the target position which would

make it more difficult to determine whether a given detection is a new or repeat individual.

Figure 4.9 shows the reprojected points onto two individual urchins. Each individual was observed in eight images. All the points land on the urchin in question with the individual at the top of frame having a reprojection error (*i.e.* the mean distance from each reprojected point to the centroid of points) of 7.8mm and the individual at the bottom of frame had a reprojection error of 38mm. Given that urchins tend to be approximately 150-300mm in diameter [36], reprojected points derived from the centre of automated detection bounding boxes are likely to fall within their body on the composite image, therefore simplifying localisation and aggregation.

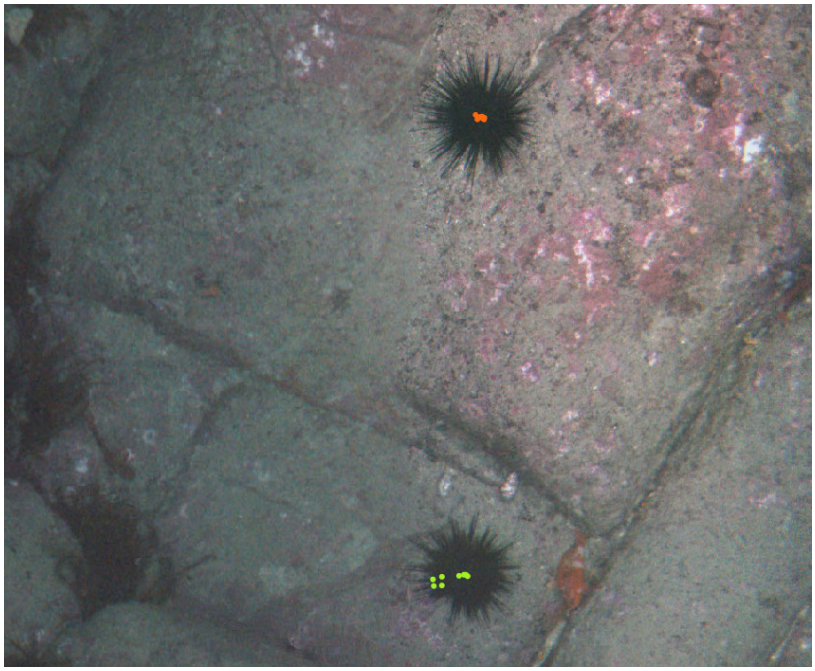


Figure 4.9 – Reprojected points from repeat observations of two individual Long-Spined Sea Urchins (*C. rodgersii*). Each individual was observed in 8 separate views. Reprojection errors for individuals at the top and bottom of frame were 7.8mm and 38mm respectively.

To verify this, we will observe urchins in the unprocessed images, reproject their positions onto the photomosaic and verify that these reprojected locations also contain an urchin. Figure 4.10 shows a region of a photomosaic containing urchins. The

urchins detected in the unprocessed images on the right of the figure have been projected onto the mosaic at the left. Additionally, the locations are projected back onto each image viewing the point to verify that the reprojection process is valid across all views of the scene. This allowed the repeat observations of the same individual to be clustered. The number of unique individuals could then be estimated from the unprocessed images. These individual's detections can then be reprojected onto the photomosaic. If the reprojection process is accurate, it would be expected that the reprojected points on the mosaic contain an individual urchin. This is the case for the points shown in Figure 4.10. This means that density estimates from photomosaic will match those made from the image detections.

Figure 4.10 – Urchin observations in photomosaic (**left**) and unprocessed AUV images (**right**). The original tagged individuals are marked with green flags, and the reprojections of those tags are shown in blue. The images were processed sequentially, so urchins detected in the first images are reprojected onto the future images. All urchins tagged in an image were correctly reprojected onto the future images so repeat tags were not needed. All individuals tagged in the unprocessed images can be seen in the photomosaic suggesting that these detections can be used to estimate the density of individuals.

To further test the reprojection, five regions in the mosaic were selected (Figure 4.11). All images capturing the central point of the region were identified and estimates of

the urchin density were made from the images and the photomosaic. These are shown in Table 4.2. All the individuals observed in images were accurately reprojected onto an individual in the photomosaic, resulting in identical counts of urchins using the two methods at all five points. This demonstrates that accurate reprojections of individuals detected in unprocessed AUV images are possible, and these detections can be used to produce density estimates.

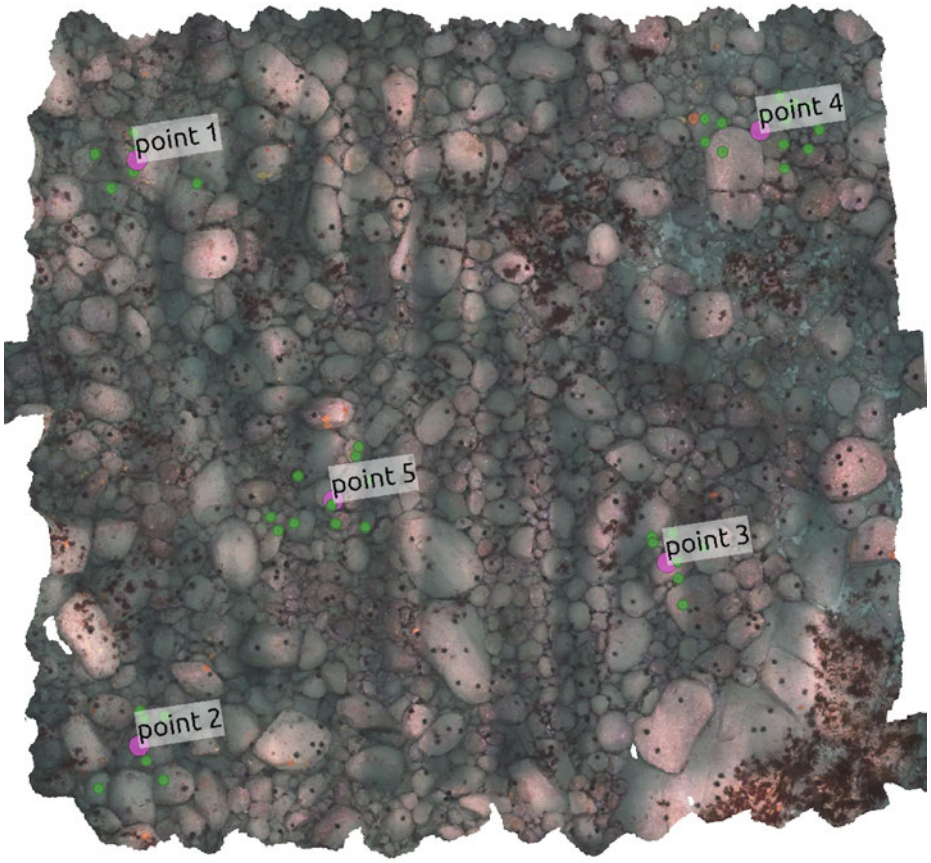


Figure 4.11 – Reprojected positions of urchins observed in unprocessed AUV images (**green**). Urchins observed in images capturing views of the five points (**magenta**) were reprojected onto the photomosaic at the points shown in **green**. Every individual detected in the unprocessed images was reprojected onto a unique urchin in the photomosaic, thereby validating the reprojection process.

Table 4.2 – Counts of urchins made in the underlying images and directly in the photomosaic. For each of the 5 points shown in Figure 4.11, the number of images capturing the plot are shown. Urchins in each image are tagged and reprojected onto the photomosaic and back onto the remaining views. The number of unique individuals are counted from the image detections, and these and each individual's position on the mosaic is observed to confirm that they are also present in the photomosaic. The number of individuals counted in the images and the photomosaic were the same for all five points. This demonstrates that detections made in images can be used to estimate the density of benthic targets.

Point	**# Views**	**# Urchins in Images**	**# Urchins in Mosaic**
1	7	7	7
2	9	6	6
3	8	7	7
4	11	14	14
5	12	11	11

4.7 Summary

This chapter described the approach for estimating the abundance and spatial distribution of stationary and slow moving benthic targets, using the urchin *Centrostephanus rodgersii* as an example. Automated detection allows for rapid analysis of photomosaic where high densities of targets are present and facilitate low cost observation of large seafloor extents. While measurements of urchin density using the automated detections tended to underestimate compared to manual counts, the pipeline was able to observe similar temporal trends in the abundance of *C. rodgersii* in urchin barrens as manual estimates from imagery from the literature.

In the next chapter we will explore how the approach can be modified in order to estimate the distribution and abundance of benthic fauna which do move during the course of the survey image collection. While in this chapter we replicated traditional surveys by estimating target abundance using virtual quadrats, the next chapter will also demonstrate how the unique ability to observe spatial distributions of targets can be exploited to better understand how the target organisms utilise the habitat.

Chapter 5

Estimating Abundance and Spatial Distribution of Mobile Benthic Fauna

5.1 Introduction

We have established that the spatial distribution of targets that do not move during the course of an SfM survey can be estimated by detecting individuals in the composite image. This chapter will shift the focus to mobile benthic targets where this is no longer the case.

AUVs present an attractive option for observing mobile benthic fauna. As discussed, they allow rapid benthic surveys without the restrictions on depth and time associated with SCUBA diving. They facilitate simultaneous observation of mobile fauna and their habitat, and can be used to localise individuals relative to that habitat. This is not possible using catch based methods which are traditionally used for studies of mobile benthic fauna.

As discussed in Chapter 3, movement of targets during a survey makes estimates of abundance and distribution made from the composite image unreliable. This means

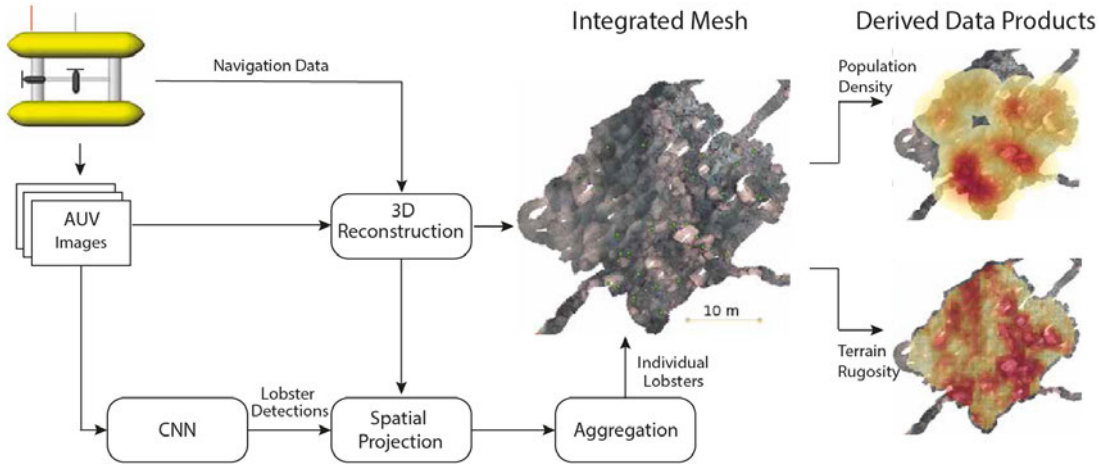


Figure 5.1 – Overview of the pipeline for automated observation of the spatial distributions of mobile benthic fauna. Images and navigation information are captured by the AUV. This information is used to generate detailed, texture-mapped 3D digital models of the environment. A CNN is trained to identify lobsters and individual sightings are integrated into the 3D map. The resulting integrated map can be used to estimate population density and relationships with underlying habitat characteristics (*e.g.* structural complexity).

that mobile targets need to be observed in the underlying source imagery to increase the likelihood that all observed individuals are considered. Figure 5.1 outlines the proposed pipeline. This involves making detections of individuals in the AUV images, projecting these detections onto the reconstruction, and aggregating repeat detections to estimate the density and distribution of individuals.

Analysing the unprocessed images results in a much greater volume of data compared to direct observation in the composite images. This make manual analysis infeasible, so Section 5.3 describes training and implementation of a automated detection pipeline to facilitate sustainable deployment of the proposed pipeline.

Since individuals are no longer being observed directly in the georeferenced photomosaic, the detections need to be localised by reprojecting them into 3D space using the reconstruction and the vehicle's navigation solution. The vehicle is equipped with stereo cameras which could be used to triangulate the 3D position of individuals appearing in both frames. Stereo approaches are often used for observing mobile fauna and can give highly precise and accurate estimates of the size and position of observed

targets [92, 93, 27], but require locating corresponding points in the two frames. To automate this process would require detection of the target in both frames, finding the corresponding detections (in the case of multiple objects being detected simultaneously), and precisely determining the pixel location of the same feature on the target in the two frames. The proposed approach in Section 5.4 only requires detection in a single image, and small deviations in the projected pixel (*e.g.* within a detection bounding box) are unlikely to result in large errors when estimating the 3D position of the target.

Localised observations of individuals can be used to generate a model of their fine-scale (*i.e.* within reef) habitat preferences which are prohibitively time-consuming to collect using traditional, *in situ* observational techniques. Such models can be used to identify specific habitats which provide refuge for the target fauna, allowing for targeted deployment of management (*e.g.* protected areas) with maximum benefit.

Since multiple overlapping views are required to perform an SfM survey, targets detected in each image do not necessarily represent a unique individual since a single individual is likely to be imaged on multiple occasions. Section 5.5 presents an approach for associating these repeat detections to individuals using an understanding of the species' locomotory behaviour and the vehicle's movement during the survey. This allows reporting the most likely populations given a set of localised detections and a model of the targets' movement. This approach is validated against simulated mobile targets, then applied to real detections of *J.edwardsii* individuals from imagery collected by the AUV *Sirius*.

5.2 Observing Southern Rock Lobster Using Benthic Imaging AUVs

The southern rock lobster (*Jasus edwardsii*) is an economically, culturally, and ecologically important organism supporting valuable commercial and recreational fisheries in southern Australia and New Zealand [163]. Understanding habitat preferences, and

estimating population sizes and characteristics of *J. edwardsii* is critical to ensure its fisheries are properly managed and sustainable into the future.

J. edwardsii is known to predate upon long-spined sea urchin (*Centrostephanus rodgersii*) in their expanded southern range [106, 163] and removal of lobster by fisheries has been shown to contribute to the formation of urchin barrens [125]. Therefore, management of *J. edwardsii* populations also forms an important aspect in the control of the invasive urchins, and has been proposed as a method for encouraging recovery of urchin barren regions in Tasmania [36, 107].

Rock lobsters are able to move along the seafloor propelled by their legs. Studies have recorded walking speeds of various species of lobster at $\mathcal{O}(10^{-3} - 10^{-2})\text{ms}^{-1}$ [80, 148, 137]. This movement is likely to result in distortions (see Figure 3.4), shadowing (see Figure 3.5), or absence of individuals in the complete composite mosaic generated from benthic image sequences. Therefore, better estimates of their abundance and distribution can be made by observing individuals in the unprocessed AUV images, rather than the photomosaic.

5.3 Automated Detection of Southern Rock Lobster in AUV imagery

The imagery used to demonstrate the pipeline for observing the spatial distribution and abundance of *J. edwardsii* at reef scales was collected by the AUV *Sirius* in the Huon Marine Park of the south-eastern coast of Tasmania. During this survey, benthic image pairs were collected and used to generate a 3D reconstruction and 2D top-down composite image of the seafloor. A portion of this survey composed of over 10,000 images covering approximately 30m x 30m of seafloor area will be used in this section. Manual detection of individual lobsters in this number of images would be extremely time consuming and would likely result in a processing bottleneck in routine surveys. This section describes an automated detection pipeline which allows for efficient analysis of the AUV imagery.

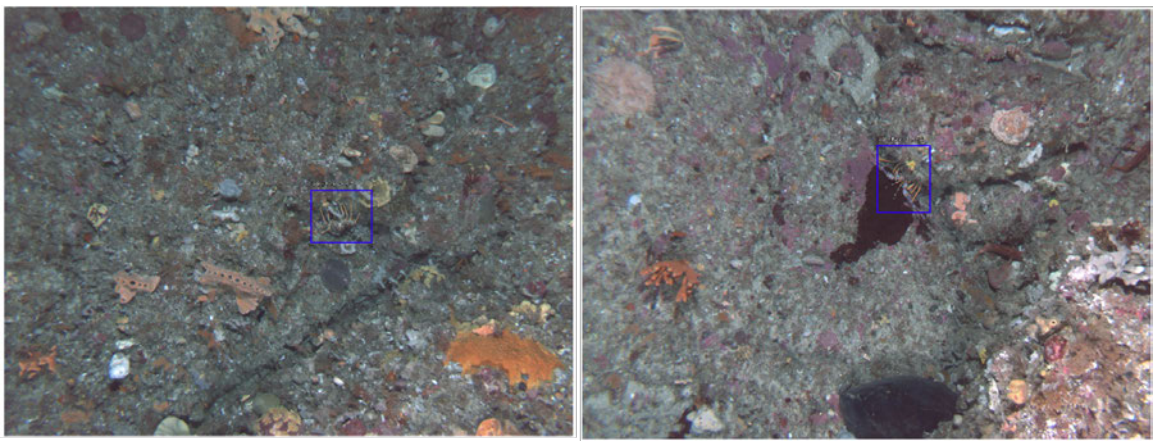


Figure 5.2 – Automated detections of southern rock lobster (*Jasus edwardsii*) individuals in two images collected by the AUV *Sirius* in the Huon Marine Park, Tasmania. Lobsters are cryptic and cover only a small part of the image fields of view making manual detection difficult. However, the automatic detection algorithm performs well (Figure 5.4).

5.3.1 Training the Lobster Detector

As in Section 4.4, the RetinaNet object detection algorithm was used to automate detection of target individuals. However, in this case the model was trained and deployed on the unprocessed AUV images rather than the photomosaic. To train the network, bounding box annotations were manually made on a set of 186 instances of *J. edwardsii* observed in randomly selected images collected by the AUV *Sirius* during a survey of the Huon Marine Park, south of Tasmania in 2013.

Unlike the urchins, lobsters are observed in a variety of orientations and do not present a uniform cross section in images (Figure 5.2). Therefore, the one-click annotation system using a uniform box size (Section 4.4.1) could not be used. Instead the script was modified to allow boxes to fit each individual target to be manually annotated. Where only a portion of the individual was visible - often the antennae and/or the head emerging from an overhanging reef section - the box was drawn around the visible parts.

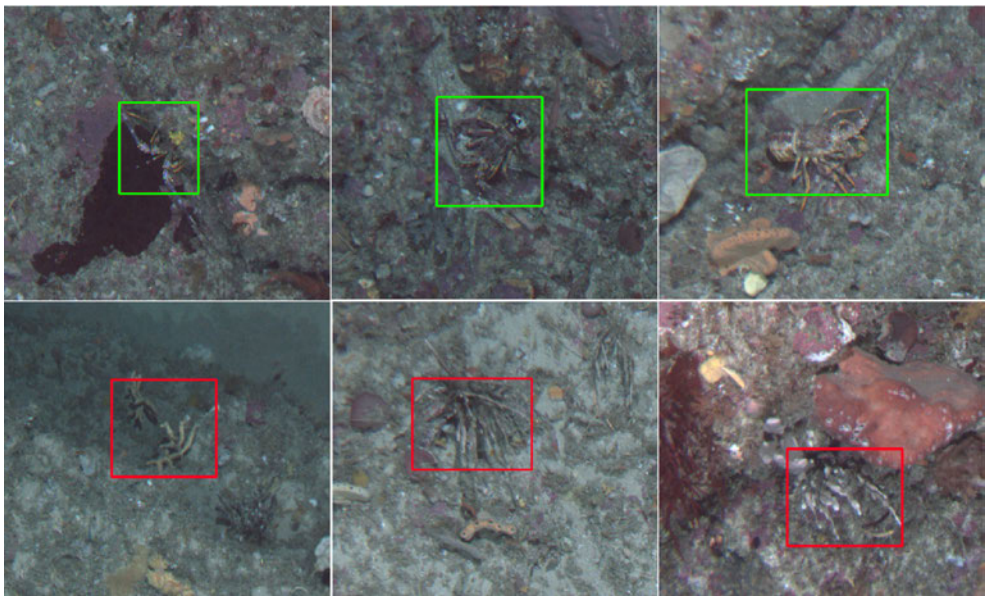


Figure 5.3 – Examples of correct **(green)** and false positive **(red)** automated detections of lobsters in images captured by the AUV *Sirius* in southeastern Tasmania. The successful detections had classification scores of 0.93, 0.99, 0.98 from left to right, and the false positives had scores of 0.80, 0.76, and 0.94. The algorithm is able to capture individuals even when only their antennae are visible (see top left) which is often how lobsters are observed on reefs. Note that the images are enlarged and detections centred to show detail, the full images containing detections are shown in Figure 5.2.

5.3.2 Evaluating Performance

To evaluate the performance of the trained lobster detector, a sample of 435 instances in which the model gave a positive prediction with a confidence of 50% or greater was analysed. Each detection was verified to determine whether or not a lobster was present in the proposed bounding box. From these detections the receiver operating characteristic (ROC) was generated (Figure 5.4). The ROC curve plots the false positive rate (FPR) against the true positive rate (TPR) for varying detection thresholds.

A "perfect" model has a fpr of zero and a tpr of 1. The detection threshold used was 0.82, and was chosen by finding the point on the ROC curve which minimises the geometric distance to (FPR = 0, TPR = 1). The model appeared to separate

true and false positives well, with a high area under curve of 0.92, meaning that a randomly selected true positive detection is 92% more likely to have a higher detection score than a randomly selected false positive. Lobsters were detected in successive overlapping frames with high classification scores. False positives occurred on certain types of corals and sponges (see Figure 5.3), however they tended to give higher classification scores sporadically and did not appear to be consistently detected in successive frames. Therefore, detections can be made more robust by considering the sequence of images and only using positive detections that occur in multiple sequential frames.

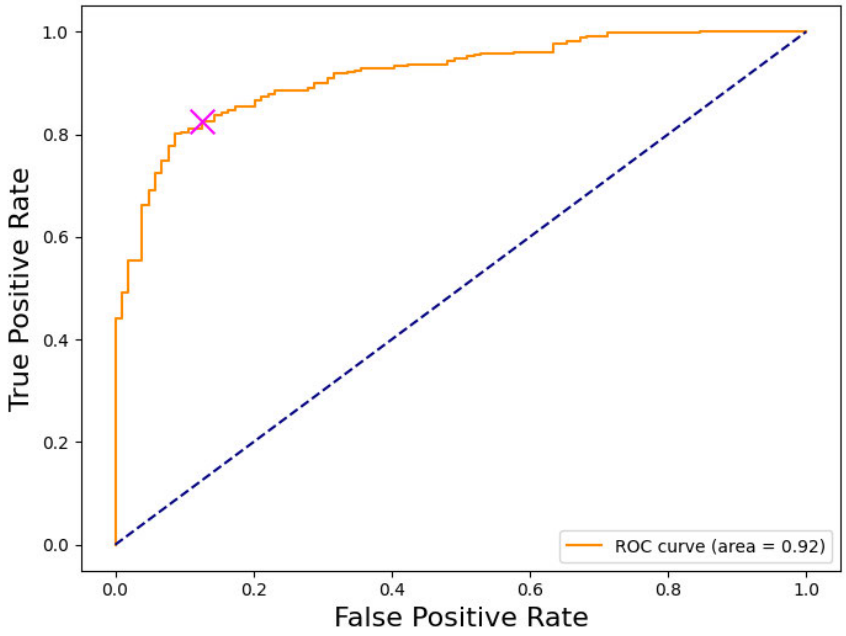


Figure 5.4 – Receiver operating characteristic (ROC) curve for automated detection of southern rock lobster using the RetinaNet object detection framework as tested on 435 individual automated detections. The ROC curve is used to determine the detection threshold used throughout this section. The point along the curve closest to (0,1) (**magenta**), gives the detection threshold used throughout this section.

We now have a model which can reliably automatically detect individual lobsters in AUV images. This allows for analysis of the large volume of images produced in a typical survey identifying the location within frames where individual lobsters

appear. The next step in performing a spatial census of these animals is to localise the detections within the surveyed reef.

5.4 Localisation of Detected Individuals

Recall that in Section 3.4, it was shown that the intersection point between a ray passing through a benthic target's pixel location in an image, and the seafloor, gives the target's approximate 3D position (see Figure 3.6). Therefore, by finding the intersection point between rays passing through detected lobsters in the raw images and the 3D mesh generated from those images, we can estimate the position of detected individuals.

Many SfM software packages can perform this projection from the source images onto the 3D reconstruction. The general geometric solution is also presented in Section 3.4.1. This allows localisation of detected individuals using only the image in which they were detected, the camera pose at that frame, the 3D mesh, and the camera calibration parameters.

This process allows individuals detected in the underlying images to be projected onto the final 3D surface and 2D projections of the seafloor. Figure 5.5 shows an example of the automated detection of lobsters projected onto the mosaic. Unlike for the urchins where the reprojection errors were relatively small, the projected points from a single lobster appear to show some variation between views. This is in part due to the movement of the lobster, and variations in the shape and exact location of the detection bounding box in different cameras. It may also be caused by the higher structural complexity if this site compared to the urchin barren in the previous chapter. A more complex 3D reconstruction is more prone to errors in identifying the exact mesh triangle the target lies in.

The reprojected locations of the full set of automated detections are shown in Figure 5.6. Now that we have localised detections for individuals observed in the images, the question then becomes one of determining how many individuals are actually

present in this area of seafloor.

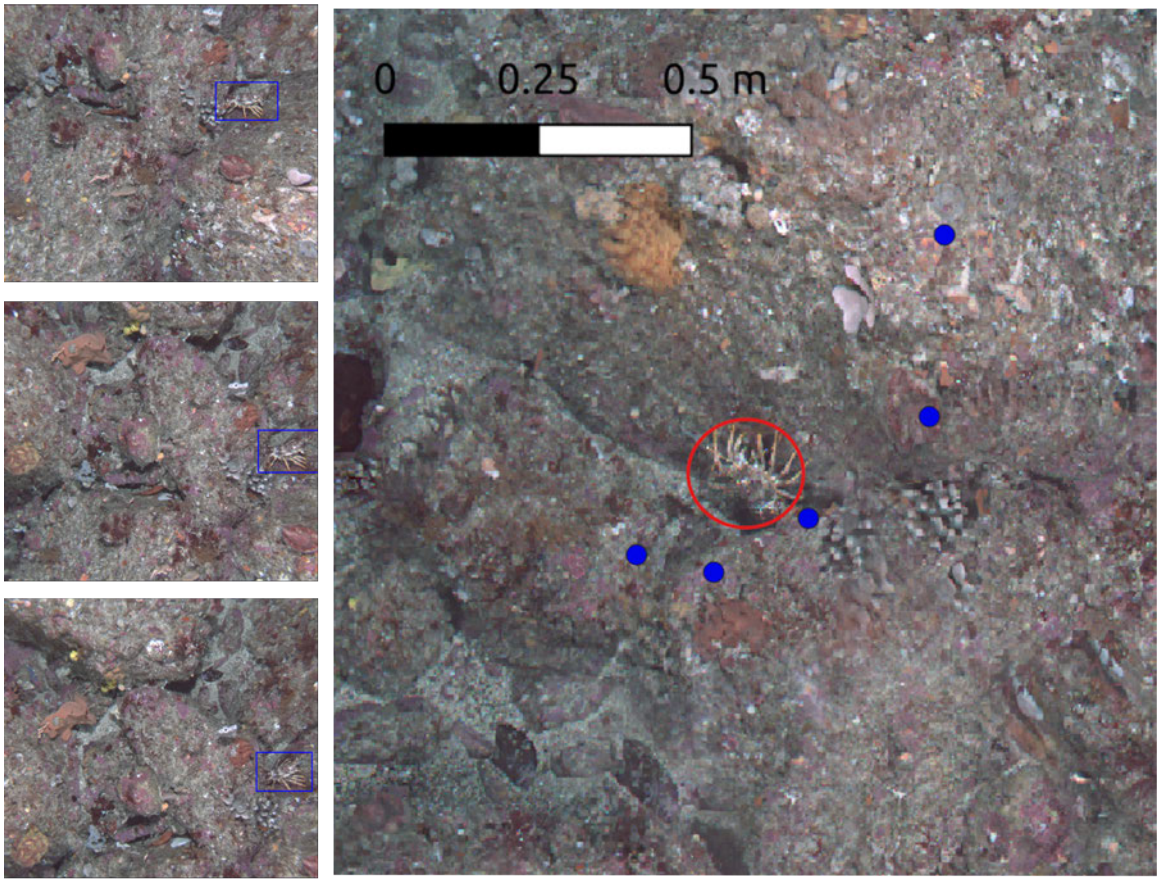


Figure 5.5 – Distribution of the reprojections of repeated detections of the same individual southern rock lobster in a AUV captured image sequence. The individual lobster was detected by the automated algorithm in five images (three of which are shown on **left**). Each detection was reprojected onto the mosaic (**right**) by finding the intersection point of a ray passing through the centre of the detection box and the reconstructed seafloor surface (Section 3.4.1).

5.5 Aggregating Localised Detections for Spatial Census of Mobile Targets

One of the primary aims of this thesis is to develop a method for estimating abundance of benthic targets - including mobile benthic fauna - using reef-scale composite

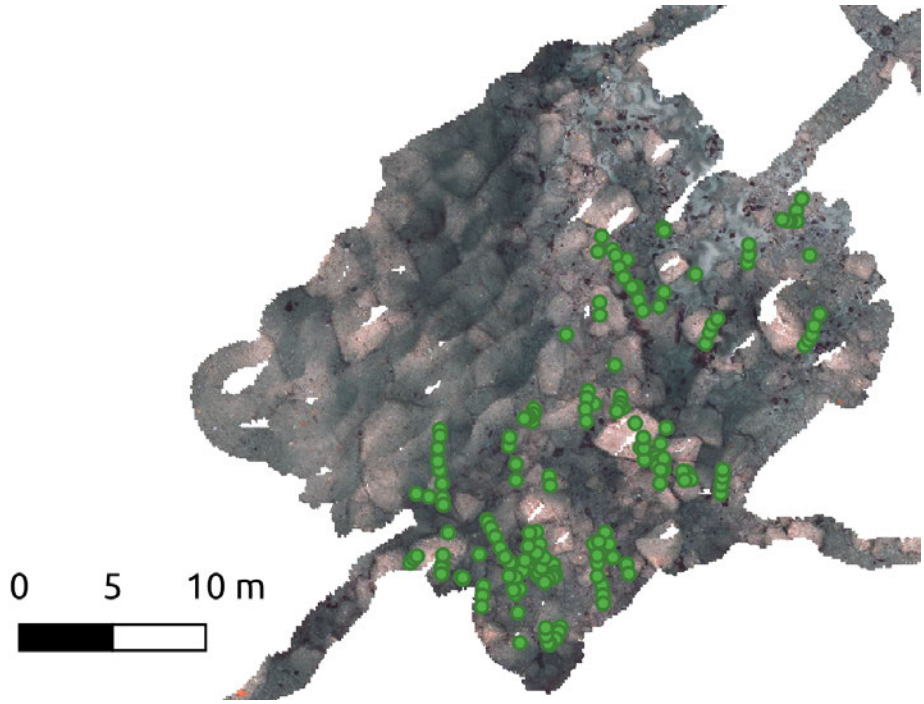


Figure 5.6 – Photomosaic generated from images collected by the AUV *Sirius* in the Huon Marine Park, Tasmania. 171 localised detections of *Jasus edwardsii* are shown in green. The detections were made in the unprocessed images and reprojected onto the photomosaic so likely represent repeat observations of individual lobsters.

imaging techniques. The methods in Section 5.3 facilitate automated detection of mobile benthic targets and localisation of their positions in 3D. However, since these detections are made in the unprocessed AUV images which capture overlapping sections of seafloor, each individual lobster may be observed in multiple frames. Using these detections to estimate abundance is therefore likely to result in an overestimate of the actual abundance. An approach for associating detections to individuals is required to generate accurate estimates of the abundance of targets in the surveyed area.

Section 3.5 discussed how approaches to this problem can range from simple clustering of detections, to more involved approaches which apply information about the detections and the behaviour of the target species to associate detections to specific individuals. In this section we will apply two approaches at either ends of this spec-

trum. The first is a simple distance based clustering approach, and the second is a modified tracking approach.

5.5.1 Distance based aggregation

A simple approach is to group adjacent detections using a clustering approach. The density based spatial clustering of applications with noise (DBSCAN) algorithm [60] provides an efficient way of performing this operation. Unlike many clustering algorithms (*e.g.* K-means clustering [128]), DBSCAN does not require the number of clusters to be determined, which is suitable for our application since we are trying to estimate the number of individuals in the survey area.

Figure 5.7 shows projected automated detections of lobsters and proposed locations of individuals estimated by clustering the detections using DBSCAN.

While this clustering approach is simple to implement and does produce estimates of the abundance and distributions of organisms from the automated detections, it does not consider the timing of the detections, the survey parameters, and the likely rate of movement of detected individuals which can all be used to exclude impossible or unlikely associations between detections to individuals and therefore produce a better estimate of the actual distribution and abundance of targets.

The sequence and timing of detections can help determine their most likely source, *e.g.* two adjacent detections may be more likely to have come from a single target if they were made closer in time before the target had the opportunity to move away from the initial location and other targets to move into it. Applying information about the sequence of detections and an understanding of how targets may move, is likely to provide better estimates of the distribution of individuals from which a set of detections is made. The remainder of this section will develop a modified tracking approach which utilises this information for estimating the abundance of mobile benthic targets.

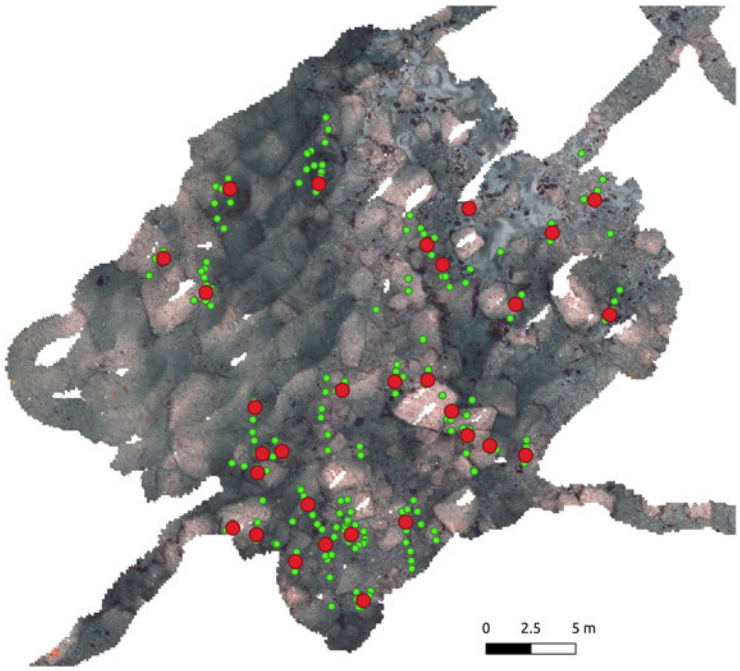


Figure 5.7 – Positions of reprojected detections of *J. edwardsii*, made in the unprocessed AUV images used to generate this reef-scale photomosaics (**green**). The DBSCAN clustering algorithm was used to filter repeat detections, and the proposed locations of individual lobsters are also shown (**red**)

5.5.2 Using survey characteristics and target species behaviour for detection association

Tracking multiple moving objects from observations can be framed as a data association problem. Mobile robots track features they observe from (often noisy) sensor data, and use these detections to estimate their own path [75, 216]. Data association is often applied in this context to aid in navigation and mapping by field robots. In our application, as we make new observations we'd like to determine whether an observed target is a repeat detection of a previously seen individual and should be assigned that existing track, or if they are a new individual and should be assigned to a new track.

Multiple hypothesis tracking (MHT) is an established technique for predicting the as-

sociations of observations to moving individuals. It is traditionally applied to acoustic or laser scan data [23] however it is equally applicable to tracking individuals from detections in images. In typical applications, associations between observations and target tracks are scored by matching the modelled motion of the object and the position of the observation, as well as by comparing the appearance of the target.

Here we will modify the MHT approach presented by Blackman and Popoli [22] and Kim *et al.* [110] to apply to the case of detections of mobile targets made in benthic images in order to generate and score hypotheses for the targets' distribution and abundance. This approach will be validated against simulated data, and applied to the set of localised detections of *J. edwardsii* made earlier in this chapter.

As mentioned, it is often difficult to distinguish between individuals of a given species. So the approach proposed here will rely on a model of the targets' movement to associate observations and to form hypotheses of their distribution and abundance. The MHT approach demonstrated by Blackman and Popoli [22], and Kim *et al.* [110] will be modified to accept localised detections of mobile fauna as inputs and generate ranked hypotheses for the abundance and distribution of the observed targets using a model of the animals locomotory behaviour.

Given a set of localised detections, the possible associations of detections to individuals can be generated. The likelihoods of these associations can then be used to exclude impossible associations (*i.e.* those that would require the target to move faster than it is capable of), and to rank the likelihood of global association hypotheses. These global hypotheses represent the full set of associations, and therefore describe the proposed motion tracks of each individual in the survey. By scoring the likelihood of the global hypotheses we are able to propose the most likely solution for the distribution and motion of targets within the survey area.

5.5.3 Building abundance hypotheses

The first step in the MHT process is creating the possible tracks from the set of detections. Say we have two detections, D_0 and D_1 made at timesteps t_0 and t_1

respectively. There are two possible scenarios, either the detections represent two distinct individuals, or D_1 is a repeat detection of the individual observed in D_0. This results in three potential tracks (Tr_0 - Tr_2):

$$Tr_0 : [D_0, -]$$
$$Tr_1 : [-, D_1]$$
$$Tr_2 : [D_0, D_1]$$

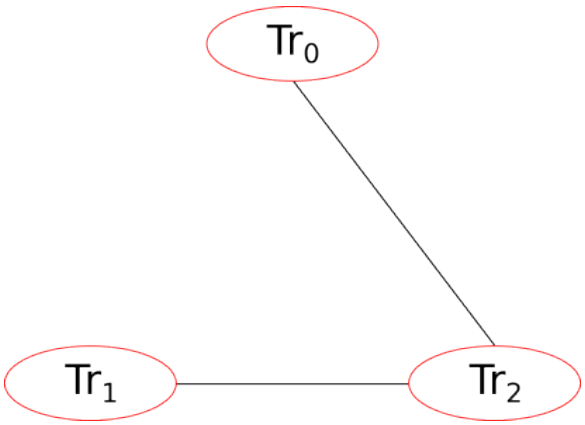


Figure 5.8 – Graph representation of the first three tracks generated in Section 5.5.3. Each track forms a node in the graph, and incompatible tracks, *i.e.* those which share a detection, are connected by edges. maximal independent set (MIS) of the graph nodes form the global hypotheses.

Tracks which share a detection are incompatible (*e.g.* Track 0 and Track 2 which both include detection D_0). By combining compatible tracks, global hypotheses can be generated. Here we are left with the two possible hypotheses, H_0 where there are two distinct individuals, and H_1 where the same individual has been detected twice. In this simple case the compatible tracks are obvious. However, when processing a large set of detections, the hypotheses can be generated using a graph representation (Figure 5.8). Set each track as a graph node, and create edges betweein tracks which share a detection (*i.e.* incompatible branches). Each abundance hypothesis we wish

to extract is a maximal set of nodes which do not share edges, *i.e.* they form a MIS of the generated vectors. Graph analysis software (*e.g.* igraph [51]) can be used to generate the network, and calculate the MISs to provide the hypotheses.

$$\begin{aligned} H_0 : Tr_0, Tr_1 &\quad : [D_0, -], [-, D_1] \\ H_1 : Tr_2 &\quad : [D_0, D_1] \end{aligned}$$

Repeating this exercise with a third detection D_2 made at the next timestep t_2 results in seven potential tracks:

$$\begin{aligned} Tr_0 &: [D_0, -, -] \\ Tr_1 &: [D_0, -, D_2] \\ Tr_2 &: [-, D_1, -] \\ Tr_3 &: [-, D_1, D_2] \\ Tr_4 &: [D_0, D_1, -] \\ Tr_5 &: [D_0, D_1, D_2] \\ Tr_6 &: [-, -, D_2] \end{aligned}$$

The graph representation of this set of tracks is shown in Figure 5.9. The compatible tracks result in the following five possible hypotheses.

$$\begin{aligned} H_0 : Tr_0, Tr_2, Tr_6 &\quad : [D_0, -, -], [-, D_1, -], [-, -, D_2] \\ H_1 : Tr_1, Tr_2 &\quad : [D_0, -, D_2], [-, D_1, -] \\ H_2 : Tr_0, Tr_3 &\quad : [D_0, -, -], [-, D_1, D_2] \end{aligned}$$

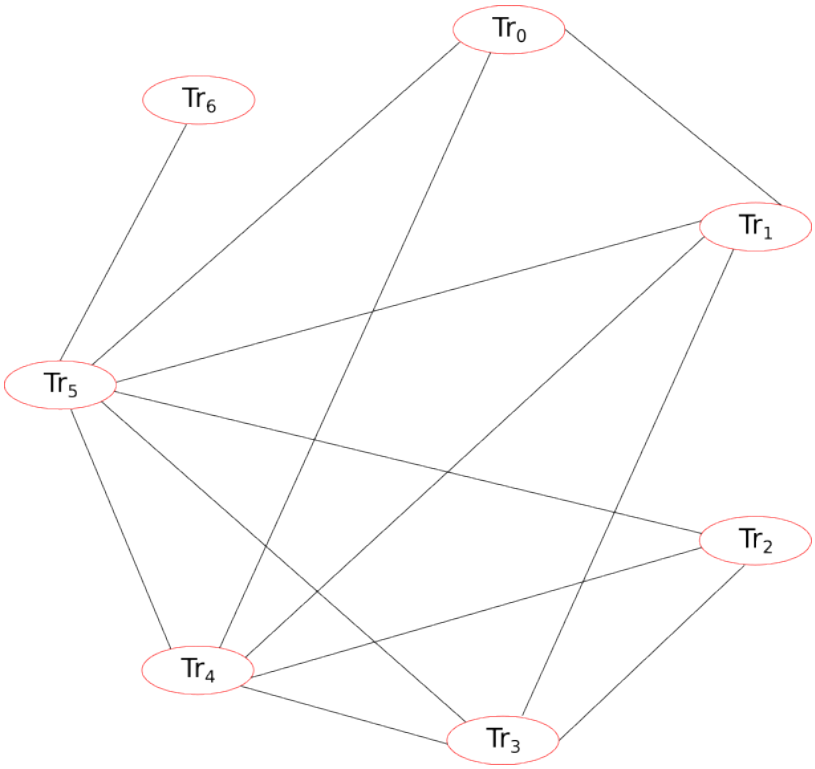


Figure 5.9 – Graph representation of the tracks generated at the next step after those shown in Figure 5.8. An additional detection was made increasing the number of potential tracks to seven, and increasing the complexity of the graph.

$$H_3 : Tr_4, Tr_6 \quad : [D_0, D_1, -], [-, -, D_2]$$
$$H_4 : Tr_5 \quad : [D_0, D_1, D_2]$$

The hypothesis space rapidly expands as the number of detections increases. A mechanism for scoring and ranking hypotheses would allow unlikely scenarios to be discarded and reduce the hypotheses space to a manageable level. Additionally, scored hypotheses could be used to determine the likeliest solution, and therefore the best estimate for the distribution of individuals in the survey.

5.5.4 Gating and scoring abundance hypotheses

In MHT, tracks are scored using the log likelihood ratio (LLR) between the hypotheses that a given detection came from that track, and that it came from the background (*i.e.* is a new individual).

Blackman and Popoli [22], and Kim *et al.* [110] demonstrated that the track score can be calculated incrementally for each timestep (t) [1]:

$$S(t) = S(t-1) + \Delta S(t)$$

The score increment $\Delta S(t)$ can be calculated as follows[2]:

$$\Delta S(t) = ln\frac{V}{2\pi} - \frac{1}{2}ln|\sum| - \frac{d^2}{2} \tag{5.1}$$

Where:

V : The observed area (*i.e.* camera footprint)

$\sum$: The covariance of the target's predicted position

d^2 : The Mahalanobis distance between the observed and expected position of the target

In order to calculate $\sum$ and d^2, we require a model that approximates the movement

[1] The original formulation [22] also includes an appearance score in the increment. However, this thesis implements only a motion score. The appearance score can be added and could use *e.g.* CNN features of the detected individuals to assist in scoring the associations between detections. However, as discussed (and can be seen in Figure 5.3), individual lobster are difficult to distinguish from one another, so doing so using learned features is also likely to be unsuccessful.

[2] If the track is being updated with a new detection (*e.g.* Track 2 in the scenario described in Section 5.5.3) this is the function to use. However, if a new track is being initialised (as in tracks 0 and 1), the score is based on the likelihood of encountering an individual. A standard approach is to assume a uniform likelihood across the survey area, however, it is also possible to use known habitat associations of the target species, *e.g.* with habitat complexity or habitat forming organisms to create a background likelihood map. However, we will continue with the standard approach here.

of the observed targets. These will allow an estimate of the target's position from a previous observation and allow gating of associations to exclude unlikely pairs (by setting a threshold value for d^2).

Modelling target movement model as a Gaussian Random Walk

Say we want to observe an organism which is known to have a maximum locomotion speed V_{max} and we make a detection of an individual at time t_0 and at a position X_0. It follows that at time t_1, the individuals position X_1 must be in the range:

$$X_0 - (t_1 - t_0) * v_{max} <= X_1 <= X_0 + (t_1 - t_0) * v_{max}$$

If a new detection falls outside this range, then it is not possible that that detection is of the same individual originally observed so that association need not be considered in the proposed solution. This process of eliminating associations is called gating.

We can model the movement of a target where individuals take a step of random length up to the maximum length ($(t_1 - t_0) * v_{max}$) at each timestep. This model of the targets' movement gives a uniform distribution of possible locations of the individual at a proceeding time step. This means that there is an equal likelihood of the individual being at any given point within the region bounded by the maximum step size from their original location. While this allows elimination of impossible associations, it does not provide a way of estimating the likelihood of a given association.

A more realistic model of the targets' movement may incorporate the assumption that an individual is more likely to take a smaller step than a larger one. This would reflect an organism that has an affinity to a particular home range which is true for many benthic species. One way to simulate this is to use a gaussian random walk, where the length of each step is randomly selected from a normal distribution centred on the target's original position. Similar approaches have been used to model the movement of a range of targets [13], *e.g.* insects [4, 49] and large herbivores [19]. This approach provides a way to score the associations of each detection, and therefore

to rank multiple competing hypotheses for the distribution of observed targets. This approach also allows for gating of associations using a confidence interval of the normal distribution.

In this model, the position X_1 of an individual at time t_1 is given by the following:

$$X_1 \sim \mathcal{N}(X_0,\, \sigma^2)$$

X_1 lies within a normal distribution centred on a mean $\mu = X_0$, its position at time t_0. The variance of the distribution, σ^2, is analogous to the typical speed of the organism. A "faster" target can travel further in a given time interval, so it follows that its possible position lies in a wider distribution.

Modelling the motion of individuals in this way provides a straightforward way of predicting the position of the individual after any number of time steps. Using the additive properties of normal distributions, the individual's possible location at t_2 forms a normal distribution with the same mean, $\mu = X_0$, and with twice the initial variance, *i.e.* $2\sigma^2$. More generally the distribution of the positions of an individual initially observed at X_0 at time step i is given by:

$$X_i \sim \mathcal{N}(X_0,\, i\sigma^2) \tag{5.2}$$

If we assume that the target of interest's movement can be approximated by a 2D plan and their position by (x, y) in some coordinate frame, then the model of the target motion allows us to calculate the covariance ($\sum$) and Mahalanobis distance (d^2):

$$|\sum| = \begin{vmatrix} i\sigma^2 & 0 \\ 0 & i\sigma^2 \end{vmatrix} = (i\sigma^2)^2 \tag{5.3}$$

$$d^2 = (X_0 - X_i)\Sigma^{-1}(X_0 - X_i)^T \tag{5.4}$$

Equation 5.1 can then be used to score each track, and therefore the generated hypotheses. As the detections are processed, tracks which diverge from the highest scoring hypotheses can be ignored, a process known as track pruning [110]. This allows the number of tracks - and therefore the number of hypotheses - to be controlled. This allows the hypothesis space to be maintained at a manageable level while reducing the likelihood that the correct track will be discarded.

5.5.5 Impact of map expansion rate and target mobility

Equation 5.1 demonstrates how the rate of map expansion, and the movement of the targets will affect our ability to reasonably estimate the abundance of targets from detections.

$$\Delta S(t) = ln\frac{V}{2\pi} - \frac{1}{2}ln|\sum| - \frac{d^2}{2}$$

We can see that the score increases as the image footprint area (V) increases, and decreases as covariance (Σ), and mahalanobis distance (d^2) increase. The variance (σ^2) is a proxy for the rate of movement of the target. Faster moving animals will have a larger σ^2 and therefore greater uncertainty in estimates of their position at future timesteps. Both Σ and d^2 will be larger for a larger σ^2 so the score - and therefore the certainty with which a detection can be associated to an individual - will be lower. A greater rate of map expansion will tend to reduce the time between detections (i in eqn. (5.1)), this will result in smaller values for Σ and result in higher scores for those detections.

This model captures the fact that faster moving fauna require a survey with a greater rate of map expansion in order to generate abundance estimates with acceptable levels of uncertainty.

5.6 Testing Aggregation Model Using Simulated Data

Validating abundance estimates made from localised detections in a reef-scale AUV survey would require either comprehensive *in situ* observations of the survey area to accurately estimate the abundance, or performing surveys in a controlled area where the abundance and distribution of targets is known. In the absence of these data, we will instead validate the MHT detection aggregation approach using simulated data. This will allow us to determine whether we can generate reasonable estimates of abundance, provided that our assumptions about the movement of the targets are sound.

Let us setup a simple scenario to test the ability of this process to accurately associate detections to individuals and estimate their abundance. Consider a 1m x 5m region of seafloor containing three individuals (I_{0-2}) of a mobile species (Figure 5.10a).

The movement of this species can be modelled using a gaussian random walk with a variance (σ^2) of 0.1m^2. This means that if an individual is observed at position X_0 at time t_0, then the probability density function of the target's position at time $t_0 + 1\text{s}$ is normally distributed with a mean position of X_0 and a variance of 0.1. By sampling this distribution the movement of the targets can be simulated (Figure 5.10c).

Let us observe these targets using images captured by a benthic imaging AUV. Assume that the vehicle images the seafloor every second with a 1m x 1m footprint (*i.e.* 1m^2) while traversing the region from left to right at a speed of $0.6\text{m}\,\text{s}^{-1}$. Targets falling inside the image footprint as the vehicle passes overhead are detected and localised (Figure 5.10d).

In this example, each individual is detected twice, giving six total detections (Table 5.1). Following the process outlined in Section 5.5.3, the possible tracks can be generated from this set of detections. The tracks can be scored using Equation 5.1, allowing gating of unlikely associations (*e.g.* between detections D_4 and D_1), and ranking of the proposed hypotheses. After gating in this scenario, we are left with

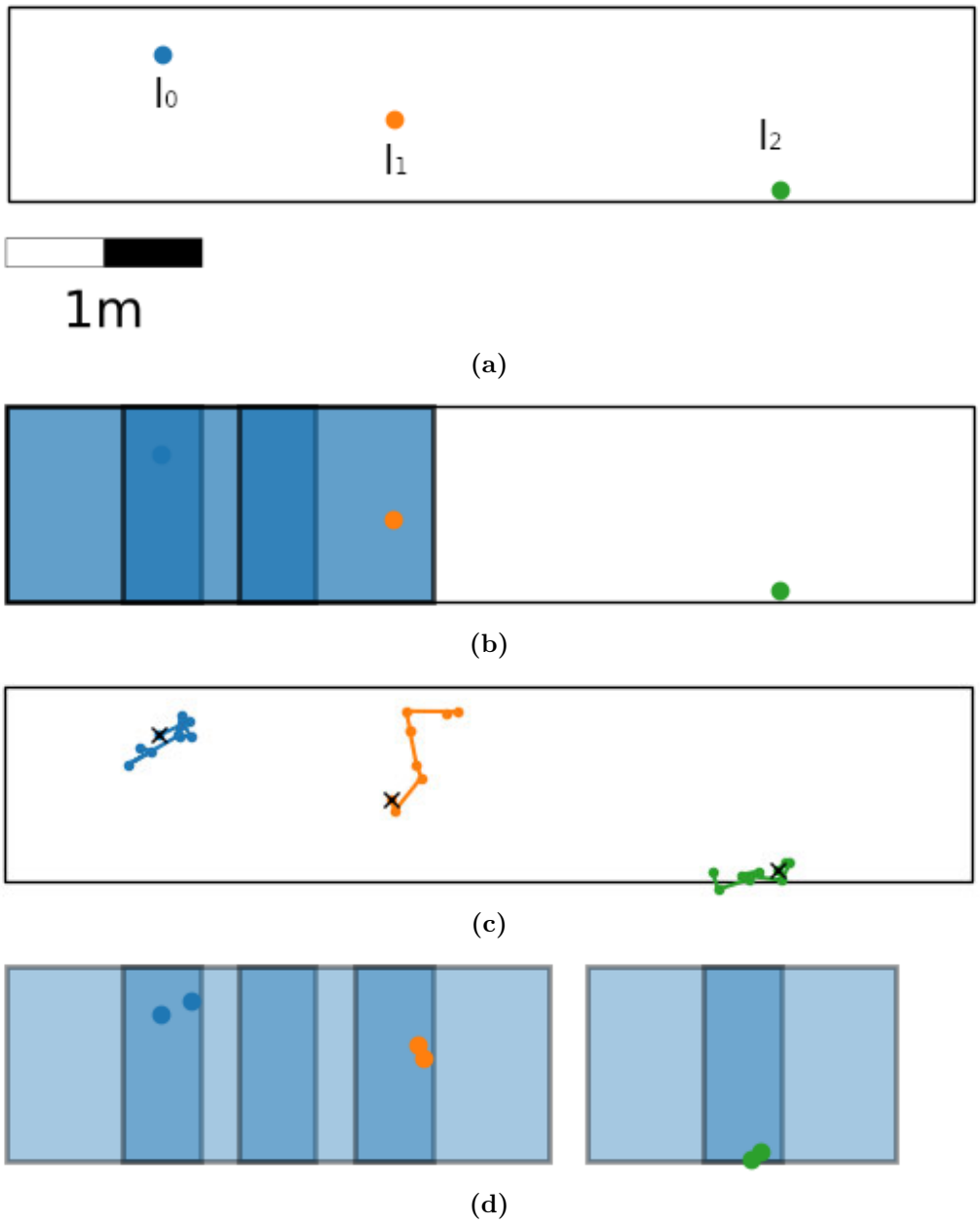


Figure 5.10 – Simulated observations of mobile benthic targets in a 1m x 5m seafloor region by a benthic imaging AUV. Three individuals are present in the survey area, (a) shows their positions at the beginning of the survey. The AUV traverses the region from left to right, the first three image footprints are shown in blue in (b). The movement of the three targets is generated using a gaussian random walk. The resulting paths are shown in (c), and the frames in which the individuals are detected are in (d).

Table 5.1 – Details of the detections of simulated mobile benthic fauna illustrated in Figure 5.10.

Detection id	Time (s)	Individual	Position (m)
D_0	0	I_0	0.800, 0.752
D_1	1	I_0	0.957, 0.823
D_2	2	I_1	2.155, 0.532
D_3	3	I_1	2.129, 0.602
D_4	5	I_2	3.901, 0.051
D_5	6	I_2	3.843, 0.006

13 tracks and 16 hypotheses. Figure 5.11 shows the number of predicted individuals and the hypothesis for each of the predicted scenarios.

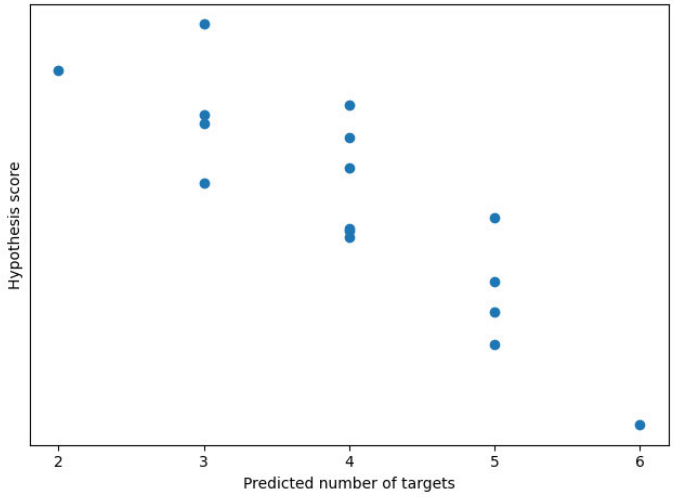


Figure 5.11 – Predicted number of targets and hypothesis scores for detected simulated mobile benthic targets shown in Figure 5.10. The simulated scenario contained three individuals, and the highest scoring hypothesis predicted this number correctly

.

The highest scoring hypothesis predicted the correct number of individuals (3), and predicted the correct associations between detections and individuals, *i.e.* detections D_0 and D_1 to individual I_0, D_2 and D_3 to individual I_1, and D_4 and D_5 to individual

I_2. The three tracks representing this are:

$$[D_0, D_1, -, -, -, -] : I_0$$
$$[-, -, D_2, D_3, -, -] : I_1$$
$$[-, -, -, -, D_4, D_5] : I_2$$

This simulation demonstrates the process of applying the modified MHT pipeline to detections of benthic targets in a simplified example. Next we will simulate a more realistic scenario and test the robustness of the model to incorrect estimates of the movement parameter, σ^2 and the background probability.

5.6.1 Robustness of MHT Model to Parameter Errors

In the previous example, the simulated AUV covered the survey region in a single pass. However, in reality multiple passes are typically required in order to adequately cover the region, and to ensure enough overlap between images to generate the 3D reconstruction.

Let's now consider a 5m x 5m seafloor area. As above, there are three individuals present and their movement can again be modelled using a gaussian random walk with $\sigma^2 = 0.1$. The region is imaged by an AUV which follows a back-and-forth path starting at the bottom left and ending at the top right. Figure 5.12 show an example of this simulation.

To test the robustness of the MHT model to poor estimates of the movement parameter for the targets σ^2 and the background probability we will generate a set of detections from the simulation over 50 runs. To evaluate the performance of the model, for each run we will determine how many incorrect hypotheses score higher than the correct one. We will also test the model's ability to estimate the correct

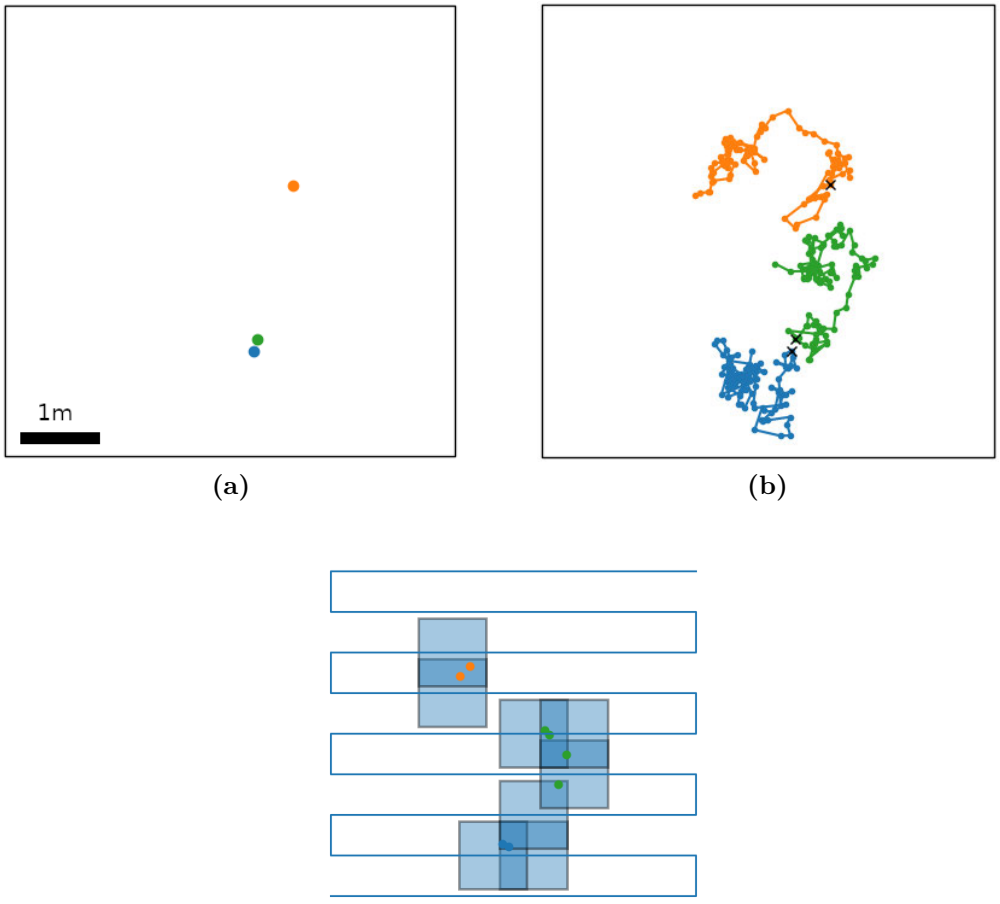


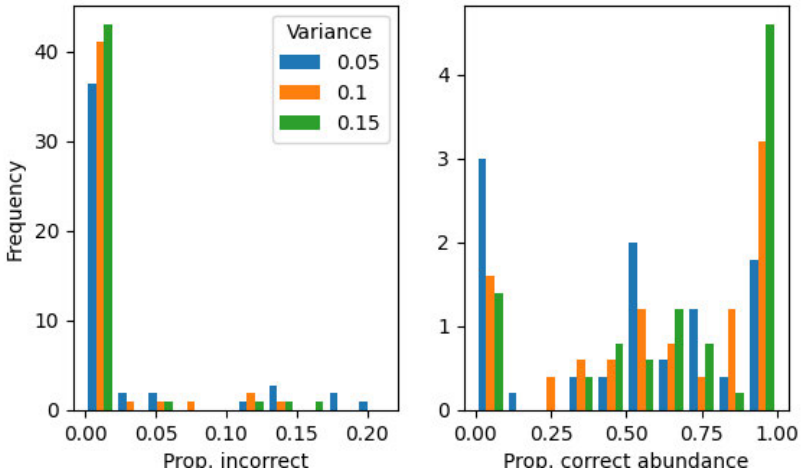
Figure 5.12 – Simulated observations of mobile benthic targets in a 5m x 5m seafloor region by a benthic imaging AUV. Three individuals are present in the survey area, (a) shows their positions at the start of the survey. The movement of the three targets is generated using a gaussian random walk. The resulting paths are shown in (b). The AUV traverses the region following a back and forth track starting at the bottom left and ending at the top right, and the frames in which the individuals are detected are shown in blue in (c).

abundance by determining how many of the 10% highest scoring hypotheses which predict an an abundance of 3 individuals (Figure 5.13). Overall the model appears to be relatively robust to incorrect values for the two parameters. Underestimating the variance appears to have an effect, particularly on the likelihood of correctly estimating the abundance. This is likely due to the model overestimating the abundance when smaller values of variance since detections are less likely to be associated if the target is slower moving.

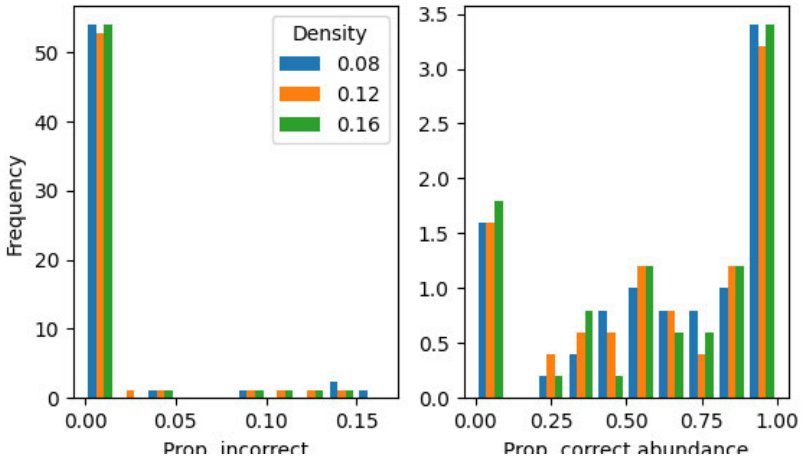
This simulation demonstrates that with accurate estimates of the movement parameters for the target individuals, detections can be correctly associated to individuals using an MHT based pipeline. It also demonstrated that reasonable abundance estimates can still be made even if using imperfect estimates of the model parameters. This means that we are able to generate estimates of abundance and distribution of targets using localised detections. Next we will apply this approach to the detections of *J. edwardsii* generated in the previous sections.

5.7 Estimating Abundance of *J. edwardsii* Using Localised Detections and MHT Aggregation Model

In this section we will use the detections of *J. edwardsii* generated from imagery collected by the AUV *Sirius* in the Huon Marine Park in southern Tasmania using the automated pipeline presented in Section 5.3. These detections will be passed to the MHT association pipeline in order to propose and score tracks, and generate ranked hypotheses for the distribution of individuals in the surveyed area. Figure 5.14 shows the process for applying the pipeline to detections of benthic targets.



(a)



(b)

Figure 5.13 – The effect of varying the variance σ^2 (a) and density (b) in the MHT pipeline. 50 simulations were run using a variance of 0.1 and density of 0.12 and hypotheses were generated using the MHT process using varying values for the two parameters. The number of incorrect, higher scoring hypotheses for each simulation run are shown on the left, and the proportion of the highest scoring hypotheses which predicted the correct abundance for the simulation are shown to the right.

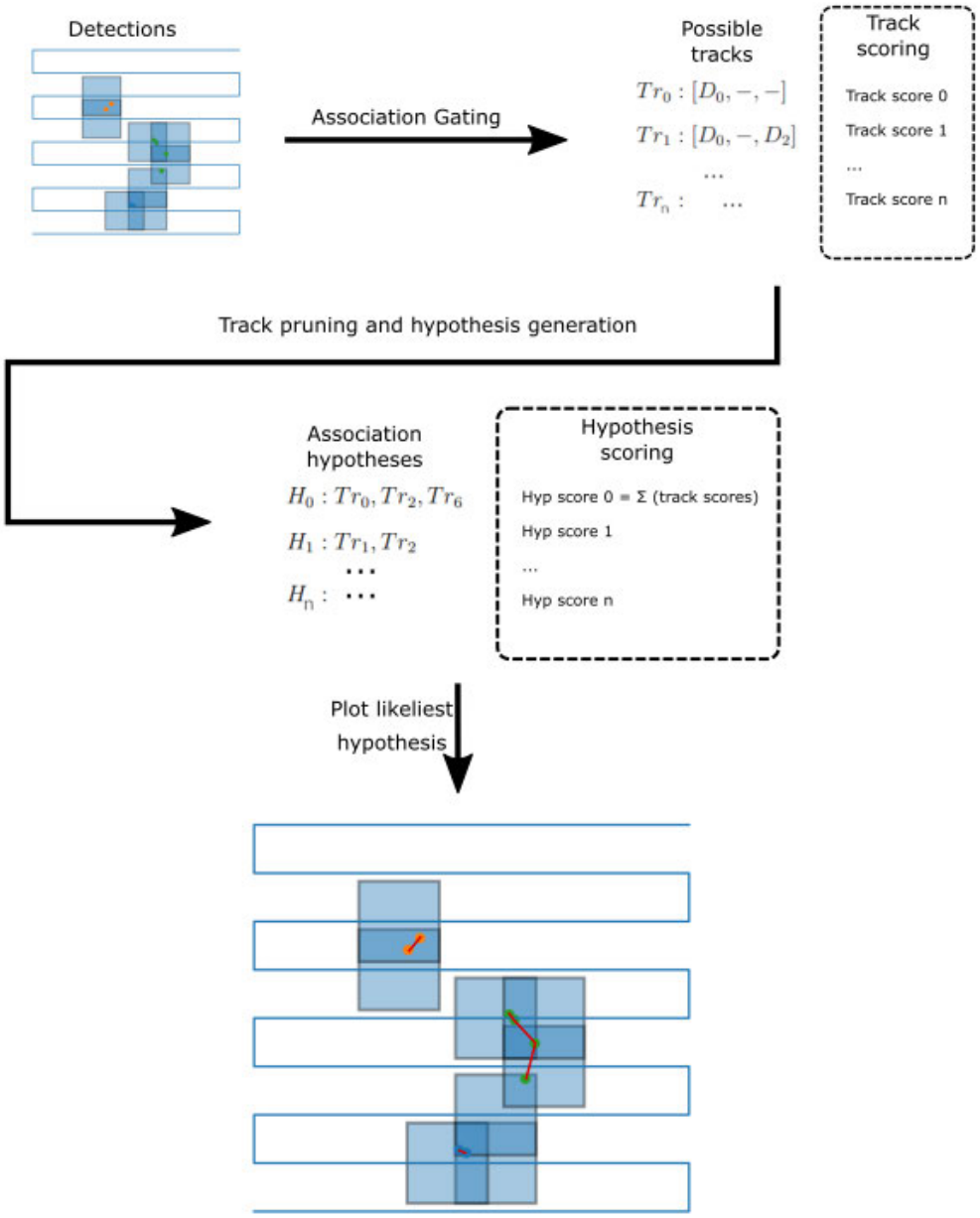


Figure 5.14 – Process diagram for applying MHT based association pipeline to detections of mobile benthic targets. After the detections are made, an understanding of the locomotory behaviour of the targets is used to discard impossible associations between detections (gating). The remaining associations are combined into possible tracks. These tracks are scored using Equation 5.1 and pruning is used to reduce the number of tracks. Compatible tracks are combined into hypotheses for the distribution of individuals. The hypothesis scores are the sum of their component tracks, and the highest scoring hypothesis represents the most likely associations. These can be plotted to show the tracks of individuals during the survey.

5.7.1 Modelling the Movement of *J. edwardsii*

To apply the association process to the localised detections of *J. edwardsii* made earlier in this chapter, a model of movement of individuals over survey timescales is needed. The first requirement is to select an appropriate value for the variance in the distribution (σ^2 in eqn 5.2). This requires an understanding of the locomotory behaviour of the target which can be gained from experiments, or from observational studies. For example, if we are interested in observing the movement of lobsters, we could look to studies of their locomotion.

Melville-Smith and Beale [137] observed the movements of released and recaptured individuals of a similar species - the western rock lobster (*Panulirus cygnus*)[3] - over a period of several days. The speeds observed ranged between 0.056 to 0.19 $\mathrm{m\,s^{-1}}$, with a mean speed of 0.12 $\mathrm{m\,s^{-1}}$.

Recall that the locomotion model introduced above gives a distribution of the target's position at a given timestep by assuming that the target takes a step of length randomly selected from a normal distribution. If we set the timestep to one second, then based on the observations in this study, the maximum expected step in one timestep is 0.19 m. Let's say the observations in that study represent 95% of the possible distribution of speeds of *P. cygnus* individuals, *i.e.* the probability of an individual moving further than 0.19 m in a second is 5%. The corresponding z-value can be calculated from the area under the curve of the standard normal distribution.

$$P(Z < z) = 0.95$$
$$z = 1.96$$

[3]Studies exploring the fine-scale (*i.e.* $<$ 1 metre) locomotion of *J. edwardsii* over AUV survey timescales (minutes-hours) could not be found so the morphologically and geographically similar *P. cygnus* has been used. This lack of studies is likely due to the difficulties of observing movement at these scales, and the methods presented in this thesis could be modified to observe these behaviours. Studies of rock lobster locomotion are typically catch based, and use recapture information over scales of weeks-years to observe larger scale movements of individuals (*e.g.* [12]).

If we also assume that the position of an individual after one second is normally distributed around its starting position (let's say $X_0 = 0$), then we can estimate the variance (σ^2) of this distribution. Let's say an individual is detected at position $X_0 = 0$, the distribution of their position after one second will have a mean (μ) of 0 so:

$$
\begin{aligned}
z &= \frac{X - \mu}{\sigma} \\
\sigma &= \frac{X - \mu}{z} \\
&= \frac{0.19 - 0}{1.96} \\
&= 0.097 \\
\sigma^2 &= 0.0094
\end{aligned}
$$

This estimate for the variance can now be used to score proposed associations between detections of *J. edwardsii* and individuals proposed using the MHT aggregation pipeline and therefore rank hypotheses for the target distribution from our survey imagery.

5.7.2 Estimating Parameters for MHT

Recall Equation 5.1, which is used to calculate the association score between two detections:

$$\Delta S(t) = ln\frac{V}{2\pi} - \frac{1}{2}ln|\sum| - \frac{d^2}{2}$$

So, we need to estimate values for:

V : The observed area (*i.e.* camera footprint)

$\sum$: The covariance of the target's predicted position

d^2 : The Mahalanobis distance between the observed and expected position of the target

During benthic surveys, the AUV *Sirius* is programmed to capture images from a fixed altitude above the seafloor. For the Huon Marine Park survey being analysed here, this results in an image footprint of approximately 1.6m x 1.3m of seafloor. Therefore, we will use a value of 2.08m^2 for the area parameter V.

The covariance and mahalanobis distance are calculated for each association pair with the variance value estimated in Section 5.7.1 using the elapsed time and distance between the candidate detections (Equations 5.3 and 5.4).

In addition to scoring the associations - which represent repeat detections of the same individual - we need to generate scores for the detection of new individuals. Recall that MHT implementations typically use a uniform likelihood across the survey area (see Footnote 2). This likelihood can be approximated using an estimate of the density of targets in the area. While this may be somewhat circular - since we are deploying the MHT process to estimate the abundance of targets in the survey region - we can use a simple approach for clustering detections (*e.g.* a distance based clustering approach as shown in Figure 5.7) to give a ballpark value for the target density. The distance based clustering estimated 30 individuals in the 30m x 30m region, giving a density of 0.03 individuals m^{-2}. Given the camera footprint area, we would therefore expect approximately 0.06 individuals in a given image, or one individual in roughly 17 images.

5.7.3 Managing the Number of Tracks

Figure 5.6 shows the 171 detections of *J. edwardsii* which we will use to estimate their abundance in the survey area. The tracks and hypotheses can be built as for the simulated data in Section 5.6.

Abundance hypotheses are formed by finding the MIS of the generated tracks. This operation is NP-hard [110], therefore limiting the number of tracks is essential to enable hypotheses to be generated [46]. As discussed above, we omit unlikely associations using the movement model for the targets by setting a threshold on the mahalanobis distance (d^2). Lobsters are known to have small home ranges [212], and are therefore unlikely to travel long distances from their original locations over the survey timescales. Therefore, we can also gate associations by setting a threshold for distance such that detections of individuals further apart than this threshold are omitted. In this application a threshold distance of 2m was used, *i.e.* detections made over 2m apart are assumed to have come from different individuals, regardless of the elapsed time.

Another approach which is widespread in applications of MHT is known as track pruning [110, 22]. Pruning involves identifying the global hypothesis at each timesteps (*i.e.* the hypothesis with the highest score) then discarding tracks which diverge from it. However, rather than removing tracks that are incompatible at the current timestep, the nodes being considered are those in the global hypothesis track several timesteps in the past. This allows pruning decisions to be delayed until further information about the candidate tracks is known, reducing the likelihood of deleting the correct track.

Applying these steps to the set of 171 detections still results in an infeasibly large number of tracks which does not allow hypotheses to be generated. Dividing the detections into subsets will allow hypotheses to be generated for each set individually. These hypotheses can then be combined to estimate the overall abundance. Figure 5.16a shows the seven clusters. This number was chosen to result in sets of detections that can be processed without the need for bespoke computational equip-

ment. The clusters were selected by manually grouping adjacent detections, and each contain between 15 and 37 detections.

5.7.4 Applying the MHT Pipeline to Detections

Applying the association pipeline to the detections of *J. edwardsii* follows the same process as outlined for the simulated data shown in Section 5.6. Each detection includes the time the image was taken, and the projected location of the detected individual.

The possible tracks are then generated using the detections in each of the seven clusters. Tracks which include an association between two detections are discarded if they do not meet the mahalanobis and distance thresholds, and valid tracks are scored using Equation 5.1. New tracks formed which do not propose an association are scored using the likelihood of encountering an individual in a frame as discussed above.

Pruning is performed at each timestep after the tracks are built. The tracks are combined to generate abundance hypotheses which are scored by summing the scores of their component tracks. The highest scoring hypothesis is set as the global hypothesis, and tracks diverging from this at 5 steps before the current timestep are discarded. Increasing the number of steps used results in a higher likelihood that the correct track will be retained as it gives more chances for associations to be confirmed. However, a greater value will result in more tracks to pass along to the next timestep, increasing the complexity of hypothesis generation. The chosen value was used to balance the tradeoff between accuracy and computational efficiency.

Once all detections have been processed, the final hypotheses are generated and scored from the remaining tracks. The predicted abundance and scores for all hypotheses in each cluster are shown in Figure 5.15. The largest score represents the proposed global hypothesis, and the track associations predicted by these are shown in Figure 5.16b. By combining the global hypotheses from each cluster we can estimate the overall

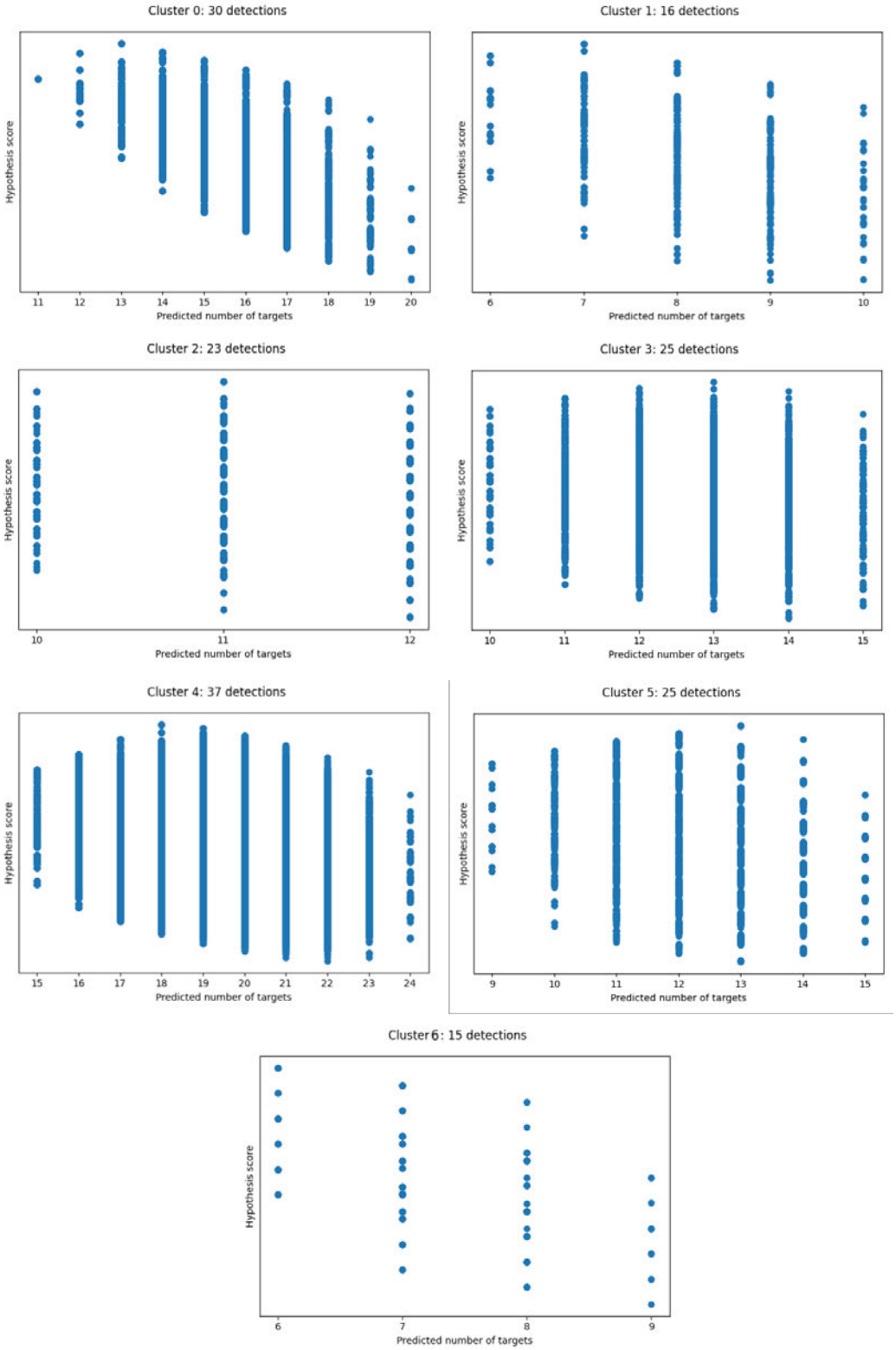


Figure 5.15 – Scores for the abundance hypotheses generated in each of the seven clusters shown in Figure 5.16a. The hypothesis score (y-axis) estimates the likelihood of the proposed associations, and the number of individuals contained in each hypothesis is shown on the x-axis. The distribution of individuals predicted by the most likely hypotheses from each cluster are shown in Figure 5.16b.

abundance of individuals on this reef section. Using the associations predicted, the total abundance estimated from this set of detections was 81 individuals.

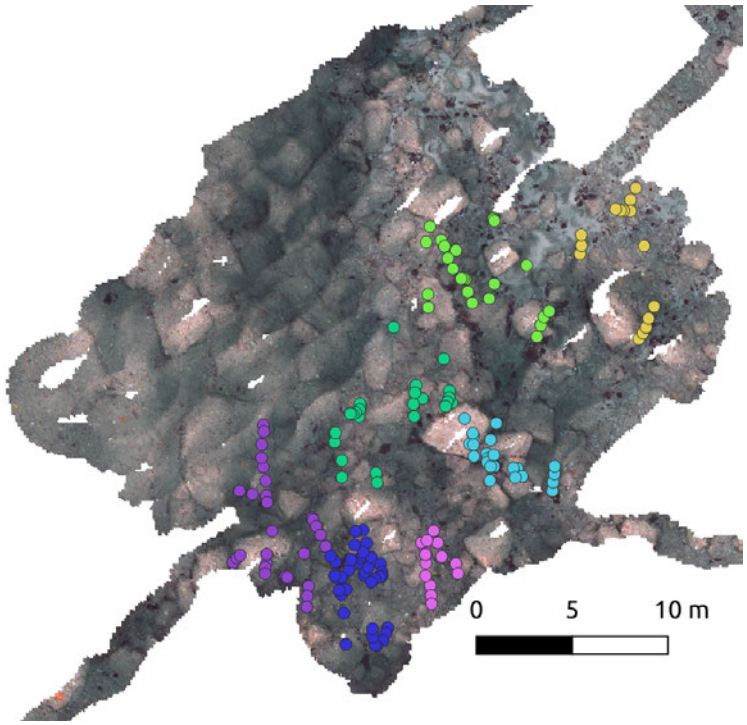
While we cannot directly validate the associations and overall abundance estimate, the previous section demonstrated that if appropriate assumptions regarding the target's movement are made then this pipeline can accurately associate detections to individuals. The model for locomotion of *J. edwardsii* was based on studies of a similar species (*P. cygnus*) in comparable environments, therefore we can assume that the generated estimate represents a reasonable approximation. Regardless of the approximation of this particular distribution of individuals, this example demonstrates how these estimates can be generated. Figure 5.15 shows scores for all hypotheses for each cluster, this shows that there are a large number of possible solutions to the aggregation problem other than the one shown. A better understanding of the movement parameters of the specific targets being observed will produce more reliable estimates and this process is likely necessary for future applications of this work.

5.8 Summary

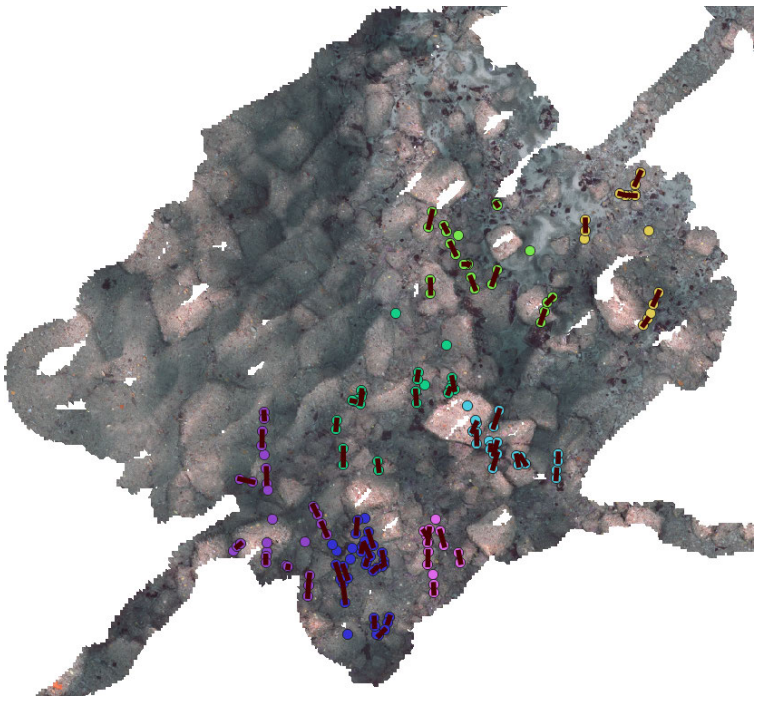
The techniques developed in this chapter facilitate estimating the spatial distribution and abundance of mobile benthic fauna over large seafloor extents. The use of automated detection allows the large numbers of images collected during seafloor to be efficiently analysed and used to generate localised detections of individual targets.

Modifying the MHT approach for tracking targets detected from a moving benthic imaging platform allows these detections to be used to estimate the abundance of targets in the survey area.

The proposed pipeline facilitates simultaneous spatial observation of mobile benthic fauna and their habitats at high resolutions and over large spatial extents that are impossible to capture using traditional techniques. These data can be used to better understand the relationship between these targets and their habitat.



(a)



(b)

Figure 5.16 – 171 Automated detections of *J. edwardsii* and proposed associations generated using an MHT based pipeline. The detections were grouped into seven clusters (a) to facilitate generating proposed association tracks and global hypotheses. The tracks from the highes scoring hypothesis for each cluster are shown in (b). The combination of these highest scoring hypotheses result in a predicted abundance of 81 individuals in the survey area.

Chapter 6

Exploiting Measures of Structural Complexity to Model Benthic Ecosystems

6.1 Introduction

In addition to the imagery used to identify individual organisms within the surveys considered to date, the data collected by our AUV systems also provides us with the opportunity to model the structural complexity of the habitats we are surveying at multiple scales and over relatively large spatial extents. This information can provide further insights into the characteristics of these environments. This work allows observation of habitat complexity and the distribution of individuals at high resolutions. Therefore these existing approaches can be applied at a scale that is difficult to achieve using existing observational techniques.

In this chapter we will consider two ways in which this structural information can be exploited in order to better model habitats that would be very difficult or impossible to achieve by other, more traditional survey methods. In the first instance we will show how the observation of the location of individual organisms can be associated with

measures of structural complexity to build models of species-habitat interactions. In the second case, we present a novel segmentation algorithm that combines structural information and imagery to segment habitat forming species that cannot be readily quantified using a pointwise (or bounding box) approach.

6.2 Modelling Species-Habitat Interactions Using Localised Detections

The methods developed so far in this thesis allow high resolution estimates of the spatial distributions of benthic features and fauna at scales which are difficult or impossible to replicate using traditional methods. Analysis of historical and future survey data provides an opportunity for unique insights into species-habitat interactions in benthic habitats. High resolution estimates of the spatial distributions of these organisms within a reef allow us to scrutinise other variables at that scale to determine how they may be influencing the fine scale spatial patterns in the distribution of individuals of the target species.

Structural complexity has been identified in many studies as an important driver of benthic community composition (see Section 2.3). AUV and other SfM based survey approaches allow for high resolution modelling of the 3D structure of the seafloor which can be used to quantify complexity across the surveyed area [69].

By pairing the 3D model generated by the AUV survey pipeline, and the spatial distributions of marine fauna estimated using the approaches presented in this thesis, it is possible to observe how fauna interact with the variation in structural complexity within a patch of reef. In this section, the impact of structural complexity on the distributions of southern rock lobster *J.edwardsii* is explored.

6.2.1 Observing Within Reef Habitat Preferences

Before we model the effect of the variation in structural complexity across the survey site on the observed distribution of *J. edwardsii*, it is useful to know whether there appears to be any difference in observed complexity in the regions where lobsters were detected and those where they weren't. This would suggest that the methods are capable of observing an impact of structural complexity on lobster distribution within the survey area.

The survey plot was divided into 3mx3m quadrats, evenly spaced on a 0.5m grid. Three metrics of structural complexity were calculated within each quadrat - rugosity, and seafloor height range and standard deviation. These values were estimated from the 3D mesh in each quadrat using the complexitymetrics software package [32]. Rugosity (see Section 2.1.1) measures the ratio between the surface area of the mesh and the planar area of the quadrat (*i.e.* 3x3m = $9m^2$), height range and standard deviation are measured using the elevation values for the seafloor from the 3D mesh.

Similarly, the three metrics were calculated in quadrats centred on 30 points where a lobster individual was detected, and a further 30 randomly selected points where lobsters were not detected. The images capturing these regions were also manually checked to ensure no lobsters were present which were missed by the automated detection algorithm. The distributions of the metrics for the three sets of quadrats are shown in Figure 6.1. The quadrats where lobsters were observed included quadrats with higher rugosity and height range than appeared in the quadrats where they were absent. Mean habitat rugosity and height range were significantly higher ($p = 0.01$, 0.03 respectively) in the lobster quadrats, suggesting lobsters were present in the more complex regions of the survey plot. This is consistent with previous observations, which generally find that lobsters prefer more complex habitats (e.g. [212, 74]).

These results show that we are able to observe non-uniform usage of habitat by fauna in AUV survey imagery using the automated census approach developed in this thesis. Next we'll employ the localised detections of lobsters and the fine spatial scale measurements of habitat complexity to build a SDM for predicting the presence

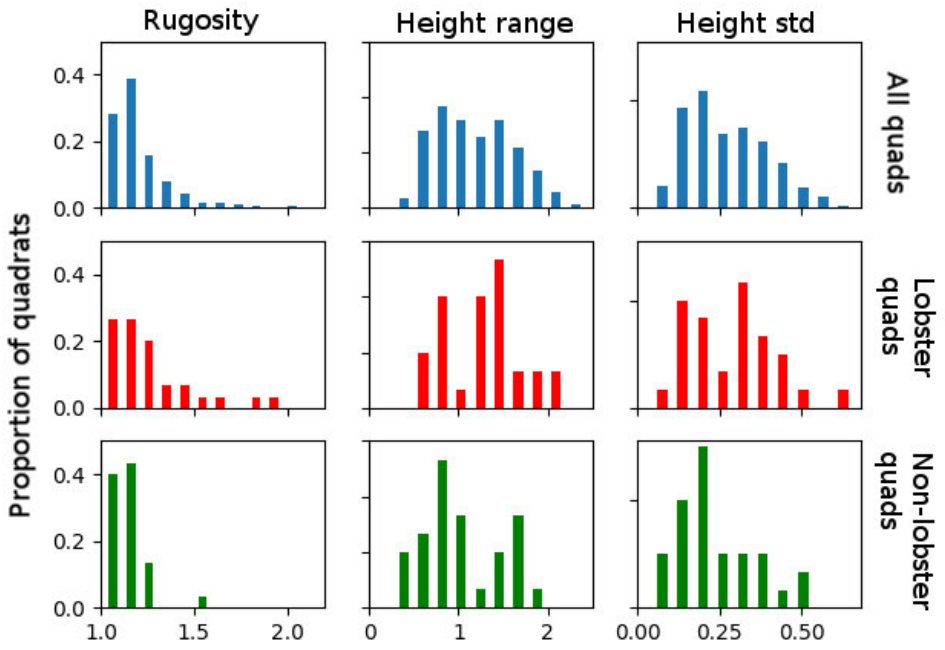


Figure 6.1 – Rugosity, height range, and height standard deviation measured from the 3D surface reconstructions in 3m x 3m quadrats centred at **(top row)** evenly spaced (0.5m) grid points across the entire mapped area, **(middle row)** the 30 lobster detections shown in Figure 5.7, and **(bottom row)** 30 randomly chosen points where lobsters were not detected. Mean rugosity and height range were measurably higher in the quadrats where lobsters were observed.

of *J. edwardsii*.

6.2.2 Modelling the Effects of Complexity on Within Reef Distributions of Benthic Fauna

A common approach for observing the effect of habitat or environmental variables on the presence of organisms is species distribution modelling. SDM is typically used to predict the presence (or absence) of plants or animals over large geographic scales using broad scale variables, *e.g.* for modelling the influence of climate data on the distribution of Joshua Trees across continental United States [78]. In marine environments, SDMs have been used in a number of studies to predict habitat associations

of sessile [102] and mobile fauna such as fish [73, 139] and rock lobster [101] using acoustic bathymmetry.

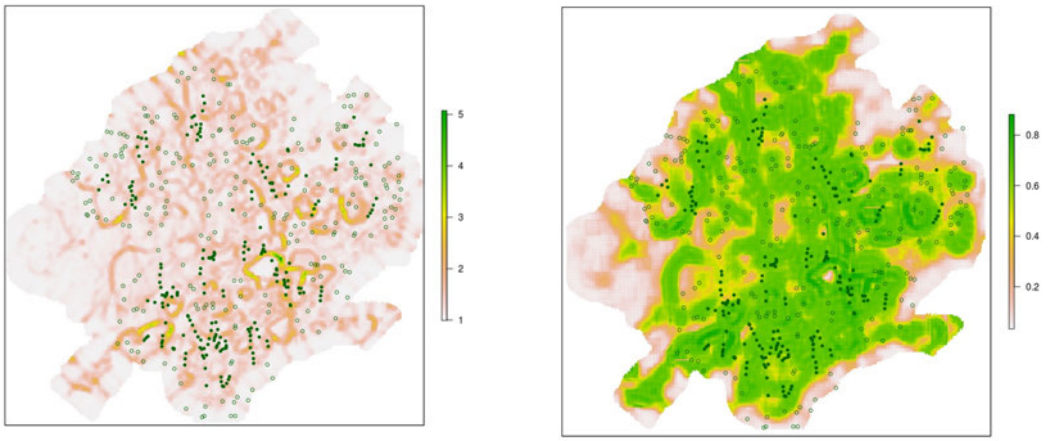
The methods presented so far in this chapter allow us to generate localised detections of mobile fauna within a reef. This information can be used to scrutinise the characteristics of the locations in which the animals have been observed to occur. As shown above (Sec. 6.2.1), we can observe non-uniform use of habitat complexity by mobile fauna using these methods. Therefore, it should also be possible to observe the extent to which the structural complexity of the habitat impacts the distribution of the observed fauna within it using SDM.

Species distribution models require three pieces of information, locations where the species of interest is present, locations where it is not present, and the value of some variable(s) at each of these points which will be used to build the model. In this section we will use the detections of *J. edwardsii* as described in Section 5.3 to generate presence and absence data, and the 3D habitat reconstruction generated using the AUV imagery to estimate the structural complexity over the surveyed area.

The reprojected points of the positive detections of *J. edwardsii* in the unprocessed AUV images were used as the lobster presence points (solid circles in Figure 6.2a). In total there were 225 detections made. Using all the detections (rather than the locations of the clusters as in the previous section) gives more data points for training as it records all the locations where any lobsters were observed. It does mean that false positive detections are included, but the performance of the detector means that these are not widespread (see Figure 5.4). For a less accurate detector, the detection threshold could be raised to reduce the instances of false positives in the presence points. To determine the absence locations, 225 random points (to match the number of presence points) were selected in the areas of the map at least 1m, but less than 3m away from any of the presence points (open circles Figure 6.2a). The habitat complexity was quantified by calculating the rugosity of the 3D mesh in virtual quadrats on a 10cm grid over the survey region. Rugosity was calculated in a range of quadrat sizes (0.6m - 1.6m). This allows us to determine whether a particular scale is a more relevant driver for predicting lobster presence in the habitat. A multi-

band raster was generated, where each layer represented the rugosity at each of the quadrats for a given quadrat size.

The values of the raster at each of the presence and absence points were used to generate a maximum entropy (MaxEnt) model (following the approach of Hijmans and Elith [99]). A 5-fold cross validation was performed to test the explanatory power of rugosity on the presence of lobsters in this plot. The model had an AUC of 0.70, and a Correlation Coefficient of 0.44. This suggests that while relatively weak, there exists a relationship between the local habitat complexity and the presence of lobster individuals at this site. The largest quadrat size (1.6m) showed the largest contribution to the relationship suggesting that complexity over a larger region may be more important in the lobsters' habitat choice. This may relate to a typical "home range" of these individuals. The trained model was then used to predict the likelihood of lobster presence based on the structural complexity at each pixel in the raster (Figure 6.2b).



(a) **(b)**

Figure 6.2 – Habitat rugosity measured in 0.6 × 0.6m virtual quadrats from a 3D habitat reconstruction of a reef in the Huon Marine Park, Tasmania. Reprojected locations of automatically detected southern rock lobster (**closed circles**), and randomly generated pseudo-absence points (**open circles**)(a). Likelihood of the presence of southern rock lobster based on a SDM using habitat complexity as the environmental variable (b).

While the model presented here only explained a portion of the spatial variation in lobster presence, the approach shows how high resolution seafloor reconstructions can be used to observe and model animal-habitat interactions at a spatial resolution and extent that would be difficult to capture using traditional observational techniques. The ability to model fine-scale habitat preferences of benthic fauna facilitates more nuanced assessments of suitable habitats than would otherwise be possible with broad-scale traditional techniques which typically decouple the observations of the target individuals and the environmental variables.

Here we have used complexity information from the AUV survey data to model its impact on rock lobster presence. However, as discussed in Section 2.3, the distribution of mobile organisms has been shown to be impacted by the presence of habitat forming organisms (*e.g.* kelp and hard corals). While AUV surveys can image these organisms, their diffuse nature mean that point or bounding-box based detection methods cannot adequately quantify their distributions within the survey imagery. The next section will describe the use of image segmentation for observation of diffuse seafloor features and how the data generated by benthic imaging AUVs can be used to improve the performance of automated segmentation models.

6.3 Mapping Habitat Forming Benthic Organisms Using Semantic Segmentation

The methods described in Chapters 4 and 5 utilise object detection algorithms to estimate the locations of stationary reef features and slow moving organisms, and faster moving mobile fauna within the reef-scale reconstruction. These animals' locations within the observed reef habitat can be approximated using points or rectangular regions, and counts or distributions of these discrete instances can be used to answer questions about their ecology (*e.g.* [196, 15, 130]).

However, to use the terminology of Adelson [2], reefs are largely composed of stuff, not things. That is, many reef features, including those of particular interest to

ecologists do not grow as discrete objects which can be easily delineated from their neighbours as individual instances. Rather, reef structures and sessile organisms grow in unique shapes, driven by *e.g.* local oceanographic conditions [193], organisms' recruitment dynamics [10], anthropogenic influences [54], and ecological interactions among neighbouring elements [33]. Valuable ecological information is contained in the shape and extent of these objects. For example, the percent cover of kelp or corals on a reef can provide a useful descriptor of a reef's condition. In fact, coral cover is the most widespread measure used to monitor change in coral reefs, particularly in response to disturbances [150]. The spatial extent, shape, and growth patterns of coral colonies can be related to the local oceanographic conditions present and can reflect ecosystems' responses to stressors [217]. Similarly, kelp cover on temperate reefs can be an indicator of overall reef health, and can be used *e.g.* to quantify the impact of over-grazing by invasive herbivores [71, 123].

Estimating the cover of kelp, corals, or other habitat forming organisms using *in situ* methods is time consuming, and is limited in it's ability to observe spatial patterns fine spatial scales over large extents. Image based methods have been shown to replicate diver estimates [199, 40], however are also tied by this same tradeoff. Reef-scale image mosaics can be used to observe the cover of these features over these extents, however manually delineating them is extremely labour intensive so is not generally feasible for routine surveys. Automating this process allows for efficient estimation of the spatial extents and distributions of habitat forming benthic organisms. Pixelwise image classification approaches also allow for more nuanced observation of the spatial interactions of structures and organisms within a reef when compared with traditional *in situ* methods which generally compress the spatial information into a percentage cover estimate over some sampling unit. This way of observing diffuse benthic features can then be used to model spatial relationships within reefs (see Chapter 5).

Manual annotation of benthic survey imagery to the pixel level is not typically used as the main analysis technique since it is highly time consuming. However, automated segmentation approaches may facilitate more widespread use of pixelwise classification, allowing a more complete view of benthic habitats. Off-the-shelf algorithms for

pixelwise classification are becoming increasingly accessible and perform well on benthic imagery. Figure 6.3 shows results from a uNet algorithm [166] trained on tiles from a photomosaic of a coral reef at Lizard Island.

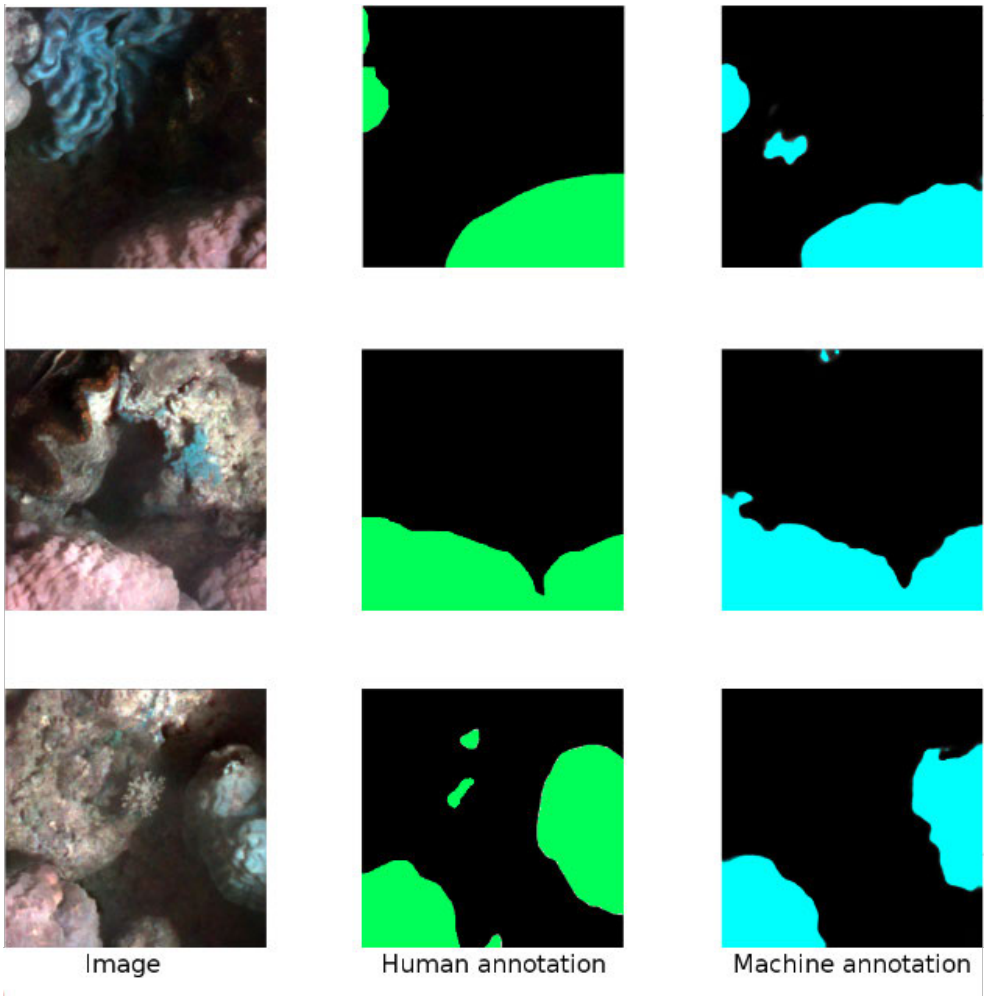


Figure 6.3 – Example outputs of the uNet segmentation algorithm trained on RGB image tiles from a reef-scale photomosaic of a reef on Lizard Island, QLD.

While the appearance of benthic imagery is important for classifying the contents, SfM surveys such as those generated by AUVs also generate a 3D reconstruction of the habitat which may also be useful for segmenting the imagery.

In this section we will explore this by developing an architecture for combining the structure and appearance information captured by 3D benthic reconstructions, and

assess the benefits of it's application for automated segmentation of reef-scale photo-mosaics.

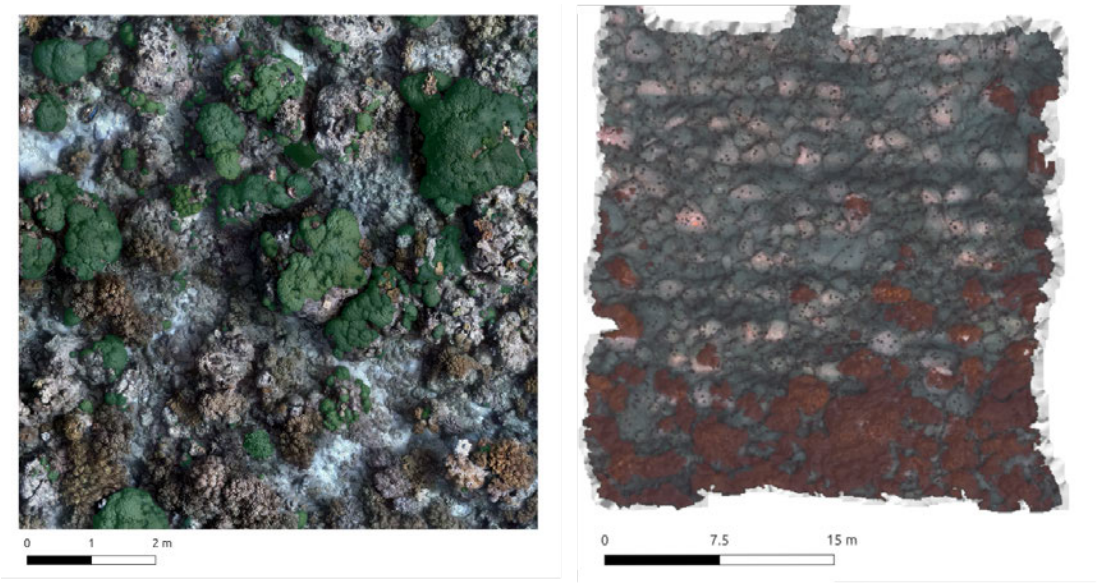


Figure 6.4 – Multi-view image reconstructions overlain with manual annotations of corals (**left**) and kelp (**right**). The coral reef mosaic was generated from imagery collected by a diver operated benthic imaging platform at Resort Reef, Lizard Island, QLD and the rocky reef by the AUV *Sirius* at St Helens Island, Tasmania. The live hard corals (**green**) were manually annotated by a coral reef ecologist, and the kelp (**red**) being much more obvious were annotated by the author. Both mosaics are overlain with a hillshade texture using the digital elevation model generated using a SfM solution from the respective survey images to illustrate the 3D structure of the two reefs. This DEM was used to train the segmentation algorithms in this section.

6.3.1 Using Habitat Structure to Aid Segmentation of Reef-Scale Imagery

Reefs are structurally diverse habitats, and the structure of individual elements are often used to identify them. For example, the physical structure of corals, their morphology, exhibits a massive variety, with different taxonomic groups displaying a wide range of growth forms. Additionally, environmental factors drive the structure of colonies, so individuals of the same species can be observed to grow in a variety

of ways in different habitats [66]. As such, the physical structure of live corals is likely to be distinct from the surrounding reef features. This structural information may therefore aid in automated segmentation of live coral from the surrounding reef. Similarly, on temperate rocky reefs, vegetation (*e.g.* kelp or seagrass) have a much different physical structure compared to bare rocks or other reef features.

Using image-based SfM reef mapping techniques, a large-scale image stitched from multiple views of the reef is captured. An estimate physical structure of the surveyed area is also produced. Such reconstruction techniques are often used to study the structural complexity of reefs, or study the growth patterns in corals and other structures. Here, this information will be used to supplement the image information to perform segmentation of the reef-scale photomosaics generated by the AUV.

In this section, the reconstruction is used in the form of a digital elevation model (DEM), where the height of the reef at each pixel is recorded, giving a 2D representation of the structure. This format allows the structural information to be provided to the segmentation algorithm in much the same way the image data is, and since the image and DEM are aligned, the colour and structural information for each pixel can be easily associated.

6.3.2 Building and Training the Segmentation Models

The models used throughout this section are based on Ronenberger's [166] uNet (see Figure 6.5). This algorithm was originally developed for segmentation of biomedical imagery, however has been widely applied to pixel-wise classification tasks in a range of applications (*e.g.* [41, 156, 11]), including for segmenting images of coral reefs [138]. It has a relatively simple network structure which lends itself to easy modification for different input data. The network consists of a series of convolutional and pooling layers, which encode the input data into a smaller feature vector, followed by a series of up convolutions layers which decode the feature vector into a segmentation mask which predicts a class label for each pixel in the input image.

This type of network can be used for both single (*e.g.* [138]) or multi-class (*e.g.* [104])

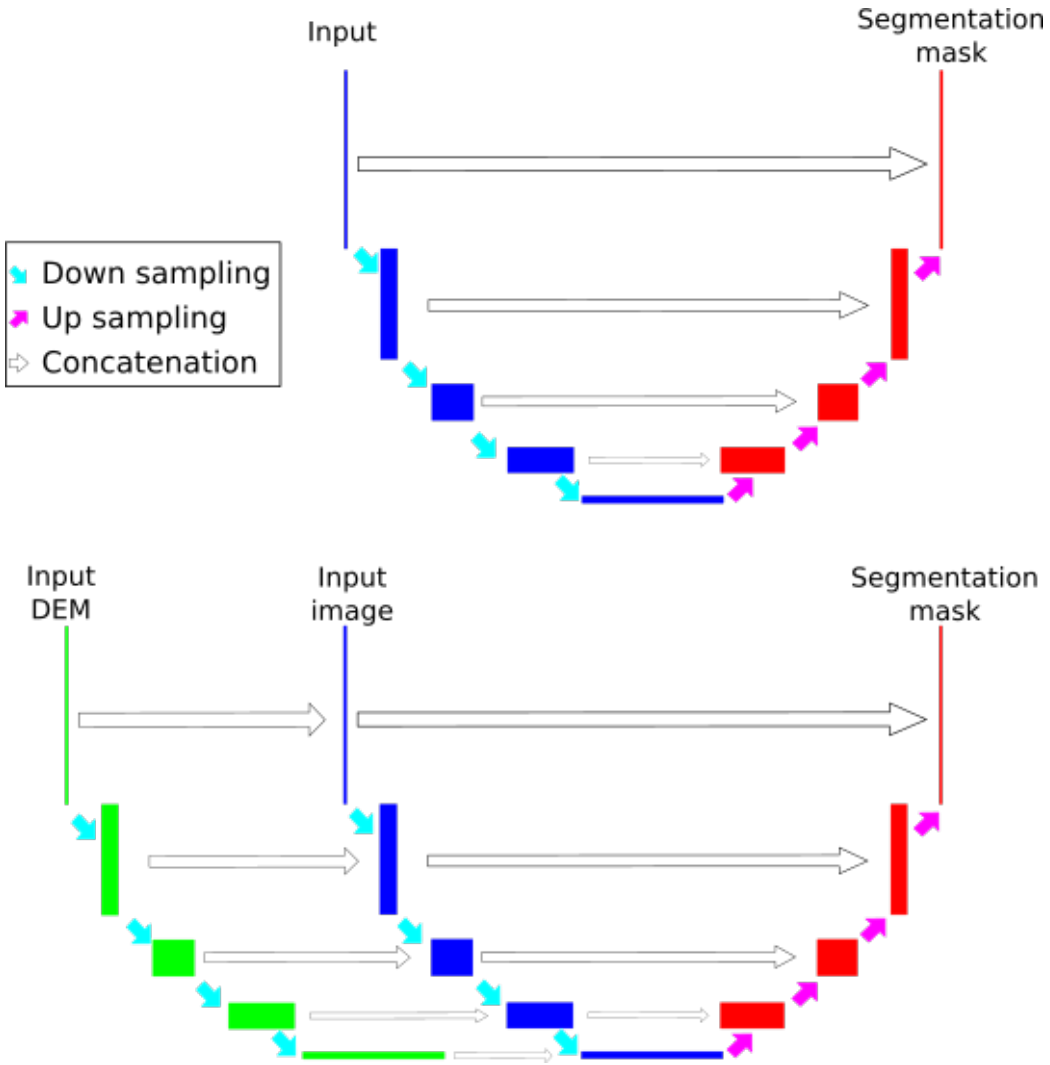


Figure 6.5 – The uNet (**top**), and wNet (**bottom**) image segmentation algorithms. In the uNet model, the input passes through an encoding arm (**blue**) of four down sampling blocks (**cyan**), and a decoding arm (**red**) of a further four up sampling blocks (**magenta**) resulting in a segmentation mask of the same dimension as the input. The output from each down sampling block is concatenated with the corresponding block on the decoding arm. The wNet model modifies this structure by adding an encoding arm (**green**) to allow for different weights to be trained for the structural information contained in a reef DEM.

problems. In the single-class case, pixels in the input image are either the target class, or they are background, *i.e.* a binary classification problem. For the analysis in the following section the problem of pixel-wise binary classification of kelp from reef-scale images is considered, and each image pixel is classified as one of two classes (kelp, or background). However, the approach can be adapted to observe multiple habitat classes, provided that sufficient annotated imagery of each class of interest are available.

The base model used in these tests is derived from the original version presented by Ronneberger *et al.* [166]. Like that model it consists of an encoding arm comprised of 5 blocks, each containing two convolutional layers with 3x3 kernels, and followed by a rectified linear unit (ReLu) and a max pooling operation. In the decoding arm, the model used here also follows that in the paper by performing a 2x2 convolution to up sample the feature map, and two further convolutions using 3x3 kernels, again each followed by ReLu units. At each block in the decoding arm, the feature map from the corresponding block in the encoding arm are concatenated. Two main modifications were made to this model in the following section. Firstly it was found that the model converged more reliably on the kelp data by using a smaller number of feature channels in each of the convolutional layers. Where the original authors use 64, 128, 256, 512, and 1024 filters for each of the down sampling blocks (and the reverse for the up sampling blocks), in this work I use 32, 64, 128, 256 and 512 filters. Secondly, after each convolutional layer (apart from the 2x2 up sampling convolutions), a batch normalisation layer is included to stabilise the model and improve the training speed [103, 170, 213, 112]. Additionally, I used the dice coefficient [53][1] to calculate the loss function during training of the models where Ronneberger *et al.* use the cross entropy loss function. The dice coefficient is more suited to imbalanced datasets, which is the case here since kelp only covers a small proportion of the whole surveyed reef (see Figure 6.4).

Imagery from a survey by *Sirius* at St Helens Island, Tasmania was used to demonstrate the segmentation approaches. The mosaic and DEM from the survey was

[1]The dice coefficient standardises segments by their area, thus improving performance in cases where the data is dominated by the background set (as is common in benthic habitats).

divided into training, test and validation sets (using a 60/20/20 split) and these sets were split into 512x512 pixel tiles. To test the ability of the network to segment kelp from the survey data, the mosaic tiles from the test set were passed to the uNet model as described above.

Two further approaches were used to incorporate the DEM into the segmentation algorithm. For the first method, the DEM tiles were appended to the corresponding mosaic tiles, forming a four channel "image" with red, green, blue and depth channels. These 4D tiles were then passed to the algorithm as the image tiles were.

For the second approach, the model architecture was modified to add a second encoding arm to the network, resulting in what I will refer to as the wNet model. This arm had the same structure as the existing arm, and the feature maps from this arm were concatenated with both the down sampling blocks in the image encoding arm, and the final up sampling arm (see bottom, Figure 6.5).

6.3.3 Comparing Segmentation Outputs

Outputs of the three segmentation models on an unseen section of the surveyed reef patch are shown in Figure 6.6. As in the coral reef example (Figure 6.3), using the image data only resulted in a model which performed well on the unseen test data.

Incorporating the reef's structural information as an added channel in the input image, as in the uNet 4D model, resulted in very poor performance on the new data, giving an output which did not resemble the manually labelled kelp regions at all. This suggests that the image and depth channels require separate weights in order to accurately classify the image pixels.

The wNet model does this by introducing an independent encoding arm for the depth image. As a result it performs much better than the uNet 4D model and results in a output which looks similar to that of the image only uNet model at first glance. Where the uNet and wNet models differ, however, is their performance in regions of the reef

with lower density of kelp. Figure 6.7) shows the intersection over union (IoU[2]) of the predicted kelp cover of the two models and the ground truth on 512x512 pixel tiles. For tiles with (ground truth) kelp cover lower than $\sim 6.5\%$, the wNet model showed improved performance. At higher cover, the two models' performance were similar. This suggests that regions of lower kelp density will benefit the most from deploying automatic segmentation of kelp using both image and structure information.

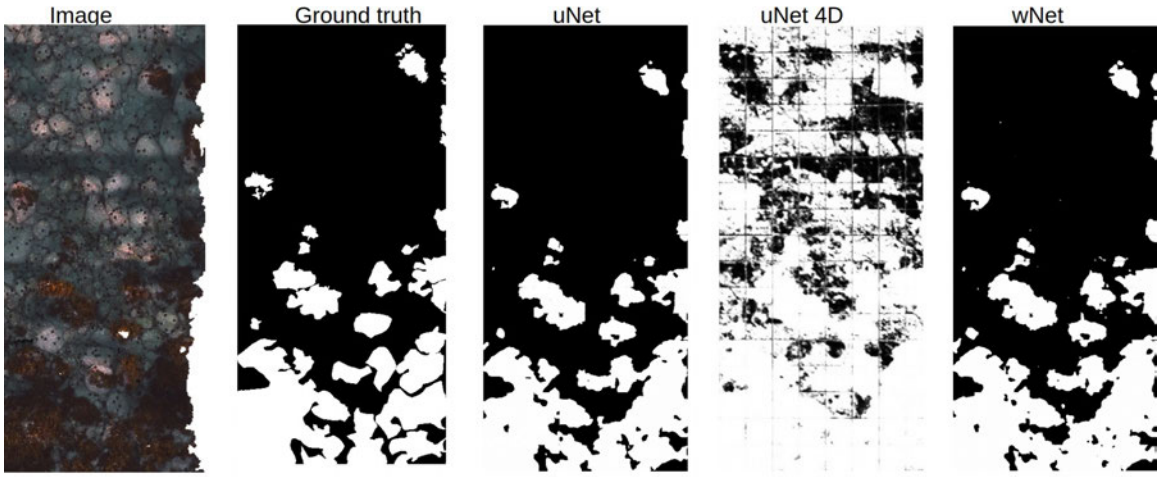


Figure 6.6 – Automated segmentation of kelp from AUV generated reef photomosaic using purely image information (**uNet**), combined image and structural information as a 4D RGBD input (**uNet 4D**), and using a purpose build modified network for the combined input data (**wNet**).

6.4 Summary

This chapter has demonstrated the use of seafloor reconstructions for providing novel observations of benthic habitats. By pairing the methods for generating within-reef spatial distributions of benthic fauna presented in this thesis with the fine-scale estimates of habitat complexity obtained from 3D reconstructions it was shown that SfM methods are an ideal source of data for modelling species-habitat interactions.

[2]IoU is a measure used to compare the output of segmentation (or object detection) models. It computes the pixel area shared by the prediction and ground truth (*i.e.* the intersection) and divides it by the sum of the pixel areas of the prediction and ground truth while only counting the overlapping area once (*i.e.* the union).

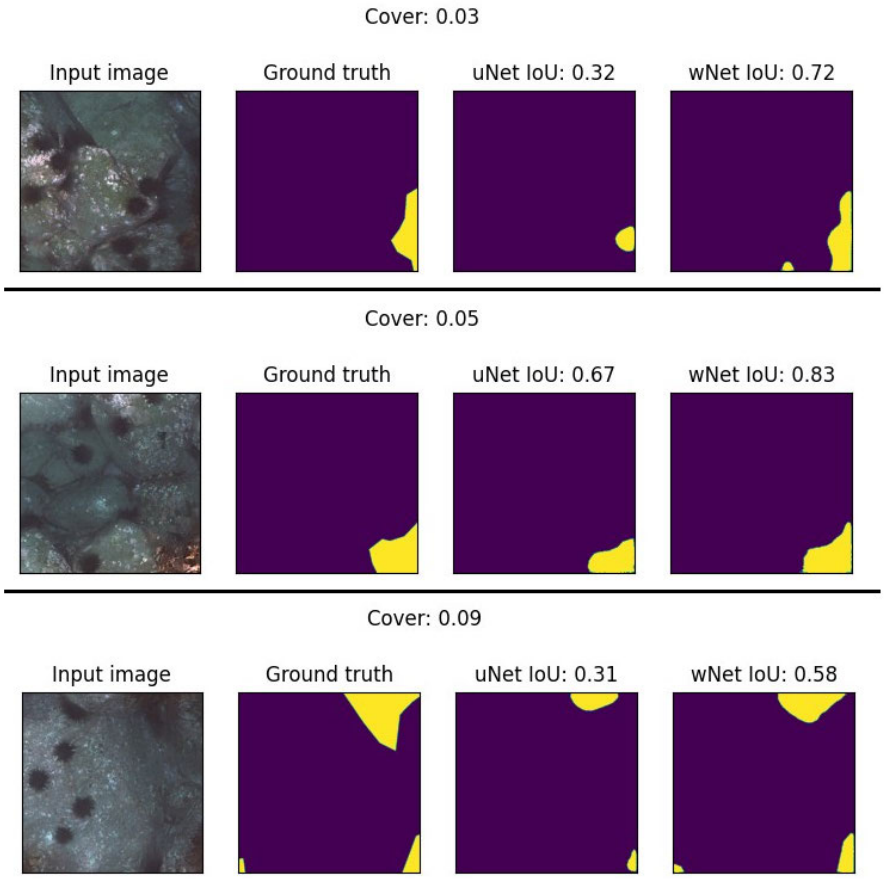


Figure 6.7 – Intersection over Union (IoU) of automated kelp segmentation in three examples of regions of reef with low kelp cover. The wNet image segmentation algorithm developed in this chapter, which utilises both the mosaic and DEM, appears to perform better in regions of low kelp cover then the uNet, which uses only the image information (see Figure 6.8). Three example images which demonstrate this are shown.

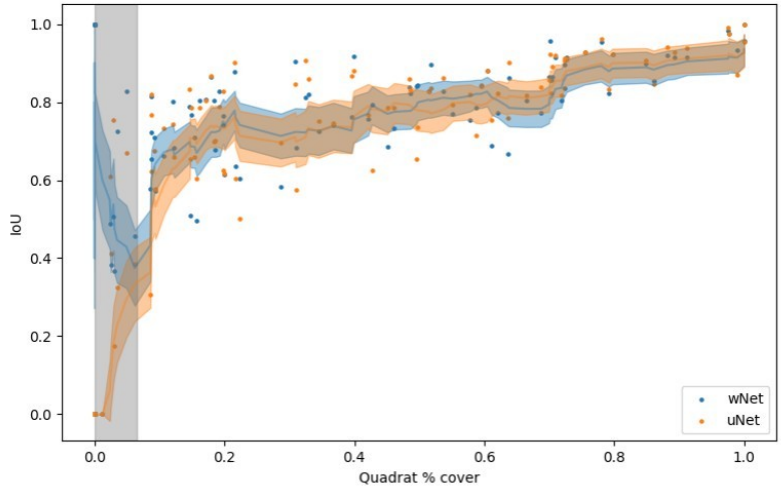


Figure 6.8 – Intersection over Union (IoU) of two kelp segmentation algorithms, wNet which uses both the image information and the DEM generated using the AUV surveys (**blue**), and uNet, which uses only the image information (**yellow**). Lines and shaded regions represent 10 point moving average and standard error respectively. The use of structural information improves performance at kelp cover below $\sim$ 6.5% (**grey shaded region**), otherwise the two networks display similar in performance

Using these data to model relationships between benthic fauna and their habitat allows observations of the interactions at fine-scales over large spatial extents which are not possible using existing observational techniques.

Similarly, the novel wNet segmentation model presented here demonstrates a further use case for benthic reconstruction data. The model incorporates both image and structural information, both derived from the same AUV survey data, while maintaining separate encoding weights for the two data types. The model was applied for segmentation of kelp in a reef-scale benthic composite image outperformed a typical method for combining these data where the image and structure information were concatenated before passing them to a segmentation model with typical architecture. Furthermore the wNet model resulted in improved segmentation performance in regions of low kelp density compared to an image only network. Low kelp density regions include those which have been subject to recent disturbances such as urchin barrens (see Section 4.2). Improved segmentation quality at low densities would allow for improved ability to detect recovery of these systems, *e.g.* in response to management activity.

The ability to generate high performing segmentation models at all densities of the target organisms can in turn be used to build improved models of the relationship between benthic fauna and their habitat in the same way that the complexity data was used here.

Chapter 7

Conclusions and Future Considerations

7.1 Introduction

This thesis aimed to facilitate routine deployment of reef-scale, composite seafloor imaging techniques for estimating the abundance and spatial distributions of benthic targets. The implementation of automated object detection allows for survey imagery to be efficiently analysed, and the detection, localisation, and aggregation pipelines developed make them robust to movement of the target individuals.

This chapter provides a summary of the primary contributions of this thesis. The details and implications of the contributions are discussed in Section 7.2, and Section 7.3 proposes directions for future expansion and applications of this work. Finally, a brief summary is presented in Section 7.4.

7.2 Summary of Contributions

The contributions of this thesis address two main hurdles to applying reef-scale composite imaging for estimating the abundance and distribution of benthic targets:

1. The large volume of images collected make routine deployment of pipelines which rely on manual analysis make infeasible due to the associated time and cost.

2. Estimates of the distribution and abundance of mobile targets made using their detections in composite images are unreliable, since the process to create them assumes a rigid, stationary scene.

7.2.1 Estimating the Abundance and Spatial Distributions of Benthic Targets Using Observations From a Moving Imaging Platform

It was demonstrated that the motion of benthic targets affects their appearance in the data products generated using SfM processes. This thesis presented a general approach which guides the steps required to generate estimates of the abundance and distribution of targets from these data based on their movement.

Stationary targets can be detected and localised in the composite images, this provides a straightforward pipeline for estimating their abundance and distribution. However due to their movement between image captures, mobile targets can be distorted, missing, or duplicated in composite images. While observations of individuals in unprocessed images can be localised, each view does not necessarily represent a unique individual and a set of images used to reconstruct the habitat likely includes many observations of individual targets. Therefore a tracking based approach which modifies MHT was developed which applies knowledge of the target's locomotory behaviour to score associations between detections and individuals. This approach generates scored hypotheses for the composition of the community of targets given a set of localised detections.

This facilitates the development of specific methods for observing target species based on their locomotory behaviour as shown for the case-studies presented.

7.2.2 Automated Spatial Mapping of Stationary and Slow Moving Benthic Targets

The simplest targets for observation using reef-scale composite images are those which do not move over the course of the survey. These targets appear undistorted in the final image products, and the data's georeferencing can be applied to localise individuals. The contribution of this work is in generating automated detections of benthic targets, demonstrating that detection models can be trained with relatively small sets of manually annotated training data.

It was shown that estimates of density made using the automated detections of slow moving benthic targets were comparable to those using manual analysis, and the pipeline was able to detect temporal trends in the density of sea urchins made using manual analysis of imagery, and direct in water observations.

This means that researchers can be confident that automated detection pipelines can be applied for observation of the abundance, density and distribution of benthic targets in reef-scale photomosaics. This addresses the challenge of budgeting sufficient time for manual analysis of photomosaics for extracting these information and the tradeoff between resolution and extent inherent to traditional *in situ* observational techniques.

7.2.3 Automated Detection and Localisation of Mobile Benthic Targets

Since mobile benthic targets cannot be reliably detected in composite images, detections need to be made in the unprocessed images then reprojected into 3D space. However, the image sets used to generate benthic reconstructions can contain tens of thousands of images. Therefore manual analysis of these data would result in a processing bottleneck, likely preventing widespread uptake of such surveys. This work applied an automated detection pipeline to these images, again demonstrating that performant models can be generated with relatively small training datasets.

Detections made in unprocessed images cannot rely on the georeferencing of the reconstruction to estimate the position of the target. Instead, the vehicle's navigation solution and the geometry of the scene were used to localise detected individuals.

Together, the presented methods allow for efficient generation of localised detections of mobile benthic fauna from benthic images and seafloor reconstruction data. This allows for the large amount of imagery collected in routine SfM or AUV surveys to be feasibly analysed facilitating sustainable implementation of the observation of mobile benthic targets in these data.

7.2.4 Segmentation of Seafloor Photomosaics Using Appearance and Structure Information

While this work was primarily concerned with detection and mapping of targets which occur as distinct individuals (*e.g.* urchins or lobsters), benthic habitats are characterised by features which cannot be discretely detected. These include habitat forming organisms such as kelp or coral, and the abiotic substrate of the seafloor itself. The ability to map such features allows a more comprehensive understanding of these ecosystems, including their effect on the distribution and abundance of discrete targets. This information can be used to model the effect of habitat features on target organisms, and these models can be used to better estimate the abundance of mobile targets from localised detections using the approached developed in this thesis.

Classification of these features can be assisted by understanding their structure as well as their appearance. SfM surveys allow observation of both over large spatial extents, therefore present a unique opportunity to apply automated segmentation using these two sources of information.

It was found that utilising both structure and appearance resulted in improved segmentation performance of kelp, particularly where it occurred at low densities. However, simply stacking the structural information with the composite image resulted in much worse performance. Instead a novel architecture was developed which learned

separate encoding weights for the two data sources resulting in the improved performance.

7.3 Future Research

The methods presented in this work allow for unique observations of benthic fauna and habitats. Routinely collecting these data and analysing them using the tools developed here provides an exciting opportunity for understanding how animals interact with their habitat at fine spatial scales. The most obvious directions for future work stemming from this thesis are to validate the abundance estimates of mobile fauna against traditional observational techniques (Section 7.3.1), applying these methods for developing models of within-reef species-habitat interactions (Section 7.3.2), and applying the presented segmentation approach to different environments (Section 7.3.3).

7.3.1 Field Validation of Abundance Estimates

In section 4.5, automated estimates of urchin density were compared with expert annotations made from the same imagery. This allowed for a direct comparison of the methods presented in this thesis with existing observations and found that the automated pipeline could detect temporal trends in urchin density.

However, no such dataset was available to validate the estimates of abundance of mobile benthic fauna. Simulated data was used to show that accurate abundance estimates can be made, provided that the assumptions about the locomotory behaviour of the targets are sound. However, validation of these methods against data collected using existing techniques is essential for widespread adoption and application of results for management of the target species and their habitat.

7.3.2 Building Fine-Scale Models of Species-Habitat Interactions

The techniques presented in this thesis allow for observations which are not possible using traditional techniques. They allow for high-resolution (*i.e.* within reef) estimates of the distributions of individuals across large spatial extents which are paired with millimetre scale habitat reconstructions and maps of habitat features. This provides a unique opportunity to observe the habitat use of benthic fauna.

A demonstration of this capability was demonstrated in 6.2.2 where a SDM was applied to predict the suitability of habitat within a rocky reef for *J. edwardsii* based on measures of structural complexity. A weak, but significant relationship between the two was observed. Further study using these approaches could develop a better understanding of habitat preferences of benthic fauna and the effect of the presence of different habitat types on the fine-scale distributions of fauna. Such an understanding could aid in management of benthic habitats, *e.g.* the planning of protected areas by identifying features which support target species.

Improved models for the relationship between different habitat features and the presence of mobile targets would also allow for better association of detections to individuals. These would allow for the background likelihood estimates to vary based on the habitat characteristics, rather than relying on a fixed value for the background likelihood of encountering an individual as used in this work.

7.3.3 Applying Structure and Appearance Based Segmentation Models

The segmentation approach developed here was applied to observe kelp on a rocky, temperate reef. Another application well suited to such an approach is surveys of coral reefs. The 3D structure of coral colonies is partially determined by their taxonomy, therefore utilising SfM derived estimates, paired with reef-scale composite images may

allow for improved classification performance. Better segmentation models would allow for more accurate estimates of coral cover and spatial distributions of colonies.

Other possible applications of this approach include observation of subsea infrastructure, and for underwater archaeology which would both benefit from improved segmentation of benthic features.

7.4 Summary

This thesis has addressed several major roadblocks which have prevented widespread implementation of SfM surveys for estimating the abundance and fine-scale spatial distribution of benthic targets in large seafloor extents. This will pave the way for routine deployment of these surveys facilitating a better understanding of benthic fauna and habitats, and improved management and protection of these ecosystems.

List of References

[1] Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A. E., and Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11):e00938.

[2] Adelson, E. H. (1996). Human Vision and Electronic Imaging. *Proceedings of SPIE - The International Society for Optical Engineering*, 2657:1–12.

[3] AgiSoft (2023). AgiSoft MetaShape Professional 2.0.

[4] Ahmed, D. A., Benhamou, S., Bonsall, M. B., and Petrovskii, S. V. (2021). Three-dimensional random walk models of individual animal movement and their application to trap counts modelling. *Journal of Theoretical Biology*, 524:110728.

[5] Alexander, T. J., Barrett, N., Haddon, M., and Edgar, G. (2009). Relationships between mobile macroinvertebrates and reef structure in a temperate marine reserve. *Marine Ecology Progress Series*, 389:31–44.

[6] Allison, E. H., Perry, A. L., Badjeck, M. C., Neil Adger, W., Brown, K., Conway, D., Halls, A. S., Pilling, G. M., Reynolds, J. D., Andrew, N. L., and Dulvy, N. K. (2009). Vulnerability of national economies to the impacts of climate change on fisheries. *Fish and Fisheries*, 10(2):173–196.

[7] Andrew, N. L. (1993). Spatial heterogeneity, sea urchin grazing, and habitat structure on reefs in temperate Australia. *Ecology*, 74(2):292–302.

[8] Arel, I., Rose, D. C., and Karnowski, T. P. (2010). Deep Machine Learning - A New Frontier in Artificial Intelligence Research. *IEEE Computational Intelligence Magazine*, 5(4):13–18.

[9] Auster, P. J., Malatesta, R. J., and Larosa, S. C. (1995). Patterns of microhabitat utilization by mobile megafauna on the southern New England (USA) continental shelf and slope. *Marine Ecology Progress Series*, 127(1-3):77–85.

[10] Babcock, R. and Mundy, C. (1996). Coral recruitment: Consequences of settlement choice for early growth and survivorship in two scleractinians. *Journal of Experimental Marine Biology and Ecology*, 206(1-2):179–201.

[11] Baheti, B., Innani, S., Gajre, S., and Talbar, S. (2020). Eff-UNet: A novel architecture for semantic segmentation in unstructured environment. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2020-June:1473–1481.

[12] Barrett, N., Buxton, C., and Gardner, C. (2009). Rock lobster movement patterns and population structure within a Tasmanian Marine protected area inform fishery and conservation management. *Marine and Freshwater Research*, 60(5):417–425.

[13] Bartumeus, F., Da Luz, M. G., Viswanathan, G. M., and Catalan, J. (2005). Animal search strategies: A quantitative random-walk analysis. *Ecology*, 86(11):3078–3087.

[14] Beijbom, O., Edmunds, P. J., Kline, D. I., Mitchell, B. G., and Kriegman, D. (2012). Automated annotation of coral reef survey images. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1170–1177.

[15] Bell, J. and Galzin, R. (1984). Influence of live coral cover on coral-reef fish communities. *Marine Ecology Progress Series*, 15:265–274.

[16] Bell, S. S. (1985). Habitat complexity of polychaete tube-caps: influence of architecture on dynamics of a meioepibenthic assemblage. *Journal of Marine Research*, 43(3):647–671.

[17] Bender, A. (2013). *Autonomous exploration of large-scale benthic environments*. PhD thesis, The University of Sydney.

[18] Bennett, S., Halford, A. R., Choat, J. H., Hobbs, J.-P. A., Santana-Garcon, J., Ayling, A. M., Harvey, E. S., and Newman, S. J. (2018). Geography and island geomorphology shape fish assemblage structure on isolated coral reef systems. *Ecology and Evolution*, 8(12):6242–6252.

[19] Berthelot, G., Saïd, S., and Bansaye, V. (2021). A random walk model that accounts for space occupation and movements of a large herbivore. *Scientific Reports*, 11(1):14061.

[20] Bewley, M. S., Douillard, B., Nourani-Vatani, N., Friedman, A., Pizarro, O., and Williams, S. B. (2012). Automated species detection: An experimental approach to kelp detection from sea-floor AUV images. *Australasian Conference on Robotics and Automation*, 2012:1–10.

[21] Blackall, L. L., Wilson, B., and Van Oppen, M. J. H. (2015). Coral-the world's most diverse symbiotic ecosystem. *Molecular Ecology*, 24(21):5330–5347.

[22] Blackman, S. and Popoli, R. (1999). *Design and Analysis of Modern Tracking Systems*. Artech House Radar Library, Boston, 1st edition.

[23] Blackman, S. S. (2004). Multiple hypothesis tracking for multiple target tracking. *IEEE Aerospace and Electronic Systems Magazine*, 19(1):5–18.

[24] Bodenmann, A., Thornton, B., and Ura, T. (2017). Generation of High-resolution Three-dimensional Reconstructions of the Seafloor in Color using a Single Camera and Structured Light. *Journal of Field Robotics*, 34(5):833–851.

[25] Boldt, J. L., Williams, K., Rooper, C. N., Towler, R. H., and Gauthier, S. (2018). Development of stereo camera methodologies to improve pelagic fish biomass estimates and inform ecosystem management in marine waters. *Fisheries Research*, 198(2018):66–77.

[26] Bouguet, J. Y. (2008). Camera calibration toolbox for Matlab.

[27] Boutros, N., Shortis, M. R., and Harvey, E. S. (2015). A comparison of calibration methods and system configurations of underwater stereo-video systems for applications in marine ecology. *Limnology and Oceanography: Methods*, 13(5):224–236.

[28] Brock, J. C., Wright, C. W., Clayton, T. D., and Nayegandhi, A. (2004). LIDAR optical rugosity of coral reefs in Biscayne National Park, Florida. *Coral Reefs*, 23(1):48–59.

[29] Brown, C. J., Smith, S. J., Lawton, P., and Anderson, J. T. (2011). Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine, Coastal and Shelf Science*, 92(3):502–520.

[30] Brunetti, A., Buongiorno, D., Trotta, G. F., and Bevilacqua, V. (2018). Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. *Neurocomputing*, 300:17–33.

[31] Bryson, M., Ferrari, R., Figueira, W., Pizarro, O., Madin, J., Williams, S., and Byrne, M. (2017). Characterization of measurement errors using structure-from-motion and photogrammetry to measure marine habitat structural complexity. *Ecology and Evolution*, 7(15):5669–5681.

[32] Bryson, M., Porter, A., and Figueira, W. (2018). complexitymetrics.

[33] Buddemeier, R. and Kinzie, R. (1976). Coral growth. *Oceanography and Marine Biology: An Annual Review*, 14:183–225.

[34] Buhl-Mortensen, L., Vanreusel, A., Gooday, A. J., Levin, L. A., Priede, I. G., Buhl-Mortensen, P., Gheerardyn, H., King, N. J., and Raes, M. (2010). Biological structures as a source of habitat heterogeneity and biodiversity on the deep ocean margins. *Marine Ecology*, 31(1):21–50.

[35] Burns, J., Delparte, D., Gates, R., and Takabayashi, M. (2015). Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs. *PeerJ*, 3:e1077.

[36] Byrne, M. and Andrew, N. (2013). Centrostephanus rodgersii. In *Developments in Aquaculture and Fisheries Science*, volume 38, chapter 17, pages 243–256. Elsevier.

[37] Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698.

[38] Cappo, M. (2010). *Development of a baited video technique and spatial models to explain patterns of fish biodiversity in inter-reef waters*. Phd thesis, James Cook University.

[39] Cappo, M. a., Harvey, E. B., Malcolm, H. C., and Speare, P. a. (2003). Potential of Video Techniques To Monitor Diversity , Abundance and Size of Fish in Studies of Marine Protected Areas. *Aquatic Protected Areas-what works best and how do we know*, (2003):455–464.

[40] Carleton, J. H. and Done, T. J. (1995). Quantitative video sampling of coral reef benthos: large-scale application. *Coral Reefs*, 14(1):35–46.

[41] Chen, C., Bai, W., Davies, R. H., Bhuva, A. N., Manisty, C. H., Augusto, J. B., Moon, J. C., Aung, N., Lee, A. M., Sanghvi, M. M., Fung, K., Paiva, J. M., Petersen, S. E., Lukaschuk, E., Piechnik, S. K., Neubauer, S., and Rueckert, D. (2020). Improving the Generalizability of Convolutional Neural Network-Based Segmentation on CMR Images. *Frontiers in Cardiovascular Medicine*, 7:1–15.

[42] Chow, C. F. Y., Bolton, C., Boutros, N., Brambilla, V., Fontoura, L., Hoey, A. S., Madin, J. S., Pizarro, O., Torres-Pulliza, D., Woods, R. M., Zawada, K. J. A., Barbosa, M., and Dornelas, M. (2023). Coral settlement and recruitment are negatively related to reef fish trait diversity. *Coral Reefs*, 42(2):519–533.

[43] Clarke, M. E., Whitmire, C., Fruh, E., Anderson, J., Taylor, J., Rooney, J., Ferguson, S., and Singh, H. (2010). Developing the SeaBED AUV as a tool for conducting routine surveys of fish and their habitat in the Pacific. In *2010 IEEE/OES Autonomous Underwater Vehicles, AUV 2010*.

[44] Collier, J. S. and Brown, C. J. (2005). Correlation of sidescan backscatter with grain size distribution of surficial seabed sediments. *Marine Geology*, 214(4):431–449.

[45] Collins, D. L., Langlois, T. J., Bond, T., Holmes, T. H., Harvey, E. S., Fisher, R., and McLean, D. L. (2017). A novel stereo-video method to investigate fish–habitat relationships. *Methods in Ecology and Evolution*, 8(1):116–125.

[46] Cong, S. and Hong, L. (1999). Computational complexity analysis for multiple hypothesis tracking. *Mathematical and Computer Modelling*, 29(9):1–16.

[47] Cresswell, K., Hartmann, K., Gardner, C., and Keane, J. (2019). Tasmanian Longspined Sea Urching Fishery Assessment 2018/19. Technical report, Institute for Marine and Antarctic Studies, University of Tasmania.

[48] Crickenberger, S. (2016). Predicting a range shift and range limits in an introduced tropical marine invertebrate using species distribution models. *Hydrobiologia*, 763(1):193–205.

[49] Crist, T. O., Guertin, D. S., Wiens, J. A., and Milne, B. T. (1992). Animal Movement in Heterogeneous Landscapes: An Experiment with Eleodes Beetles in Shortgrass Prairie. *Functional Ecology*, 6(5):536.

[50] Crossland, J. (1976). Fish trapping experiments in northern new zealand waters. *New Zealand Journal of Marine and Freshwater Research*, 10(3):511–516.

[51] Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695:1–9.

[52] Dayoub, F., Dunbabin, M., and Corke, P. (2015). Robotic detection and tracking of Crown-of-Thorns starfish. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 2015, pages 1921–1928. IEEE.

[53] Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302.

[54] Dodge, R. E. and Rimas Vaisnys, J. (1977). Coral populations and growth patterns: responses to sedimentation and turbidity associated with dredging. *Journal of Marine Research*, 35(4 , Nov.1977):715–730.

[55] Doszpot, N., McWilliam, M., Pratchett, M., Hoey, A., and Figueira, W. (2019). Plasticity in Three-Dimensional Geometry of Branching Corals Along a Cross-Shelf Gradient. *Diversity*, 11(3):44.

[56] Dunn, D. C. and Halpin, P. N. (2009). Rugosity-based regional modeling of hard-bottom habitat. *Marine Ecology Progress Series*, 377(Leslie 2005):1–11.

[57] Durden, J. M., Bett, B. J., Schoening, T., Morris, K. J., Nattkemper, T. W., and Ruhl, H. A. (2016). Comparison of image annotation data generated by multiple investigators for benthic ecology. *Marine Ecology Progress Series*, 552:61–70.

[58] Edwards, C. B., Eynaud, Y., Williams, G. J., Pedersen, N. E., Zgliczynski, B. J., Gleason, A. C., Smith, J. E., and Sandin, S. A. (2017). Large-area imaging reveals biologically driven non-random spatial patterns of corals at a remote reef. *Coral Reefs*, 36(4):1291–1305.

[59] Elith, J. and Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):677–697.

[60] Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 226–231. Elsevier.

[61] Ferrari, R., Figueira, W. F., Pratchett, M. S., Boube, T., Adam, A., Kobelkowsky-Vidrio, T., Doo, S. S., Atwood, T. B., and Byrne, M. (2017). 3D photogrammetry quantifies growth and external erosion of individual coral colonies and skeletons. *Scientific Reports*, 7(1):1–9.

[62] Ferrari, R., McKinnon, D., He, H., Smith, R., Corke, P., González-Rivero, M., Mumby, P., and Upcroft, B. (2016). Quantifying Multiscale Habitat Structural Complexity: A Cost-Effective Framework for Underwater 3D Modelling. *Remote Sensing*, 8(2):113.

[63] Filbee-Dexter, K. and Scheibling, R. E. (2014). Sea urchin barrens as alternative stable states of collapsed kelp ecosystems. *Marine Ecology Progress Series*, 495:1–25.

[64] Floater, M. S. (2015). Generalized barycentric coordinates and applications. *Acta Numerica*, 24:161–214.

[65] Flukes, E. B., Johnson, C. R., and Ling, S. D. (2012). Forming sea urchin barrens from the inside out: An alternative pattern of overgrazing. *Marine Ecology Progress Series*, 464:179–194.

[66] Foster, A. B. (1979). Phenotypic plasticity in the reef corals Montastraea annularis (Ellis & Solander) and Siderastrea siderea (Ellis & Solander). *Journal of Experimental Marine Biology and Ecology*, 39(1):25–54.

[67] Friedlander, A. M. and Parrish, J. D. (1998). Habitat characteristics affecting fish assemblages on a Hawaiian coral reef. *Journal of Experimental Marine Biology and Ecology*, 224(1):1–30.

[68] Friedman, A. (2013). *Automated interpretation of benthic stereo imagery.* PhD thesis, The University of Sydney.

[69] Friedman, A., Pizarro, O., Williams, S. B., and Johnson-Roberson, M. (2012). Multi-Scale Measures of Rugosity, Slope and Aspect from Benthic Stereo Image Reconstructions. *PLoS ONE*, 7(12):1–14.

[70] Friedman, A., Steinberg, D., Pizarro, O., and Williams, S. B. (2011). Active learning using a variational dirichlet process model for pre-clustering and classification of underwater stereo imagery. *IEEE International Conference on Intelligent Robots and Systems*, 2011:1533–1539.

[71] Gagnon, P., Johnson, L. E., and Himmelman, J. H. (2005). Kelp patch dynamics in the face of intense herbivory: Stability of Agarum clathratum (Phaeophyta) stands and associated flora on urchin barrens. *Journal of Phycology*, 41(3):498–505.

[72] Galaiduk, R., Figueira, W. F., Kingsford, M. J., and Curley, B. G. (2013). Factors driving the biogeographic distribution of two temperate Australian damselfishes and ramifications for range shifts. *Marine Ecology Progress Series*, 484:189–202.

[73] Galaiduk, R., Radford, B. T., and Harvey, E. S. (2018). Utilizing individual fish biomass and relative abundance models to map environmental niche associations of adult and juvenile targeted fishes. *Scientific Reports*, 8(1):9457.

[74] Galparsoro, I., Borja, Á., Bald, J., Liria, P., and Chust, G. (2009). Predicting suitable habitat for the European lobster (Homarus gammarus), on the Basque continental shelf (Bay of Biscay), using Ecological-Niche Factor Analysis. *Ecological Modelling*, 220(4):556–567.

[75] Garulli, A., Giannitrapani, A., Rossi, A., and Vicino, A. (2005). Mobile robot SLAM for line-based environment representation. In *Proceedings of the 44th IEEE Conference on Decision and Control*, volume 2005, pages 2041–2046. IEEE.

[76] Gleason, H. A. (1920). Some Applications of the Quadrat Method. *Bulletin of the Torrey Botanical Club*, 47(1):21.

[77] Goatley, C. H., González-Cabello, A., and Bellwood, D. R. (2016). Reef-scale partitioning of cryptobenthic fish assemblages across the Great Barrier Reef, Australia. *Marine Ecology Progress Series*, 544:271–280.

[78] Godsoe, W., Strand, E., Smith, C. I., Yoder, J. B., Esque, T. C., and Pellmyr, O. (2009). Divergence in an obligate mutualism is not explained by divergent climatic factors. *New Phytologist*, 183(3):589–599.

[79] Goetze, J. S., Bond, T., McLean, D. L., Saunders, B. J., Langlois, T. J., Lindfield, S., Fullwood, L. A., Driessen, D., Shedrawi, G., and Harvey, E. S. (2019). A field and video analysis guide for diver operated stereo-video. *Methods in Ecology and Evolution*, 10(7):1083–1090.

[80] Golet, W. J., Scopel, D. A., Cooper, A. B., and Watson, W. H. (2006). Daily patterns of locomotion expressed by American lobsters (Homarus americanus) in their natural habitat. *Journal of Crustacean Biology*, 26(4):610–620.

[81] González-Rivero, M., Bongaerts, P., Beijbom, O., Pizarro, O., Friedman, A., Rodriguez-Ramirez, A., Upcroft, B., Laffoley, D., Kline, D., Bailhache, C., Vevers, R., and Hoegh-Guldberg, O. (2014). The Catlin Seaview Survey - kilometre-scale seascape assessment, and monitoring of coral reef ecosystems. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 24(S2):184–198.

[82] González-Rivero, M., Harborne, A. R., Herrera-Reveles, A., Bozec, Y.-M., Rogers, A., Friedman, A., Ganase, A., and Hoegh-Guldberg, O. (2017). Linking fishes to multiple metrics of coral reef structural complexity using three-dimensional technology. *Scientific Reports*, 7:1–15.

[83] Gordon, S., Aston, E., Lechene, M., Harianto, J., Bray, P., Figueira, W., Gonzalez Rivero, M., and Ferrari Legorreta, R. (2023). Reef Restoration and Adaption Program (EcoRRAP). Standard Operational Procedure Number 1: Overview and in-field workflow. Technical Report 14.

[84] Graham, N. A. J. and Nash, K. L. (2013). The importance of structural complexity in coral reef ecosystems. *Coral Reefs*, 32(2):315–326.

[85] Gratwicke, B. and Speight, M. R. (2005). The relationship between fish species richness, abundance and habitat complexity in a range of shallow tropical marine habitats. *Journal of Fish Biology*, 66(3):650–667.

[86] Greene, L. E. and Alevizon, W. S. (1989). Comparative Accuracies of Visual Assessment Methods for Coral Reef Fishes. *Bulletin of Marine Science*, 44(2):899–912.

[87] Harman, N., Harvey, E. S., and Kendrick, G. A. (2003). Differences in fish assemblages from different reef habitats at Hamelin Bay, south-western Australia. *Marine and Freshwater Research*, 54(2):177–184.

[88] Harris, C. and Stephens, M. (1988). A Combined Corner and Edge Detector. In *Proceedings of the 4th Alvey Vision Conference*, volume 15, pages 147–152.

[89] Hartley, R. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2nd edition.

[90] Harvey, E., Fletcher, D., and Shortis, M. (2001). A comparison of the precision and accuracy of estimates of reef-fish lengths determined visually by divers with estimates produced by a stereo-video system. *Fishery Bulletin*, 99(1):63–71.

[91] Harvey, E., Fletcher, D., and Shortis, M. (2002). Estimation of reef fish length by divers and by stereo-video. A first comparison of the accuracy and precision in the field on living fish under operational conditions. *Fisheries Research*, 57(3):255–265.

[92] Harvey, E. S., Goetze, J., McLaren, B., Langlois, T., and Shortis, M. R. (2010). Influence of Range, Angle of View, Image Resolution and Image Compression on Underwater Stereo-Video Measurements: High-Definition and Broadcast-Resolution Video Cameras Compared. *Marine Technology Society Journal*, 44(1):75–85.

[93] Harvey, E. S. and Shortis, M. (1995). A system for stereo-video measurement of sub-tidal organisms. *Marine Technology Society Journal*, 29(4):10–22.

[94] He, N., Fang, L., and Plaza, A. (2020). Hybrid first and second order attention Unet for building segmentation in remote sensing images. *Science China Information Sciences*, 63(4):1–12.

[95] Healey, D. and Hovel, K. A. (2004). Seagrass bed patchiness: Effects on epifaunal communities in San Diego Bay, USA. *Journal of Experimental Marine Biology and Ecology*, 313(1):155–174.

[96] Helmuth, B. S. T. (1998). Intertidal Mussel Microclimates: Predicting the Body Temperature of a Sessile Invertebrate. *Ecological Monographs*, 68(1):51.

[97] Hernández-Landa, R. C., Barrera-Falcon, E., and Rioja-Nieto, R. (2020). Size-frequency distribution of coral assemblages in insular shallow reefs of the Mexican Caribbean using underwater photogrammetry. *PeerJ*, 2020(4):1–30.

[98] Hewitt, J. E., Thrush, S. F., Cummings, V. J., and Turner, S. J. (1998). The effect of changing sampling scales on our ability to detect effects of large-scale processes on communities. *Journal of Experimental Marine Biology and Ecology*, 227(2):251–264.

[99] Hijmans, R. J. and Elith, J. (2013). Species Distribution Modeling with R. *R Cran Project*, 2013:1–78.

[100] Hill, J. and Wilkinson, C. (2004). Methods for ecological monitoring of coral reefs. *Australian Institute of Marine Science, Townsville*, page 117.

[101] Hovey, R. K., van Niel, K. P., Bellchambers, L. M., and Pember, M. B. (2012). Modelling deep water habitats to develop a spatially explicit, fine scale

understanding of the distribution of the western rock lobster, Panulirus cygnus. *PLoS ONE*, 7(4).

[102] Ierodiaconou, D., Monk, J., Rattray, A., Laurenson, L., and Versace, V. (2011). Comparison of automated classification techniques for predicting benthic biological communities using hydroacoustics and video observations. *Continental Shelf Research*, 31:S28–S38.

[103] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *32nd International Conference on Machine Learning, ICML 2015*, 1:448–456.

[104] Isaka, S., Kawanaka, H., Aronow, B. J., and Prasath, V. B. (2019). Multi-Class Segmentation of Lung Immunofluorescence Confocal Images Using Deep Learning. *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, pages 2362–2368.

[105] Jalal, A., Salman, A., Mian, A., Shortis, M., and Shafait, F. (2020). Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecological Informatics*, 57(2020):1–13.

[106] Johnson, C., Ling, S., Ross, J., Shepherd, S., and Miller, K. (2005). Establishment of the long-spined sea urchin (Centrostephanus rodgersii) in Tasmania: first assessment of potential threats to fisheries. Technical report, Fisheries Research and Development Corporation (Australia).

[107] Johnson, C. R., Ling, S. D., Sanderson, J. C., Frusher, S., Gardner, C., Hartmann, K., Jarman, S., Little, R., and Marzloff, M. P. (2007). Rebuilding Ecosystem Resilience : Assessment of management options to minimise formation of ‘ barrens ’ habitat by the long - spined sea urchin (Centrostephanus rodgersii) in Tasmania. Technical report, Fisheries Research and Development Corporation, Institute for Marine and Antarctic Studies, NRM North.

[108] Johnson-Roberson, M., Pizarro, O., Williams, S. B., and Mahon, I. (2010). Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. *Journal of Field Robotics*, 27(1):21–51.

[109] Kenny, A. J., Cato, I., Desprez, M., Fader, G., Schüttenhelm, R. T., and Side, J. (2003). An overview of seabed-mapping technologies in the context of marine habitat classification. *ICES Journal of Marine Science*, 60(2):411–418.

[110] Kim, C., Li, F., Ciptadi, A., and Rehg, J. M. (2015). Multiple hypothesis tracking revisited. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4696–4704.

[111] Kinzel, P. J., Legleiter, C. J., and Nelson, J. M. (2013). Mapping River Bathymetry With a Small Footprint Green LiDAR: Applications and Challenges 1. *JAWRA Journal of the American Water Resources Association*, 49(1):183–204.

[112] Kohler, J., Daneshmand, H., Lucchi, A., Hofmann, T., Zhou, M., and Neymeyr, K. (2020). Exponential convergence rates for Batch Normalization: The power of length-direction decoupling in non-convex optimization. *AISTATS 2019 - 22nd International Conference on Artificial Intelligence and Statistics*.

[113] Komyakova, V., Munday, P. L., and Jones, G. P. (2013). Relative importance of coral cover, habitat complexity and diversity in determining the structure of reef fish communities. *PLoS ONE*, 8(12):1–12.

[114] Kovalenko, K. E., Thomaz, S. M., and Warfe, D. M. (2012). Habitat complexity: Approaches and future directions. *Hydrobiologia*, 685(1):1–17.

[115] Krocker, R. and Schenke, H. W. (2006). Comparison of predicted and measured bathymetry. *Acta Acustica united with Acustica*, 92(1):159–161.

[116] Kwasnitschka, T., Hansteen, T. H., Devey, C. W., and Kutterolf, S. (2013). Doing fieldwork on the seafloor: Photogrammetric techniques to yield 3D visual models from ROV video. *Computers & Geosciences*, 52:218–226.

[117] Lange, I. D., Molina-Hernández, A., Medellín-Maldonado, F., Perry, C. T., and Álvarez-Filip, L. (2022). Structure-from-motion photogrammetry demonstrates variability in coral growth within colonies and across habitats. *PLOS ONE*, 17(11):e0277546.

[118] Larson, E. R., DiStefano, R. J., Magoulick, D. D., and Westhoff, J. T. (2008). Efficiency of a Quadrat Sampling Technique for Estimating Riffle-Dwelling Crayfish Density. *North American Journal of Fisheries Management*, 28(4):1036–1043.

[119] Lassau, S. A., Hochuli, D. F., Cassis, G., and Reid, C. A. (2005). Effects of habitat complexity on forest beetle diversity: Do functional groups respond consistently? *Diversity and Distributions*, 11(1):73–82.

[120] Li, J., Skinner, K. A., Eustice, R. M., and Johnson-Roberson, M. (2017). WaterGAN: Unsupervised Generative Network to Enable Real-time Color Correction of Monocular Underwater Images. *IEEE Robotics and Automation Letters*, 3(1):387–394.

[121] Limmon, G. V., Masdar, H., Muenzel, D., Shalders, T. C., Djakiman, C., Beger, M., Jompa, J., and De Brauwer, M. (2023). A cause for hope: largely intact coral-reef communities with high reef-fish biomass in a remote Indonesian island group. *Marine and Freshwater Research*, 74(6):479–490.

[122] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollar, P. (2020). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327.

[123] Ling, S. and Keane, J. (2018). Resurvey of the longspined sea urchin (Centrostephanus rodgersii) and associated barren reef in Tasmania. Technical Report December, Institute of Marine and Antarctic Studies, University of Tasmania.

[124] Ling, S. D. (2008). Range expansion of a habitat-modifying species leads to loss of taxonomic diversity: A new and impoverished reef state. *Oecologia*, 156(4):883–894.

[125] Ling, S. D., Johnson, C. R., Frusher, S. D., and Ridgway, K. R. (2009). Overfishing reduces resilience of kelp beds to climate-driven catastrophic phase shift. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52):22341–22345.

[126] Ling, S. D., Mahon, I., Marzloff, M. P., Pizarro, O., Johnson, C. R., and Williams, S. B. (2016). Stereo-imaging AUV detects trends in sea urchin abundance on deep overgrazed reefs. *Limnology and Oceanography: Methods*, 14(5):293–304.

[127] Liu, G., Heron, S. F., Mark Eakin, C., Muller-Karger, F. E., Vega-Rodriguez, M., Guild, L. S., de la Cour, J. L., Geiger, E. F., Skirving, W. J., Burgess, T. F., Strong, A. E., Harris, A., Maturi, E., Ignatov, A., Sapper, J., Li, J., and Lynds, S. (2014). Reef-scale thermal stress monitoring of coral ecosystems: New 5-km global products from NOAA coral reef watch. *Remote Sensing*, 6(11):11579–11606.

[128] Lloyd, S. P. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137.

[129] Luckhurst, E. and Luckhurst, K. (1978). Analysis of the Influence of Substrate Variables on Coral Reef Fish Communities. *Marine Biology*, 323:317–323.

[130] MacArthur, R. H. (1958). Population Ecology of Some Warblers of Northeastern Coniferous Forests. *Ecology*, 39(4):599–619.

[131] MacLeod, N., Benfield, M., and Culverhouse, P. (2010). Time to automate identification. *Nature*, 467(7312):154–155.

[132] Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Kendrick, G., and Fisher, R. (2016). Automatic annotation of coral reefs using deep learning. In *OCEANS 2016 MTS/IEEE Monterey*, pages 1–5. IEEE.

[133] Mahmood, A., Bennamoun, M., An, S., Sohel, F. A., Boussaid, F., Hovey, R., Kendrick, G. A., and Fisher, R. B. (2019). Deep Image Representations for Coral Image Classification. *IEEE Journal of Oceanic Engineering*, 44(1):121–131.

[134] Marrable, D., Tippaya, S., Barker, K., Harvey, E., Bierwagen, S. L., Wyatt, M., Bainbridge, S., and Stowar, M. (2023). Generalised deep learning model for semi-automated length measurement of fish in stereo-BRUVS. *Frontiers in Marine Science*, 10(June):1–11.

[135] Mattison, J. E., Trent, J. D., Shanks, A. L., Akin, T. B., and Pearse, J. S. (1976). Movement and feeding activity of red sea urchins (Strongylocentrotus franciscanus) adjacent to a kelp forest. *Marine Biology*, 39(1):25–30.

[136] McLean, D. L., Langlois, T. J., Newman, S. J., Holmes, T. H., Birt, M. J., Bornt, K. R., Bond, T., Collins, D. L., Evans, S. N., Travers, M. J., Wakefield, C. B., Babcock, R. C., and Fisher, R. (2016). Distribution, abundance, diversity and habitat associations of fishes across a bioregion experiencing rapid coastal development. *Estuarine, Coastal and Shelf Science*, 178:36–47.

[137] Melville-Smith, R. and Beale, N. E. (2009). Movement patterns of individual migrating western rock lobster, Panulirus cygnus, in Western Australia. *New Zealand Journal of Marine and Freshwater Research*, 43(5):1095–1102.

[138] Mizuno, K., Terayama, K., Hagino, S., Tabeta, S., Sakamoto, S., Ogawa, T., Sugimoto, K., and Fukami, H. (2020). An efficient coral survey method based on a large-scale 3-D structure model obtained by Speedy Sea Scanner and U-Net segmentation. *Scientific Reports*, 10(1):1–11.

[139] Monk, J., Ierodiaconou, D., Harvey, E. S., Rattray, A., and Versace, V. L. (2012). Are we predicting the actual or apparent distribution of temperate marine fishes? *Plos One*, 7(4):e34558.

[140] Morris, K. J., Bett, B. J., Durden, J. M., Huvenne, V. A., Milligan, R., Jones, D. O., McPhail, S., Robert, K., Bailey, D. M., and Ruhl, H. A. (2014). A new method for ecological surveying of the abyss using autonomous underwater vehicle photography. *Limnology and Oceanography: Methods*, 12(NOV):795–809.

[141] Murphy, H. and Jenkins, G. (2010). Observational methods used in marine spatial monitoring of fishes and associated habitats: a review. *Marine and Freshwater Research*, 61(2):236–252.

[142] Nanni, L., Ghidoni, S., and Brahnam, S. (2017). Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognition*, 71:158–172.

[143] Nanninga, G. B., Saenz-Agudelo, P., Zhan, P., Hoteit, I., and Berumen, M. L. (2015a). Not finding Nemo: limited reef-scale retention in a coral reef fish. *Coral Reefs*, 34(2):383–392.

[144] Nanninga, G. B., Saenz-Agudelo, P., Zhan, P., Hoteit, I., and Berumen, M. L. (2015b). Not finding Nemo: limited reef-scale retention in a coral reef fish. *Coral Reefs*, 34(2):383–392.

[145] Nilsson, H. C. and Rosenberg, R. (2000). Succession in marine benthic habitats and fauna. 197:139–149.

[146] Oakley-Cogan, A., Tebbett, S. B., and Bellwood, D. R. (2020). Habitat zonation on coral reefs: Structural complexity, nutritional resources and herbivorous fish distributions. *PLoS ONE*, 15(6).

[147] of Australia, C. (2023). Win-win under our oceans: Climate related marine invasive species. Technical Report November, The Senate of The Commonwealth of Australia, Canberra.

[148] O'Grady, D. F., Jury, S. H., and Watson III, W. H. (2001). Use of a treadmill to study the relationship between walking, ventilation and heart rate in the lobster Homarus americanus. *Marine and Freshwater Research*, 52:1387–1394.

[149] Olah, C., Mordvintsev, A., and Schubert, L. (2017). Feature Visualization: How neural networks build up their understanding of images. *Distill*, 2(11).

[150] Osborne, K., Dolman, A. M., Burgess, S. C., and Johns, K. A. (2011). Disturbance and the dynamics of coral cover on the Great Barrier Reef (1995-2009). *PLoS ONE*, 6(3).

[151] Padilla, R., Netto, S. L., and Da Silva, E. A. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. *International Conference on Systems, Signals, and Image Processing*, 2020-July:237–242.

[152] Pais, M. P. and Cabral, H. N. (2018). Effect of underwater visual survey methodology on bias and precision of fish counts: a simulation approach. *PeerJ*, 6(7):e5378.

[153] Palma, M., Magliozzi, C., Casado, M. R., Pantaleo, U., Fernandes, J., Coro, G., Cerrano, C., and Leinster, P. (2019). Quantifying coral reef composition of recreational diving sites: A structure from motion approach at seascape scale. *Remote Sensing*, 11(24):1–21.

[154] Parsons, D. F., Suthers, I. M., Cruz, D. O., and Smith, J. A. (2016). Effects of habitat on fish abundance and species composition on temperate rocky reefs. *Marine Ecology Progress Series*, 561:155–171.

[155] Pedersen, N. E., Edwards, C. B., Eynaud, Y., Gleason, A. C. R., Smith, J. E., and Sandin, S. A. (2019). The influence of habitat and adults on the spatial distribution of juvenile corals. *Ecography*, 42(10):1703–1713.

[156] Pedraza, A., Gallego, J., Lopez, S., Gonzalez, L., Laurinavicius, A., and Bueno, G. (2017). Glomerulus Classification with Convolutional Neural Networks. In Valdés Hernández, M. and González-Castro, V., editors, *Medical Image Understanding and Analysis. MIUA 2017*, volume 723 of *Communications in Computer and Information Science*, pages 839–849. Springer International Publishing, Cham.

[157] Phillips, K., Boero Rodriguez, V., Harvey, E. S., Ellis, D., Seager, J., Begg, G., and Hender, J. (2008). Assessing the operational feasibility of stereo-video and evaluating monitoring options for the Southern Bluefin Tuna Fishery ranch sector. Technical report, Australian Government: Department of Agriculture, Fisheries and Forestry.

[158] Pizarro, O., Friedman, A., Bryson, M., Williams, S. B., and Madin, J. (2017). A simple, fast, and repeatable survey method for underwater visual 3D benthic mapping and monitoring. *Ecology and Evolution*, 7(6):1770–1782.

[159] Polunin, N. V. and Roberts, C. M. (1993). Greater biomass and value of target coral-reef fishes in two small Caribbean marine reserves. *Marine Ecology Progress Series*, 100(1-2):167–176.

[160] Price, D. M., Robert, K., Callaway, A., Lo lacono, C., Hall, R. A., and Huvenne, V. A. I. (2019). Using 3D photogrammetry from ROV video to quantify cold-water coral reef structural complexity and investigate its influence on biodiversity and community assemblage. *Coral Reefs*, 38(5):1007–1021.

[161] Priede, I., Bagley, P., Smith, A., Creasey, S., and Merrett, N. (1994). Scavenging deep demersal fishes of the Porcupine Seabight, north-east Atlantic: observations by baited camera, trap and trawl. *Journal of the Marine Biological Association of the United Kingdom*, 74(3):481–498.

[162] Puig-Pons, V., Muñoz-Benavent, P., Espinosa, V., Andreu-García, G., Valiente-González, J. M., Estruch, V., Ordóñez, P., Pérez-Arjona, I., Atienza, V., Mèlich, B., de la Gándara, F., and Santaella, E. (2019). Automatic Bluefin Tuna (Thunnus thynnus) biomass estimation during transfers using acoustic and computer vision techniques. *Aquacultural Engineering*, 85:22–31.

[163] Punt, A. E. and Kennedy, R. B. (1997). Population modelling of Tasmanian rock lobster, Jasus edwardsii, resources. *Marine and Freshwater Research*, 48(8):967.

[164] Remmers, T., Grech, A., Roelfsema, C., Gordon, S., Lechene, M., and Ferrari, R. (2024). Close-range underwater photogrammetry for coral reef ecology: a systematic literature review. *Coral Reefs*, 43(1):35–52.

[165] Ridgway, K. R. (2007). Long-term trend and decadal variability of the southward penetration of the East Australian Current. *Geophysical Research Letters*, 34(13):1–5.

[166] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351:234–241.

[167] Rubbens, P., Brodie, S., Cordier, T., Destro Barcellos, D., Devos, P., Fernandes-Salvador, J. A., Fincham, J. I., Gomes, A., Handegard, N. O., Howell, K., Jamet, C., Kartveit, K. H., Moustahfid, H., Parcerisas, C., Politikos, D., Sauzède, R., Sokolova, M., Uusitalo, L., Van den Bulcke, L., van Helmond, A. T. M., Watson, J. T., Welch, H., Beltran-Perez, O., Chaffron, S., Greenberg, D. S., Kühn, B., Kiko, R., Lo, M., Lopes, R. M., Möller, K. O., Michaels, W., Pala, A., Romagnan, J.-B., Schuchert, P., Seydi, V., Villasante, S., Malde, K., and Irisson, J.-O. (2023). Machine learning in marine ecology: an overview of techniques and applications. *ICES Journal of Marine Science*, 80(7):1829–1853.

[168] Sadovy, Y. (2005). Trouble on the reef: The imperative for managing vulnerable and valuable fisheries. *Fish and Fisheries*, 6(3):167–185.

[169] Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., and Harvey, E. S. (2016). Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography: Methods*, 14:570–585.

[170] Santurkar, S., Tsipras, D., Ilyas, A., and Madry, A. (2018). How does batch normalization help optimization? *Advances in Neural Information Processing Systems*, 2018-December(NeurIPS):2483–2493.

[171] Schoening, T., Bergmann, M., Ontrup, J., Taylor, J., Dannheim, J., Gutt, J., Purser, A., and Nattkemper, T. W. (2012). Semi-automated image analysis for the assessment of megafaunal densities at the Artic deep-sea observatory HAUSGARTEN. *PLoS ONE*, 7(6):1–14.

[172] Schramm, K. D., Harvey, E. S., Goetze, J. S., Travers, M. J., Warnock, B., and Saunders, B. J. (2020). A comparison of stereo-BRUV, diver operated and remote stereo-video transects for assessing reef fish assemblages. *Journal of Experimental Marine Biology and Ecology*, 524:1–12.

[173] Schroff, F., Kalenichenko, D., and Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823. IEEE.

[174] Seiler, J., Williams, A., and Barrett, N. S. (2012). Assessing size, abundance and habitat preferences of the Ocean Perch Helicolenus percoides using a AUV-borne stereo camera system. *Fisheries Research*, 129-130:64–72.

[175] Shields, J., Pizarro, O., and Williams, S. B. (2020). Towards Adaptive Benthic Habitat Mapping. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9263–9270. IEEE.

[176] Shortis, M. R., Ravanbakhsh, M., Shafait, F., and Mian, A. (2016). Progress in the automated identification, measurement, and counting of fish in underwater image sequences. *Marine Technology Society Journal*, 50(1):4–16.

[177] Siddiqui, S. A., Salman, A., Malik, M. I., Shafait, F., Mian, A., Shortis, M. R., and Harvey, E. S. (2017). Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. *ICES Journal of Marine Science*, 75(1):374–389.

[178] Sigler, M. F. (2000). Abundance estimation and capture of sablefish (Anoplopoma fimbria) by longline gear. *Canadian Journal of Fisheries and Aquatic Sciences*, 57(6):1270–1283.

[179] Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In *3rd International Conference on Learning Representations, ICLR*, pages 1–14.

[180] Singh, H., Armstrong, R., Gilbes, F., Eustice, R., Roman, C., Pizarro, O., and Torres, J. (2004). Imaging Coral I: Imaging Coral Habitats with the SeaBED AUV. *Subsurface Sensing Technologies and Applications*, 5(1):25–42.

[181] Skinner, K. A., Zhang, J., Olson, E. A., and Johnson-Roberson, M. (2019). UWStereoNet: Unsupervised Learning for Depth Estimation and Color Correction of Underwater Stereo Imagery. In *International Conference on Robotics and Automation (ICRA)*, pages 7947–7954. IEEE.

[182] Smale, D. A., Kendrick, G. A., Harvey, E. S., Langlois, T. J., Hovey, R. K., Van Niel, K. P., Waddington, K. I., Bellchambers, L. M., Pember, M. B., Babcock, R. C., Vanderklift, M. A., Thomson, D. P., Jakuba, M. V., Pizarro, O., and Williams, S. B. (2012). Regional-scale benthic monitoring for ecosystem-based fisheries management (EBFM) using an autonomous underwater vehicle (AUV). *ICES Journal of Marine Science*, 69(6):1108–1118.

[183] Smith, R., Self, M., and Cheeseman, P. (1988). A stochastic map for uncertain spatial relationships. In *Proceedings of the 4th international symposium on Robotics Research*, pages 467–474. MIT Press.

[184] Soviany, P. and Ionescu, R. T. (2018). Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction. *Proceedings - 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2018*, pages 209–214.

[185] Spampinato, C., Giordano, D., Di Salvo, R., Chen-Burger, Y.-H. J., Fisher, R. B., and Nadarajan, G. (2010). Automatic fish classification for underwater species behavior understanding. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams - ARTEMIS '10*, pages 45–50. ACM.

[186] Spyksma, A. J. P., Miller, K. I., and Shears, N. T. (2022). Diver-generated photomosaics as a tool for monitoring temperate rocky reef ecosystems. *Frontiers in Marine Science*, 9(July):1–19.

[187] Stoner, A. W., Ryer, C. H., Parker, S. J., Auster, P. J., and Wakefield, W. W. (2008). Evaluating the role of fish behavior in surveys conducted with underwater vehicles. *Canadian Journal of Fisheries and Aquatic Sciences*, 65:1230–1243.

[188] Sun, Y., Wang, X., and Tang, X. (2013). Deep convolutional network cascade for facial point detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3476–3483.

[189] Tan, M., Langenkämper, D., and Nattkemper, T. W. (2022). The Impact of Data Augmentations on Deep Learning-Based Marine Object Classification in Benthic Image Transects. *Sensors*, 22(14).

[190] Thornton, B., Bodenmann, A., Pizarro, O., Williams, S. B., Friedman, A., Nakajima, R., Takai, K., Motoki, K., Watsuji, T. o., Hirayama, H., Matsui, Y., Watanabe, H., and Ura, T. (2016). Biometric assessment of deep-sea vent megabenthic communities using multi-resolution 3D image reconstructions. *Deep-Sea Research Part I: Oceanographic Research Papers*, 116:200–219.

[191] Tolimieri, N., Clark, M., Singh, H., and Goldfinger, C. (2008). Evaluating the SeaBED AUV for Monitoring Groundfish in Untrawlable Habitat. In *Marine Habitat Mapping Technology for Alaska*, pages 129–142.

[192] Tolimieri, N., Clarke, M., Clemons, J., Wakefield, W., and Powell, A. (2020). The abundance and habitat use of demersal fishes on a rocky offshore bank using the ROPOS remotely operated vehicle. *Deep Sea Research Part I: Oceanographic Research Papers*, 157:1–12.

[193] Torres-Pulliza, D., Dornelas, M. A., Pizarro, O., Bewley, M., Blowes, S. A., Boutros, N., Brambilla, V., Chase, T. J., Frank, G., Friedman, A., Hoogenboom, M. O., Williams, S., Zawada, K. J. A., and Madin, J. S. (2020). A geometric basis for surface habitat complexity and biodiversity. *Nature Ecology & Evolution*, 4(11):1495–1501.

[194] Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, 203(1153):405–426.

[195] Underwood, A. J. (1996). *Experiments in Ecology*. Cambridge University Press.

[196] van Gemert, J. C., Verschoor, C. R., Mettes, P., Epema, K., Koh, L. P., and Wich, S. (2015). Nature Conservation Drones for Automatic Localization and Counting of Animals. In Agapito, L., Bronstein, M. M., and Rother, C., editors, *Lecture Notes in Computer Science*, volume 8925 of *Lecture Notes in Computer Science*, pages 255–270. Springer International Publishing.

[197] Ventura, D., Bruno, M., Jona Lasinio, G., Belluscio, A., and Ardizzone, G. (2016). A low-cost drone based application for identifying and mapping of coastal fish nursery grounds. *Estuarine, Coastal and Shelf Science*, 171:85–98.

[198] Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 2018.

[199] Waddington, K. I., Meeuwig, J. J., Evans, S. N., and Bellchambers, L. M. (2010). Assessment of the benthic biota of a deep coastal ecosystem by remote and in situ sampling techniques. *Marine and Freshwater Research*, 61(10):1164.

[200] Wallace, J., Kok, H., Buxton, C., and Bennett, B. (1984). Inshore small-mesh trawling survey of the Cape south coast. part 1. Introduction, methods, stations and catches. *South African Journal of Zoology*, 19(3):154–164.

[201] Walther, D., Edgington, D. R., and Koch, C. (2004). Detection and tracking of objects in underwater video. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1.

[202] Warnock, B., Harvey, E. S., and Newman, S. J. (2016). Remote drifted and diver operated stereo-video systems: A comparison from tropical and temperate reef fish assemblages. *Journal of Experimental Marine Biology and Ecology*, 478:45–53.

[203] Wiens, J. A. (1989). Spatial Scaling in Ecology. *Functional Ecology*, 3(4):385.

[204] Williams, S., Pizarro, O., How, M., Mercer, D., Powell, G., Marshall, J., and Hanlon, R. (2009a). Surveying noctural cuttlefish camouflage behaviour using an AUV. In *2009 IEEE International Conference on Robotics and Automation*, pages 214–219. IEEE.

[205] Williams, S. B., Pizarro, O., Jakuba, M., and Barrett, N. (2010). AUV benthic habitat mapping in south eastern Tasmania. *Field and Service Robotics*, 62:275–284.

[206] Williams, S. B., Pizarro, O., Mahon, I., and Johnson-Roberson, M. (2009b). Simultaneous Localisation and Mapping and Dense Stereoscopic Seafloor Reconstruction Using an AUV. In *International Symposium on Experimental Robotics*, volume 54, pages 407–416.

[207] Williams, S. B., Pizarro, O. R., Jakuba, M. V., Johnson, C. R., Barrett, N. S., Babcock, R. C., Kendrick, G. A., Steinberg, P. D., Heyward, A. J., Doherty, P. J., Mahon, I., Johnson-Roberson, M., Steinberg, D., and Friedman, A. (2012). Monitoring of Benthic Reference Sites: Using an Autonomous Underwater Vehicle. *IEEE Robotics & Automation Magazine*, 19(1):73–84.

[208] Willis, T. J. and Babcock, R. C. (2000). A baited underwater video system for the determination of relative density of carnivorous reef fish. *Marine and Freshwater Research*, 51(8):755–763.

[209] Willis, T. J., Millar, R. B., and Babcock, R. C. (2000). Detection of spatial variability in relative density of fishes: Comparison of visual census, angling, and baited underwater video. *Marine Ecology Progress Series*, 198:249–260.

[210] Wilson, S. K., Fulton, C. J., Depczynski, M., Holmes, T. H., Noble, M. M., Radford, B., and Tinkler, P. (2014). Seasonal changes in habitat structure underpin shifts in macroalgae-associated tropical fish communities. *Marine Biology*, 161(11):2597–2607.

[211] Wilson, S. K., Graham, N. A., and Polunin, N. V. (2007). Appraisal of visual assessments of habitat complexity and benthic composition on coral reefs. *Marine Biology*, 151(3):1069–1076.

[212] Wynne, S. P. and Côté, I. M. (2007). Effects of habitat quality and fishing on Caribbean spotted spiny lobster populations. *Journal of Applied Ecology*, 44(3):488–494.

[213] Yang, G., Pennington, J., Rao, V., Sohl-Dickstein, J., and Schoenholz, S. S. (2019). A mean field theory of batch normalization. *7th International Conference on Learning Representations, ICLR 2019*.

[214] Yoshimura, K. and Motokawa, T. (2010). Bilaterality in the regular sea urchin Anthocidaris crassispina is related to efficient defense not to efficient locomotion. *Marine Biology*, 157(11):2475–2488.

[215] Young, M. A., Ierodiaconou, D., Edmunds, M., Hulands, L., and Schimel, A. C. G. (2016). Accounting for habitat and seafloor structure characteristics on southern rock lobster (Jasus edwardsii) assessment in a small marine reserve. *Marine Biology*, 163(6):141.

[216] Yousif, K., Bab-Hadiashar, A., and Hoseinnezhad, R. (2015). An Overview to Visual Odometry and Visual SLAM: Applications to Mobile Robotics. *Intelligent Industrial Systems*, 1(4):289–311.

[217] Zawada, K. J., Madin, J. S., Baird, A. H., Bridge, T. C., and Dornelas, M. (2019a). Morphological traits can track coral reef responses to the Anthropocene. *Functional Ecology*, 33(6):962–975.

[218] Zawada, K. J. A., Dornelas, M., and Madin, J. S. (2019b). Quantifying coral morphology. *Coral Reefs*.

[219] Zhang, J., Skinner, K. A., Vasudevan, R., and Johnson-Roberson, M. (2019). DispSegNet: Leveraging Semantics for End-to-End Learning of Disparity Estimation From Stereo Imagery. *IEEE Robotics and Automation Letters*, 4(2):1162–1169.

[220] Zhou, W., Berrio, J. S., Worrall, S., and Nebot, E. (2019). Automated Evaluation of Semantic Segmentation Robustness for Autonomous Driving. In *IEEE Transactions on Intelligent Transportation Systems*, pages 1–12. IEEE.

[221] Zyner, A., Worrall, S., Ward, J., and Nebot, E. (2017). Long short term memory for driver intent prediction. In *IEEE Intelligent Vehicles Symposium, Proceedings*, number Iv, pages 1484–1489. IEEE.