Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Enhancement, Summarization and Analysis of Underwater Videos of Nephrops Habitats

A dissertation submitted to the University of Dublin
for the degree of Doctor of Philosophy

Ken Sooknanan
Trinity College Dublin, November 2014

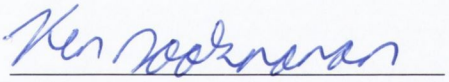
Signal Processing and Media Applications
Department of Electronic and Electrical Engineering
Trinity College Dublin

Declaration

I hereby declare that this thesis has not been submitted as an exercise for a degree at this or any other University and that it is entirely my own work.

I agree that the Library may lend or copy this thesis upon request.

Signed,



Ken Sooknanan

November 22, 2014.

To my family, friends and Bayesian Theorem.

Abstract

Harvesting the commercially significant lobster, Nephrops Norvegicus, is a multi-million dollar industry in Europe. Stock assessment is essential for maintaining this activity but it is conducted by manually inspecting hours of underwater surveillance videos. The motivation for this thesis is to improve this manual inspection process by exploring object recognition techniques for detecting these burrows automatically. As the visibility in these videos is generally poor, the recognition aspect of this system is combined with two additional preprocessing steps of image enhancement and content summarization. These three techniques are discussed separately in the thesis.

In the first step of image enhancement, the radial degradations (vignetting) associated with the illumination distribution of the light source and the absorption from water in these images are corrected. To perform this correction a novel image enhancement technique is developed that uses ideas from the vignetting and underwater correction literature. In this technique, a new degradation model is derived to take into account the spatial deteriorations in each colour channel that occur outside of the light beam footprint on the sea floor. Unlike current techniques in the vignetting literature, this model does not restrict the shape of the degradations to being circular and located at the image center, but instead follow a general elliptical shape and center of the light beam footprint. Novel techniques are also developed for estimating the parameters for the model, which use the attenuation from corresponding points across multiple frames. Correction is performed by attenuating pixel values according to the gain field parameterized by the model. When evaluated against a state of the art vignetting technique, the method achieves superior results.

In the content summarization chapter of this thesis, the tedious manual process that scientists endure by inspecting thousands of video frames is reduced to the scanning of a single image. This particular image is a mosaic that is created by aligning and rendering all of the video frames together. These mosaics are a useful analysis tool as they offer a wide area view of the surveyed seabed area, which makes it easy for scientists to spot spatial relationships among burrows. To align the video frames, a Bayesian framework for registration is developed that uses the burrows (blobs) in these images for feature matching. Once aligned, the overlapping regions among these frames are rendered with a new technique that uses the estimate of the light beam center to capture image details of well-lit regions in the generated mosaic. Experiments performed in this chapter show these new alignment and rendering techniques achieve improved results when compared to four state of the art systems from the literature.

The final algorithm presented involves identifying burrows automatically from the generated video mosaics. Using mosaics for this application improves on the existing video based technique by summarizing the results in a single image as opposed to inspecting thousands of video frames. Recognition in this system is performed by first detecting candidate objects and then classifying

them into burrow and non-burrow classes. For object detection, a novel segmentation technique is developed that targets the characteristic dark appearance of burrows. To decipher the burrows from the objects detected, a new feature set is developed for this application that is motivated by a current scientific description of Nephrop burrows. Two well established classifiers (KNN and SVN) are explored for burrow classification with this feature set. Experiments performed in this chapter show that the proposed system is more accurate and robust than a previous system developed by Lau et al.

Acknowledgments

As my PhD journey draws to a close, I would like to thank God almighty and the persons whom I had the pleasure of working with, such as my supervisors, members of the Sigmedia Lab, and various marine scientists. My supervisors included: Anil Kokaram, Naomi Harte (SB), David Corrigan, Gary Baugh, and James Wilson. Some of the marine scientists that I collaborated with include: Jennifer Doyle (Galway), Colm Lordan (Galway), Alessandro Ligas (Belfast), Adrian Weetman (Scotland), and other scientists throughout the world from SGNEPS. The past and present members of the SIGMEDIA group whom made it a friendly and enlightening working environment are: Francois, Andrew, Felix, Yunfeng, Ian, Aibhe, Eoin, Colm, Finnian, Roisin, Kangyu, Luca, Sam, Marcin, Liam, Natasha, Steven, Mohamed, Dan, Dee, Robbie, Nora, Brian, Frank, Jenny and Bernie (my Irish mother).

I am also grateful to the persons whom I met playing various sports in Ireland: football, squash, triathlon, surfing, tennis, running, pool, hiking, dancing, cycling, rifle, karate, American football, kayaking, and cricket. I enjoyed my weekly football matches with the lads from the department: Shane, John, Conor, Geoff, Sean, Liam, Cormac, Edna, Dave, Francois, and Solar etc. The on and off morning runs that I had with my German friend, Isabell was also quite enjoyable. My weekly squash matches were also great fun with the persons from the squash club: Elvy (best coach ever), Ana, Dervla, Tony, Robert, John, Mark, Isabell, Rhona, Dominic, Brian, Kevin, Peter and many more. The American football matches organized by the Kingstons were also great fun, thanks: Peter Kingston, John Kingston, Kate, Sid, Ronan, Naoise, and many others. Among all of these activities however, I enjoyed the cricket club the most, as these guys treated me as though I was family: Robert G., Peter K., Andy K., Jittin T. (who had a crush on Lucy C.), Kevin O., Ad K., Garin H., Nick R., Richard K., John R., Rohan K., Sid G., Steve D., Pranav G., Asrar K., Alok, Jennifer G., Jessica K., Rebecca O., Alanna C., Anchal, Ronan S., Naoise B., Sam J., Fin O., Cal M., Steve A., Nittin R., Mikey P., Amit C., Vishnu M., Luke J., Fred C., Max M., James W., Anil, K., Raj, Rajan, Hammad, and about 1000 more persons whom I had the joy of meeting during my 3 year duration as the 4ths captain at Trinity College.

I would like to thank my friends and family from back home in Trinidad for emailing and calling me on many occasions. Some of the friends whom kept in contact with me include: Sanjay B., Rayshad A., Marrisa B., Shiva Mahadeo, Shiva Maharaj, Arnold R., Trisha P., Akash P., Cathy R., and many more. Some of family members that supported me include: Auntie Pinky, Neil, Trevor, Dale, Karren, Hazroon, Mitra, Jorden, Kyle, Andy, Vicky, Moi and Scott.

I would like to express my gratitude to Science Foundation Ireland for funding my research, and my two examiners Noel O'Connor and Rozenn Dahyot for making my viva quite pleasant.

Lastly, I would like to give an extraordinary amount of thanks to my mother, Hazroon Sooknanan, for putting time aside from her busy schedule to call me every week and give me an endless supply of encouragement, understanding and support; you are magnificent mom!

List of Acronyms

DoG Difference of Gaussian

DWT Discrete Wavelet Transform

eSMEM error-based Split and Merge Expectation Maximization

EM Expectation Maximization

GMM Gaussian Mixture Model

GVF Gradient Vector Flow

LoG Laplacian of Gaussian

LSE Least Square Error

LTM Long Term Memory

mRNP messenger Ribonucleotide Protein

MAP Maximum A Posteriori

MRF Markov Random Fields

PDF Probability Density Function

PSF Point Spread Function

ROI Region of Interest

SMEM Split and Merge Expectation Maximization

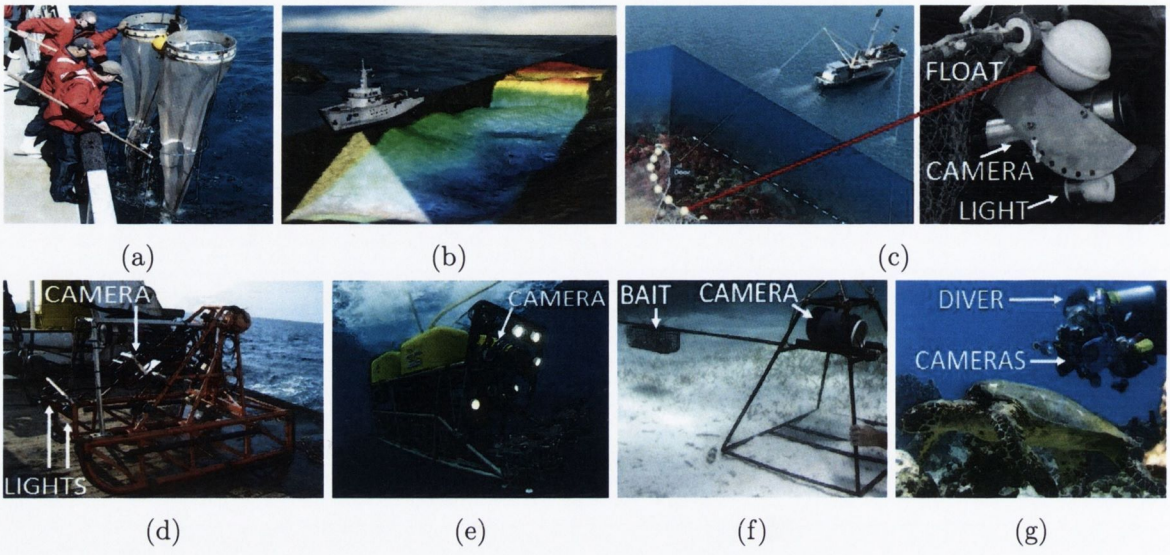
SVM Support Vector Machine

Contents

1 Introduction

The oceans cover about two-thirds of the surface of the earth and are home to most of the living organisms on the planet. These marine organisms help to regulate the climate as a result of their significant contributions to the oxygen cycle [14]. Apart from this vital role, the ocean is a valuable resource for many of our everyday demands such as food, medicine, raw materials e.g. oil and gas etc. To maintain and understand how these resources impact on critical matters such as climate conditions, scientists have been continuously performing oceanic studies. Like most scientific research, these studies are based on data collection and analysis. One of the effective methods utilized for collecting data on various oceanic habitats and their associated marine life, is the use of exploration surveys. The information gathered in these surveys is then used in a number of applications such as performing population censuses [12], archeology [69], geological mapping [11] [47], and assessing the biological environment [12].

There are three common types of exploration surveys: i) biological sampling, ii) acoustic, and iii) video. In biological sampling, nets are used to collect a variety of species at scattered locations, as seen in Figure 1.1 (b). In each haul, characteristics of the species caught, such as their quantity, weight, height, sex, and maturity etc., are recorded for stock assessment and management. In acoustic surveys (see Figure 1.1 (a)), echoes from sound waves are used for detecting moving organisms in the ocean, which are then physically sampled using trawl nets to verify the particular species. Although acoustics and biological sampling methods give acceptable results, they have significant drawbacks. One main drawback is the physical trauma and even death of the species caught during biological sampling encounter, which unfortunately can add to the decline of endangered species such as sharks [7]. Another drawback with these



(a) (b) (c) (d) (e) (f) (g)

Figure 1.1: a) Fish egg surveys that use fine mesh nets, b) acoustic seabed mapping survey, and examples of underwater imagery survey equipment that has camera and lights attached to c) Trawl Nets, d) Sled, e) ROV, f) stationary cage with bait, and g) a human diver.

methods is they cannot detect some species such as lobsters and crabs that mostly reside in burrows in the seabed.

To get around the drawbacks of biological and acoustic sampling methods, scientists began to use digital imagery as another form of data collection and analysis. However, because of the poor visibility underwater, capturing this type of imagery is not an easy task, and depends a lot on the type of data required and the terrain. For instance, when surveying large flat areas, like in Nephrops surveys [12], video recordings are made with cameras mounted on equipment such as sleds or trawl nets that are dragged along the sea floor, as shown in Figure 1.1 (c) and (d). If the terrain is however not flat, more robust options such as Remotely Controlled Vehicles (ROV) with cameras can be employed (see Figure 1.1 (e)), which are also useful for taking physical samples. Another from of capture is the use of divers with hand held cameras (see Figure 1.1 (f)), which are used in delicate terrain such as archaeological sites or coral reefs, to keep damage to a minimum. For dangerous scenarios such as capturing sharks however, cameras are mounted in protective cages, and left stationary on the sea floor with small pieces of bait to attract the respective species. A sample image of this bait set up is shown in Figure 1.1 (f).

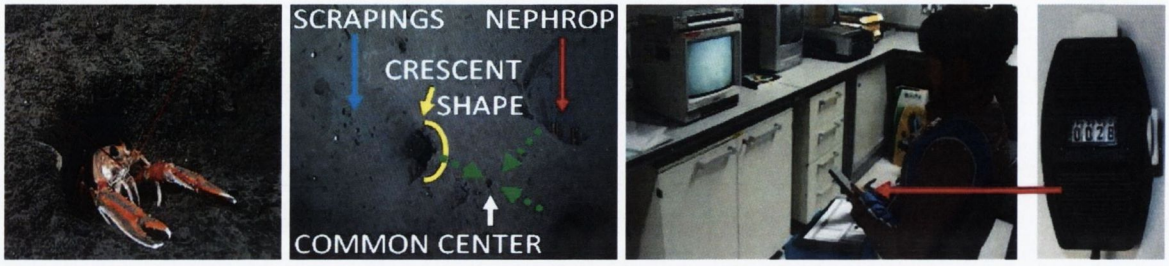
After the video surveys are completed, the recordings are then manually analyzed by scientists [12]. In some cases however, this analysis can be time consuming and tedious due to four main reasons. First, the recordings can be quite long, i.e. from a couple of minutes in archaeological surveys [69] to hours in shark surveys [7], which may cause user fatigue when reviewing. Secondly, the quantity of information to be extracted can also be quite large. For example, seabed mapping surveys are used only for observing sediment composition [47], which is

Figure 1.2: The main harvesting grounds (red dots) for Nephrops throughout Europe.

readily assessed, but for performing population census of Nephrops, users are required to count thousands of burrows [12]. The third reason why manual analysis can be quite tedious and error prone is due to the poor visibility conditions present in these videos, due to absorption and backscattering properties of the water medium [69]. Apart from the quantity and quality of the data, there is also the problem of resolving different results obtained from separate scientists, which can sometimes involve painstakingly repeating the entire analysis procedure.

To solve some of problems associated with manual analysis, research is being pursued in various areas of video DSP and computer vision [69] [47]. Some of these areas include i) image enhancement [69], ii) content summarization [26], and iii) content analysis [47]. An image enhancement technique being actively researched is correcting the colour degradations that occur in this environment due to the absorption from water molecules [69]. In the area of content summarization, tools are being developed to generate large area views or mosaics of the surveyed seabed area [26] [65], and automatically extract video clips when scientifically interesting events occur such as changes along the seabed [47]. For the area of content analysis, researchers are developing techniques to automatically detect specific items of interest from the survey videos such as mammals [51], lobsters [46], and crabs [55] etc.

This thesis is concerned with improving the manual video analysis procedures that are associated with performing population census of the particular lobster, *Nephrops Norvegicus* (hereafter referred to by genus alone). This species of lobster grows up to 25cm long, is pink-orange in colour (see Figure 1.3 (a)), and lives in burrows in muddy type seabeds throughout the Atlantic and Mediterranean Seas [52]. These burrows may have multiple entrances, in which case they are referred to as complexes [12]. Nephrops are commonly referred to as the Dublin Bay prawn, and are the most important commercial crustacean in Europe, with approximately 60,000 tons being



(a) (b) (c)

Figure 1.3: A Nephrop, along with its b) four characteristic features in a 3-burrow complex, which marine scientists use to identify and count them in c) the current manual analysis process with mechanical tallies.

caught annually [12]. To maintain this multi-million dollar [71] industry, underwater surveys are performed yearly to monitor the species stock at its main harvesting grounds throughout Europe, as shown in Figure 1.2. These surveys are performed by pulling sleds or trawl nets along the sea floor with high intensity lights and cameras attached. Using the recorded videos, stock assessment is then performed by marine scientists manually counting the species burrow complexes, with hand held mechanical tally counters, as shown in Figure 1.3 (c). However, identifying these complexes is not an easy task for scientists because of the following reasons.

1. The high intensity lights used in these surveys only provides uniform illumination within their beam footprint on the sea floor. Outside of this footprint region, the light degrades drastically (see Figure 1.5), making object recognition in these areas quite difficult.

2. The field of view in these videos is narrow (see sample frames in see Figure 1.6), which makes spotting spatial relationship among neighboring burrows tedious and error prone.

3. There are thousands of other burrows from other creatures which are not Nephrops. To discriminate between these various burrows, scientists search for features that are characteristic of Nephrops (see Figure 1.3 (b) for visual illustrations), such as:

   - Crescent shaped burrow entrances. Other species such as Calocaris Macandreae have burrow entrances that are usually circular in shape [52].
   - The presence of sediment ejecta or claw marks around the entrance, which the creature make when entering and exiting the burrow.
   - The presence of Nephrops, as they are very territorial in nature [52], their presence in particular burrows usually indicate their place of dwelling.
   - A pattern of multiple entrances focusing towards a common centrum. This arrangement is commonly referred to as a complex, and is formed as the creature creates several interconnected burrow entrances below the sea floor.

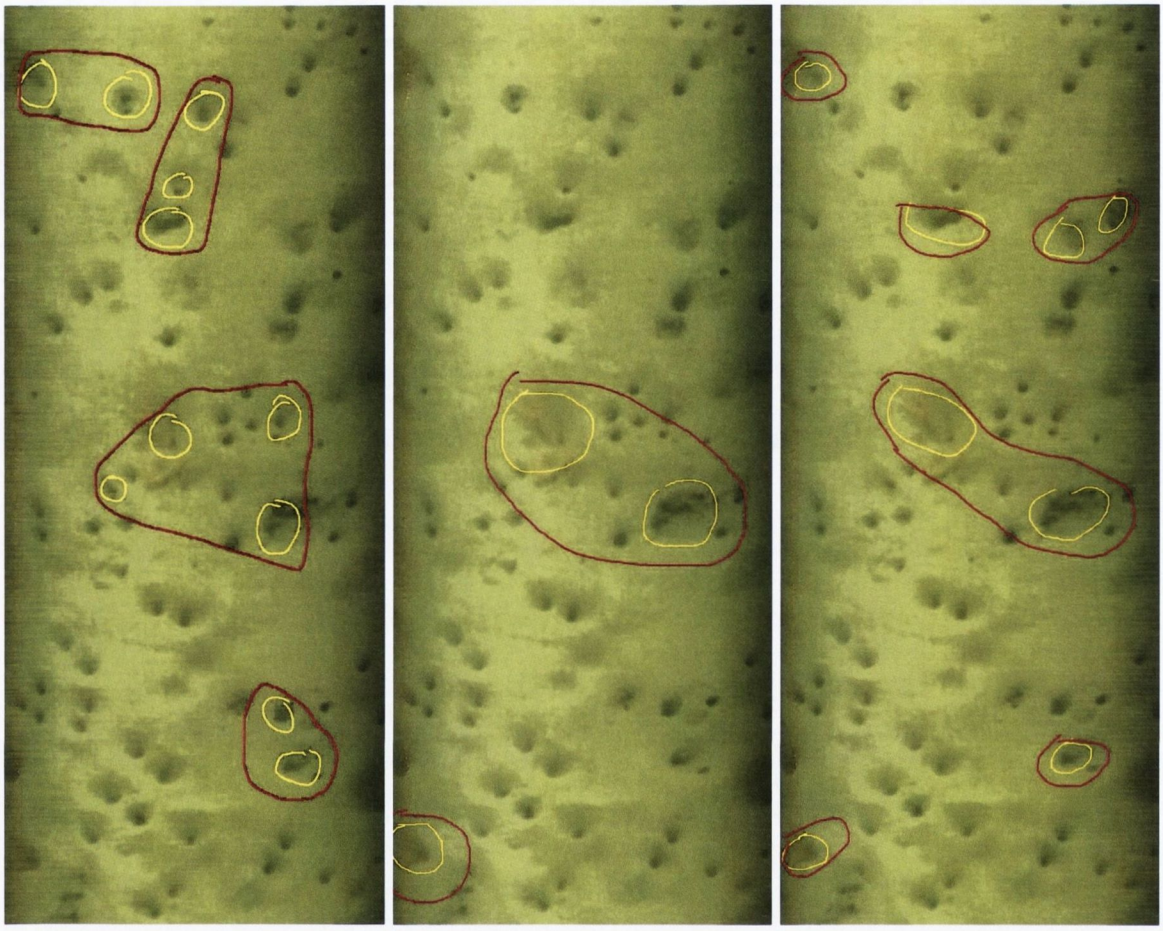


Figure 1.4: Manual selections of Nephrops burrows (yellow) and their corresponding complexes (red), obtained from scientists: (Left) Adrian Weetman from the Marine Laboratory in Scotland, (Middle) Alessandro Ligas from the Biosciences Institute in Belfast, (Right) Jennifer Doyle from the Marine Institute in Galway

4. Inter-assessor variability. Although the characteristic features of Nephrops burrows (as described above) do help in their identification [52], the procedure is still subjective among human assessors. This subjectivity often leads to discrepancies in the complex counts obtained from different scientists, which can be difficult to resolve for long sequences. An example of this discrepancy is shown in Figure 1.4, which is a underwater video mosaic with selections of Nephrops burrows (yellow) and their corresponding complexes (red) that was independently obtained from three scientists.

To improve this tedious and error prone manual analysis procedure for monitoring Nephrops habitats, two strategies are proposed in this thesis that use the underwater survey videos. The first strategy uses post-processing techniques to improve the visual quality of these videos, so

Figure 1.5: (Left) Original image from a seabed video, and (right) its corrected version using the image enhancement technique developed in this work.

that manual inspection can be performed more accurately. To accomplish this task, two novel applications are developed for performing image enhancement and content summarization. The image enhancement application corrects the illumination degradations in these videos, and the content summarization application then combines these corrected video frames to create a large area view or mosaic of the surveyed seafloor area. If however, the seafloor in the particular video is well illuminated, then the mosaics are generated from the original video frames, without performing any correction. These mosaics are useful analysis tools as their wide area view of the seabed make it easy for scientists to spot large spatial relationships among burrows.

The second strategy proposed in this thesis for improving analysis of Nephrops habitats involves the use of machine learning classification systems to automatically recognize the two most scientifically important objects in these videos, burrows and lobsters. This object recognition or content analysis application builds on the previous two visual quality improvement applications by automatically annotating the identified objects in the generated video mosaic. The key advantage of using mosaics to perform this annotation is that the results can easily be verified by scanning a single image. These annotations can also be used as a reference to easily resolve burrow count discrepancies among different scientists. The exact relationship among these three applications in relation to the original video sequence and their layout in the thesis is summarized in Figure 1.7. Sample results obtained from the image enhancement application are shown in Figure 1.5, and Figure 1.6 show results obtained from the summarization and object recognition applications.

These applications were developed in collaboration with the Marine Institute in Galway, Ireland, where Nephrops surveys are performed by pulling a sled along the sea floor attached with high intensity lights and cameras, as shown in Figure 1.1 (d). The cameras on these sleds are mounted approximately vertical to the seafloor, which result in mainly translational

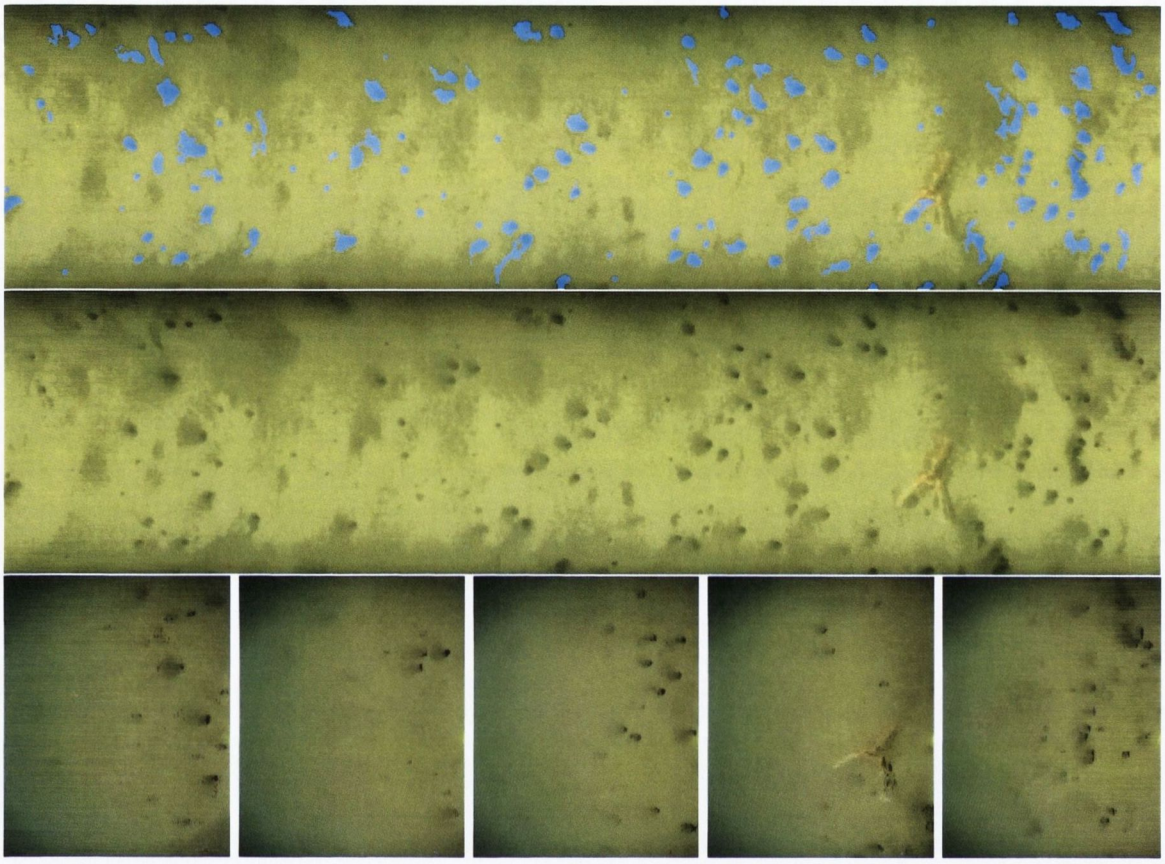


Figure 1.6: Samples of original video frames (bottom) used to generate the Mosaic in the middle, and the detected burrows (blue) shown in the top, obtained from the summarization and recognition applications developed in this work.

camera motion in the recorded videos. These types of survey videos are mainly analyzed in this work. The analysis is performed mainly using the scientifically accredited software, Matlab [54], which was installed on a desktop personal computer with 4.0 GB of RAM and a 2.33 GHz Intel processor. It should be noted that all of the applications developed in this research are created for offline usage, and hence computational efficiency is not of a critical importance.

An outline for the remainder of this thesis, along with a list of the major contributions and publications made with respect to the proposed visual quality improvement and object recognition strategies, are summarized in the following sections.

1.1 Thesis Outline

This thesis is comprised of four main sections. First, is a literature review on the visual quality improvement and object recognition strategies developed in this research. In the next two

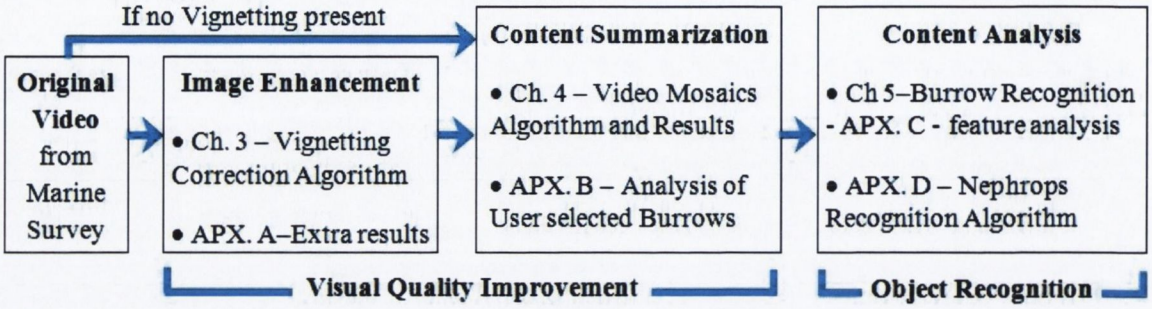


Figure 1.7: Relationship among the Image enhancement, Content Summarization, and Analysis applications developed in this work, and their layout in the thesis.

sections, details on the development, implementation, and experiments performed on these two strategies are presented respectively. The exact relationship and layout of these two strategies throughout the thesis are summarized in Figure 1.7. Lastly, conclusions drawn from the various experiments performed in this work are presented. Further details on each of these sections along with the main contributions made in each chapter/appendix are given below.

1.1.1 Literature Review

Chapter 2 : Enhancement, Summarization and Analysis of Video: A Review
A review of the literature relevant to the algorithms developed in this thesis is presented in this chapter. These relevant areas include image enhancement, content summarization, and object recognition. Also given in this review, are reasons for choosing the particular approaches used in this work.

1.1.2 Visual Quality Improvement Strategy

Chapter 3 : Improving Underwater Visibility Using Vignetting Correction
In this chapter the derivation of the new image enhancement technique is shown, that combines ideas from the vignetting [45] [91] and underwater correction literature [69]. This technique models the degradations in these types of images from the illumination distribution of the light source, absorption from the water medium, and vignetting from the camera lens. The parameters for this model are estimated using corresponding points from multiple frames. Once these parameters are estimated, a correction mask is created to amplify the degraded image details. This correction mask takes into account the minimal degradations that occur within the light beam footprint on the sea floor, by not amplifying the image details within this region. When evaluated, this new image enhancement method achieves improved results when compared to the state of the art vignetting correction technique introduced by Kim & Pollefeys [45].

Main Contributions:

1. A new degradation model is derived that
   (a) accounts for deteriorations from: i) the light source, ii) absorption from the water medium, and iii) vignetting from the camera lens.
   (b) does not restrict the shape and center of the degradations to being circular and centered at the image center
2. The procedure for estimating the parameters for this new model, which is a linear approach based on point correspondences.
3. The correction procedure, which takes into account the footprint region of the light beam on the sea floor where minimal or no degradation occurs. This procedure involves estimating the extent of the footprint region and incorporating it into a correction mask where pixels within this region are not boosted.

Appendix A : Supplementary Results for Chapter 3

In this section, additional results are presented from testing the proposed and the previous state of the art vignetting correction technique by Kim & Pollefeys [45] on marine survey videos with a large variety of seabed types. Analysis of these results show both algorithms enhance visibility in most cases, with the proposed method achieving substantially improved results to the previous technique by Kim & Pollefeys [45].

Chapter 4 : Mosaics From Marine Videos

In this chapter the algorithms developed for generating the wide area view or mosaic of the sea floor from the survey videos are given. These algorithms involve first aligning all of the video frames and then rendering their overlapping regions. For alignment, corresponding features (blobs) in each respective frame are matched. Once aligned, the overlapping regions are then rendered by using the vignetting center to select well lit regions from individual frames. When evaluated, the performance of the proposed method achieves improved results when compared to four state of the art mosaicking algorithms using synthetic and real data.

Main Contributions:

1. Improving the blob based image alignment technique developed by Matas et al. [53] by
   (a) using the difference of Gaussians image to robustly detect blobs in unevenly lit regions.
   (b) combining feature matching [9] and an exhaustive searching [79] method in the homography estimation procedure.
2. A new rendering technique that uses the estimated vignetting center in a weighting function to capture well lit image details in the generated mosaic.

3. A system to cross reference sections in the generated mosaics with the original video.

Appendix B : Analyzing the Selection of Nephrops Burrow Complexes from Different Scientists

This section investigates the usage of mosaics in the current Nephrops analysis procedure. It accomplishes this by examining the Nephrops complexes that were selected independently by three marine scientists using both video and mosaics. The examination shows there are significant inconsistencies among the users in the: i) video counts, and ii) the selected clusters in the mosaics, and iii) their corresponding burrow members. Two key observations are made from this investigation. First, is that the complex counts obtained from the mosaics are generally greater than the corresponding video, this possibly implies the improved visibility and field of view does help scientists to spot complexes more easily. Secondly, the inconsistencies among the counts highlight the selection of Nephrops complexes is error prone.

Main Contributions:

1. The Nephrops burrow complex selection procedure among scientists is error prone.
2. The improved visibility and field of view in mosaics does help scientists spot complexes more easily.
3. Mosaics can be used to identify the inconsistencies among the different scientists, which could improve the accuracy of the procedure in the future.

1.1.3 Object Recognition Strategy

Chapter 5 : Burrow Recognition Using Mosaics

Using the generated mosaics, burrow recognition is now performed. This chapter provides details of the various steps undertaken in creating and testing the proposed burrow recognition system. The main steps include object detection, feature extraction, and classification. Object detection is performed by targeting dark regions in the mosaic using segmentation and shape modeling techniques. Features are then extracted from these candidate regions, which are then used to classify them. Two supervised learning schemes, a k-Nearest Neighbor (KNN) and a Support Vector Machine (SVM), are used to perform this classification. The performance of this system shows improved results when compared to the state of the art burrow detection technique developed by Lau et al. [46].

Main Contributions:

1. Using mosaics to perform burrow recognition, which improves visibility, and simplifies the tedious video inspection process to the browsing of a single image.
2. A novel object detection procedure that
   (a) uses the difference of Gaussians image to detect objects robustly in unevenly lit areas

(b) targets the dark contrasting characteristic of burrows

(c) uses segmentation [6] and shape modeling [60] procedures to obtain the entire object region

3. A new feature set for burrow recognition that is motivated by its current scientific description [36]

4. The use of supervised classification schemes (SVM, KNN) for this application. These schemes use training data from a large variety of burrow and non-burrow objects found in these videos to aid in the classification process.

Appendix C : Supplementary Analysis for Chapter 5

To optimize the supervised machine learning algorithms utilized in chapter 5 for automatic burrow recognition, an optimum set of features have to be selected. This selection was made from among the newly developed set in this work, and the existing set proposed by Lau et al. [46]. To accomplish this task, two feature selection procedures are examined: i) PCA (Principal Component Analysis [18]), and ii) exhaustive search (i.e. examining all possible feature combinations [18]). The analysis using PCA is given in this section, along with the performance results (recall, precision and classification error) from some of the feature combinations from the exhaustive search procedure. These two sets of results are used to complement the full feature selection discussion given in chapter 5.

Appendix D : Detecting Nephrops Using Mosaics

This section outlines the preliminary work that was performed for automatically identifying the actual Nephrops creatures in the generated video mosaics. To accomplish this task, a recognition system is developed (similar to the one in chapter 5 for burrow recognition) with three main steps of object detection, feature extraction, and classification. Object detection is first performed by targeting regions in the mosaics with the same bright pink-orange colour characteristics of Nephrops using a novel Bayesian segmentation technique. Features are then extracted from these candidate regions, which are then used to classify them. Two supervised learning schemes, a k-Nearest Neighbor (KNN) and a Support Vector Machine (SVM), are used to perform this classification. The performance of this system shows improved results when compared to the state of the art Nephrops detection technique proposed by Lau et al. [46].

Main Contributions:

1. Using mosaics to perform Nephrops recognition, which simplifies verifying the results to the browsing of a single image.

2. A novel object detection procedure that

(a) targets objects with the same bright pink-orange colour of Nephrops.

(b) uses a segmentation [6] scheme to obtain most of the object regions.

3. A new feature set for Nephrops recognition that is motivated by their current scientific description [36].

4. The use of supervised classification schemes (SVM, KNN) for this application. These schemes use training data from a large variety of Nephrops and non-Nephrops objects found in these videos to aid in the classification procedure.

1.1.4 Conclusions

Chapter 6 : Conclusions
In the final chapter of the thesis, the conclusions made from the various experiments performed in the image enhancement, summarization and object recognition chapters are highlighted, and a discussion on future work is given.

1.2 Publications

The work developed in this thesis was published/presented at international conferences, Study Groups, and in scientific reports as listed below.

1.2.1 Conference Papers

[**76**] K. Sooknanan, A. Kokaram, G. Baugh, J. Wilson, N. Harte and D. Corrigan. Improving Underwater Visibility Using Vignetting Correction. In *Visual Information Processing and Communication III at IS&T/SPIE Electronic Imaging Conference 2012*, Burlingame, California, USA, February 2012, pp. 1 - 8.

[**78**] K. Sooknanan, A. Kokaram, G. Baugh, J. Wilson, N. Harte, and D. Corrigan. Indexing and Selection of Well-Lit Details in Underwater Video Mosaics Using Vignetting Estimation. In*IEEE International Conference on Oceans (OCEANS'12)*, Yeosu, Republic of Korea, May 2012, pp. 1 - 7.

[**77**] K. Sooknanan, A. Kokaram, J. Doyle, J. Wilson, N. Harte, and D. Corrigan. Mosaics For Burrow Detection in Underwater Surveillance Video. In *IEEE International Conference on Oceans (OCEANS'13)*, San Diego, California, USA, September 2013, pp. 1 - 6.

1.2.2 Study Group Presentations

1. Oral presentation at "Study Group on Nephrops Surveys (SGNEPS 2012)", Acona, Italy, 2012.

2. Oral presentation at "Study Group on Nephrops Surveys (SGNEPS 2013)", Barcelona, Spain, 2013.

1.2.3 Scientific Reports

1. Report of the Workshop and training course on Nephrops burrow identification, International Council for the Exploration of the Sea (ICES) Living Resources Committee, H. C. Andersens Boulevard 44-46, DK-1553 Copenhagen V., Denmark, March 2012.

2. Report of the Workshop and training course on Nephrops burrow identification, International Council for the Exploration of the Sea (ICES) Living Resources Committee, H. C. Andersens Boulevard 44-46, DK-1553 Copenhagen V., Denmark, November 2013.

2

Enhancement, Summarization and Analysis of Video: A Review

With the relentless increase in digital media, manual analysis can sometimes overwhelm the user. To assist with this problem, research is being pursued in the area of video content analysis (VCA). In this area, algorithms are explored for detecting objects and events of interest in video. The automated results obtained from these systems are usually presented in a summarized form to the user such as key frames [48], and video skims [74]. The detection process is usually performed by mapping low-level visual features such as colour, motion, and shape to the particular high level event or object to be detected. In some circumstances however, this mapping is not fixed. For example, the colours extracted from an image when an event such as a goal is scored, depends on multiple factors such the viewing position of the camera. Thus, some human interpretation is needed when trying to describe these high-level semantic events using low-level features. This is what is known as the semantic gap [67], and is one of the main obstacles faced when building VCA systems to detect high-level semantic events or objects. To overcome this issue, the objects of interest and their corresponding search domains are restricted in most high level VCA systems.

These restrictions are observed in the VCA systems of many popular broadcast sports. For example, to extract baseball and snooker highlights, authors Chang et al. [63] and Rea et al. [67] developed algorithms to search for specific items among the broadcast videos of the corresponding sport. In these two cases low level features such as colour and texture are statistically modeled using Hidden Markov models to determine the particular camera view and corresponding high level events of interest. Sections of the video where these detected events

occurred are then presented as video skims.

Detecting objects and events of interest in underwater video can be more challenging than in other environments because of the poor visibility in this environment. To solve this issue most VCA systems geared towards underwater video employ image enhancement techniques. For example, to recognize specific objects from underwater survey videos, Bazeille et al. [4] first corrected the colour degradations associated with this environment due to the absorption from the water medium. Then a colour-based segmentation procedure is used for detecting objects which are then classified based on their size and colour features.

Following these general ideas of: i) incorporating image enhancement, and ii) restricting the search item to a specific domain, a VCA system geared towards underwater video is proposed in this work. Specifically, the system is to search for burrows from underwater survey videos of Nephrops habitats. To help explain the design choices made in creating this system, a brief overview on the current burrow analysis procedure is now given. Afterwards, the relevant literature with regards to image enhancement, summarization and object recognition will be given. In each of these reviews, examples concerned with underwater and non-underwater videos are discussed.

2.1 Burrow Analysis

Nephrops Norvegicus is the most valuable lobster in Europe, with estimated annual landings of some 60,000 tons [12]. This species of lobster grows up to 25cm long, is pink-orange in colour and lives in burrows in muddy type seabeds throughout the Atlantic and Mediterranean Seas [52]. These burrows may have multiple entrances, in which case they are referred to as complexes [12], examples of which are shown in Figure 2.1. Evidence of the structure of these complex systems was researched by Marrs et al. [52]. In their research, visual inspections were conducted by divers, and also a resin was poured into some of these complexes to verify their interlinking tunnels beneath the sea floor. Examples of some of these resin structures obtained in their research are shown in Figure 2.2

To maintain the multi-million dollar [71] industry involved with the harvesting of this species of lobster, underwater surveys are performed yearly to maintain its stock. These surveys are performed by pulling sleds or trawl nets along the sea floor with high intensity lights and cameras attached. Using the recorded videos, stock assessment is then performed by marine scientists counting the species burrow complexes manually. Scientists perform this process by clicking mechanical tally counters while inspecting the captured video playing at its recorded speed of 25 frames per second, on an 18 inch television screen.

Identifying these complexes from the videos is not an easy task for three main reasons. First, the visibility in some of these videos are very poor due to the uneven illumination distribution from the light sources used in these surveys, as shown in Figure 2.3. Secondly the range of view in these videos is very narrow, which makes it difficult to spot spatial relationships among

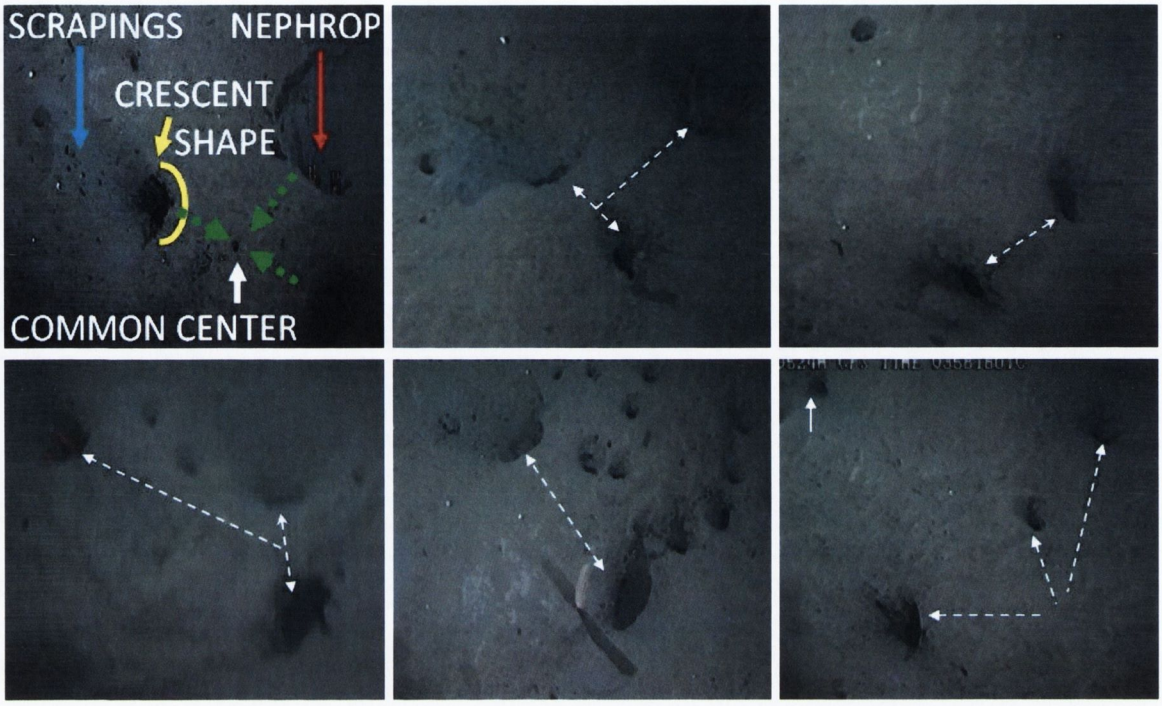


Figure 2.1: Samples of different Nephrop complex systems. The white dashed-arrows indicate the burrows belonging to the respective complex. (images obtained from the Marine Institute in Galway)

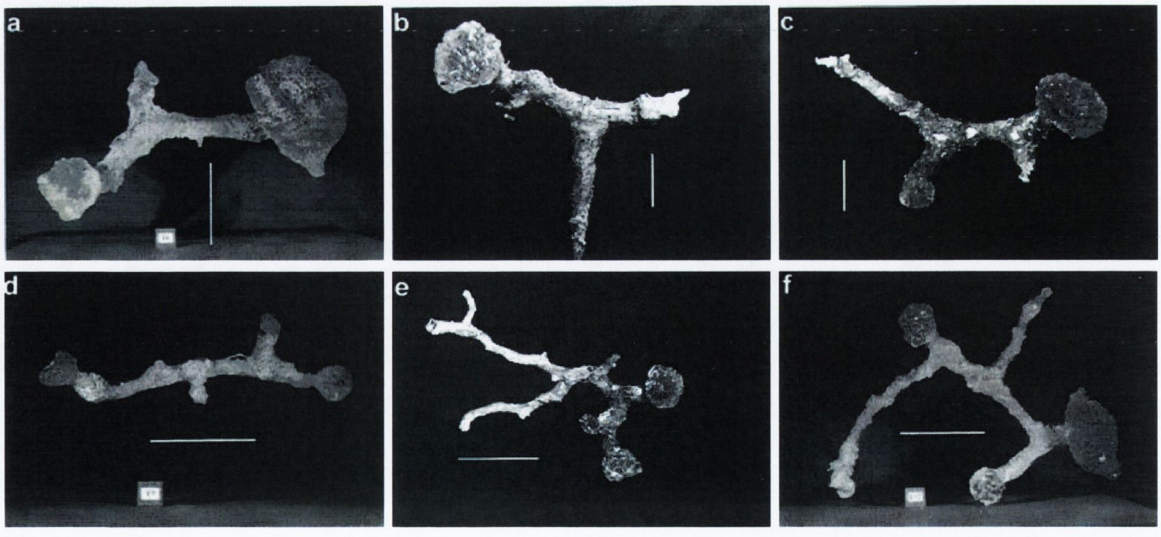


Figure 2.2: Samples of resin structures obtained from different Nephrop complex systems. (images obtained from the Marine Institute in Galway)

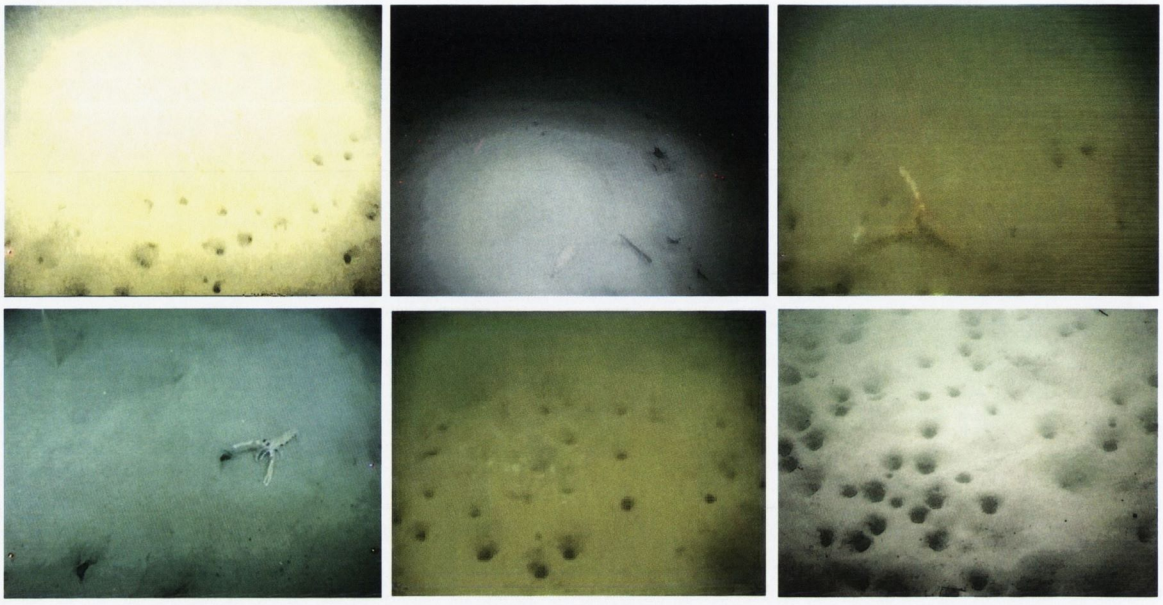


Figure 2.3: Sample frames from different survey videos of resin Nephrop habitats. (images obtained from the Marine Institute in Galway)

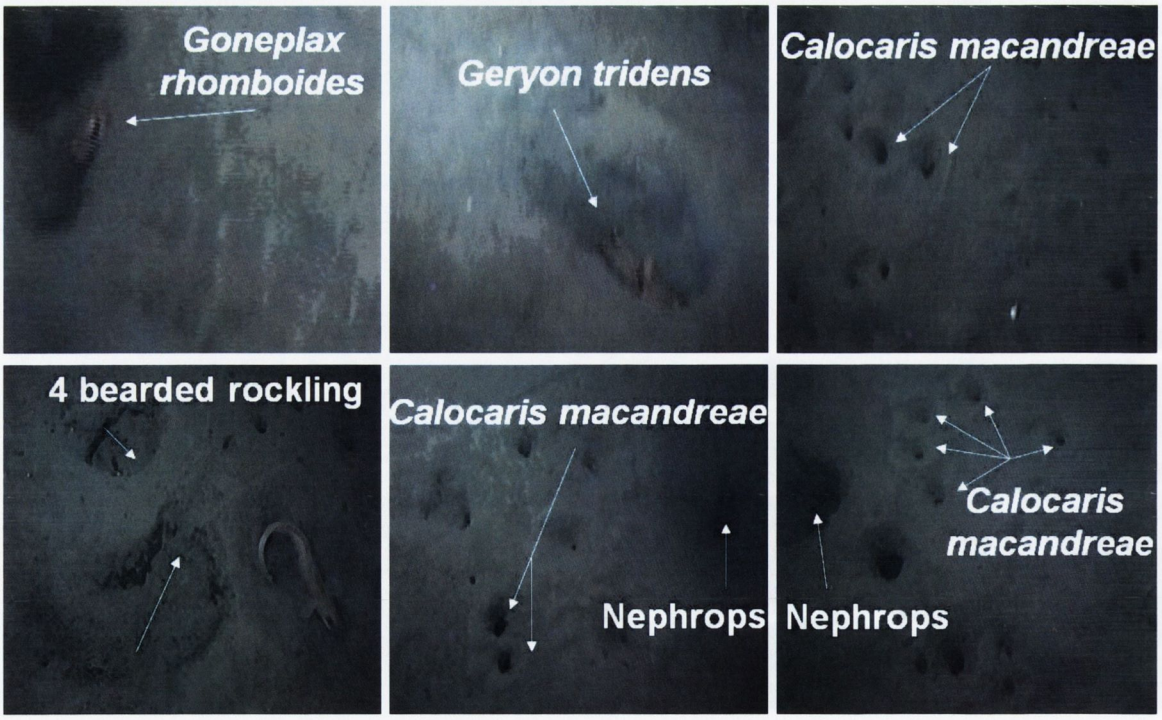


Figure 2.4: Samples of burrows of different species that are present in the survey videos of Nephrops habitats. (images obtained from the Marine Institute in Galway)

neighboring burrows. Finally, the presence of thousands of other burrows from different creatures in these videos. Sample burrows from some of the different species are shown in Figure 2.4. As seen, burrows of other species do resemble that of Nephrops, which makes this procedure very subjective.

To distinguish Nephrops burrows from other creatures, marine scientists search for four characteristic features:

1. Crescentic shaped entrances, other species such as Calocaris Macandreae have burrow entrances that are usually circular in shape [52].
2. Presence of sediment ejecta or claw marks around the entrance, which the creature makes when entering and exiting the burrow.
3. Presence of Nephrops themselves in the burrows.
4. A pattern of multiple entrances focusing towards a common centre.

These features are shown in the sample Nephrop complexes in Figure 2.1. Because of the difficulties involved in this analysis, the counting is performed by multiple scientists separately and then their counts are compared. If the error among these counts lie within an error of 20%, their average value is taken as the actual count, otherwise outliers are eliminated or the entire procedure is repeated [36]. One of the main issues faced by scientists in this procedure is they cannot verify if identical complexes are counted from a total tally count alone. In the past they tried to resolve this problem by using video annotation software, but it proved too tedious a process to label thousands of frames [36].

2.2 Image Enhancement

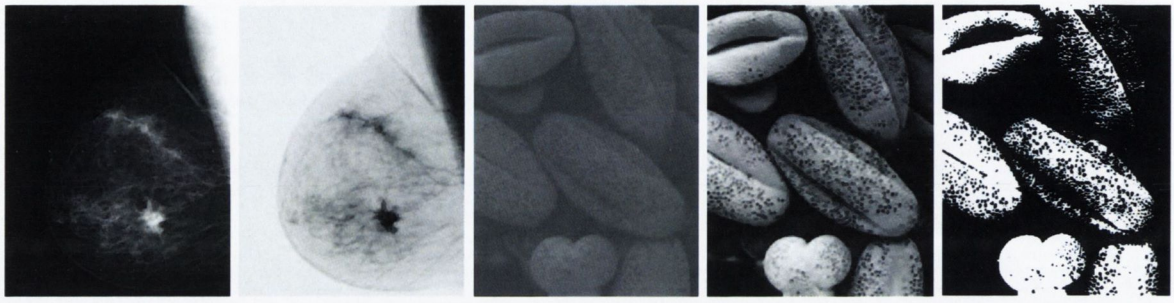
Image enhancement means improving the perception of information in images. It involves using a transformation, T to alter the pixel intensities at location $\mathbf{x}$ in the original image, $I(\mathbf{x})$ to values in the enhanced image, $G(\mathbf{x})$, given by:

$$G(\mathbf{x}) = T(I(\mathbf{x})) \tag{2.1}$$

The majority of the existing image enhancement techniques can be categorized into either frequency domain or spatial methods. In frequency domain methods, the image is first transformed into the frequency domain, altered and then transformed back to the image domain, whereas for spatial techniques, the raw image pixels are altered in the image domain.

2.2.1 Enhancement of Non-Underwater Images

For enhancing non-underwater images, some examples of frequency domain methods are denoising processes which use low and high pass filtering [29]. For spatial domain methods, explanations for three of the commonly used techniques are as follows.



(a) (b) (c) (d) (e)

Figure 2.5: Examples of image enhancement techniques. Original mammogram in (a) enhanced using its digital negative in (b). The original rice grain image in (c) enhanced using (d) histogram equalization, and (e) thresholding. (Images taken from [66])

Digital Negative [41]. Computed by taking the negative of an image: $N(\mathbf{x}) = 255 - I(\mathbf{x})$.

Thresholding [29]. Useful for quickly isolating objects from the background when their image intensities are distinctively different, as shown in Figure 2.5 (e). A threshold value accomplishes this task by:

$$G(\mathbf{x}) = \begin{cases} 1 & I(\mathbf{x}) > \text{threshold} \\ 0 & I(\mathbf{x}) \leq \text{threshold} \end{cases}$$

Histogram equalization [66]. This one of the most common techniques used for contrast enhancement. It works by scaling the image intensities with the transformation: $T(I) = (P - 1)c(I)$, where P is the number of gray levels and $c(I)$ is the normalized cumulative histogram of the original image.

Sample results from using these techniques are shown in Figure 2.5. In most cases, the choice of the image enhancement method utilized depends on the specific application. For example, to enhance medical mammogram images involves taking the negative of the image [66]. For taking everyday pictures, most cameras have built in histogram equalization functions, for contrast enhancement.

Unfortunately, for the underwater images of this proposed work, these global enhancement methods are not applicable in most cases. This is because of the radial degradations present in these images due to the illumination distribution of the light source on the sea floor. Correcting these spatial degradations may result in the colour space of the corrected image being distorted when using techniques such as histogram equalization, as shown in Figure 2.6. In the vignetting literature [3, 24, 43, 45, 91] however, similar spatial degradations are corrected, as seen in Figure 2.7. In this body of work a spatial degradation model is used to correct these radial degradations, which can preserve the colour space in the corrected image, as shown in Figure 2.6. Because

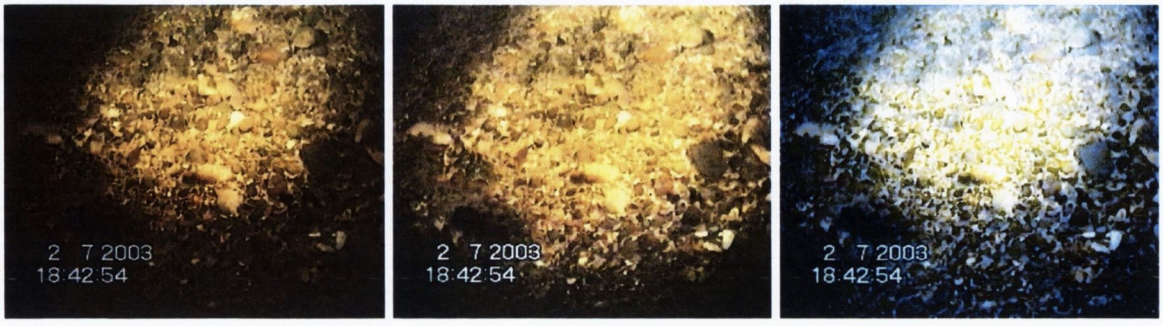


Figure 2.6: (Left) Original degraded underwater image, with the corresponding corrections obtained from the vignetting correction technique of Kim & Pollefeys [45] (Middle), and from using histogram equalization (Right).

of the close relationship this body of work has with the proposed work, a brief review of the literature in this area is now presented.

2.2.1.1 Vignetting

Vignetting is the radial fall off in image brightness from the image center that occurs in natural images. One of the main causes of this phenomenon is due to the refraction and absorption of light from the camera lens. These radial degradations, $T(\mathbf{r})$, can be corrected with the cosine fourth law [17], or the common polynomial spatial model given by:

$$T(r(\mathbf{x})) = 1 + \sum_{n=1}^{D} B_n r^{2n} r(\mathbf{x}) \tag{2.2}$$

where D and B_n are the number and value of the respective attenuation coefficients used, and $r(\mathbf{x}) = \sqrt{(\mathbf{x} - \mathbf{x_c})^2}/0.25\sqrt{(rows)^2 + (cols)^2}$, is the normalized radii at image position $\mathbf{x}$, with respect to the image center, $\mathbf{x_c}$. The two assumptions of this model are that the degradations are circular in shape and centered at the image center. Using this spatial model, the image formation process is given as:

$$G(\mathbf{x}) = f(kT(\mathbf{x})I(\mathbf{x})) \tag{2.3}$$

where $G(\mathbf{x})$ and $I(\mathbf{x})$ are the measured and undegraded image intensities, and $f()$ and k are the camera response function and exposure (i.e. shutter durations) setting. Thus, to correct vignetting the parameters, $T(r(\mathbf{x}))$, $f()$ and k should be estimated.

The most straight forward approach of estimating these parameters is to compare the degraded image with an identical image captured with uniform illumination. As it is difficult to capture real scenes this way, the degradations from a white piece of paper can be used [3, 43]. Using these two images, the parameters are then estimated using the intensity ratios between

their point correspondences at different radii locations. The main drawback of this approach is that it cannot be used to correct images captured from unknown cameras.

Zheng et al. [91] developed a more general approach for correcting the vignetting with the use of a single image. Their method involves first segmenting regions in the image with similar colour and texture properties. Then using the intensity gradient along these segmented regions at different radii to estimate the vignetting parameters. As perfect segmentations might not occur initially due to the dark regions on the image periphery, it is performed iteratively with the correction procedure until convergence occurs. Although good results are obtained from this technique as shown in Figure 2.7, it is not suited for underwater videos because the illumination between these frames is sometimes very unstable which may result in erroneous segmentations.

As there are multiple images of the same scene in a video, instead of using region segmentations, vignetting can be removed by analyzing the attenuations of overlapping regions from different frames [24, 45]. In this approach the intensities of point correspondences from overlapping regions of images taken at different exposures are used to recover the vignetting and camera response functions. Using this information, Goldman and Chen [24] introduced a non-linear approach for estimating these parameters simultaneously, while Kim & Pollefeys [45] came up with a linear solution to estimate them separately. They accomplished this by decoupling the vignetting and response estimation procedures by using the intensity ratio of the point correspondences. Once decoupled, the response function is estimated using a weighted least squares method [42] with corresponding image intensities at the same radii, taken at different

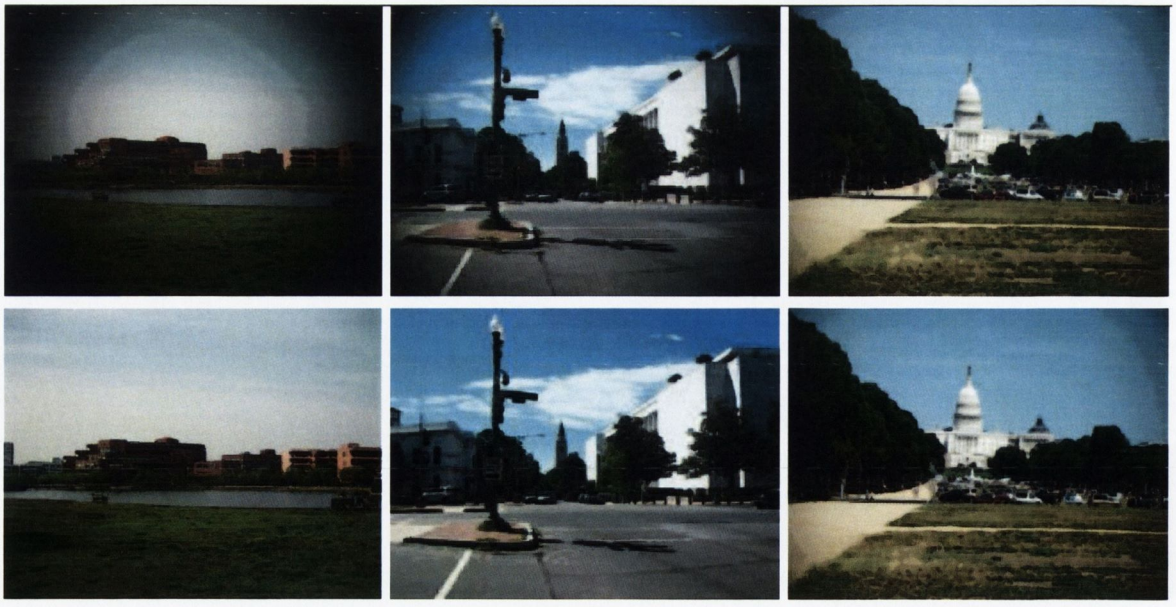


Figure 2.7: (Top row) Degraded images and (Bottom row) and corrections obtained using the technique developed by Zheng et al. [91]. (Images taken from [91])

exposures. Then, the vignetting function is recovered in a similar manner using corresponding image intensities at different radii.

In general, although these vignetting techniques are robust, they have two main drawbacks when it comes to correcting the radial degradations in underwater images. First is they do not account for the colour degradations that occur in this environment. Secondly, their model assumes the degradations are circular in shape and centered at the image center, which may not be the case. In these underwater images the shape and center of the degradations follow the illumination distribution of the light source on the sea floor as seen in Figure 2.6.

With the some of the most relevant methods used for enhancing natural images now discussed, a review of the techniques used in the underwater literature is now given.

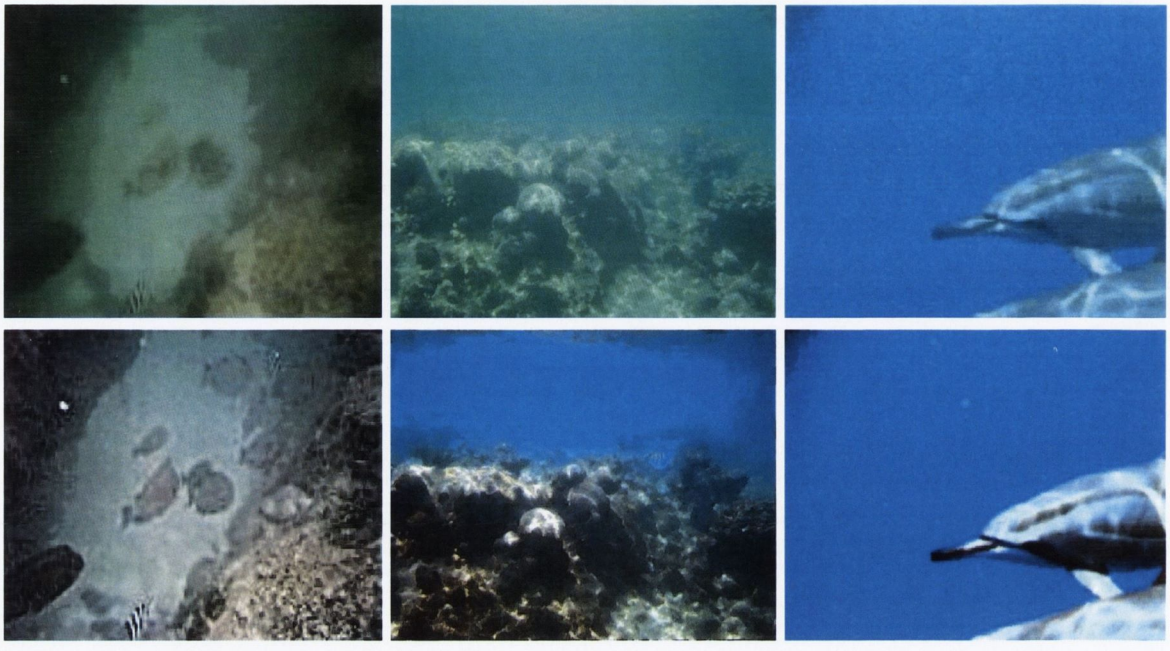
2.2.2 Enhancement of Underwater Images

For underwater imagery some of the techniques used to enhance natural images may not be applicable. This is because of the additional degradations that occur in this environment due to the absorption from the water medium. These degradations are not uniform across the visible spectrum, and can hence distort the colour space in the captured images. The red channel in particular is absorbed more than the blue and green, which is why a blue/green contrast is observed in most underwater images [8]. These colour imbalances depend on various factors such as the distance of the respective objects from the camera, the illumination distribution of the light sources, and the organic composition of the water itself [56].

It is important to correct these various degradations for applications such as underwater robot tracking [4], seabed mapping [47], 3D reconstruction [70], and coral health analysis [34]. There are even applications in the defense sector such as detecting sea mines [81], where it is important to correct these colour degradations. Most of the existing techniques for enhancing underwater images can be grouped into two categories: Non parametric, and those that use a degradation model. A review of the existing methods in each of these categories are now presented.

2.2.2.1 Non-Parametric methods

in this first category of methods, the degradations due to the absorption of water is not catered for directly. Instead, enhancement is performed using global and local techniques such as histogram equalization [21] and colour mapping [84]. For example, Garcia et al. [21] introduced a frequency based method called homomorphic filtering for enhancing underwater images. In their method unwanted low frequency components at the image periphery are removed by high pass filtering [29]. Then the contrast of the resulting image is enhanced with a similar technique to histogram equalization [66]. Other authors such as Bazeille et al. [5] verify this approach works well with their test set, samples of which are shown in Figure 2.8 (a). The main drawback from this approach is the colour space in the corrected image can become distorted, as shown



(a) (b) (c)

Figure 2.8: Examples of corrections obtained using Non-Parametric methods. (Top row) Captured underwater images and the corrections obtained using the techniques developed by (a) Garcia et al. [21], (b) Torres-Mendez [84], and (c) Iqbal et al. [38]. (Images taken from [21,38,84]).

in Figure 2.6.

This colour distortion can be prevented to a certain extent by restricting the colour ranges of the corrected images with known values. Torres-Mendez [84] performed this task by mapping the colour space from known image patches onto the degraded images. This is a supervised learning method that uses a Markov Random Field (MRF) belief propagation system to perform the mapping. While impressive results are obtained from their test cases, as shown in Figure 2.8 (b), it has a drawback in that the user has to manually select the training images.

To escape the tedious process of selecting training images manually, Iqbal et al. [38] came up with an unsupervised learning method. In this method the colour ranges of the corrected image are limited to the ranges from the captured images. With this limitation, correction is performed using a global contrast enhancement technique similar to histogram equalization [66]. To compensate for colour imbalances during the enhancement process, the ranges for the red and green channels are set to the limits of the blue channel, where degradations are assumed to be minimal. In spite of the good results obtained from their test images, as shown in Figure 2.8 (c), colour distortions can still possibly occur in the corrected images when the ranges for the red and green channels are naturally very different from that of the blue channel.

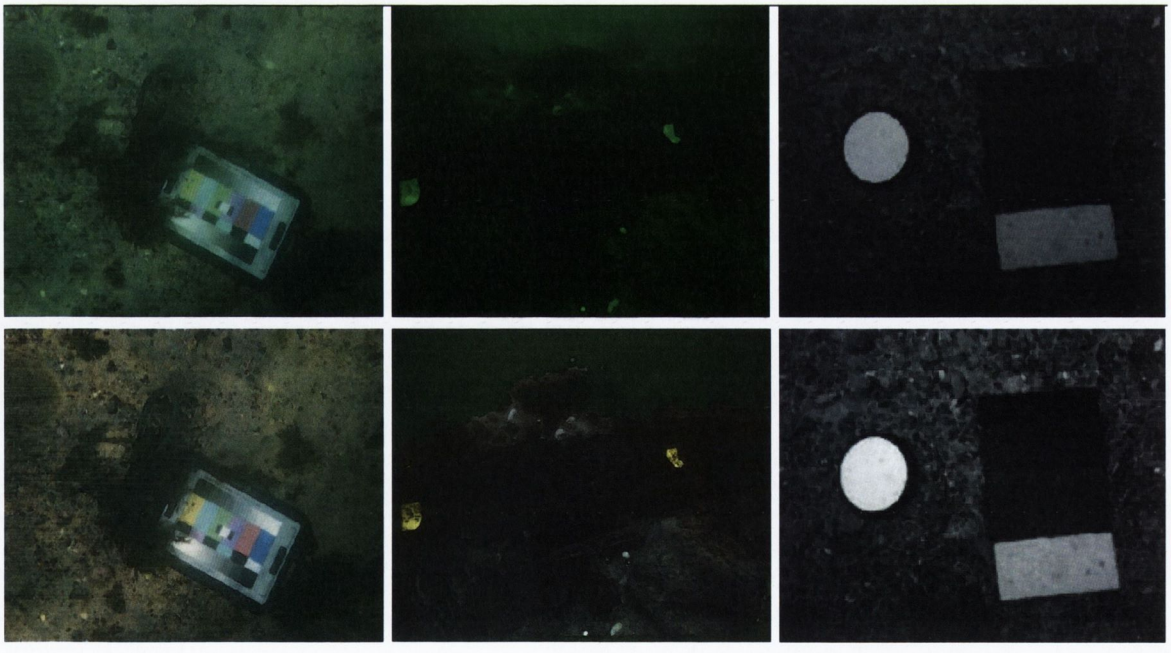
2.2.2.2 Model-Based Techniques

To correct the colour imbalances more effectively, many authors model the illumination deteriorations that occur in this environment due to the absorption from the water medium [8, 69, 70]. The model used in most cases is the Beer-Lambert law [56], given as:

$$G_\lambda(z(\mathbf{x})) = I_\lambda(z_o(\mathbf{x})) \exp -(n_\lambda z(\mathbf{x})) \tag{2.4}$$

Using the above expression, the undegraded image, $I_\lambda(z(\mathbf{x}))$ is recovered from the captured image $G_\lambda(z(\mathbf{x}))$, with knowledge of the depth $z(\mathbf{x})$ between the camera lens and each object at image location $\mathbf{x}$, and the attenuation coefficient, n_λ, for each colour channel, $\lambda = \{R, G, B\}$.

Although these methods achieves reasonable results, as shown in Figure 2.9, the attenuation coefficients, n_λ, first need to be estimated before they can be used. This is accomplished with the use of either special equipment, and/or user input. Examples of this equipment include the use of special polarizer lens [69], and spectrometers [1, 8]. Also, depth estimates with a-priori knowledge of the colour of specific objects such as the McBeth colour charts [8], limestones [70], and colour plates [1] are used. Another drawback of these methods is that they assume uniform illumination and degradation for all objects at the same depth. This is not true in some cases where the illumination degradations also depend on the spatial location of the object with respect



(a) (b) (c)

Figure 2.9: Examples of corrections obtained using Model-based methods. (Top row) Captured underwater images and the corrections obtained using the techniques developed by (a) Bongiorno et al. [8], (b) Sedlazeck et al. [70], and (c) Ahlen et al. [1]. (Images taken from [1, 8, 70]).

to the footprint of the light beam on the sea floor, as seen in Figure 2.6.

2.2.3 Scope for a new enhancement model

An overview of some of the existing enhancement techniques used for natural and underwater images have been presented in this section. The limitations of these techniques with respect to enhancing underwater images fall into three categories. First, are those that use non-parametric methods, which cannot account for the degradations that occurs in this environment due to absorption from the water medium [21, 38, 84]. As a result of this drawback, the colour space in corrected images can be distorted using these methods. The second category of enhancement techniques use the Beer-Lambert law to model and correct these degradations. The main drawback of these methods is they do not account for the spatial deteriorations due to the illumination distribution of the light source on the sea floor. Lastly, are the vignetting correction techniques that can cater for these spatial deteriorations to a certain extent, but do not take into account the colour degradations that occur due to absorption from the water medium.

The limitations of these techniques show there is need for a degradation model that caters for the spatial degradations due to the illumination distribution from the light source, and also absorption from the water medium. In this work, these limitations are addressed with the derivation of a new degradation model. Additionally, this new model improves on the existing techniques reviewed in three ways.

1. The shape and central location of the deteriorations are not restricted to being circular and centered at the image center, as in existing vignetting techniques [24, 45, 91]. Instead, the degradations are allowed to have an elliptical shape with central location at the center of the light beam footprint on the sea floor.

2. Depth estimates for performing correction or special equipment for initial calibration are not required as in existing underwater techniques [1, 8, 70]. Instead the system uses the attenuation from point correspondences to estimate all of the respective model parameters. These parameters are continuously updated throughout the sequence to account for degradation changes that may occur due to height variations of the light source.

3. It can account for the footprint region of the light beam on the sea floor where minimal or no degradation occurs. It accomplishes this task by estimating the extent of the footprint region and then incorporating it into a correction mask where pixels within this region are not boosted.

Further details on this new degradation model, along with performance comparisons with the state of the art vignetting technique by Kim & Pollefeys [45], will be given in the next chapter.

2.3 Content Summarization

Automatic summarization involves locating and extracting the most important or relevant information in a sequence and presenting it in a simple and concise manner for the user to understand. Good summarization can save users valuable time in applications with vast amounts of data. Similar to the layout of the previous section, a review of some of the techniques used for summarizing natural video is first given, followed by a review of techniques used for underwater video.

2.3.1 Summarization of Natural Video

For summarizing natural video, there are two main techniques: key frames [93] [48], and video skims, which are described as follows.

2.3.1.1 Key Frames

In the key frame technique a single [93] or multiple frames [95] are extracted from each shot in the video, which is considered to best represent it. Here a video shot is an unbroken sequence of frames captured from a single camera [48]. Some authors such as Liu et al. [48] argued that key frames are the most simple and effective method of summarizing video. In their method it is assumed the most relevant frames show a pattern of motion acceleration followed by deceleration. Using the motion vectors from MPEG B-frames to calculate the average motion magnitude and direction, consecutive frames showing this pattern are grouped together. From this group a key frame is then selected as the frame with the maximum motion.

2.3.1.2 Video Skims

A video skim is a collection of the most essential video segments (with matching audio) from the original sequence. They are commonly referred to as a summary sequence [74]. Generating video skims is not straightforward, as an understanding of the footage is required in order to extract the most essential parts. In other words this procedure is domain dependent. For example, to extract the essential baseball and snooker highlights, authors Chang et al. [63] and Rea et al. [67] developed algorithms to search for specific items among the broadcast videos of the corresponding sport. In these two cases low level features such as colour and texture are statistically modeled using Hidden Markov Models to determine the particular camera view and corresponding high level events of interest.

2.3.2 Summarization of Underwater Videos

Most of the literature [22, 46, 47, 55, 64] show underwater survey videos are summarized with respect to important scientific data and events. For example, Lebart et al. [47] extracted video clips showing when the composition of the seabed changed. In their method changes in seabed

composition are mapped to when large colour and texture variations occur in the video. For colour the mean intensities in red and blue channels are used, and for texture the variance of the gray-level gradient is used. These features are monitored using k-means clustering along 11×11 image patches of the video frames, and seabed changes are detected when a large change in their centroid values occur.

Another application where summarization is useful is analyzing the behavior of sea creatures such as lobsters [64]. Pons et al. [64] introduced a method for accomplishing this task from a static scene using background subtraction [32]. In their method a background image is first created from the static scene using pixels whom intensity values vary within a defined margin. Then, objects are detected in each frame after they are subtracted from this background image, by thresholding the image intensities. Video frames which contain objects with a significant size are classified as interesting frames for the Marine biologist to further analyze.

For analyzing hours of marine surveillance videos, automatic analysis and summarization is also very useful. Gebali et al. [22] developed an algorithm to search for salient events in these videos by analyzing the temporal changes in image intensities. Once a large change in these temporal values is detected the relevant video frames are extracted and a video abstract is created. Other techniques for summarizing the content of these videos involve searching for specific objects of interest such as crabs [55] and lobsters [46]. Mehrnejad et al. [55] developed an algorithm to detect crabs by searching for objects with similar colours. Once detected the number of these objects is displayed to the user for further analysis. Lau et al. [46] searched for lobsters in these videos based on there characteristic bright appearance and size features. Their method of summarization involved the number of lobsters detected and an indexing system showing the respective video frames where they can be found.

Another useful summarization method is the use of video mosaics. A video mosaic is a large image created by combining all of the video frames. This combination is performed by first aligning all of the respective frames on a common coordinate system, and then rendering their overlapping regions. A mosaic is a useful analysis tool for scientists for these survey videos as it provides the user with a wide view of the entire area captured in the video. It is used in such applications as seabed mapping [47] and navigation [84], and various aerial [39], satellite [39], and underwater surveys [26] [65].

2.3.3 Choice of Summarization Method

From the techniques discussed for summarizing natural and underwater videos, mosaics are chosen for use in this application for two main reasons. First, a wide area view of the seabed would be a useful analysis tool for scientists, as these videos suffer from a small field of view due to them being recorded in close proximity to the sea floor. Secondly, as burrows are located in most of the image frames, accurate verification of the results obtained from the automated system would entail tediously scanning over the entire video. But as the generated mosaic

contains all of the burrows present in the video, this tedious process is reduced to the scanning of a single image. The improved visibility seen when comparing the example mosaic with original frames in Figure 2.10, illustrate the advantages of using this method of summarization for this application. Because of the relevance of this particular summarization method, a review of some of the existing techniques is now given.

2.3.4 Mosaicking Techniques

Early mosaic generation was performed manually [62] [33]. In these cases the images are few, and motion is mainly translational so that overlapping regions can be easily identified and aligned by visual inspection. Once aligned, the overlapping region from the respective frame with the best quality is selected and placed in the respective section in the existing mosaic. High quality mosaics can be created this way, but the motion has to be mainly translational so that users can manually cut and paste the overlapping regions easily. This manual procedure can become quite tedious if there are many images.

To improve on this manual process, automated techniques were developed that used feature [9, 26, 65] or pixel [73] matching procedures to align the images. A pixel matching method was developed by Shum et al. [73], where patches from consecutive frames are to align then alignment. Then, the overlapping regions are combined using a weighted average scheme. The main drawback of this approach is that the radial degradations that occur in underwater video sequences can cause alignment errors when matching pixels only.

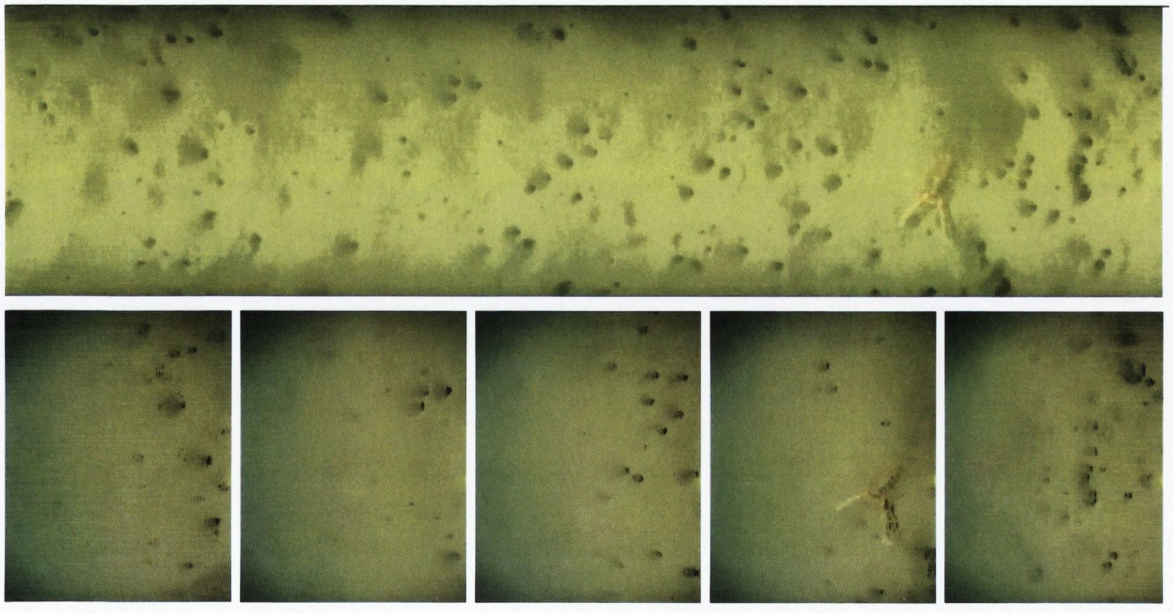


Figure 2.10: (top) Mosaic generated from 200 frames, samples of which are shown in bottom row.

A more robust solution for underwater images would be to use feature matching for alignment. Gracias et al. [26] matched corner features [30] for image alignment and combined the overlapping regions by using the median intensity value. Qing-Zhong et al. [65] also matched these features for image alignment, but for rendering the overlapping regions they used a weighted average method. These approaches have two main drawbacks when applied to unevenly lit images. First, the corner features [30] are not fully robust to illumination variations, and thus the registration accuracy of the system may degrade under these circumstances. Secondly, using statistical techniques such as the median and mean values might not capture the best image details among the overlapping regions.

These alignment and rendering issues are addressed to some extent by the technique proposed by Brown et al. [9]. In their method SIFT features [50] are used for matching, which are robust to a degree of illumination variations. For rendering, a weighting function is used to select sections of the overlapping regions from different frames [9,62]. For this method of combination, selection is made from the particular frame where the overlapping regions is located closest to the image center [9], as it is here where the best image quality is perceived to exist. The selected regions are then blended [10] into the existing mosaic to eliminate any image seams. This method can produce good results in most cases, but still has two weak points. First, for underwater survey videos of the sea floor, the best image details are located within the well lit regions of the frame, which is not necessarily the image center. Secondly, for very blurry images it is difficult to extract SIFT features.

2.3.5 Scope for new work

An overview of some of the existing mosaicking algorithms for combining non-underwater and underwater images has been presented. For image alignment, these algorithms can be grouped into two categories: those that use pixel matching and those that match features. For unevenly lit scenes, feature matching can be more robust than global pixel matching schemes. A limitation of using corner and SIFT features in these algorithms is that they are not accurately extracted from blurry images.

Likewise, for rendering image details from overlapping regions, the algorithms presented here can be grouped into two classes: those that use statistical methods, and those that select sections from different frames. The statistical methods which involve using the median or a weighted average value from the entire set, may not capture the best image details from unevenly lit sequences. The other class of rendering which involves selecting sections from different frames could capture good image details, but the sections would have to be selected from well lit regions.

The limitations of these algorithms presented here show the alignment and rendering stages needed to generate high quality mosaics from underwater survey videos can be improved using two key contributions. First, to cope with image blur, the use of matching object regions as opposed to matching points might be a more robust approach. Secondly, to improve the

rendering process for these unevenly lit sequences, the position of the light beam on the sea floor can be used for selecting good image details from overlapping frames.

Figure 2.10 shows a section of a mosaic generated from 200 frames of an actual underwater Nephrops survey video, using the proposed mosaicking algorithm with these two key contributions. As can be seen, visibility is improved tremendously, and good data is captured. Further details of this proposed mosaicking algorithm along with more results and comparisons to previous state of the art methods are given in chapter 4.

2.4 Object Recognition

Object recognition in computer vision is the task of locating and identifying specific objects in images or video. Early attempts of object recognition mainly used geometric models that comprised of simple shapes such as lines and circles [57]. These models would predict a 3D model from its 2D projection using edge information. But, due to large view-point and illumination changes, the edge information required for these systems could not be reliably extracted, which limited their scope. A comprehensive review on these early geometric-based recognition systems can be found in the article written by Mundy [57].

Advancement of feature descriptors and pattern recognition techniques [50] solved some of the view-point and illumination problems faced by geometric based systems, and led to appearance based recognition systems [82] [59]. In these systems, images are searched exhaustively (using every pixel) for patches that match a template of the respective object. An example of this type of recognition system is the fast template matching technique developed by Omachi et al. [59]. In this technique, the image is first roughly searched for matching patches of the template at different locations, widths and heights. Then, if similar patches are found, a more detailed match is then performed between them and the corresponding template at the respective size. Although in some circumstances appearance based methods give good results, their main downfall is that a large quantity of templates are needed to robustly represent an object under different circumstances such as changes in light, colour, shape, size etc. To solve the template representation problem faced by appearance based methods, feature based approaches were developed. In these approaches objects are first detected and then recognized based on their characteristic features.

Similar to the layout of the previous section, a review of some of techniques used for object recognition in natural video is first given, followed by a review of techniques used for underwater video.

2.4.1 Object Recognition in Non-Underwater Video

In the non-underwater video literature, some of the techniques utilized in video surveillance and rock identification applications are found to be closely related to the proposed work. The

respective techniques from each application are described below.

2.4.1.1 Video Surveillance

This application involves monitoring video of a static scene for moving objects. It is widely used for security applications [32], and for analysis of marine animals [64]. In most cases moving objects are detected using a background subtraction technique [32] [64], where a background (reference) image of the static scene is first created and then subsequent frames are subtracted from it. The background image can be created using techniques such as Gaussian Mixture Modeling [32], or by capturing the pixels that vary within a defined margin [64]. After background subtraction is performed, moving objects can be detected using edge detection [32] [64]. The idea from these applications of using background subtraction to highlight regions of interest will be incorporated into the proposed detection procedure. In detail, an approximate homogeneous sandy background image is first created by heavily blurring the original mosaic, then after subtracting it from a lightly blurred version, candidate burrow regions are highlighted. Figures 2.11 (a-d) illustrate this procedure. As outlined in [29], there exist many other segmentation techniques that could be applied to these images for object detection such as thresholding, edge detection, and clustering etc. However, this background subtraction method is chosen for this application as it performed effectively on the data sets used in this research. More details on this object detection technique are given in Chapter 5.

2.4.1.2 Rock Identification

This application is widely used for analyzing space exploration images captured from Mars [19], and for navigation purposes [13]. On a sandy background, as shown in Figure 2.12 (a), dark coloured rocks resemble the situation of burrows on the muddy sea floor. The obvious difference between these two objects is burrows are below the ground level while rocks are above. This key difference is used for rock detection in some algorithms [20]. Other techniques used for detecting candidate rock regions are the use of edges [13], or segmentation based on similar textural patterns [19], [16]. These candidate regions are then classified as rocks based on their extracted shape features such as their best-fit ellipse [19], or peak values from a circular Gabor filter [16]. The idea from some of these applications of using the characteristic circular shape of rocks for their recognition will be explored in the proposed feature selection process. Specifically, a circularity-fit feature is developed for examining the general shapes of burrows.

2.4.2 Object Recognition in Underwater Video

In the water literature, some of the techniques utilized in recognition of marine species applications are found to be closely related to the proposed work. The respective techniques are described below. After, a review of the existing burrow recognition application is given.

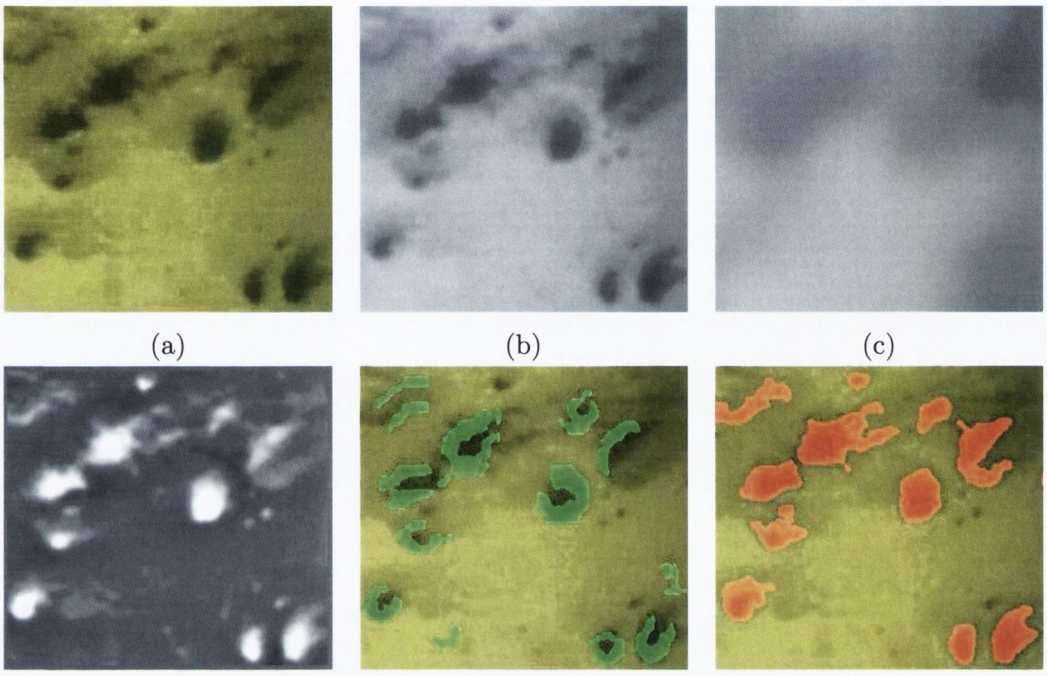


Figure 2.11: Object detection comparison. a) Original, d) targeted dark regions obtained by subtracting a lightly blurred grayscale image in (b) from a heavily blurred (background) version in (c), and candidate objects detected using technique by: (e) Lau et al. [46], and (f) this work.

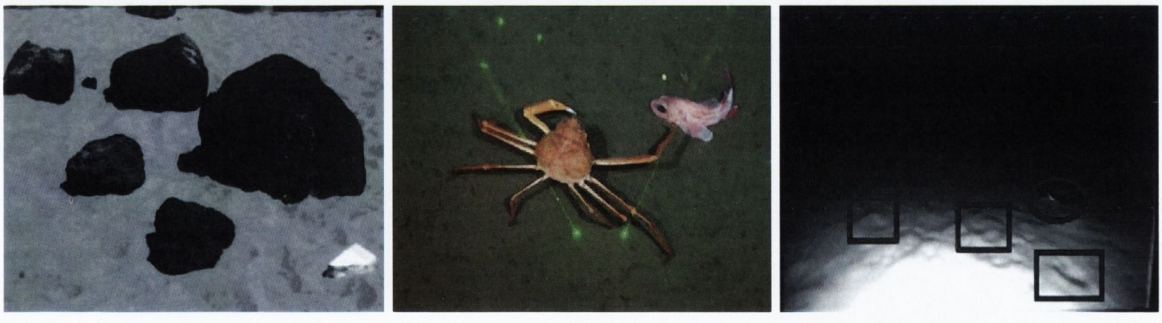


Figure 2.12: Sample images from related objection applications involving (a) Rocks [13], (b) Crabs [55], (c) Lobsters (circle) and Burrows (squares) from Lau et al. [46].

2.4.2.1 Recognition of Marine Species

Apart from the importance of burrow detection for Nephrops assessment [12], attempts have been made to identify actual marine animals such as mammals [51], lobsters [46], and crabs [55]. Sample images utilized in some of these applications are shown in Figures 2.12 (b) and (c). These algorithms involve two main steps: detection and classification. Candidate regions are first detected by targeting characteristic features of the particular species such as their brightness [46], colour [55], and shape [51]. The common method utilized for targeting characteristic features is thresholding [46] [55] [51].

In the second step of these algorithms, the candidate regions detected are classified using features such as size [46] and length [51], or in some cases matched against a template [55] [51]. The idea of using characteristic features of the particular item of interest for its initial detection, will be incorporated into the proposed detection process. Specifically, the dark contrasting regions that are characteristic of burrows in these type of videos will be targeted.

2.4.2.2 Recognition of Marine Burrows and Lobsters

The exact problem explored in this thesis of recognizing burrows from underwater videos of Nephrops habitats has been tackled recently by authors Lau et al. [46]. In their approach, candidate burrow and lobster objects are first detected directly in each video frame using edges, and then classified. A decision tree framework is used to perform this classification using size, shading, and textural features. The visual challenges associated with the uneven lighting in these videos, as seen in Figure 2.12 (c), are addressed by performing object detection in a gray scale contrast space, where the influence of absolute brightness has minimal effect. This space is created in a similar manner to the background subtraction method described in the video surveillance section above, by taking the difference of two differently blurred gray scale versions of the respective image. The results obtained from this system are summarized in two ways. First, the number of the respective objects detected throughout the entire sequence is stated. Secondly, for the user to verify these results, they are highlighted in each frame using a Graphical User Interface, as shown in Figure 2.13.

2.4.3 Scope for new work

Although acceptable results are obtained with the previous approach by Lau et al. [46], it has three main drawbacks. First, the use of edges for detecting objects not only produces incomplete segmentations, but also detects many other objects on the sea floor, as seen in Figure 2.11 (e). Secondly, using a strict set of rules to perform classification may have worked well for their data set, but might not be applicable to other data sets. Lastly, verifying their video based results via their GUI still involves the tedious manual inspection of thousands of frames. These drawbacks show their recognition procedure can be improved.

The system proposed in this thesis attempts to solve these drawbacks using four key con-

Figure 2.13: GUI from Lau et al. [46] burrow recognition system. (Image taken from [46])

tributions. First, mosaics are explored for detecting and summarizing the automated results, which would reduce the time spent tediously inspecting thousands of frames to the scanning of a single image. Secondly, a novel object detection method is developed to specifically target dark burrow-like regions. In this method, segmentation and shape modeling techniques are employed, to capture most of the object regions. An example of the complete segmentations obtained with this method in comparison to that from the previous method by Lau et al. [46], is shown in Figures 2.11 (e) and (f). The third contribution is a new feature set for this application that is motivated by a current scientific description of Nephrop burrows [36]. As a result of these features meaningful descriptions marine scientists easily relate to them and can use their values to further analyze the data. Lastly, to get around the problem of using strict rules for classifying objects with a large diversity in size and shape features, supervised learning classification schemes are explored. These schemes use training data from a large variety of burrow and non-burrow objects found in these videos to aid in the classification process, and can be updated with new data to adopt to most situations.

As the proposed system uses a feature based classification system, brief literature on it is given in the next section. In this review the specific choices made in the design of this system

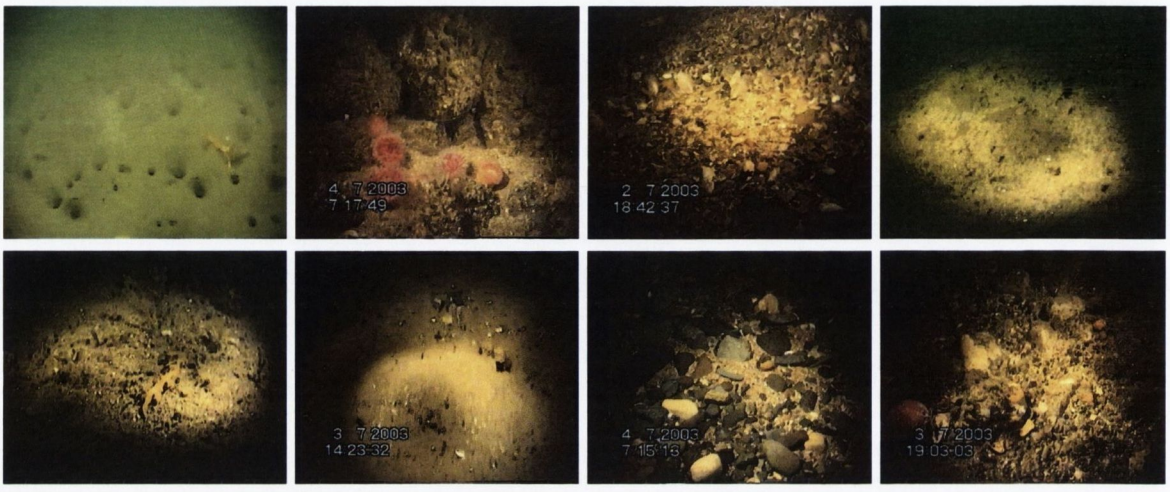


Figure 2.14: Sample images from the various videos collected for this work.

are explained.

2.4.4 Feature Based Classification Systems

For this application, because it would be impossible to represent all of the different shapes and sizes of the various burrows found on the sea floor using previous techniques such as template matching [59], a feature based approach is used. These approaches usually involve a five stage design cycle of: i) Data Collection, ii) Object Detection and Grouping, iii) Feature Choice and Extraction, iv) Classification Model Selection, v) Training and Evaluation, and vi) Optimization. A short background on each of these stages together with the choices made for this work, is now given.

2.4.4.1 Data Collection

Collection of sufficiently sized and suitable dataset is the keystone of any classification or object recognition algorithm. The dataset should encompass the potential variability of the data in the real world. Having a representative dataset allows the engineer to obtain an intuition about what features and classification algorithms may be suitable for the given task. For the burrow counting work discussed in this thesis, 22 2000-3000 frame sequences taken from surveys of Nephrop habitats form the dataset. For the testing of the image enhancement algorithm, 7 additional videos are used. These videos are ideal for testing as they have a wide range of: i) degradations, ii) colour, and iii) texture properties, due to their different seabed sediments such as sand, rocks, shells, clay etc. Sample images from some of these videos are shown in Figure 2.14.

2.4.4.2 Object Detection and Grouping

These first step in the operation of these systems is to detect a preliminary set of objects that will then be classified. The choice of the detection method, depends on the application. For example in surveillance applications involving a fixed background, background subtraction followed by thresholding is very effective for detecting foreground objects [32] [64]. For applications where the background is not fixed, such as detecting lobsters in underwater videos [46], detecting parts of the objects via their edges can be used. If the application requires detection of larger regions of the object, such as in rock detection [19], [16], segmentation techniques [72] based on characteristic features are usually used.

For this application, segmentation based on the characteristic dark appearance of burrows, is used for object detection. This technique is chosen for two reasons. First, these images are usually very blurry, and detecting parts of the objects with techniques such as edges might not be effective in some cases. Second, more scientifically important features can be extracted from the entire region, such as its burrow diameter, animal claw mark region, dark entrance area etc. This developed technique, as detailed in chapter 5, gives more complete segmentations when compared to using edges from the previous burrow recognition application [46], as shown in Figure 2.11.

2.4.4.3 Feature Choice and Extraction

The choice of distinguishing features is the next critical item for creating an efficient recognition system. The method utilized for extracting these features is also important, as it would determine if the system can identify the same object with variations in illumination, geometric, scale etc. Prior knowledge of the item of interest, the test set, and characteristics of other items in it, usually provide good clues for which features would be best suited. Examples of features used by different applications include colour features using mean RGB values for detecting seabed changes [47], textural features using wavelets for medical image retrieval [87], and shape features using best fit ellipses for rock detection [20]. For this application a new feature set is developed that is motivated by a current scientific description of Nephrop burrows [36]. Some of these features include the burrow diameter, animal claw marks, dark entrance area etc. Along with these new features, the existing features developed for this application by Lau et al. [46], are also examined. Further details of these features and their method of extraction are given in chapter 5.

2.4.4.4 Classification Model Selection

The task of the classifier component in the recognition system is to use the set of features and assign the particular object to a class. There are two main categories of classifiers, supervised and unsupervised. In unsupervised procedures [18], classifiers learn automatically from the inputed samples what categories to separate objects into, using techniques such as clustering etc. For

supervised classification, labeled samples are used to train the system to assign objects to the respective classes. Some of the different models among the supervised procedures include logistic regression, linear discriminant analysis, the k-Nearest Neighbor (KNN) classifier, decision trees, Support Vector Machines (SVMs) and Neural Networks.

Examples of applications that use different classifiers include using an SVM for categorizing masses in mammograms [49], the use of a KNN classifier for estimating the age of a person from time stamped facial images [90], and a Gaussian Mixture Model (GMM) for speaker identification from audio clips [44]. For this application, two supervised learning classifiers, the KNN and SVM are used for two reasons. First, as there is labelled training data available, there is no need to venture into unsupervised procedures. Secondly, as it is not known if the selected features would follow a particular pattern, classifiers that are non-parametric (KNN), and use linear discriminant functions (SVM), are explored. A brief introduction into these two classifiers is now given, along with a method for combining their results [94]. Further details for these systems can be found in [18].

KNN. This classifier works by assigning query objects to the class of objects that occur most frequently among their respective k-nearest neighbors. The search metric used for this application is the Euclidean distance, z, among the n features between the query object, $\mathbf{q}$, and each object from the training data set, $\mathbf{t}$, given by: $z^2 = \sum_n (\mathbf{q_n} - \mathbf{t_n})^2$.

SVM. In this classifier, features are mapped into a higher dimensional space where the object is then classified using decision margins or hyperplanes. This mapping is performed by a kernel function, which is usually either linear, polynomial, or a (Gaussian) radial basis function. In the higher dimensional space, a hyperplane is found that acts as a decision boundary between the object classes. The boundary is chosen so that the distance between it and the training examples is maximized. To perform the various experiments in this work, the SVM classifier from the scientifically accredited software, Matlab [54], is used as a black box. With this software all of the default settings are used with two exceptions. First, the kernel function is set to Gaussian Radial Basis Function, as it is popularly used in the literature [94] [89] [83] [58]. Second, the `kktviolationlevel`, is set to 0.05, to allow 5% of the data violate the Karush-Kuhn-Tucker (KKT) conditions. This particular setting is necessary as there would not be a clear decision boundary that perfectly separates the burrow and non-burrow objects, and by alowing a small fraction of the data to violate these conditions, a more effective boundary would be calculated.

Hybrid KNN-SVM. In this work the possibility of combining these two classification systems to improve the overall system performance is explored. To perform this combination the technique by Mei et al. [94] is employed. This technique has two assumptions. First, it is assumed the KNN is best suited to classify objects whom k-nearest neighbors are all of the same class. Secondly, a optimal training set for the SVM and KNN systems can be achieved with objects whom k-nearest neighbors are all of the same class. These two assumptions is generalized in the

overall combination with three steps as follows.

1. Prune the training set for both systems with the KNN by only retaining objects whom k-nearest neighbors are of the same class.

2. For the test set, use the KNN to classify objects whom k-nearest neighbors are of the same class.

3. Use the SVM to classify the remaining objects.

2.4.4.5 Training and Evaluation

Training, in supervised learning schemes, is the process of using manually labeled sample objects with known classes to influence the decision of the classifier. While evaluation is the process of measuring the performance of the classifier by comparing the items it detected against ground truth values. The training and evaluation data sets for the proposed burrow recognition system are obtained from ten mosaics generated from ten 2000-3000 frame (576×712) sequences of actual underwater surveillance videos. Each of these mosaics contained an average of 712 burrows and 3562 non-burrow objects, which were manually labeled by two experts from the Marine Institute, Galway, Ireland. From all of these labeled sequences, an optimal subset is selected for training by evaluating the performance of the system on a constant test set, further details of which are given in Chapter 5.

To perform this evaluation, three metrics are used: i) Precision, ii) Recall, iii) Classification Error, along with the use of the Receiver Operating Characteristic (ROC) space, which are described as follows:

Precision Precision is the fraction of retrieved items that are relevant. A high value in this metric is desired as it implies the system is detecting a relatively low number of false alarms, N_F, compared to the number items that it correctly detects, N_C.

$$Precision = \frac{N_C}{N_C + N_F} \times 100\% \tag{2.5}$$

Recall. Recall is the ratio of the number of items correctly detected to the total number relevant items. A high value in this metric is also desired as it implies the system returns most of the relevant results.

$$Recall = \frac{N_C}{N_C + N_M} \times 100\% \tag{2.6}$$

Where N_M are the number of items missed.

Classification Error. This is the ratio of the number of items that are misclassified, N_L to the total number of items, N_T. A low value in this metric is desired as it implies the system correctly classifies most of items inputted.

$$Error = \frac{N_L}{N_T} \times 100\% \tag{2.7}$$

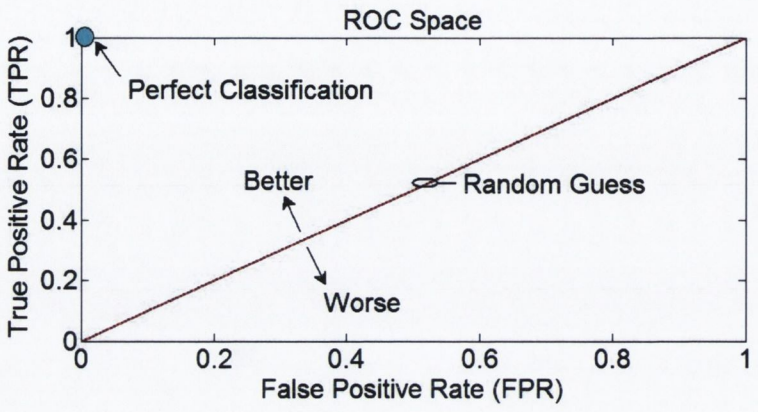


Figure 2.15: ROC space showing the point of perfect classification (blue), the line of no-discrimination (red), and regions where the respective classifier is performing better and worse than random guessing.

Receiver Operating Characteristic. The Receiver Operating Characteristic (ROC) space is a useful tool for analyzing the performance of a binary classifier. They are created by plotting the true positive rate (TPR) or Recall versus the number of false positives divided by the total number of true negatives (FPR = false positive rate). In these plots a point in the upper left corner or coordinate (0,1) of the ROC space, is desired as it implies perfect classification. A completely random guess would give a point along a diagonal line (the so-called line of no-discrimination) from the left bottom to the top right corners (regardless of the positive and negative base rates). Points above and below this line imply the classifier is performing better or worse than a random guess, as shown in Figure 2.15.

2.4.4.6 Optimization

Optimization in classifiers is well known to be a difficult problem. Domain knowledge is often brought to bear in selection of parameters and model orders. A particularly troublesome problem is overfitting [18]. This can lead to the situation in which a classifier performs well on the training data but not on other sets. Strategies for complexity reduction and for avoiding the overfitting problem include reducing: i) feature dimensionality [85], and ii) the amount of training data [88]. In this work a common training set if used for comparing the performance of the KNN and the SVM. In addition, the training set is pruned by removing points which are never used by the KNN in classification of the training set.

There are many techniques for reducing the dimensionality of the feature space. Clearly one approach is to exhaustively test every possible feature subset and select the one that yields the best possible result [75]. For large feature sets this is difficult and alternative approaches have been developed e.g. hill climbing [75]. PCA is one of the most popular techniques for dimensionality reduction. It works by transforming the data into a different space expressed by

the most significant eigenvectors that could be used to represent the training data [86]. Face recognition [85] has benefited a great deal by PCA analysis. In this proposed work, because the entire feature set being examined is relatively small (i.e. only 14), exhaustive searching analysis in the current feature space, along with PCA in the orthogonal space, are examined to obtain an optimal subset of features.

2.5 Summary

Three applications are developed in this work in the areas of image enhancement, content summarization, and object recognition. The image enhancement application involves correcting the illumination degradations in these videos to improve visibility for the scientists. The content summarization application improves the field of view for scientists by generating large area views or mosaics of the surveyed seabed area from the recorded video. Lastly, the recognition application provides a reference for the manual counts obtained from scientists by automatically detecting all of the burrows in the generated mosaic. In this chapter, a review of the literature in each of these areas is presented, which highlight the choices made in this work.

In the image enhancement review, it is highlighted that a new spatial model is needed to account for the illumination distribution of the light source and the colour degradations in this environment due to absorption from water. The two most relevant bodies of work is found in the vignetting and underwater colour literature. The vignetting correction techniques can cater for the spatial deteriorations associated with the illumination distribution of the light source to a certain extent, but they do not take into account the colour degradations that occur due to absorption from the water medium. Whereas the underwater enhancement techniques can correct the colour degradations due to absorption from the water medium using the Beer-Lambert law. However, they cannot account for the spatial deteriorations associated with the illumination distribution of the light source on the sea floor.

In the summarization review two items are highlighted. First, mosaics are best suited for content summarization in this application, as marine scientists need to view all of the burrows on the surveyed sea floor to verify their results. The other summarization techniques such as key frames and video skims cannot facilitate this need as they only consider subsets of frames from the videos. The second item highlighted is the need for improved alignment and rendering procedures for generating high quality mosaics from unevenly lit images. The existing image alignment algorithms use either pixel or feature point matching. The pixel matching techniques are not robust to illumination variations. While for blurry images, some of the feature points utilized such as SIFT may not be extracted efficiently. For rendering image details from overlapping regions, existing algorithms use either statistical methods (such as or median and weighted mean), or select image sections from separate frames. The statistical methods which involve using the median or a weighted average value from the entire set, may not capture the best image details from unevenly lit sequences. While the image selection

methods use a weighting function to select image sections from the particular frame where the overlapping regions is located closest to the image center, as it is here where the best image quality is perceived to exist. However, for underwater survey videos of the sea floor, the best image details are located within the well lit regions of the frame, which is not necessarily the image center.

Lastly, in the object recognition review, it is highlighted that the existing video-based burrow recognition application can be improved by using:

1. Mosaics for detecting and summarizing the automated results would reduce the time spent tediously inspecting thousands of frames to the scanning of a single image.

2. Segmentation techniques for object detection, more complete object regions can be obtained as opposed to using edges.

3. A new feature set that is motivated by a current scientific description of Nephrop burrows, which would be easy for marine scientists to relate to.

4. Supervised learning classification schemes as opposed to using a strict set of rules, can improve how the system generalizes on different data sets.

In the next three chapters, the image enhancement, content summarization and object recognition applications developed in this work, will be discussed respectively. In each of these chapters full details of the algorithms used and comparisons with state of the art techniques will be given. Afterwards, the conclusions made from the various experiments performed in each of these chapters will be highlighted, with a discussion on future work.

3

Improving Underwater Visibility Using Vignetting Correction

Good visibility is critical in underwater surveys to allow marine scientists to perform accurate analysis. With sufficient illumination above sea level, the human visualization range can span kilometers, but underwater this range reduces to a mere couple of meters [8], even in crystal clear water. This massive reduction is caused by the strong light absorption and backscattering properties of water molecules, which attenuate the light signal and causes blurred vision. The backscattering property diffuses the light in all directions thereby blurring visibility, while the absorption property reduces visibility by soaking up the light signal. This absorption also causes colour distortion as it does not occur uniformly across the visible spectrum [70] [8]. The red channel in particular is absorbed more than the blue and green, which is why a blue/green contrast is observed in some underwater images. These two properties attenuate natural light to such a large extent that the visibility at depths of 50-100m, where Nephrops surveys are typically performed, is practically zero.

To overcome this low visibility problem, high intensity artificial lights are used in these surveys, which are mounted on sleds or trawl nets with video cameras and pulled along the sea floor. However, due to limitations of power and available mounting positions on the survey apparatus, only a few lights can be used. Typically for the trawl net apparatus, one light is used, whereas for sleds, 6-8 lights are used. To maximize visibility, the lights are mounted to focus on the same spot, which in some cases results in a distinctive footprint of their light beams on the sea floor, as shown in Figure 3.1. Within the footprint region the illumination remains

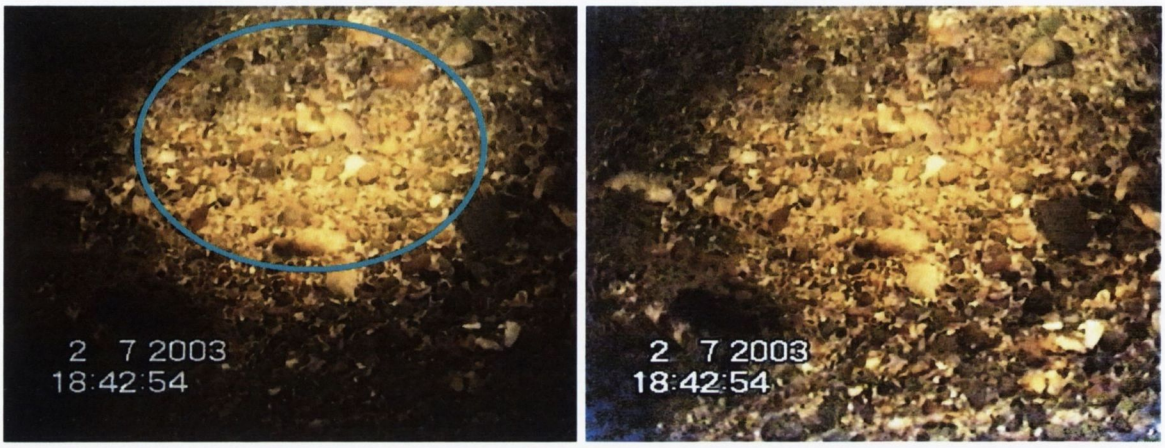


Figure 3.1: (Left) Original degraded image with footprint region circled in blue. (Right) Correction from proposed method.

relatively constant, but beyond its boundary, it degrades radially. Apart from this spatial discontinuity in illumination, deterioration in colour also occurs in these images as a result of the water medium absorption properties. These combined effects make manual analysis of video recordings difficult due to the limited field of view. Automated or computer assisted analysis is also a challenge because the i) shape, ii) position and iii) extent of the lighting footprint along with the degradations change throughout the sequence. Some of the causes for these changes in degradation are due to alterations in the height and focusing position of the camera and lights, as they are pulled along the seabed on the associated survey apparatus.

There are two relevant bodies of research literature that correct similar degradations. The first body is the underwater colour correction literature, which is derived from models of light propagation in water [69] [70] [8]. In these models, the deterioration in each colour channel is expressed as an exponential decay of the distance from the camera. These systems however do need initial calibration, which is usually performed with the use of either special equipment, and/or user input. Examples of this equipment include the use of special polarizer lens [69], and depth estimates together with a-prior knowledge of the colour of objects such as the McBeth colour chart [8], limestones [70] etc. Although this approach is widely used by the underwater community, it cannot be applied directly to this particular problem as there is no special equipment available or a-priori knowledge of the colours of objects with depth estimates to calibrate the system. Also, it assumes uniform illumination and degradation for objects at the same depth, which is not true in this case, as the illumination is uneven and the degradations also depend on the spatial location of the object with respect to the light beam footprint.

The second relevant body of work is the vignetting literature, which is geared towards correcting the degradations from camera lens in natural images [91] [45] [92]. Here the degradations are caused by the camera lens focusing the incoming light onto the center of the sensor array or

CCD, which causes the intensity in the captured image to decrease gradually from the center, in a radial fashion. To correct this phenomenon, distortion functions are used to model the vignetting from the lens and also the effect from the camera response function. Here the vignetting is usually assumed to be continuous, circular in shape, centered at the image center, and is approximately the same in all frames. The parameters for these models are estimated using either single or multiple images. Zheng et al. [91] utilized the gradient along the radial direction from a single image for their vignetting parameter estimation, whereas for the camera response, it is assumed to be known. Kim & Pollefeys [45] developed a robust method to estimate the vignetting and response functions independently and linearly based on the attenuations experienced among point correspondences between image pairs. In this method, attenuations at different radii are used to recover the vignetting function, whereas for the response function, attenuations at the same radii, taken at different exposures, are utilized. Because this methodology of using attenuations among point correspondences to estimate the vignetting function is independent of variations in image texture, it is incorporated into the proposed technique. In addition to using point correspondences, the proposed correction technique is limited to the region outside of the lighting footprint, as within this region the visibility of the image is good. Although this approach models the spatial degradations, it does not take into account the different deterioration in each colour channel due to absorption from the water medium.

As the vignetting and underwater colour correction literature solve different parts of this problem, a new solution is proposed in this work, which incorporates those basic ideas. Within the framework of this new solution, four key contributions are made. First, a new degradation model is derived that takes into account the deterioration in each colour channel, outside of the light footprint, due to absorption by water. In addition the model takes into account the illumination distribution of the light source, vignetting from the camera lens, and the footprint region. Secondly, the degradations are not restricted to being circular and located at the image center, as in typical vignetting functions, but instead follow the general elliptical shape and center of the light footprint. Third, the system does not require any special equipment or a-priori knowledge of i) the colour of objects, ii) depth measurements, or iii) camera response functions, for calibration. Instead all of the various parameters are estimated automatically using the attenuation from corresponding points across multiple frames. Lastly, to address the changes in degradation that occur from the movement of the survey apparatus, the model parameters are continuously updated throughout the sequence.

In the next section the new degradation model is derived. Afterwards, details of the correction methodology adopted for use with this new model is presented. Then, an evaluation of the proposed work using both synthetic and real data is then presented, along with comparisons with a state of the art technique for vignetting removal developed by Kim & Pollefeys [45]. Lastly, a discussion of the most relevant and interesting results obtained from the experiments performed are given in the conclusion, with suggestions for future improvement.

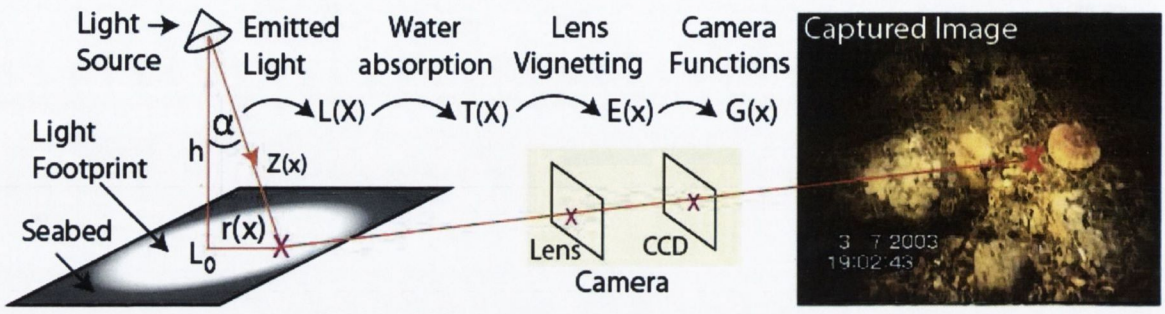


Figure 3.2: Image formation process. The light source emits scene radiance, $L(\mathbf{x})$, which the water absorbs a fraction and transmits, $T(\mathbf{x})$ the remainder to the seafloor and then through the camera lens where the vignetting transforms it to image irradiance, $E(\mathbf{x})$, which is then transformed to image intensity, $G(\mathbf{x})$, by the camera response and exposure functions.

3.1 Degradation Model

A new degradation model is developed to correct the uneven lighting distribution in underwater images from seabed surveys. This model takes into account the: i) illumination distribution from the light fixture, ii) absorption through the water medium, and iii) vignetting from the camera lens. Figure 3.2 illustrates how these various degradations are involved in the image formation process. In the literature there are state of the art methods that model each of these elements separately. But these models involve complex equations with many tuning parameters. For instance, to model the illumination distribution of a light source requires six parameters [61], colour deterioration in water uses seven parameters [69] [70], vignetting uses nine parameters [91] [45], and the camera response seven parameters [45]. As combining these equations will result in a huge ill posed solution with many different parameters to estimate, simplification is needed.

The simplification of all these degradations is accomplished in each stage of the image formation process, as shown in Figure 3.2, as follows. In the first stage, the radial fall off, $M(\mathbf{x})$, in the scene radiance, $L(\mathbf{x})$, is assumed to be primarily due to the light source illumination distribution on the seabed. Assuming the light fixture used in these surveys behaves as a point source and radiates light uniformly in all directions, this radial fall off can be modeled using the cosine cubed law [37], as:

$$L(\mathbf{x})/L^o(\mathbf{x}) = M(\mathbf{x}) = \cos^3 \alpha_M(\mathbf{x}) \tag{3.1}$$

where $\alpha_M(\mathbf{x})$ is the angle between the normal to the surface at the center of the degradation, and the image point $\mathbf{x}$ under consideration, and $L^o(\mathbf{x})$ is the scene radiance at the center of the degradation (i.e. $\alpha_M(\mathbf{x}) = 0$). Moving on in the simplification process, the scene radiance emitted from the light source is now further degraded as it pases through the water medium and a fraction is absorbed. The transmittance, $T(\mathbf{x})$, or fraction of light that passes through the medium has been effectively modeled in the literature [56], [70], [69], [8] using the Beer-Lambert

law, given by:

$$L_\lambda(z(\mathbf{x}))/L_\lambda(z_o(\mathbf{x})) = T(\mathbf{x}) = \exp -(n_\lambda z(\mathbf{x})) \tag{3.2}$$

where n_λ is an attenuation coefficient used for each colour channel, $\lambda = [r, g, b]$. $L_\lambda(z(\mathbf{x}))$ is the radiance at depth $z(\mathbf{x})$ from the light source underwater, and $L_\lambda(z_o(\mathbf{x}))$ is the corresponding radiance value taken just under the surface of the water. Assuming the light fixture behaves as a point source, then from Figure 3.2, the change in depth at image locations $r_l(\mathbf{x})$, with respect to the center of the degradation, is derived as: $\triangle z(\mathbf{x}) = \sqrt{h^2 + r_l(\mathbf{x})^2} - h = h\sqrt{1 + (r_l(\mathbf{x})/h)^2} - h$. Substituting this derivation into equation 3.2, and using the Taylor series expansion for $\sqrt{1 + (r_l(\mathbf{x})/h)^2}$, of $\sum_{n=0}^{\infty} \binom{0.5}{n}((r_l(\mathbf{x})/h)^2)^n$, and ignoring the higher order terms, $(n > 1)$, as typically $(r_l(\mathbf{x})/h) < 1$, gives:

$$L_\lambda(\mathbf{x})/L^o_\lambda(\mathbf{x}) = T(\mathbf{x}) \approx \exp -(n_\lambda r_l(\mathbf{x})^2/2h) \tag{3.3}$$

In the next stage of the image formation process, the transmitted scene radiance, $T(\mathbf{x})$, passes through the camera lens, and undergoes further degradation, which is commonly referred to as vignetting, $N(\mathbf{x})$, and is transformed to image irradiance, $E(\mathbf{x})$. This degradation can be modeled effectively using the cosine fourth law [17], as:

$$L(\mathbf{x})/L^o(\mathbf{x}) = N(\mathbf{x}) = \cos^4 \alpha_N(\mathbf{x}) \tag{3.4}$$

similar to the cosine cubed law in equation 3.1, $\alpha_N(\mathbf{x})$ is the angle between the normal to the surface at the center of the degradation, and the image point $\mathbf{x}$ under consideration. Then, in the final stage of the image formation process, this image irradiance, $E(\mathbf{x})$, undergoes further adjustments by the camera exposure setting, k, and nonlinear radiometric response function, $f()$. Combining the degradations from each stage, using equations 3.1, 3.3, 3.4, with these camera functions, the recorded image intensity, $G(\mathbf{x})$, is mathematically modeled as:

$$G_\lambda(\mathbf{x}) = f(kL^o_\lambda(\mathbf{x})M(\mathbf{x})T(\mathbf{x})N(\mathbf{x})) = f(kL^o_\lambda(\mathbf{x})\cos^3 \alpha_M(\mathbf{x})[\exp -(n_\lambda r_l(\mathbf{x})^2/2h)]\cos^4 \alpha_N(\mathbf{x})) \tag{3.5}$$

This equation is now further simplified in three steps. First the camera response $f()$ is generalized as a gamma function γ, which is a reasonable assumption as Grossberg and Nayar [27] have shown that some real world response functions are very similar to gamma curves. Secondly, it is assumed that $\alpha_M(\mathbf{x}) = \alpha_N(\mathbf{x}) = \alpha(\mathbf{x})$, which is practical as the light source illumination distribution is assumed to be the dominant degradation in this case. Lastly, to obtain an expression for the overall degradation, $B_\lambda(\mathbf{x})$, the entire equation is divided by the unattenuated image intensity, $I_\lambda(\mathbf{x}) = f(kL^o_\lambda(\mathbf{x}))$, to give:

$$\frac{G_\lambda(\mathbf{x})}{I_\lambda(\mathbf{x})} = B_\lambda(\mathbf{x}) = \frac{[kL^o_\lambda(\mathbf{x})\cos^7 \alpha(\mathbf{x}) \exp -(\frac{n_\lambda r_l(\mathbf{x})^2}{2h})]^\gamma}{[kL^o_\lambda(\mathbf{x})]^\gamma} = \left[\cos^7 \alpha(\mathbf{x}) \exp\left(\frac{n_\lambda r_l(\mathbf{x})^2}{7h}\right)^{-\frac{7}{2}}\right]^\gamma \tag{3.6}$$

Three substitutions are now made in the above expression. First, $\cos \alpha(\mathbf{x}) = h/\sqrt{r(\mathbf{x})^2 + h^2} = (1 + (r(\mathbf{x})/h)^2)^{-1/2}$, which is obtained from Figure 3.2. Where $r(\mathbf{x})^2 = [\mathbf{x} - \mathbf{c}]^T \mathbf{V} [\mathbf{x} - \mathbf{c}]$

is the radius from image point $\mathbf{x}$ on the contour of an ellipse to its center $\mathbf{c} = [c_x, c_y]^T$, with covariance matrix $\mathbf{V} = [v_1, v_2; v_2, v_3]$ capturing the extent of the degradation and the shape of the respective isophotes. Then, assuming this radii and the one in the water absorption degradation, $T(\mathbf{x})$, have approximately the same shape, the second substitution of $r(\mathbf{x})^2 = or_l(\mathbf{x})^2$ can be made, where o is a constant of proportionality among the two radii. Lastly, the substitution of $\exp(n_\lambda or(\mathbf{x})^2/7h) = 1 + (n_\lambda or(\mathbf{x})^2/7h)$, is made using the Taylor series expansion where higher terms ignored as $n_\lambda or(\mathbf{x})^2/7h \ll 1$. Making these substitutions in the equation above, and taking logarithms on either side gives:

$$\ln(B_\lambda(\mathbf{x})) = \ln\left[\left(1 + \left(\frac{r(\mathbf{x})}{h}\right)^2\right)^{-\frac{7}{2}}\left(1 + \frac{n_\lambda or(\mathbf{x})^2}{7h}\right)^{-\frac{7}{2}}\right]^\gamma = -\frac{7\gamma}{2}\ln\left[1 + a_\lambda r(\mathbf{x})^2 + \frac{on_\lambda r(\mathbf{x})^4}{7h^3}\right] \tag{3.7}$$

where $a_\lambda = (1/h^2) + n_\lambda o/7h$, is a constant at a particular height, h. Then, as usually $r(\mathbf{x}) \ll 1$, the above expression is further simplified by setting the third term to zero, and substituting the Taylor series expansion of $\ln[1 + a_\lambda r(\mathbf{x})^2] = a_\lambda r(\mathbf{x})^2$, to give:

$$B_\lambda(\mathbf{x}) \approx \exp -(3.5\gamma a_\lambda r(\mathbf{x})^2) \tag{3.8}$$

In the above expression, the constants, γ and a_λ can be estimated accurately using images taken of the same scene with different camera exposures [45], and at measured depths [69], h respectively. But as these additional calibration data are not available, a sensible approach would be to combine the two constants with the $\mathbf{V}$ parameters, and then estimate their combined effect, at a particular height, h, for each channel. The major problem with this approach however, is because of visibility challenges in these images, the $\mathbf{V}$ estimates for each channel may differ largely in overall shape. To prevent this problem from occurring, a common set of $\mathbf{V}$ parameters is used, and the attenuation differences among each channel are solved as a factor, m_λ, to these values. For the common parameters, the effective estimates from the green channel, $\mathbf{V} = \mathbf{V_g} a_\lambda \gamma$, is chosen as degradation is usually least in this channel. This approximates the effective degradation for a light source at height h above the sea floor, due to its illumination distribution, absorption from water, and vignetting from the camera lens, as:

$$G_\lambda(\mathbf{x})/I_\lambda(\mathbf{x}) = B_\lambda(\mathbf{x}) \approx \exp -(3.5 m_\lambda r(\mathbf{x})^2) \tag{3.9}$$

where $r(\mathbf{x})^2$ contains the effective degradation and shape parameters of the green channel, $\mathbf{V_g}$, and $m_\lambda = \{m_r, 1, m_b\}$ are attenuation coefficients for each channel, that represent their effective degradation with respect to the green channel. The key advantage of this proposed method, compared to state of the art methods from the vignetting [45] [92] and underwater colour correction literature [69], [70], is that it does not require estimates of: i) exposure setting k, ii) response functions γ, iii) height values h, and iv) actual attenuation coefficients for each channel, n_λ. Instead, the images are left at their respective exposure setting, and the net effect of the remaining parameters are incorporated into the $\mathbf{V}$ and m_λ estimates. These estimates are

continuously updated every 2-3 frames throughout the video sequence, to account for changes in the overall degradation that may result due to alterations in height, h, and position of the light beam on the sea floor.

3.2 Underwater Vignetting Correction

Using the expression in equation 3.9 for the effective degradation, $B_\lambda(\mathbf{x})$, the unattenuated image intensity, $I_\lambda(\mathbf{x})$, is recovered by estimating the parameters $\mathbf{c}$, $\mathbf{V}$, and m_λ. To accomplish this task $I_\lambda(\mathbf{x})$ is decoupled from the proposed degradation model by taking the ratio $A(\mathbf{x})$, of observed intensities $G_1(\mathbf{x})$, $G_2(\mathbf{x})$ from corresponding points $\mathbf{x_1 x_2}$ in consecutive frames 1 and 2. Dropping the notation λ for clarity, this gives:

$$\ln(A(\mathbf{x})) = \ln\left(\frac{G_2(\mathbf{x})}{G_1(\mathbf{x}+\mathbf{w}(\mathbf{x}))}\right) = \frac{-7m}{2}(r_2^2(\mathbf{x}) - r_1^2(\mathbf{x}+\mathbf{w}(\mathbf{x}))) \tag{3.10}$$

where $r_1(\mathbf{x}+\mathbf{w}(\mathbf{x}))$, $r_2(\mathbf{x})$, and $\mathbf{w}(\mathbf{x})$ are the corresponding positions and motion flow of image point $\mathbf{x}$ in frames G_1, G_2 respectively. With this simplified expression the required parameters are estimated and hence correction is performed. The steps undertaken to obtain the point correspondences, estimate parameters, and perform correction, will now be given.

3.2.1 Correspondence

Because of the inevitable inaccuracies associated with poor visibility underwater, the full set of correspondences using the global motion flow, $\mathbf{w}(\mathbf{x})$ ($G_2(\mathbf{x}) = G_1(\mathbf{x}+\mathbf{w}(\mathbf{x}))$), is initially obtained. This flow is obtained using either the method developed by Spindler and Bouthemy [79], or the method proposed in this thesis in Section 4.3. The method by Spindler and Bouthemy [79] is a hierarchical gradient based approach, while the other uses point correspondences. Both methods are developed for obtaining $\mathbf{w}(\mathbf{x})$ with an affine motion model, from underwater video sequences with respect to the sea-bottom area. It should be noted however, the method proposed in this thesis does show more robustness in the Nephrops environments, compared to the technique by Spindler and Bouthemy [79], from the tests performed in chapter 4.

Once the full set of corresponding points is obtained, a robust set is selected based on i) intensity, ii) motion, and iii) attenuations. For the intensity feature, underexposed and saturated points are removed by limiting the intensity range (25 to 235). In addition, only points with significant motion (> 4 pixels), and attenuation ($0.95 > A(\mathbf{x})$ AND $A(\mathbf{x}) > 1.05$), are selected.

3.2.2 Parameter Estimation

Given the proposed degradation model, the shape, central location, and colour correction parameters are needed to perform correction. With the robust set of point correspondences, each parameter is estimated as follows:

3.2.2.1 Shape and Central Location

Using the robust set of point correspondences from the green channel, the central location $\mathbf{c} = [c_x, c_y]^T$, and covariance matrix $\mathbf{V} = [v_1, v_2; v_2, v_3]$ capturing the shape and extent of the degradations on the sea floor are now estimated. Proceeding in a Bayesian fashion, the posterior $p(\mathbf{c},\ \mathbf{V}|G_1, G_2)$ is maximized with respect to the parameters $\theta = [c_x, c_y, v_1, v_2, v_3]$, to give:

$$p(\theta|G_1, G_2) \propto p_l(G_1, G_2|\theta)p_\theta(\theta) \tag{3.11}$$

The likelihood $p_l(\cdot)$ is derived directly from eq.(3.10), where the colour correction parameter is set as $m_g = 1$, and Gaussian priors, $p_\theta(\cdot)$ are used for each parameter as:

$$p_l(G_1, G_2|\theta) \propto \exp -\left[\frac{(\ln(G_2/G_1) + 3.5(r'^2_2 - r'^2_1))^2}{2\sigma_e^2}\right]; \quad p_\theta(\theta) \propto \exp -\left[\frac{\sum_{\mathbf{x}}(\theta - \theta_0)^2}{2\sigma_\theta^2}\right] \tag{3.12}$$

Initialization with Well-Lit Region

Good initial estimates for these parameters, θ_0, are obtained using the well lit region. To estimate this region, the global attenuation field of all corresponding points, $A(\mathbf{x}) = G_2(\mathbf{x})/G_1(\mathbf{x} + \mathbf{w}(\mathbf{x}))$, is used. To obtain the underlying degradation pattern, this field is sorted in increasing order of attenuations, $A_s(\mathbf{x})$, by inverting the fractional values, given by:

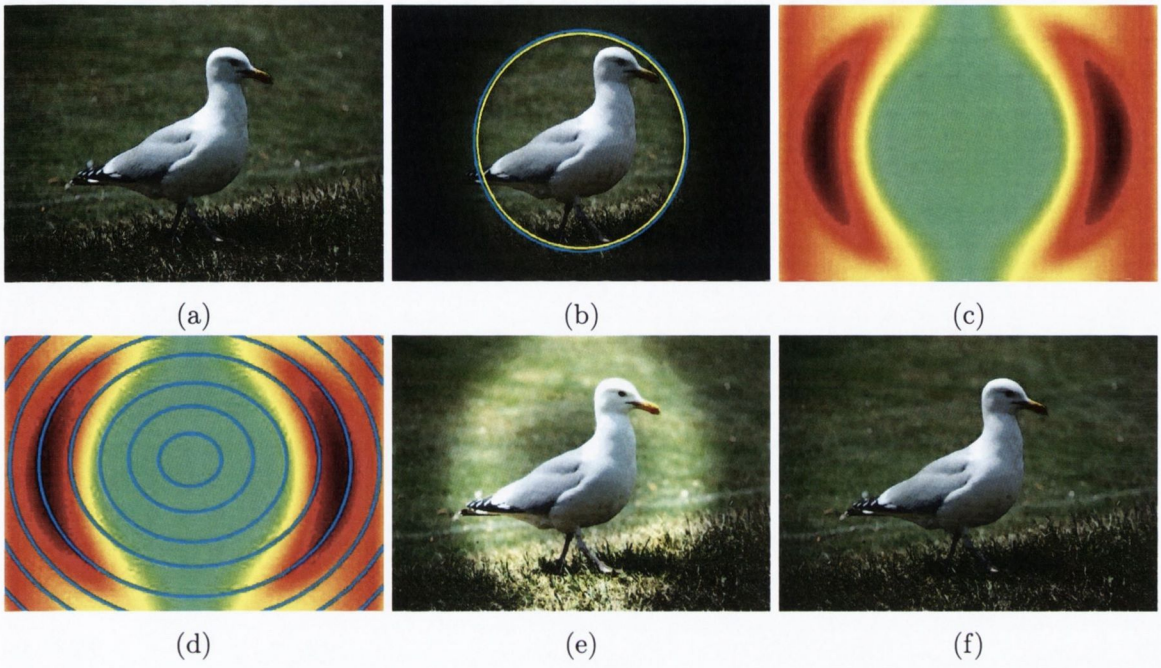
$$A_s(\mathbf{x}) = \begin{cases} 1/A(\mathbf{x}) & A(\mathbf{x}) < 1 \\ A(\mathbf{x}) & A(\mathbf{x}) \geq 1 \end{cases} \tag{3.13}$$

Figure 3.3(c) shows a sample image of this sorted gain field. Using $A_s(\mathbf{x})$, a binary map, $P(\mathbf{x}) = \{0, 1\}$ of the well-lit region is estimated as the pixel values whose attenuation values are less than the median value, μ_a, of the sorted gain field i.e. $P(\mathbf{x}) = A_s(\mathbf{x}) < \mu_a$. With this binary map, the center and shaping matrix parameters are initialized as the first and second order central moments, given by:

$$\mathbf{c_0} = \frac{\sum_{\mathbf{x}} \mathbf{x}P(\mathbf{x})}{\sum_{\mathbf{x}} P(\mathbf{x})}; \quad \mathbf{V_0^{-1}} = \frac{\sum_{\mathbf{x}}(\mathbf{x} - \mathbf{c_0})(\mathbf{x} - \mathbf{c_0})^T P(\mathbf{x})}{\sum_{\mathbf{x}} P(\mathbf{x})} \tag{3.14}$$

As only a section of the image is used (i.e. points where $P(\mathbf{x}) = 1$), the initial shape estimates of v_1 and v_3 are usually smaller than the actual values. To compensate for this, these two values are scaled using the ratio of the total image points to the number of points used. These boosted values along with the rest of the initial estimates are assumed to be similar in value to the actual $\mathbf{c}$ and $\mathbf{V}$ parameters. Because of this similarity, σ_x^2 and σ_y^2 for the central location components of the prior terms are set to 15% the image width and height respectfully, and each shape component, σ_v^2 is set to 50% of its value. Using these settings, the likelihood term, σ_e^2 is set to twice the variance of the respective parameter.

Optimization Strategy. Given the functions of likelihood and prior, the posterior expression for $p(\mathbf{c},\ \mathbf{V}|G_1, G_2)$ in equation 3.11 is non-linear in the parameters. To simplify the process, the



(a) (b) (c)

(d) (e) (f)

Figure 3.3: (a) Original image, (b) Artificial degradation, with simulated (yellow) and recovered (blue) footprints. (c) Sorted global attenuation field, $A_s(\mathbf{x})$, and (d) with concentric bands (blue) overlaid on it, for estimating the extent of the footprint region. (e) Corrections obtained from (e) using the proposed model in equation 3.9 only, and (f) with incorporating the footprint region into the correction procedure.

parameters are estimated alternately using the well known Iterated Conditional Modes (ICM) algorithm [6]. Hence, $\mathbf{V}$ is estimated by maximizing $p(\mathbf{V}|\mathbf{G_1}, \mathbf{G_2}, \mathbf{c})$, holding $\mathbf{c}$ at its current estimate. Then $\mathbf{c}$ is estimated by maximizing $p(\mathbf{c}|\mathbf{G_1}, \mathbf{G_2}, \mathbf{V})$, holding $\mathbf{V}$ at its current estimate. The conditionals from the posterior in eq. 3.12, for each parameter in this iterative scheme also take on Gaussian forms, given by:

$$p(\mathbf{c}|G_1, G_2, \mathbf{V}) \propto p_l(G_1, G_2, |\mathbf{c}, \mathbf{V})p_{\mathbf{c}}(\mathbf{c}); \quad p(\mathbf{V}|G_1, G_2, \mathbf{c}) \propto p_l(G_1, G_2|\mathbf{c}, \mathbf{V})p_{\mathbf{v}}(\mathbf{V}) \tag{3.15}$$

These conditionals were then differentiated w.r.t. the relevant unknown, set to zero, and solved, using the robust method of Singular Value Decomposition (SVD) [86]. Then lastly, estimates for $\hat{\mathbf{c}}$ and $\hat{\mathbf{V}}$ are iteratively optimized as:

$$\mathbf{c}^{\hat{n+1}} = \arg\max_{\mathbf{c}} p(\mathbf{c}|G_1, G_2, \mathbf{V}^n); \quad \mathbf{V}^{\hat{n+1}} = \arg\max_{\mathbf{v}} p(\mathbf{V}|G_1, G_2, \mathbf{c}^{n+1}) \tag{3.16}$$

The optimization is performed until the percentage change between the $(n+1)^{th}$ and n^{th} estimates for each parameter is less than 5%.

3.2.2.2 Colour Correction

The remaining colour correction parameters, m_r and m_b, for the red and blue channels respectively, are now estimated (separately) in proportion to the degradation of the green channel. This estimation is performed in a Bayesian fashion, where the posterior $p(m_\lambda|G_1, G_2, \mathbf{c}, \mathbf{V})$ is maximized with respect to each of the parameters m_λ, to give:

$$p(m_\lambda|G_1, G_2, \mathbf{c}, \mathbf{V}) \propto p_L(G_1, G_2, \mathbf{c}, \mathbf{V}|m_\lambda)p_r(m_\lambda) \tag{3.17}$$

The likelihood $p_L(\cdot)$ is derived directly from equation (3.10), while Gaussian priors, $p_r(\cdot)$ are used for each parameter (m_r, m_b) as:

$$p_L(G_1, G_2, c, V|m_\lambda) \propto \exp -\left[\frac{(\ln(G_2/G_1) + 3.5m_\lambda(r'^2_2 - r'^2_1))^2}{2\sigma_L^2}\right]$$

$$p_r(m_\lambda) \propto \exp -\left[\frac{\sum_{\mathbf{x}}(m_\lambda - {m_\lambda}^0)^2}{2\sigma_\lambda^2}\right]$$

where ${m_\lambda}^0$ is estimated from the previous frame in the video sequence, which is assumed to be similar in value to this current estimate. To enforce this similarity, the variances σ_L^2 and σ_λ^2 are set as 0.2 and 0.1 respectively.

3.2.3 Correction Procedure

Given the parameters $\mathbf{c}$, $\mathbf{V}$ and m_λ, the degraded image can now be corrected. In some cases however, using only the proposed model in equation 3.9 causes the pixels within the light beam footprint region of these images to be over amplified, as shown in Figure 3.3 (e). This is because this region is well lit and contains minimal or no degradation. To address these cases, the extent of the footprint region, r_f, is first estimated and then incorporated into a gain field, $C(\mathbf{x})$, where only the intensity surrounding this region is amplified. Cases when this compensation would improve the correction procedure are identified when the median value of the actual attenuation in the footprint region, η_a, is significantly less than the corresponding predicted value from the model, η_p, i.e. $(\eta_p - \eta_a) > 0.05$. If the actual attenuation is approximately equal to, or greater than the predicted value, then amplifying the entire image using the predicted model would be sufficient, and is performed in this manner. The estimation of the light footprint and the creation of the gain field are explained in the following sections.

3.2.3.1 Light Footprint

Using the sorted global attenuation field, $A_s(\mathbf{x})$, the extent of the footprint region, r_f, is estimated as the largest elliptical isophote radius, r_f, where the median attenuation value, η_f, is less than the threshold value, $T_f = 1.08$. Here, the median attenuation value is calculated from corresponding points within a number (y=20) of concentric bands that have the same shape $\mathbf{V}$ and central location, $\mathbf{c}$, of the overall degradations, as seen in Figure 3.3 (d). This estimation is

based on the footprint region key features of having: i) minimal degradation, and ii) the same shape $\mathbf{V}$ and iii) central location $\mathbf{c}$, as the attenuations.

3.2.3.2 Gain Field Creation

A gain field, $C_\lambda(\mathbf{x})$, is created to correct the uneven illumination at each image location, $\mathbf{x}$. For degradations that do not require their footprint region to be corrected, the values within r_f are set to 1, as negligible degradation occurs here, whereas outside it is set to $1/B_\lambda(\mathbf{x})$, to correct the apparent vignetting effect. In practice however, there are a few image points outside the footprint region that do not follow the degradation model $B_\lambda(\mathbf{x})$, and are driven into saturation after $C_\lambda(\mathbf{x})$ is initially applied. In most cases, these minor anomalies are due to suspended particles located much closer to the light source than the sea floor. To cater for this problem, $C_\lambda(\mathbf{x})$ is set to 1 at these locations. Then lastly, to simulate the smooth transition out of the footprint region, the gain field is smoothed with a Gaussian filter, g_σ ($\sigma = 10$). The three creation steps just described, can be mathematically summarized, in numerical order, as:

$$C_\lambda(\mathbf{x}) = \begin{cases} 1 & r(\mathbf{x}) \leq r_f \\ 1/B_\lambda(\mathbf{x}) & r(\mathbf{x}) > r_f \end{cases}; \; C_\lambda(\mathbf{x}) = \begin{cases} 1 & G_\lambda(\mathbf{x})C_\lambda(\mathbf{x}) > 255 \\ C_\lambda(\mathbf{x}) & \text{otherwise} \end{cases}; \; C_\lambda(\mathbf{x}) = C_\lambda(\mathbf{x}) * g_\sigma; \tag{3.18}$$

With the gain field created, the unattenuated image, $I_\lambda(\mathbf{x})$, is now recovered as $I_\lambda(\mathbf{x}) = G_\lambda(\mathbf{x})C_\lambda(\mathbf{x})$. In these surveys however, the shape, central location and footprint region of the actual degradation may change due to sudden movements of the survey apparatus as it is being pulled along the sea floor. To cater for these changes, along with erroneous point correspondences, the parameter estimation process is performed at each consecutive image pair and monitored over a 5 frame period. In this period, the gain field is only updated if percentage difference among the estimated parameters is significant ($> 5\%$), in at least 3 of the frames.

Hence, the proposed correction algorithm is summarized as follows:

1. Obtain point correspondences from consecutive frames using the global motion flow.
2. Use corresponding image intensities of green channel to estimate the degradation shape, $\mathbf{V}$, and center $\mathbf{c}$.
3. With $\mathbf{V}$ and $\mathbf{c}$, estimate of colour degradation in the red, m_r, and blue, m_b, channels.
4. Create gain field, $C_\lambda(\mathbf{x})$. If compensation for the well-lit footprint region is required, incorporate it into $C_\lambda(\mathbf{x})$, else just use the proposed model in equation 3.9.
5. Apply gain field to degraded image $G_\lambda(\mathbf{x})C_\lambda(\mathbf{x})$.

3.3 Results

The accuracy of the proposed method is now evaluated using simulated and real data. In this evaluation the results obtained are also compared to the state of the art vignetting estimation technique proposed by Kim & Pollefeys [45], noting that their work is directed at still and not moving images. The procedure used for creating the ground truth is described in the following section.

3.3.1 Ground Truth Creation

Eight ground truth test sequences are created for experimentation, each being 200 frames long with a size of 500×500 pixels for each frame. As mainly translational camera motion is observed in the actual underwater videos, only this type of motion is utilized in the creation of these ground truth sequences. The translational motion is a result of the camera being mounted approximately vertical in the survey apparatus, which is then pulled along the sea floor to record its contents. Although a minor amount of perspective distortion is observed in some

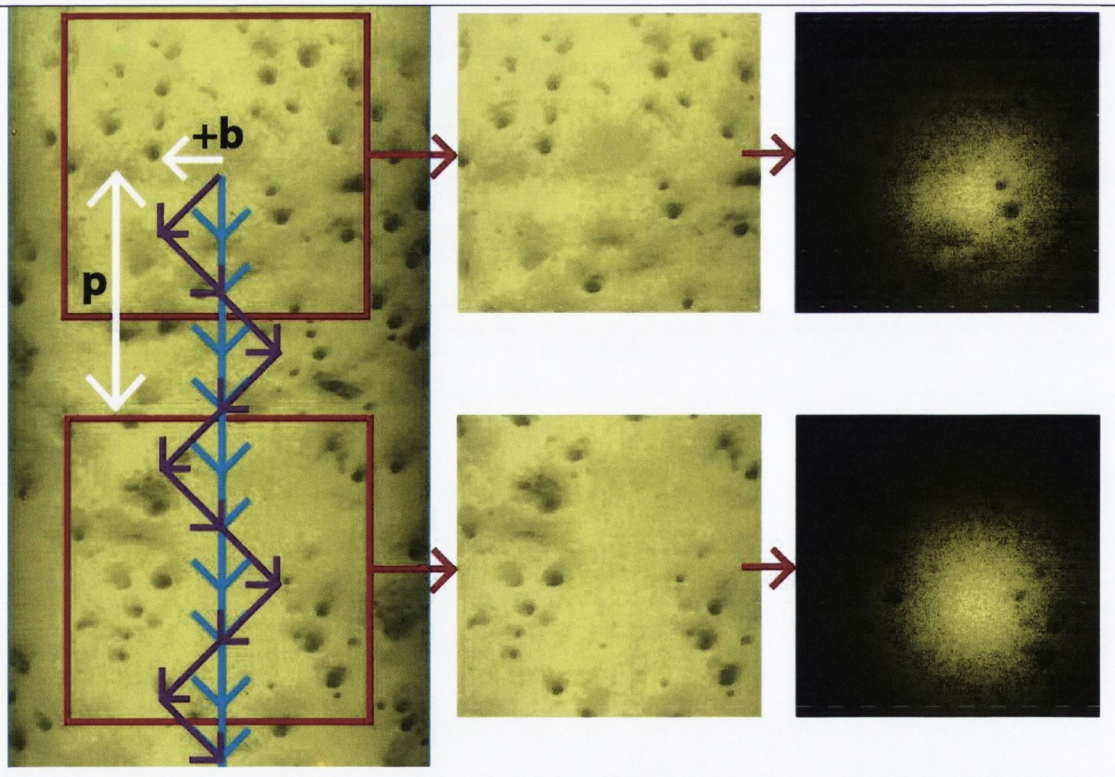


Figure 3.4: Creating two ground truth sequences by selecting a subsection (redbox) from the mosaic on the left and translating it at: i) constant motion (blue arrows), and with a symmetrical triangle wave (purple arrows), with period, p, and amplitude, $\pm b$ pixels. Then artificially degrading the generated frames shown in the middle, to give degraded versions on the right.

Figure 3.5: Sections of the two ground truth mosaics created by joining random images. Using these mosaics test sequences are generated in a similar manner shown in Figure 3.4.

videos, this type of motion is not taken into account as it is insignificant in most cases. To simulate this translational camera motion, each test sequence is created by first extracting a subregion of 500×500 pixels from an existing mosaic, and then panning it along the entire mosaic with a specific trajectory. Four mosaics were used. Two mosaics are generated from existing real footage (muddy and stony seabed) using the process in the next chapter. See Figure 3.4 for some example frames from these sequences. The next two mosaics are made by stitching together a selection of natural images, as shown in Figure 3.5.

To examine if direction of motion has any impact on the performance of the algorithm, two test sequences are created from each mosaic, as illustrated in Figure 3.4. The motion in the first sequence follows a symmetrical triangle function that has a amplitude of b pixels and is periodic at length p pixels about the constant motion value of the second sequence. The translations used in both sequences are relatively small, so that there are sufficient corresponding points between

Mosaic type	Muddy Seabed		Stoney Seabed		Random Pictures-1		Random Pictures-2	
Mosaic Size	6800×800		4000×900		600×4800		530×5600	
Motion	d(0,30)	$\triangledown(60, 40)$	d(0,-15)	$\triangle(30, 20)$	d(20,0)	$\triangleright$ $(40, 15)$	d(-25,0)	$\triangleleft$ $(50, 15)$
Test Video	1	2	3	4	5	6	7	8

Table 3.1: Information about the mosaics and motion vectors used in generating the ground truth videos, as shown in Figure 3.4. $\lambda(p, b)$ represents motion of a symmetrical triangle wave, with period of length p pixels, and amplitude, $\pm b$ pixels, moving in a direction $\lambda = \{\triangle(N), \triangledown(S), \triangleright (E), \triangleleft (W)\}$. While d(x,y) is constant horizontal (x), and vertical (y) motion.

consecutive frames among each sequence.

Lastly, to make these sequences more realistic, additive Gaussian noise: $Z(\mathbf{x}) = \sigma N(0,1)$, is added to each frame. Where $Z(\mathbf{x})$ is a pseudo random value at image location $\mathbf{x}$, that is drawn from the standard normal distribution $N(0,1)$, and $\sigma = 10$ is the respective noise level. In all, eight sequences were generated and Table 3.1 gives the source mosaic and motion parameters used to generate the sequences.

To test the sensitivity of the parameter estimation process presented in this chapter, each synthetic sequence was degraded by the function $B_o(\mathbf{x})$ as defined in equation 3.9. Tests were carried out varying one parameter at a time as follows

1. Camera Response Function (Section 1.3.2)
2. Shape σ_x, σ_y (Section 1.3.3)
3. Location $\mathbf{x_c}, \mathbf{y_c}$ (Section 1.3.4)
4. Footprint r_f (Section 1.3.5)
5. Colour m_λ (Section 1.3.6)

In order to be fair in comparing to the work of Kim & Pollefeys [45], the base degradation, $B_o(\mathbf{x})$, is set to be circular in shape, $\mathbf{V_o} = \{140, 0, ; 0, 140\}$, centered at the image center $\mathbf{c_o} = \{250, 250\}$, and has no response function, $\gamma_o = 1$. This is because the work by Kim & Pollefeys [45] assumes a much simpler vignetting model, where degradations are circular in shape and centered at the image center. These comparisons will now be discussed for each experiment, followed by an examination using actual underwater video sequences.

3.3.2 Camera Response Functions

Grossberg and Nayar [27] have shown the response functions of common brands of digital cameras are different. As our marine surveys are performed with different brands of cameras, it is

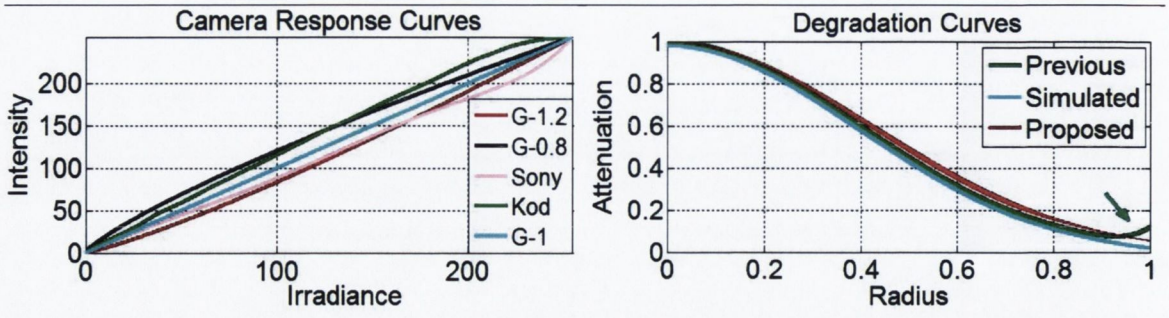


Figure 3.6: (Left) Camera response curves: Sony-DXC-9000, Kodak-kai1010CD, and $\gamma = \{0.8, 1.2, 1\}$, in pink, green, black, red and blue. (Right) Estimated degradation curves from proposed (red), and previous (green) method [45], for common degradation, $B_t(\mathbf{x})$ (blue).

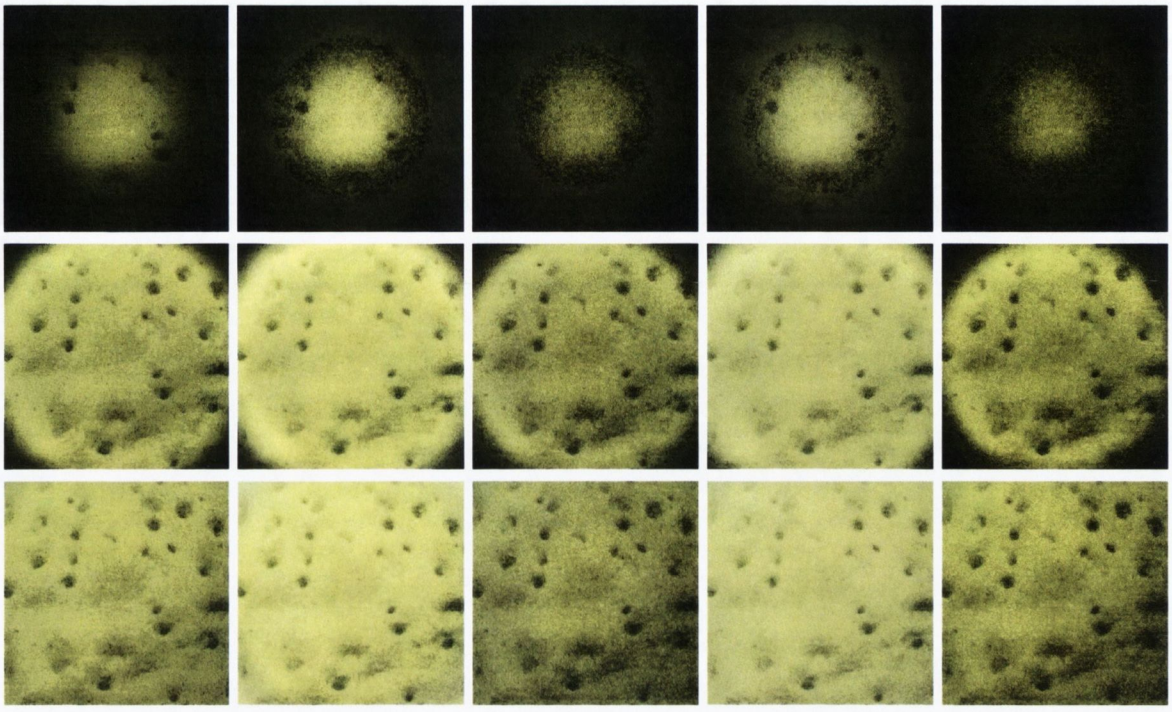


Figure 3.7: The top row shows example frames from degraded sequences with different camera response functions. The middle and bottom rows show respectively, results from correction algorithms using previous work [45] and the new algorithm presented in this thesis. The camera response functions simulated in each column are (from left to right) NONE, Kodak, γ=1.2, Sony, γ=0.8.

important to examine the impact various response functions would have on the performance of the proposed algorithm. To perform this examination, five sets of the ground truth videos are degraded with $B_o(\mathbf{x})$, and modified with a different camera response function. The response functions used are from two common brands, the: i) Sony-DXC-9000 and ii) Kodak-kai1010CD, along with gamma curves 0.8, 1.2, and 1 (i.e. $B_o(\mathbf{x})$). Intensity versus irradiance plots of these functions are shown in Figure 3.6, which are obtained from the database created by Grossberg and Nayar [27]. Using the output after the common degradation as irradiance values, the final intensity values for the test data is obtained via bilinear interpolation from these plots.

Samples of the degradation from each camera response, experienced by the same image (from test video-2), along with the corrections from the previous [45], and proposed methods, are shown in Figure 3.7. As can be seen, the proposed method is able to recover the entirety of the original image while the previous work by Kim & Pollefeys [45] is unable to recover the extremities well. In addition the previous work suffers from a ringing effect caused by overestimation of the attenuation due to the use of higher order polynomial terms to approximate the vignetting effect. Additional evidence of this overestimation is shown (highlighted with a green arrow) in

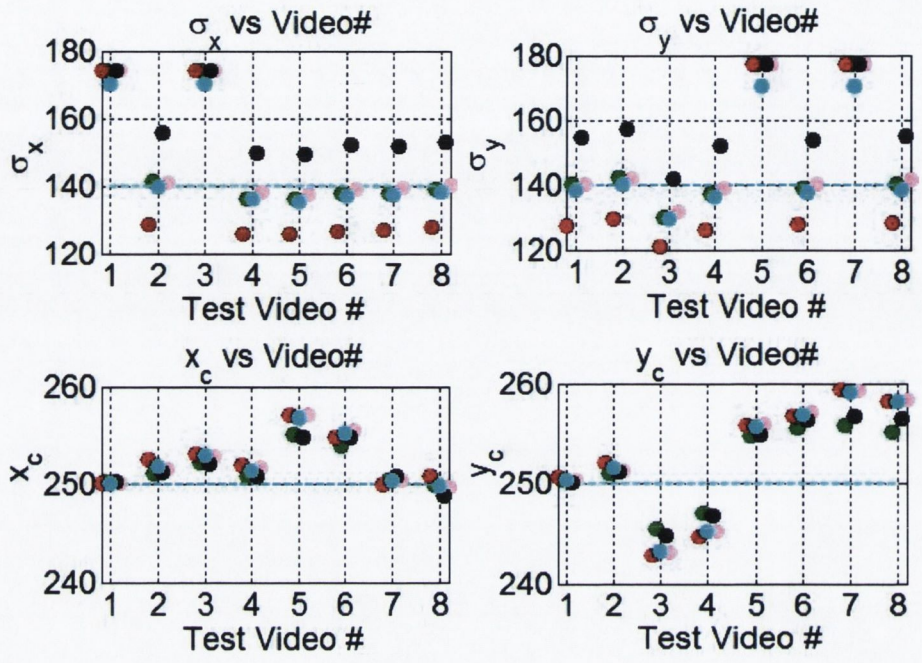


Figure 3.8: Parameter estimation results for the proposed algorithm applied to each test video (along horizontal axis). The result obtained from each camera response are colour coded: Sony (pink), Kodak (green), $\gamma = 0.8$ (black), $\gamma = 1.2$ (red), NONE (blue). The blue horizontal line shows the ground truth parameter value, distance from that line indicates the extent of the error in parameter estimation.

a plot of the attenuation curves obtained from each method in Figure 3.6. Another interesting observation made from these corrections, is they are all moderately different in appearance. Examining these differences, a pattern is observed, where brighter corrections correspond to response functions that amplify the input radiance ($\gamma = 0.8$ and Kodak-kai1010C), and darker corrections correspond to functions that reduce the input radiance values ($\gamma = 1.2$ and Sony-DXC-9000). These direct correspondences show the algorithms provide a corrected version in relation to the modifications performed by the respective camera response function.

Plots of the mean parameter estimates and PSNR values obtained are shown in Figures 3.8 and 3.10, and additional degraded ($\gamma = 1$ set) test images, together with the corrections obtained from the proposed and previous [45] methods, are given in Figure 3.9. From analyzing these results, three key observations are noted. First, is that motion in both x and y directions is needed for the shape parameters to optimize fully. Evidence of this fact is shown in the outlying σ_x and σ_y values obtained from the test videos where only horizontal (5,7) or vertical (1,3) motion are present. As a result of these outliers, the corrected images were darker, and the PSNR values are less than those obtained from corresponding sequences with motion in both directions. This issue however seems to have little effect of on the previous method, as the

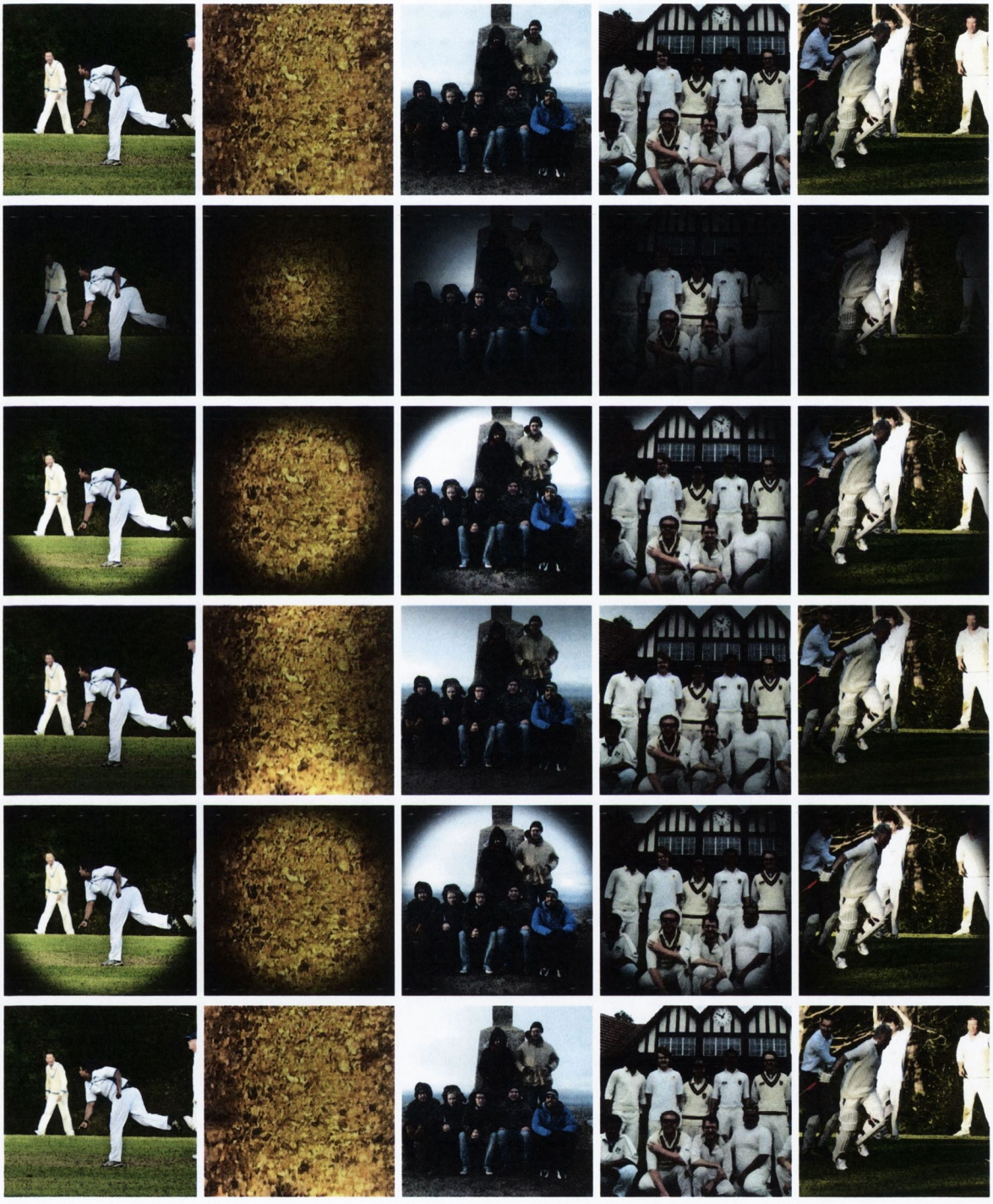


Figure 3.9: (1^{st} row) Original and (2^{nd} row) degraded images, along with corrections from (3^{rd} row) previous [45], and (4^{th} row) proposed methods, with motion in either x or y directions, and corresponding corrections with motion in both directions in (5^{th} row) and (6^{th} row) respectively.

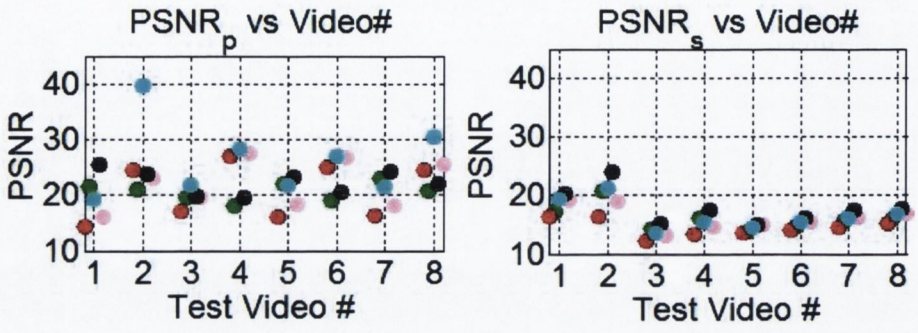


Figure 3.10: PSNR between the original, undegraded sequence and the degraded sequence for each test video and for each camera response. The result obtained from each camera response are colour coded: Sony (pink), Kodak (green), $\gamma = 0.8$ (black), $\gamma = 1.2$ (red), NONE (blue). Left shows the PSNR for the proposed algorithm and the right the PSNR for the previous work.

corrected images obtained from the test videos with and without motion in both directions look identical, as seen in fifth and third rows in Figure 3.9. The reason for this robustness in the previous method is because their parameters are estimated in one dimension along the radii, as opposed to the proposed method where they are estimated along the x and y directions. This key robustness is however not reflected in the PSNR values, as the values obtained from the previous method are generally lower than those from the proposed method, due to the residual peripheral degradations phenomenon described earlier.

The second key observation is consistent positive and negative offsets in the shape parameter estimates obtained from the gamma curves of 0.8 and 1.2. These consistencies correspond to the boosting and reduction in radiance values performed by the respective gamma curves, which show the proposed method can account for the changes from response curves that are perfect exponential functions. Good center estimates are also obtained, which average 2% in error, and similar values are obtained for each camera response. The low errors in these estimates are due to the minimal attenuation that occurs in the vicinity of the center location, which make it difficult to pin point the exact location. Whereas the similar values obtained implies the response function has minimal impact on the estimation of this parameter. The last key observation is that the best center estimates are obtained from the first two videos in all of the test sets. The main difference in these two videos, compared to the rest, is they both have very little colour and texture variations. This shows that apart from motion, colour and texture variations also affect the performance of the algorithm, to a small extent. As a result of these good center estimates, the second test video gave the highest PSNR value (for $\gamma = 1$).

3.3.3 Shape

The robustness of the proposed method to correct different elliptical shaped degradations is now examined. To mimic the oval shaped degradations seen in the actual data, the test data for this

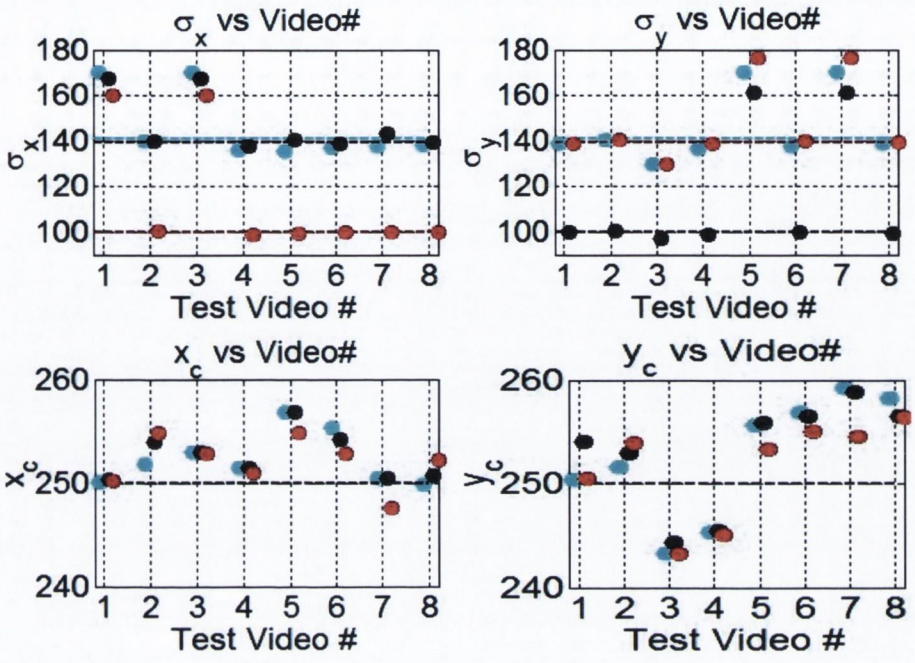


Figure 3.11: Parameter estimation results for the proposed algorithm applied to each test video (along horizontal axis). The result obtained from each shape degradation are colour coded: elliptical horizontal (black), elliptical vertical (red), and circular (blue). The horizontal lines show the ground truth parameter values, which are colour coded to correspond to the colour of the respective shape degradation.

experiment is created by applying vignetting functions to the ground truth videos with shapes that are elliptically: i) horizontal, $\mathbf{V} = \{140, 0; 0, 100\}$, and ii) vertical $\mathbf{V} = \{100, 0; 0, 140\}$. For these two test sets, the rest of parameters are kept constant by using the base center and response functions values, $\mathbf{c_o}$, and γ_o. Plots of the mean parameter estimates and PSNR values are shown in Figures 3.11 and 3.12, and degraded test images and the corresponding corrections obtained from the proposed and previous [45] methods, are given in Figure 3.13.

The results show similar trends observed in the previous experiment (section 1.3.2). Again the videos 1,3 showing motion in one direction prevent the estimation of shape to be as accurate as the in the other sequences. This consistency and low errors imply the proposed algorithm can estimate elliptical and circular shaped degradations effectively, and with similar efficiency. Another observation made is the minor fluctuations among the center estimates and PSNR values obtained for each different degradation. These minor fluctuations show the shape of the degradation does have a small influence on the performance of the algorithm. Another interesting item observed is the saturated regions in the corrected images from the previous method for the elliptically shaped degradations, as seen in the third row in Figure 3.13. These saturated regions are however removed (see 4th row in Figure 3.13), and higher PSNR values

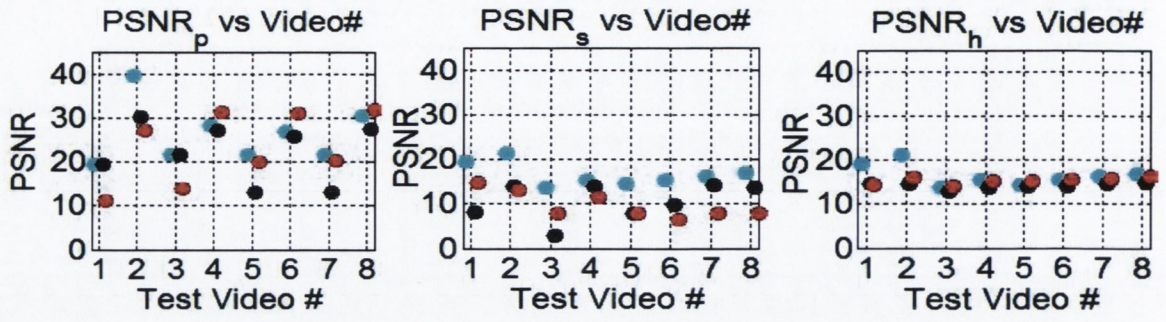


Figure 3.12: PSNR between the original, undegraded sequence and the degradad sequence for each test video and for each shape degradation. The result obtained from shape degradation are colour coded: elliptical horizontal (black), elliptical vertical (red), and based degradation which is circular (blue). Left shows the PSNR for the proposed algorithm, the middle is the original result from the previous work by Kim & Pollefeys [45], and the left is their result after incorporating the shape estimates, $\mathbf{V}$ from the proposed method.

are obtained, when parameter estimation and correction of their method is performed along the radii values of the proposed method. As these radii use the shape ($\mathbf{V}$) estimates, this illustrates the need for accurate shape estimates when correcting these degradations.

3.3.4 Center Location

The central location for these degradations may not necessarily be the image center, but instead is the point on the sea floor where all of the light sources are focused towards. In this experiment, the robustness of the proposed algorithm to correct degradations with different central locations will be examined. The test data for this experiment is created by degrading three sets of the ground truth videos with functions that have center locations at the i) at the image center ($B_o(\mathbf{x})$) $\mathbf{c} = \{250, 250\}$ ii) bottom left $\mathbf{c} = \{180, 310\}$, and iii) bottom right $\mathbf{c} = \{310, 310\}$. For these three test sets, the rest of parameters are kept constant by using the base degradation shape and response function values, $\mathbf{V_o}$, and γ_o. Plots of the mean parameter estimates and PSNR values are shown in Figures 3.14 and 3.15, and degraded test images and the corresponding corrections obtained from the proposed and previous [45] methods, are given in Figure 3.16.

Analysis of these results show the average error for the center estimates for the bottom left and right test sets are 1.3% and 1.0% more, than the average value of 2.0% obtained from the base degradation (blue in Figure 3.14), which is located at the image center. These minor errors show the proposed algorithm can estimate the center location quite effectively for these type of degradations. It is also observed that the different central locations examined have minimal impact on the shape estimates, as almost identical values are obtained to the base degradation. Because of these similar errors among the paramater estimates, the PSNR values obtained from for each test set are also very similar. Another interesting observation is the saturated regions

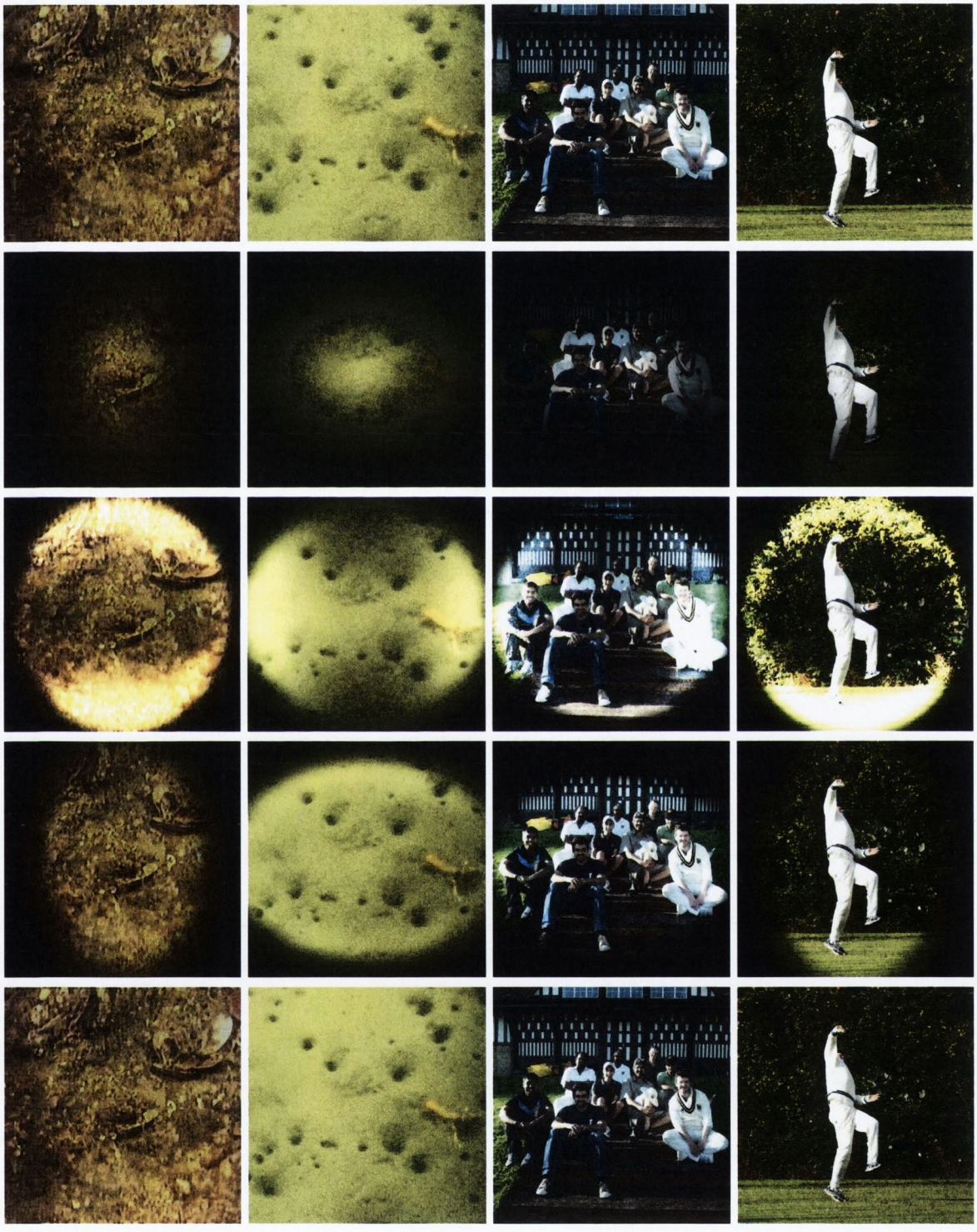


Figure 3.13: (1^{st} row) Original and (2^{nd} row) degraded images, along with corrections from previous method [45] using (3^{rd} row) circular radii, and (4^{th} row) radii with shape from $\mathbf{V}$ estimate, and lastly (5^{th} row) from proposed method with motion in both x and y directions.

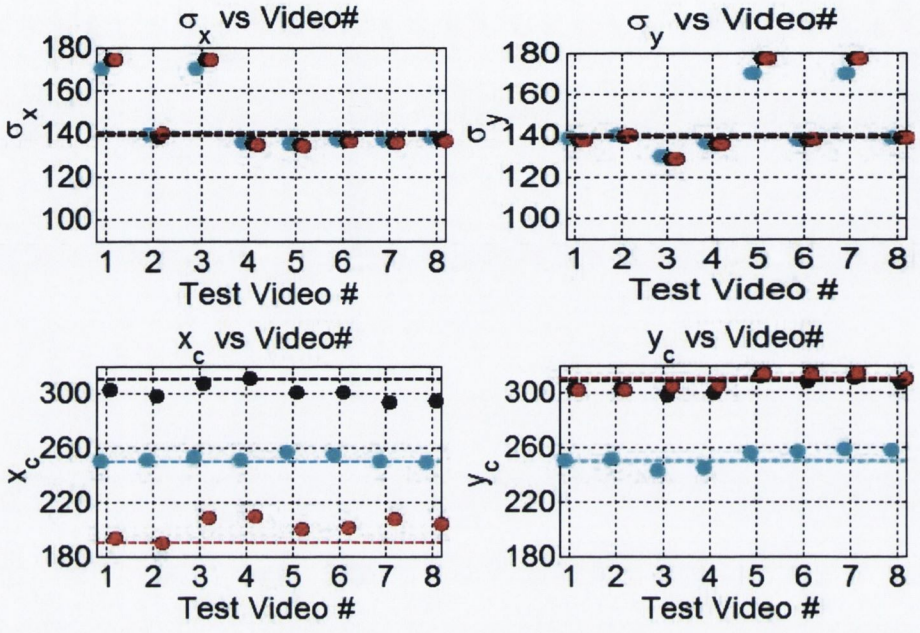


Figure 3.14: Parameter estimates obtained from the proposed algorithm for each video and from each test treatment. Each test treatment with respective central location is colour coded as: bottom left (red), bottom right (black), and at the image center (i.e. the base degradation $B_o(\mathbf{x})$) is blue. The horizontal lines show the ground truth parameter values, which are colour coded to correspond to the colour of the respective treatment.

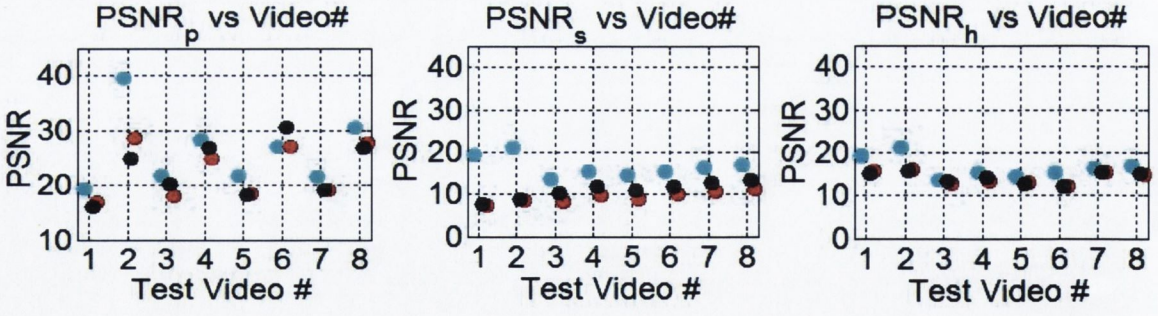


Figure 3.15: PSNR between the original, undegraded sequence and the degradad sequence for each video and from each test treatment. Each test treatment with respective central location is colour coded as: bottom left (red), bottom right (black), and at image center (blue). Left shows the PSNR for the proposed algorithm, the middle is the original result from the previous work by Kim & Pollefeys [45], and the left is their result after incorporating the center estimates, $\mathbf{c}$.

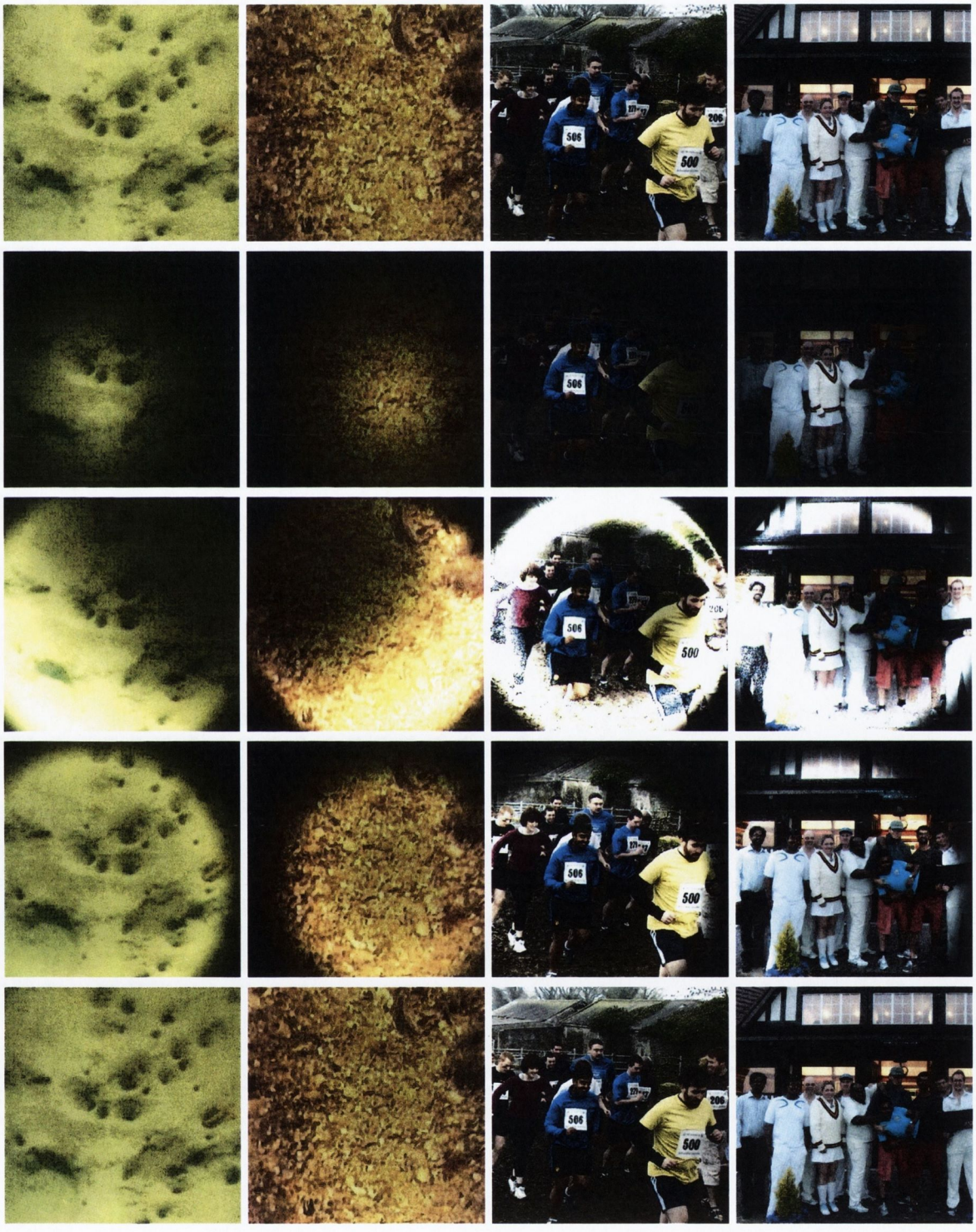


Figure 3.16: (1^{st} row) Original and (2^{nd} row) degraded images, along with corrections from previous method [45] using (3^{rd} row) circular radii, and (4^{th} row) radii with shape from **c** estimate, and lastly (5^{th} row) from proposed method with motion in both x and y directions.

obtained from the previous method by Kim & Pollefeys [45], for the degradations whom center locations are not the image center. In an identical fashion to the previous experiment, these saturated regions are eliminated when the parameter estimation and correction procedures of their method are performed along the radii of the proposed method that uses the $\mathbf{c}$ estimates. The elimination of these saturated regions show that good center estimates are needed for correcting these type of degradations effectively.

3.3.5 Footprint

As mentioned in the introduction of this chapter, the intensity of the light beams used in some of these surveys are so great that they create a footprint region on the sea floor where almost no degradation occurs. The importance of incorporating this region into the correction procedure, and how effective the proposed algorithm is at estimating it will be examined in this experiment. To perform this examination three sets of the ground truth videos are degraded with footprint regions at $r_f = 0.4$, $r_f = 0.6$, and $r_f = 0$ (i.e. the base degradation, $B_o(\mathbf{x})$). The first region at $r_f = 0.4$ is chosen because this is the typical size seen in the underwater survey videos. While the second instance at $r_f = 0.6$ is used to examine if a more extreme case would have any impact on the performance of the proposed algorithm. In the three test sets, the rest of parameters are kept constant by using the base degradation center, shape and response functions values, $\mathbf{V_o}$, $\mathbf{c_o}$ and γ_o. Plots of the mean parameter estimates and PSNR values are shown in Figures 3.17 and 3.19, and degraded test images and the corresponding corrections obtained from the proposed and previous [45] methods, are given in Figure 3.18.

Analysis of these results show the proposed algorithm is able to detect and estimate the extent of the footprint regions accurately in all of the test sets used. It is also observed that these footprint regions have minimal impact on the shape and center estimates, as similar values are obtained to the base degradation (blue), which does not have any footprint region. However, a small deteriorating pattern in the shape estimates is noticed with increasing footprint radii, as seen in top left most plots in Figure 3.17. For example, with no footprint region the error in the shape estimates average 2.2%, but with regions of size $r_f = 0.4$ and $r_f = 0.6$, this value increases to 3.5% and 4.8% respectively. The deterioration in these estimates can be explained by the rapid change in degradation that occurs when transiting out of the footprint region, which increases with the size of this region. Another interesting observation is the over amplification artifacts that occur within the footprint region when it is not incorporated into the proposed framework, as seen in the fourth row in Figure 3.18. These artifacts also are also seen in the corrections obtained by the previous method by Kim & Pollefeys [45], along with a degrading pattern in PSNR values with increasing footprint size. These over amplifications show the use of the footprint estimate is very important for correction these type of degradations efficiently.

3.3.6 Colour Degradation

The efficacy of the proposed algorithm to correct colour degradations is now examined. The test data for this experiment is created by degrading the red channel in three sets of ground truth videos with colour degradation parameters, $m_r = \{0.8, 0.8, 1\}$. To examine if the footprint region has any influence, the first test set is given a footprint region at $r_f = 0.4$. The rest of parameters are kept constant by using the center, shape and response function from the base degradation, $\mathbf{V_o}$, $\mathbf{c_o}$ and γ_o. Plots of the mean parameter estimates and PSNR values are shown in Figures 3.20 and 3.21, and degraded test images and the corresponding corrections obtained from the proposed and previous [45] methods, are given in Figure 3.22.

Examination of these results show the colour degradations have minimal impact on the shape, center, and footprint estimates, as similar values are obtained to the base degradation, which does not have any colour degradation. The footprint region does however affect the colour estimates, as only the test set that contained it gave an average error of 5.27% in m_r, while the other two sets without it are accurately estimated. The rise in error is because the intensity values of the red channel, outside of the footprint region in this degradation, are too low to optimize the parameter effectively. As a result of these erroneous estimates, the degradations are not fully corrected, as seen in the fourth row of Figure 3.22. A larger percentage of these

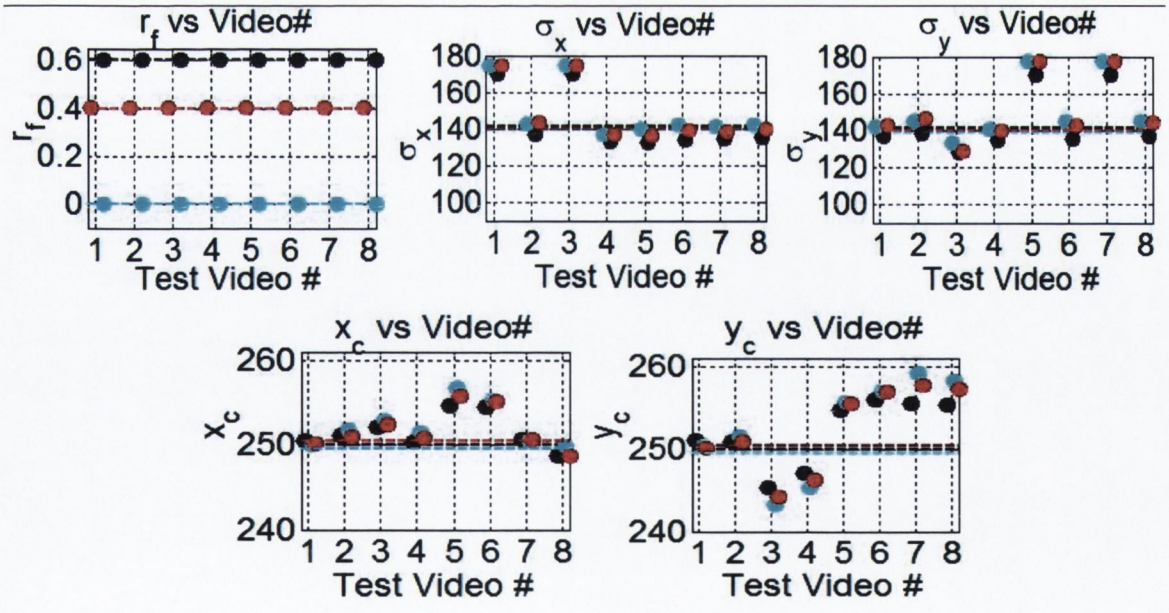


Figure 3.17: Parameter estimates obtained from the proposed algorithm for each video and from each test treatment. Each test treatment with respective footprint is colour coded as: $r_f = 0.6$ (black), $r_f = 0.4$ (red), and $r_f = 0$ (i.e. the base degradation $B_o(\mathbf{x})$) is blue. The horizontal lines show the ground truth parameter values, which are colour coded to correspond to the colour of the respective treatment.

Figure 3.18: (1^{st} row) Original and (2^{nd} row) degraded images, along with corrections from the (3^{rd} row) previous method [45], and the proposed method (4^{th} row) without and (5^{th} row) with footprint compensation (from test videos with motion in both x and y directions).

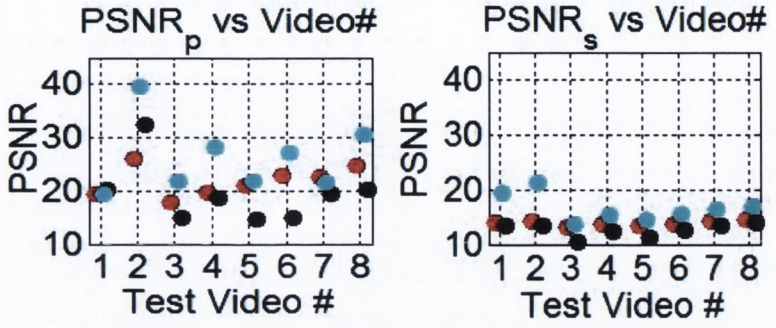


Figure 3.19: PSNR between the original, undegraded sequence and the degradad sequence for each video and from each test treatment. Each test treatment with respective footprint is colour coded as: $r_f = 0.6$ (black), $r_f = 0.4$ (red), and $r_f = 0$ (i.e. the base degradation $B_o(\mathbf{x})$) is blue. Left shows the PSNR for the proposed algorithm, and the right is the result from the previous work by Kim & Pollefeys [45].

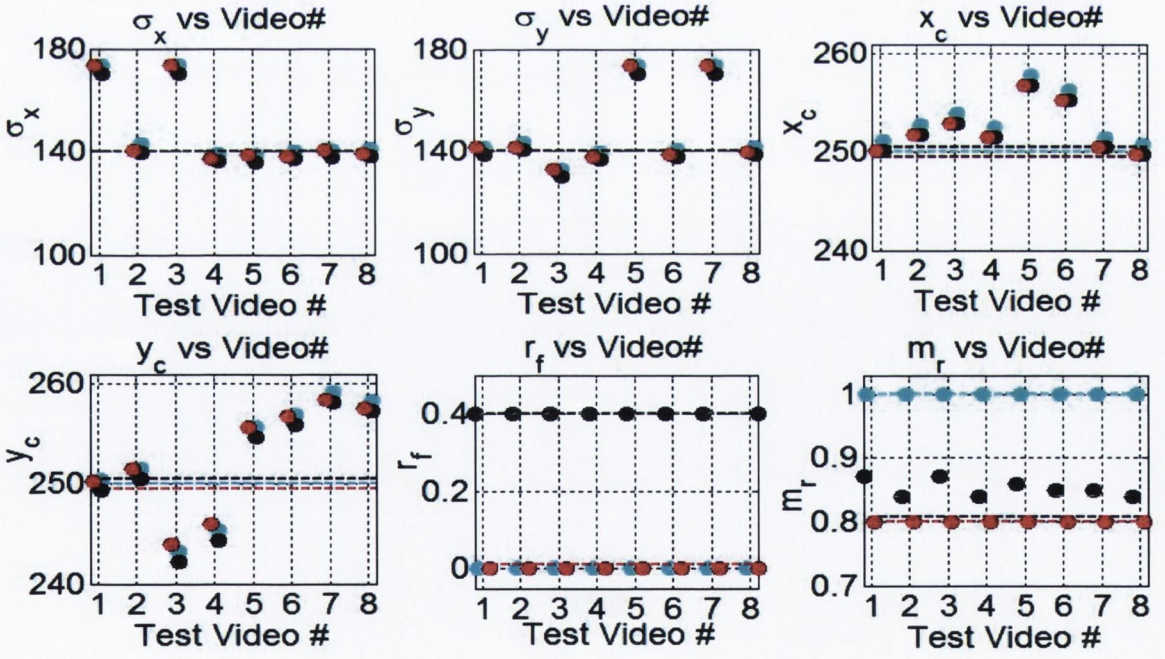


Figure 3.20: Parameter estimates obtained from the proposed algorithm for each video and from each test treatment. Each test treatment with respective red channel degraded is colour coded as: $m_r = 0.8$ with footprint at $r_f = 0.4$ (black), $m_r = 0.8$ (red), and $m_r = 1$ (i.e. the base degradation $B_o(\mathbf{x})$) is blue. The horizontal lines show the ground truth parameter values, which are colour coded to correspond to the colour of the respective treatment.

residual colour degradations are observed in the results obtained from the previous method by Kim & Pollefeys [45], in the third row of Figure 3.22. From these results obtained, it can be concluded that the proposed algorithm can estimate colour degradations similar to those used in this experiment accurately, but in cases of severe degradation that involve footprint regions, its accuracy may deteriorate.

3.3.7 Actual Degraded Videos

Having showed the proposed algorithm can effectively correct the simulated degradations in the previous experiments, it is now tested on twenty six actual underwater survey videos of the seabed. These videos are of PAL format, and have a wide variety of degradations. Twenty of these videos are surveys from different Nephrops habitats, which when corrected will be used to generate mosaics and perform content analysis on in later chapters of the thesis. While the remaining six videos are from a wide variety of underwater environments, which are diverse in colour and texture characteristics due to their different seabed sediments such as sand, rocks, shells, clay etc. Additional characteristics of these test videos are given in Table 3.2, and sample images along with results obtained from the previous and proposed methods are shown in Figures 3.23, 3.24, and 3.24, with additional sets in Appendix A.

Analysis of these results show the proposed algorithm can rectify the colour and illumination degradations in actual underwater survey videos and hence improve their visibility significantly. In contrast, the previous method does produce similar results for specific degradations that are approximately circular in shape, centered close to the image center, and with no or minimal footprint region, as shown in the Nephrops sequence in the sixth row in Figures 3.23 and 3.24. But if the degradations deviate largely from these specifications, over amplification usually occurs. This over boosting is however reduced using the hybrid scheme discussed in section 1.2.3,

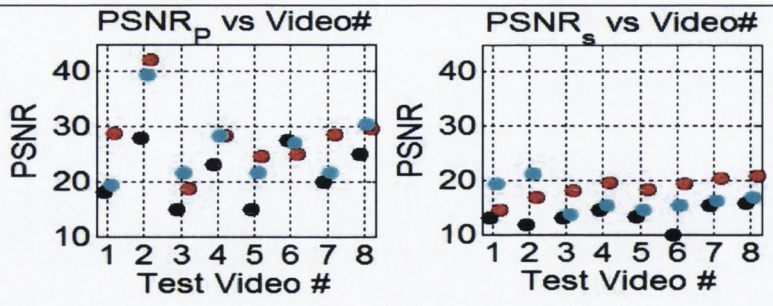


Figure 3.21: PSNR between the original, undegraded sequence and the degradad sequence for each video and from each test treatment. Each test treatment with respective red channel degraded is colour coded as: $m_r = 0.8$ with footprint at $r_f = 0.4$ (black), $m_r = 0.8$ (red), and $m_r = 1$ (i.e. the base degradation $B_o(\mathbf{x})$) is blue. Left shows the PSNR for the proposed algorithm, and the right is the result from the previous work by Kim & Pollefeys [45].

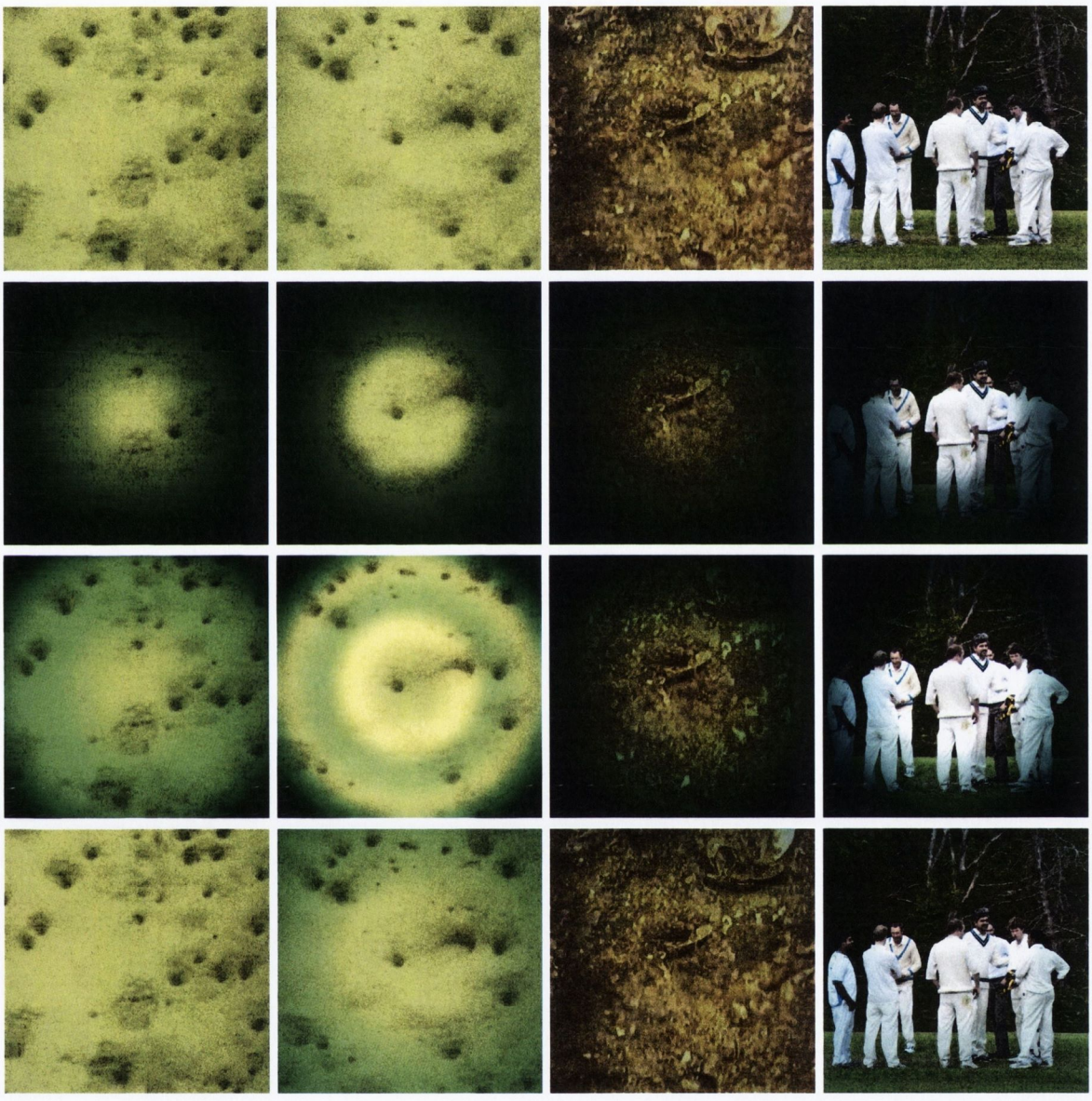


Figure 3.22: (1^{st} row) Original and (2^{nd} row) degraded images, along with corrections from the (3^{rd} row) previous [45], and (4^{th} row) proposed methods (from test videos with motion in both x and y directions).

where the shape and center location estimates from the proposed algorithm are incorporated into their correction procedure. The third and fourth columns in Figures 3.23 and 3.24 show examples of this over boosting, and its corresponding reduction when using this hybrid approach.

Inspecting these results also show three general scenarios where both the proposed and previous methods do not perform effectively. First is when the image intensities are degraded into very small values, so that amplifying them to the original values results in white noisy regions, as seen in the clay type seabed corrections in the fourth row of Figure 3.24. The second scenario is when the actual degradation deviates largely from the models, such as cases when isolated sections of negligible degradation occur outside of the footprint region. These cases usually occur when relatively tall objects on the sea floor capture greater illumination than their surrounding regions because of their closer proximity to the light source. As a result of having greater illumination, when the proposed and previous models are used, the pixels in these sections are over amplified in the corrected images, as shown in the boulder type seabed corrections in third row of Figure 3.23. The last scenario when both the proposed and previous methods do not perform effectively, is when the degradations are not symmetrical due to the light beams not pointing vertically towards the sea floor. For this case, sections of image with the greater degradations are not fully rectified, as shown in the sand wave type seabed corrections in second row of Figure 3.23.

3.3.8 Conclusion

This chapter shows that it is possible to correct illumination degradations in underwater seabed survey videos by combining ideas from the vignetting and underwater colour correction literature. Apart from this novel vignetting correction approach, three other key contributions are made that improve on the state of the art technique by Kim & Pollefeys [45]. First, is the introduction of a new degradation model that does not restrict the shape and central location of the deteriorations to being circular and centered at the image center. Instead, this model allows the deteriorations to be elliptical in shape and have a central location at any image position. In addition to these new shape and center parameters, the algorithm also parameterizes the colour

Test Video	1	2	3	4	5	6	7-26
Seabed Description	Sand-waves	Mud, weed	Mud, Sand	Shells, pebbles, rocks, cobbles	Mud, sand and gravel	Rocks, Boulders	Muddy (Nephrops)
Frames	328	654	1986	2097	2482	2778	1500 each
Sample Image	row 2 in Fig 3.23	row 4 in Fig 3.24	row 1 in Fig 3.23	row 4 in Fig. 3.23	row 3 in Fig. 3.23	row 5 in Fig. 3.23	row 6 in Fig. 3.23

Table 3.2: Characteristics of underwater survey videos used for testing

degradations that occur in this type of environment due to the properties of the water molecules. The second key contribution is the estimation procedure for these parameters, which is a linear approach that uses point correspondences, and incorporates information from previous estimates for robustness. The last key contribution is the correction method introduced that takes into account the footprint region of the light beam on the sea floor where minimal or no degradation occurs. This method involves estimating the extent of the footprint region and incorporating it into a correction mask where pixels within this region are not boosted.

From the experiments performed in this chapter, five conclusions can be made. First, the camera response function does affect the final appearance of the corrected image, and hence incorporating it into the proposed method may potentially improve results. Secondly, there needs to be motion along the x and y directions for the proposed algorithm to perform effectively, as without it the shape parameters are not fully optimized. In addition to motion, the high degradation rate that is present when transiting out of the footprint region also affects the shape estimates. Third, accurate shape, center location, colour deterioration, and footprint region estimates are very important for this application, as without them the degradations are not corrected effectively. Lastly, for degradations that are non-circular and/or not centered at the image center, the performance from the previous method by Kim & Pollefeys [45] can be substantially improved by incorporating these two variables from the proposed method into their scheme.

Although these initial results show the proposed method can enhance visibility in actual underwater sequences, it can be improved in the future by incorporating i) depth information, ii) a tilting parameter, and iii) user initialization, as follows. If depth information is available, then the radial fall off in light can be more accurately modelled with equations such as the cosine cubed law [37]. With the use of these equations the over boosting problem that occurs with relatively tall objects on the sea floor (see third row in Figure 3.24), may be avoided. Whereas a tilting parameter may be helpful for modeling the asymmetric degradation that occurs when light beams are slanted towards the sea floor, as seen in the second row in Figure 3.24. Then, for sequences with very low initial motion, the problem of efficient parameter optimization may be avoided by allowing the user to initialize these parameters. To incorporate the depth and tilting parameters effectively into the proposed model, spectral analysis at measured depths underwater, similar to the experiments performed by Bongiorno et al. [8], would have to be performed.

As a practical point of interest, the marine scientists involved in this project, Jennifer Doyle and Colm Lordan from the Marine Institute, Galway, agree that the visibility is improved in the corrected images. In the next chapter details of the video mosaic generation procedure will be presented. Here the significance of using this correction procedure prior to generating the mosaics from the video sequences will be examined.

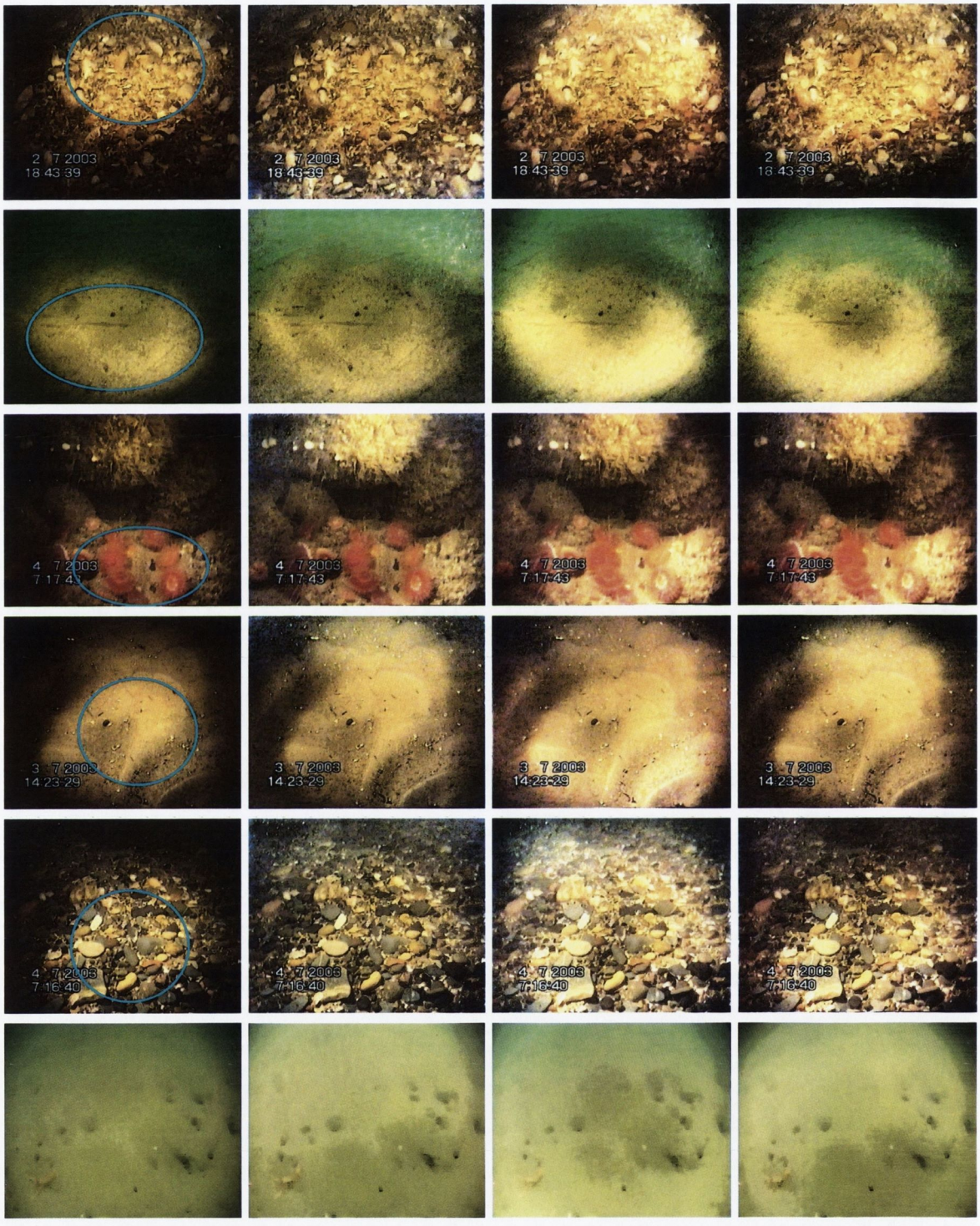


Figure 3.23: (1^{st} Column) Original degraded images with footprint superimposed in blue. (2^{nd} Column) Proposed correction. (3^{rd} Column) Kim & Pollefeys [45] original correction. (4^{th} Column) Correction after incorporating $\mathbf{c}$ and $\mathbf{V}$ estimates into Kim & Pollefeys [45] method.

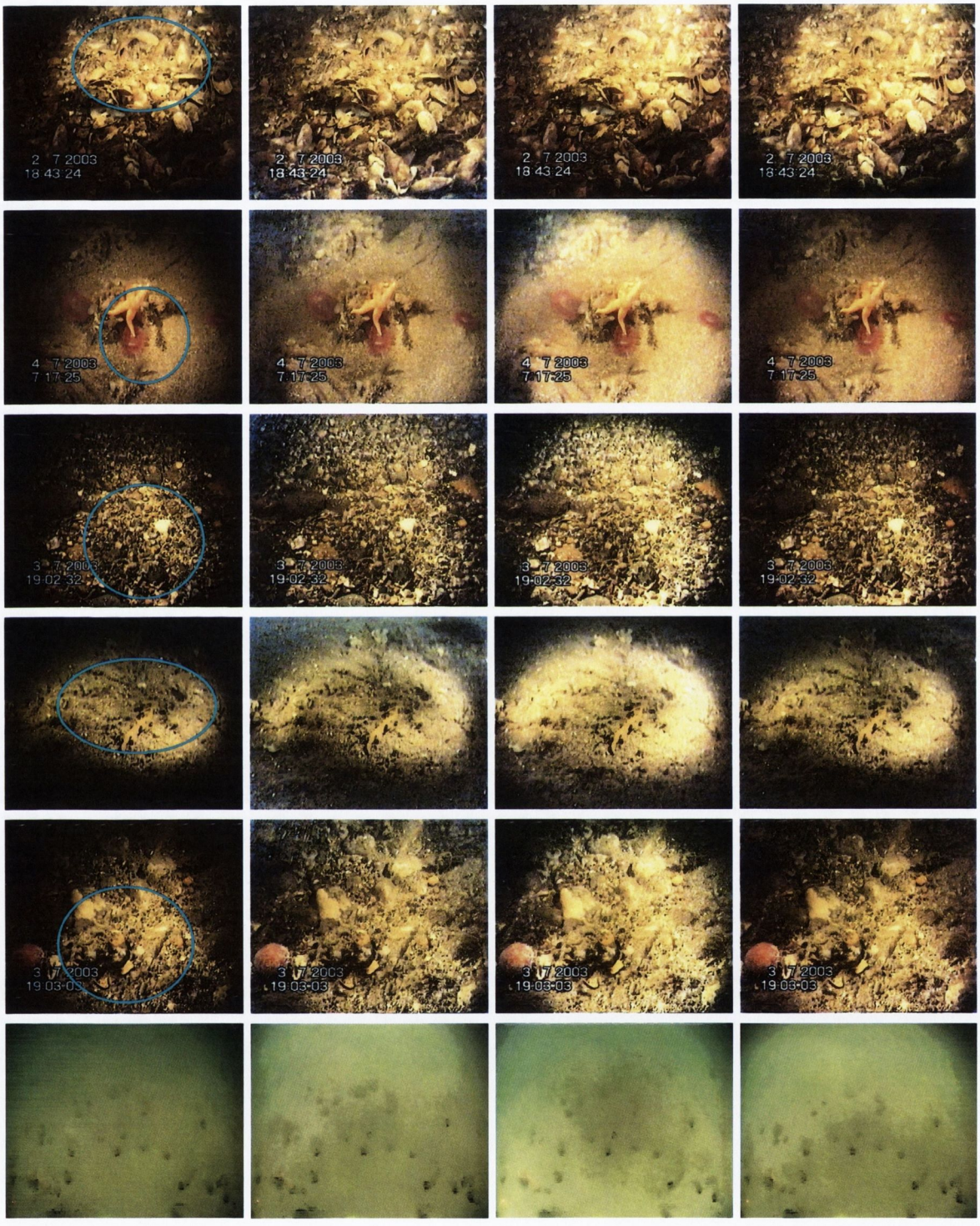


Figure 3.24: (1^{st} Column) Original degraded images with footprint superimposed in blue. (2^{nd} Column) Proposed correction. (3^{rd} Column) Kim & Pollefeys [45] original correction. (4^{th} Column) Correction after incorporating $\mathbf{c}$ and $\mathbf{V}$ estimates into Kim & Pollefeys [45] method.

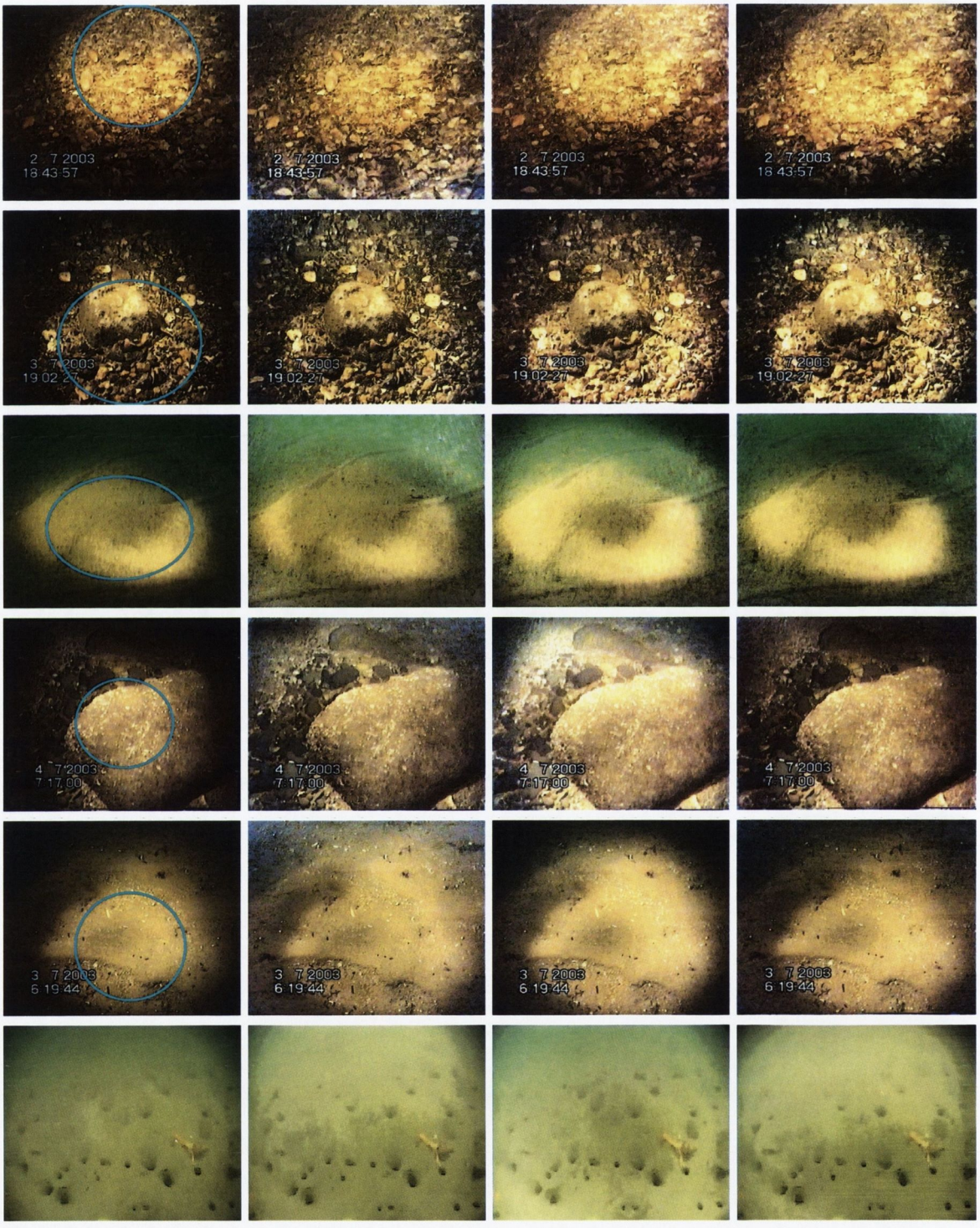


Figure 3.25: (1^{st} Column) Original degraded images with footprint superimposed in blue. (2^{nd} Column) Proposed correction. (3^{rd} Column) Kim & Pollefeys [45] original correction. (4^{th} Column) Correction after incorporating $\mathbf{c}$ and $\mathbf{V}$ estimates into Kim & Pollefeys [45] method.

4

Mosaics From Marine Videos

Although the vignetting correction does improve visibility, the types of videos considered here still suffer from a small field of view, as they are recorded in close proximity to the sea floor. Because of this restricted view, scientists sometimes encounter difficulty in spotting spatial relationships among the captured Nephrops burrows, which is vital for the species population census [12]. To solve this problem, a large area view of the surveyed sea floor can be created by aligning and rendering overlapping video frames together to form a mosaic. Apart from being a useful analysis tools for scientists, these mosaics improve visibility tremendously, as seen in Figure 4.1. In this chapter the various steps undertaken in this work to generate these high quality mosaics from the underwater survey videos will be discussed.

The creation of mosaics is a mature topic in the literature, and thus relatively little recent work was found in this domain. Most of the existing techniques [9, 26, 62, 65] describe image mosaicking as a two stage procedure comprising of first aligning the respective images (video frames) and then rendering their overlapping regions. The alignment is usually performed using either feature [9,26,65] or pixel [79] matching methods. For the Nephrops survey videos however, with uneven lighting, low texture and blurred content, these techniques may degrade in accuracy. The rendering of the mosaic using the aligned images has traditionally been performed using either the weighted mean [65] or median [26] of aligned pixels. For any of these techniques to perform well, the overlapping regions in any aligned frames should contain more or less the same information, albeit corrupted by noise. However in this case, uneven lighting violates this assumption and so can cause blurring and ghosting artifacts in the subsequent mosaics. This deterioration in image quality can be prevented by selecting sections of the overlapping

regions from different frames [9,62]. For this method of combination, selection is made from the particular frame where the overlapping regions is located closest to the image center [9], as it is here where the best image quality is perceived to exist. For these sequences however, the best image details are located within the well lit regions of the frame, which is not necessarily the image center.

To overcome the possible alignment and rendering problems described above, three key contributions are made in this work. First, is the development of a Bayesian framework for registration which takes advantage of the properties of the seabed videos. Secondly, to improve image alignment, a pixel matching technique is used to refine the results obtained from feature matching. To cope with the uneven lighting in these images, this hybrid image alignment approach is performed in the difference of Gaussians image, where the influence of absolute brightness has minimal effect. The third key contribution is the use of the center of the light beam footprint on the sea floor to capture well-lit image details in the generated mosaic. Apart from these thee key contributions, a simple system to cross reference sections of the generated mosaic with the original video frames, is also introduced to aid scientists with their analysis.

The proposed hybrid image alignment procedure is given in the next section, followed by details of the new rendering and referencing systems. Then, an evaluation of the proposed work using both synthetic and real data is then presented, along with comparisons from three state of the art mosaicking techniques, developed by Brown and Lowe [9], Gracias and Santos-Victor [26], and Li et al. [65]. Lastly, a discussion on the most relevant and interesting results obtained from the experiments performed are given in the conclusion, with suggestions for future improvement.

4.1 Underwater Video Mosaicking

4.1.1 Image Alignment

The first stage in generating a mosaic is to align and map all frames to a common reference frame, which in most cases is chosen as the first frame in the video sequence. This mapping is performed with the global registration model used by Gracias and Victor [26], given by:

$$\mathbf{H}_{r,k} = \mathbf{H}_{r,1}\Pi_{i=1}^{k-1}\mathbf{H}_{i,i+1} \tag{4.1}$$

where $\mathbf{H}_{r,k}$ is the homography mapping between the k^{th} frame, and the reference frame, $\mathbf{F}_r$. As seen from equation 4.1, to achieve this mapping the frame-to-frame homographies, $\mathbf{H}_{k-1,k}$, must first be estimated. Estimating the frame-to-frame homographies is the most crucial step in the mosaicking process, as misaligned images usually result in a deterioration in image quality in the generated mosaic. Also errors accumulate over time. This means there is a huge premium on robustness, as one failure means the subsequent frames will be misaligned with the reference frame. The key to aligning two images is to estimate a linear mapping between them that relates the location of features, u_1, in the first image to their respective location in the second image, u_2.

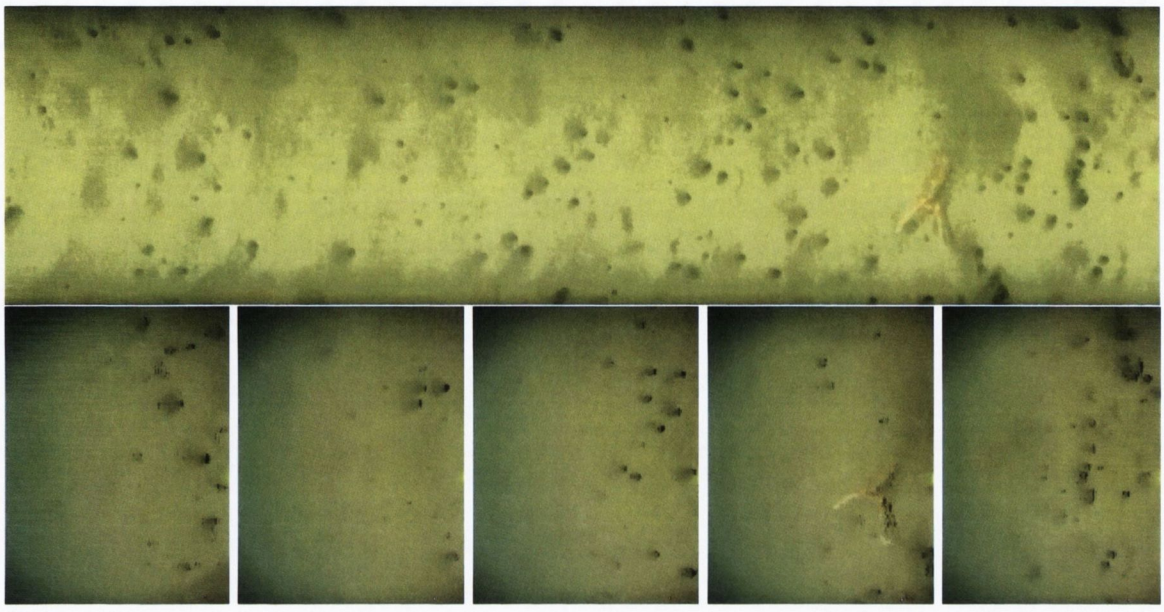


Figure 4.1: (top) Mosaic generated from 200 frames, samples of which are shown in bottom row.

In planar Euclidean geometry this mapping is referred to as a homography, $\mathbf{H}$, and is described by a non-singular 3×3 Affine transformation matrix, given by:

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \tag{4.2}$$

$$\mathbf{u_2} = \mathbf{H_{2,1}}.\mathbf{u_1}$$

Affine transformations take into account various types of motion such as translation, rotation, scaling, and shearing, and is sufficient to describe the homography between consecutive frames in this application. To estimate the homography, the common feature-based approach [9, 26, 53] is adopted, where equation 4.2 is optimized with matching features between the respective images.

The Figure 4.1 shows that the typical images from these sequences have very little strong features which would normally be used in motion estimation for mosaicking e.g. corners. To resolve this issue, the blurry burrows themselves are used as features, through the use of Maximally Stable Extremal Region (MSER) [53] features, sample extractions of which are shown in Figure 4.2. This represents a unique use for MSER features in this application. Additionally, these features are robust to illumination changes, as shown in [53]. With this feature based approach, images are aligned in three steps of: i) feature extraction, ii) matching, iii) homography estimation, which are detailed as follows.

Feature Extraction

Usually MSERs are extracted from the gray scale version of the image, but because of the uneven

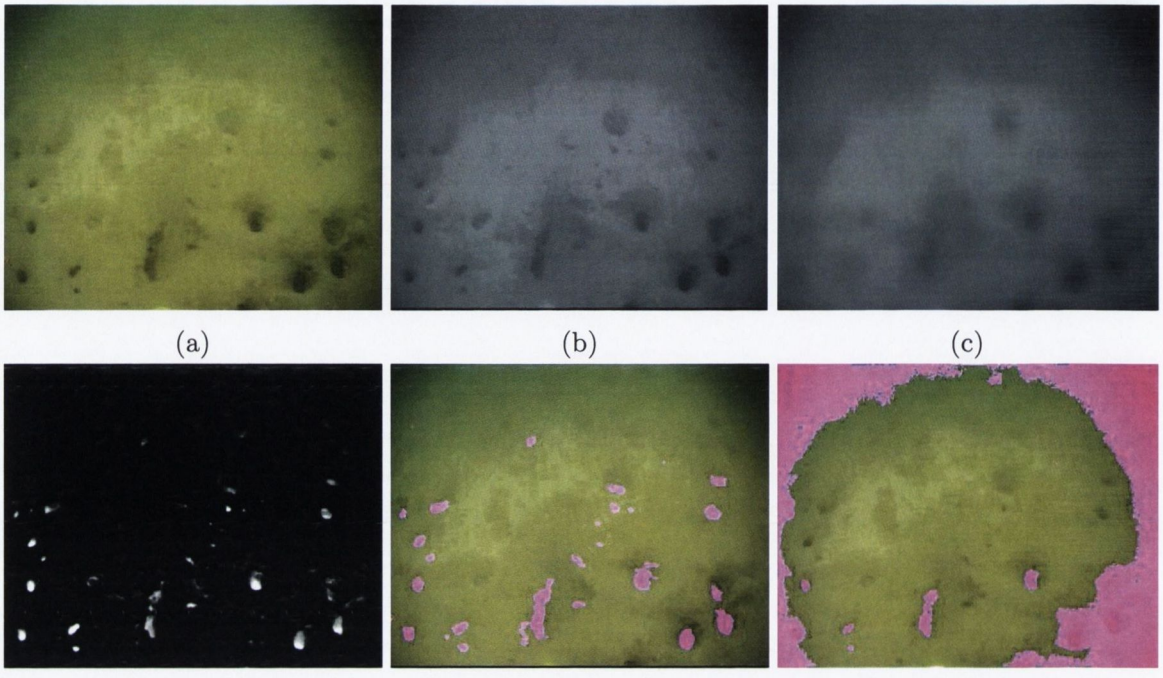


Figure 4.2: Creating the DoG image (d), to perform blob detection by subtracting a lightly blurred version (b) of the original image from a heavily burred one (c). Extracted MSER obtained from the DOG and gray scale image directly are shown in (e) and (f) respectively.

lighting in these scenes, the difference of Gaussians space is utilized, $I_g = I * G_1 - I * G_2$. Where G_1 and G_2 are two dimensional Gaussian functions with taps of 71 and 5, and corresponding variances of 30 and 2 respectively. The large value of G_1 is chosen in relation to the average burrow diameter (71 pixels), so that most of these objects would be effectively blurred out. With the burrows blurred out, an equivalent homogeneous sandy background image is created, and when the lightly blurred version ($I * G_2$) is subtracted from it, all of the burrow regions are highlighted. Using this contrast space, the MSER features are extracted using the techniques developed by Matas et al. [53]. These respective extraction steps are illustrated in Figure 4.2.

Matching

The extracted features from each MSER in the two respective frames to be aligned are now matched using the two stage process developed by Matas et al. [53]. In the first stage, each MSER in the first image is compared to its neighbors in the second image within a set radius using a voting scheme based on their features. To obtain optimum results, the search radius is set to a value of 75 pixels, as beyond this range matching objects do not usually occur in these types of videos. This range is chosen in relation to the average: i) burrow diameter (71 pixels), and ii) global motion (60 pixels) in these videos. Erroneous matching blobs with high votes are

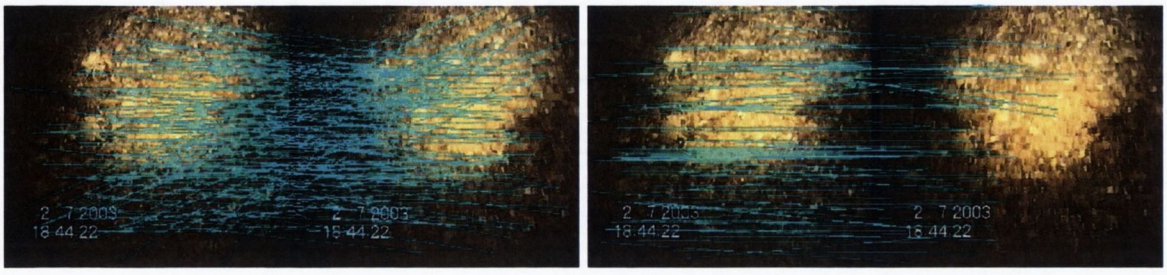


Figure 4.3: Matching Results. Original set of matching MSER features (Left) containing a large set of mismatches indicated by the crossing blue lines. Corresponding set of inliers obtained using RANSAC (Right), which are very good matches as their direction is uniform.

then eliminated in the second stage of this process using normalized cross correlation. Further details of this matching scheme can be found in [53].

Homography Estimation

The homography between the two respective frames is now estimated using their matching blobs. In this estimation a Bayesian framework is employed where the posterior, $p_o(H|u_1, u_2)$ is maximized with respect to the center of mass of the matching blobs, u_1, u_2, as follows.

$$p_o(H|u_1, u_2) \propto p_l(u_1, u_2|H)p_h(H) \tag{4.3}$$

The likelihood $p_l(\cdot)$ is derived directly from equation (4.2), and Gaussian priors, $p_h(\cdot)$ are used for each parameter as:

$$p_l(u_1, u_2|H) \propto \exp -\left[\frac{(\sum(u_2 - Hu_1)^2}{2\sigma_e^2}\right]; \quad p_h \propto \exp -\left[\frac{\sum(H - H_0)^2}{2\sigma_H^2}\right] \tag{4.4}$$

where H_o are the parameter estimates from the previous two frames, which is assumed to be very similar the current estimates. To enforce this similarity, σ_H^2 for each parameter is set to 25% of its previous value, and σ_e^2 is set to 1. Hence, with the likelihood and prior functions, the posterior given in equation 4.3 is now differentiated w.r.t. the relevant unknowns, set to zero, and solved using Singular Value Decomposition (SVD) [86].

Optimization. In practice however, the set of matching points does contain mismatches. Thus to robustly estimate the homography, the technique of RANSAC (Random Sample Consensus) [31], is now employed to select the best set of matching points, as illustrated in Figure 4.3. The idea behind RANSAC is to continually subdivide the full set of matching points into inliers and outliers based on the homography estimated with a trial set of 4 randomly sampled points at a time, until a significantly large set (i.e. $> 75\%$) of inliers is obtained. If this set is obtained, the homography is then re-estimated with it, and the process is terminated, otherwise it will keep on being repeated up to a maximum of $n = 200$ trials. In each trial, the Bayesian approach given in equation 4.3 is utilized for estimating the homography, and the inliers are set as those in which

the Euclidean distance, $d = \sqrt{(\mathbf{u_2} - \mathbf{Hu_1})^2}$, are below some threshold value, $T_d = 1$. For this application, if the maximum number of trials has occurred, and the largest set of inliers obtained is greater than a lower limit of $> 25\%$, the homography is re-estimated with this paricular set, otherwise the homography between the last pair of frames are used.

Refinement. Before the homography is estimated with the set of inliers, their matching positions are fine-tuned to half of a pixel accuracy using the block matching algorithm [40]. In this algorithm, a window of 100×100 pixels around the center of mass of each matching blob is used, and the search radius is limited to 10 pixels.

4.1.2 Rendering

With the images aligned to a common reference frame, the image details among the overlapping regions of the mosaic are now rendered to produce a seamless and uniform mosaic. The various steps undertaken in this rendering pipeline include: i) vignetting correction, ii) well lit image selection, and iii) multi-band blending. The vignetting step is detailed in the previous chapter, whereas the other two steps are explained in the next section.

4.1.2.1 Well-lit Image Selection

In practice, overlapping regions may differ because of registration errors and residual degradations after the vignetting correction is performed. To capture and preserve the best image details among these differences in the generated mosaic, overlapping regions are selected from the respective frames which are closest to the center of the light beam. This is achieved by first assigning a weighting function, $\mathbf{S}(\mathbf{x})$ to each image, and then retaining the sections with the highest weights, from the various input frames. As most of the degradations in these images follow a Gaussian like decay (as discussed in the previous chapter), a two dimensional Gaussian function is used as this weighting function, given as:

$$\mathbf{S}(\mathbf{x}) = \exp - \left[\frac{(\mathbf{x} - \mathbf{c})^2}{2\sigma^2} \right] \tag{4.5}$$

where $\mathbf{c}$ is the center of the light beam footprint, which is obtained from the vignetting correction step, and for the variance, a value of $\sigma = 160$ is used to correspond to the typical level of degradation in these images. In some instances however, the light beam is so intense that the image details within its beam footprint on the sea floor are saturated, as shown in Figure 4.10. For these cases, the user can select a center location for the weighting function that corresponds to the region in the video that contains the best quality data.

4.1.2.2 Multi-band Blending

Although selecting overlapping regions from individual frames does preserve image details, this method of combination produces unwanted seams in the generated mosaic, as shown in Figure

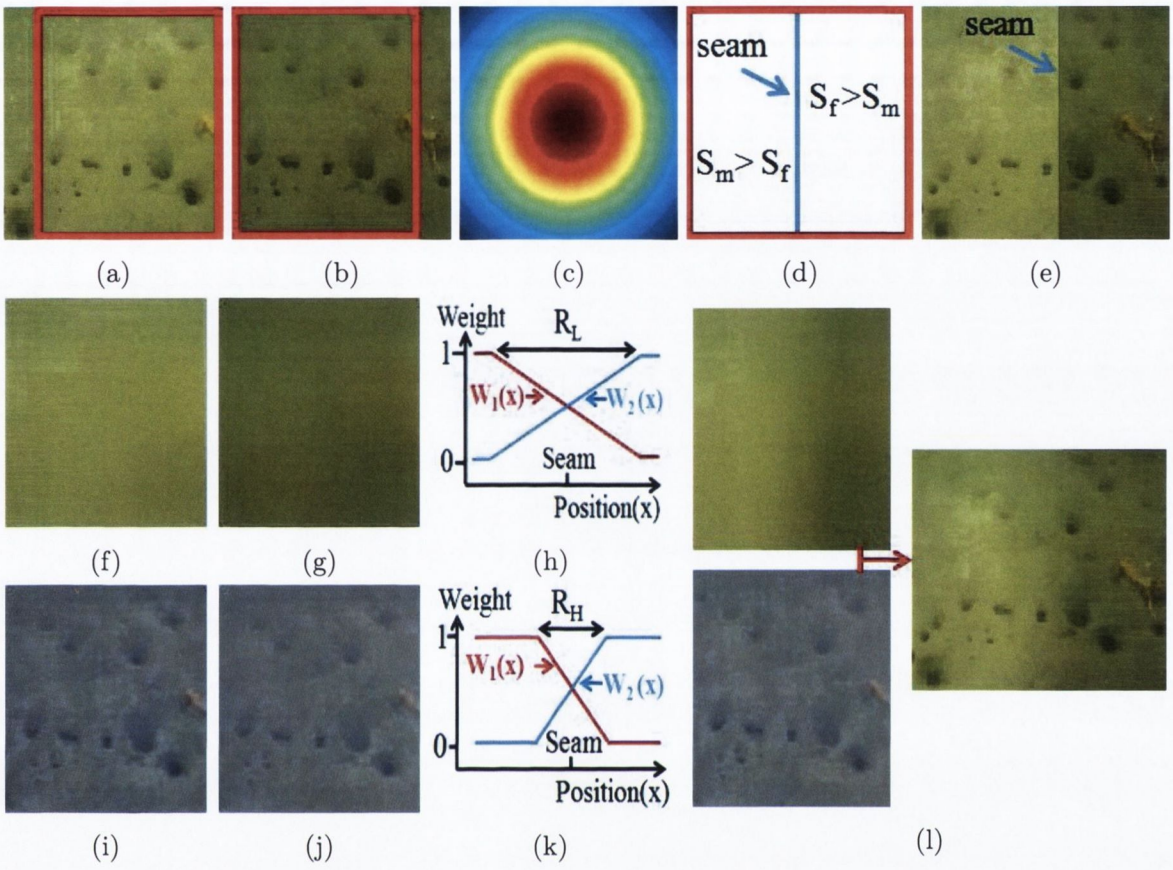


Figure 4.4: Image selection and Multi-band blending steps. First, obtain the overlapping region (red box) between the existing mosaic, (a), and new frame, (b). Then using their weights, (c), locate the image seam, (d), and perform image selection, (e). Now perform multi-band blending by combining the low and high pass versions of the overlapping regions from the mosaic (f,i), and the new frame (g,j), using their respective weighting scheme in (h) and (k). Lastly, combine the blended low and high pass versions and update the mosaic (l).

4.4. These seams occur at the added image strip boundaries, and are created as a result of their minor illumination differences. To eliminate these unwanted artifacts, the multi-band blending technique introduced by Burt and Adelson [10] is utilized. The main advantage of this technique compared to just linear blending [65], is that it blends the low frequency data over a larger range than the high frequency data, so in cases of minor misalignment, most of the high frequency details are preserved. For this application however, as the mosaics can grow very large, this blending is restricted to only two frequency bands (low and high) around a small region that is 100 pixels on either side of the respective image seam. In this case, the image seam is the common edge within the overlapping regions where the image selection weights of the two images are equal i.e. $\mathbf{S_1}(\mathbf{x}) = \mathbf{S_2}(\mathbf{x})$. With these simplifications, the remaining stages in this blending

scheme involve: i) frequency band separation, ii) generating the respective blending weights, and iii) blending each band and then reconstructing the image. These stages are described below.

Frequency Band Separation. The low frequency bands for the overlapping regions are created by blurring each with a two dimensional Gaussian filter, $g_\sigma(\mathbf{x})$, having 30 taps and a variance of 30. This large filter is used to separate the important high frequency data, such as burrows, from the low frequency sandy background, as illustrated in Figure 5. Subtracting this low frequency bands from the original regions, the corresponding high frequency bands are generated. Dropping the notation $\mathbf{x}$ for clarity, these two operations are mathematically described as:

$$\mathbf{I}_L^j = \mathbf{I}^j * g_\sigma; \quad \mathbf{I}_H^j = \mathbf{I}^j - \mathbf{I}_L^j; \tag{4.6}$$

where $\mathbf{I}_L^j$, $\mathbf{I}_H^j$, are the generated low and high pass versions of the corresponding regions, $\mathbf{I}^j$, in the mosaic ($j = m$), and the respective frame ($j = f$), to be added.

Blending Weights Generation. The blending weights for the mosaic region are first initialized to 1 at image locations $\mathbf{x}$ where its score, $\mathbf{S}_m(\mathbf{x})$, is greater than that from the frame to be added, $\mathbf{S}_f(\mathbf{x})$. This binary map is then blurred with a Gaussian filter, g_σ, given by:

$$\mathbf{W}_m(\mathbf{x}) = \begin{cases} 1 & \mathbf{S}_m(\mathbf{x}) > \mathbf{S}_f(\mathbf{x}) \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

$$\hat{\mathbf{W}}_m = \mathbf{W}_m * g_\sigma \tag{4.8}$$

Because blending is performed between two images at a time, the weights for the corresponding region in the new frame to be added to the mosaic, $\mathbf{W}_f$, is created as a mirror image to that of the mosaic by subtracting it from unity i.e. $\mathbf{W}_f = 1 - \mathbf{W}_m$. Using this weighting scheme the low frequency band is blended over a larger range than the high frequency one, by using a larger Gaussian blurring filter, g_σ. For this application, a 5 tap Gaussian with variance $\sigma^2 = 25$ is used for generating the weights for the high pass bands ($\mathbf{W}_H^m, \mathbf{W}_H^f$), whereas for the low-pass bands ($\mathbf{W}_L^m, \mathbf{W}_L^f$), the tap and variance are changed to 30 and 225 respectively.

Blending and Reconstruction. With the blending weights created, the high-pass and low-pass bands ($b = \{L, H\}$), of each image ($j = \{m, f\}$), are blended and then combined to represent the corresponding region in the mosaic, $\mathbf{I}^{m'}$, as follows:

$$\mathbf{I}^{m'} = \sum_b \sum_j \mathbf{I}_b^j \mathbf{W}_b^j \tag{4.9}$$

4.1.3 Video Referencing

These generated mosaics are useful for giving a summary of the surveyed seabed area in the recorded video. In some cases however, portions of the seabed captured are occluded by obstacles such as moving fish or floating sediments in the water etc. For these troublesome areas, users

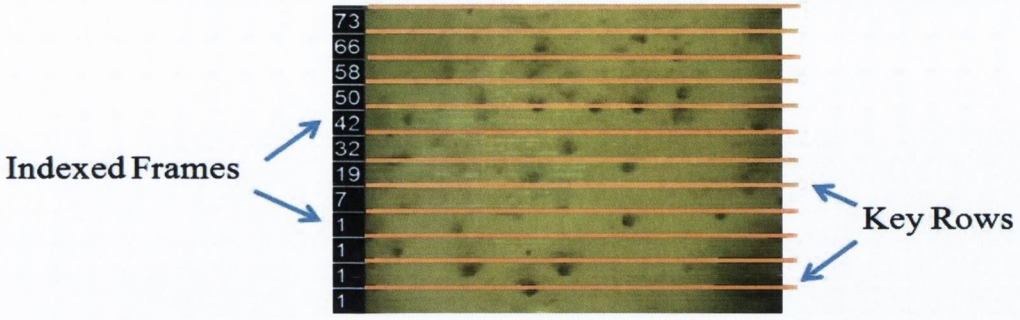


Figure 4.5: Examples of proposed indexing system.

would have to review the corresponding section in the video sequence to obtain a clearer view. To speed up this cross referencing process, it would be useful to have an indexing system that maps sections of the mosaic to their respective frames in the video sequence. In the literature, Irani and Anandan [39] solved this problem by indexing each pixel in the generated mosaic to its respective frame(s), which the user accessed with the use of a Graphical User Interface.

To simplify this process for the marine scientists, only key rows located equidistantly along the mosaic are indexed in this application. This indexing is accomplished in two steps. Firstly, each pixel in the reference row is indexed to the particular frame in the video sequence from where its respective image details are selected. Then the smallest of these frame numbers is assigned the indexing number of the respective reference row. Figure 4.5 illustrates an example of this indexing system, where the reference rows are marked as horizontal white lines, with the respective indexing number marked above it. Although in some cases this system might be slightly off due to misalignment errors, it will still guide the user to the approximate frame in the video sequence where the respective image details from that region in the mosaic are obtained.

Hence, the proposed mosaicking algorithm is summarized as follows:

1. Initialize mosaic with first frame, $\mathbf{I}_0$ and assign weights to it.
2. Read in the next frame, $\mathbf{I}_n$, and align it with existing mosaic.
   (a) Estimate homography, $\mathbf{H}_{n-1,n}$
   (b) Update Global Registration Model.
3. Perform Rendering
   (a) Estimate vignetting function, correct image $\mathbf{I}_n$ and then assign weights to it.
   (b) Select section of overlapping region from $\mathbf{I}_n$, where its weights are highest.
   (c) Eliminate seams using multi-band blending
4. Update reference rows index and repeat steps (2) to (4) until video sequence ends.

4.2 Results

The accuracy of the proposed method is now evaluated using simulated and real data, and compared to various state of the art homography estimation and mosaicking algorithms. The procedure used for creating the ground truth data is now described.

4.2.1 Ground Truth Creation

For the simulated data, two test sequences of approximately 200 frames (500×500) each, are created. These sequences are initially generated identically from an underwater Nephrops survey mosaic of size 6800×800, where a subregion of 500×500 is first selected at the beginning and then translated with constant vertical motion of 20 pixels across the entire mosaic. To examine robustness with vignetting, one of these sequences is then degraded with the proposed degradation function derived in the previous chapter, with circular shape parameters, $\mathbf{V} = \{140, 0; 0, 140\}$, no response function ($\gamma = 1$), no colour degradation ($m_\lambda = \{1, 1, 1\}$) and centered at position $\mathbf{c} = \{310, 310\}$. Lastly, to make these sequences more realistic, Gaussian noise: $Z(\mathbf{x}) = \sigma N(0, 1)$,

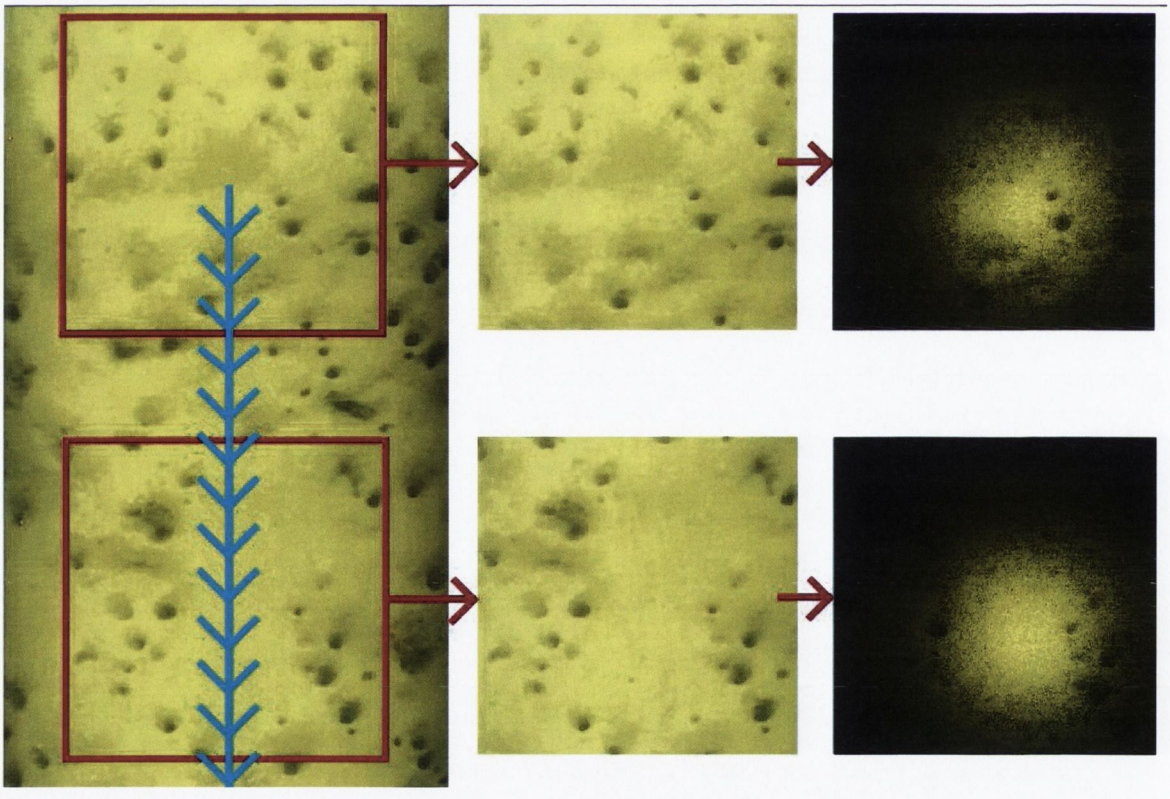


Figure 4.6: Creating the ground truth sequences by selecting a subsection (red box) from the mosaic on the left and translating it at constant motion (blue arrows). Then applying degradation to the generated frames shown in the middle to give a degraded version on the right.

is added to each frame. Where $Z(\mathbf{x})$ is a random intensity at image location $\mathbf{x}$, that is drawn from the standard normal distribution $N(0,1)$, and $\sigma = 10$ is the standard deviation. Figure 4.6 illustrates sample images of these two created sequences along with a section of the original mosaic. Using these ground truth sequences, the homography and rendering aspects from the proposed algorithm are evaluated. These evaluations will now be discussed, followed by an examination using actual underwater video sequences.

4.2.2 Homography Estimation Comparison

Correct homography estimation is a crucial step in any mosaicking process, as large errors can severely degrade the quality of the generated mosaic. The efficiency of the proposed algorithm at estimating the homography in the ground truth sequences is now examined. The results obtained are also compared to four previous techniques from the literature. The first technique is from Matas et al. [53], which uses MSER feature matching (performed in the Difference of Gaussians domain) followed by RANSAC for their estimation. The main reason for this comparison is to examine if the block matching step and the Bayesian framework in our algorithm improves accuracy, as these are the main differences between the proposed approach and this previous method. The second comparison is with the method used by Brown and Lowe [9], which uses SIFT [50] feature matching followed by RANSAC. The third method is used by both Li et al. [65] and Gracias and Santos-Victor [26], which involves the use of corner matching [30], followed by robust techniques similar to RANSAC, as detailed in their work. The last comparison is not feature based like the rest, but instead is a multi-resolution gradient based approach, which is developed for underwater applications by Spindler and Bouthemy [79].

Plots of the estimated x and y translations (T_x, T_y) obtained from each method, from the ground truth sequences with (blue) and without (red) degradations, are shown in Figure 4.7, and their corresponding average errors are listed in Table 4.1. Comparing these results to the ground truth values of $\{T_x = 0, T_y = -30\}$, two key observations are made. First, the performance of each algorithm deteriorated in the sequence with the simulated vignetting (blue plots). Examination of these deteriorations show the feature based approaches are more robust

Method	Proposed	MSER	SIFT	Corners	Motion2D
Error Seq. 1	0.35	0.79	0.61	0.81	0.24
Error Seq. 2 (with Vig.)	0.67	1.6	1.9	2.5	9.6
Error Seq. 1 and 2	0.51	1.2	1.3	1.7	4.9

Table 4.1: Average error in pixels, E, obtained from each of the various methods examined, for the two ground truth sequences. The error is calculated by : $E = \sqrt{((T_x - \hat{T}_x)^2 + (T_y - \hat{T}_y)^2)}$. Where $\hat{T}_x$ and $\hat{T}_y$ are the ground truth x and y translations of 0 and -30 respectively.

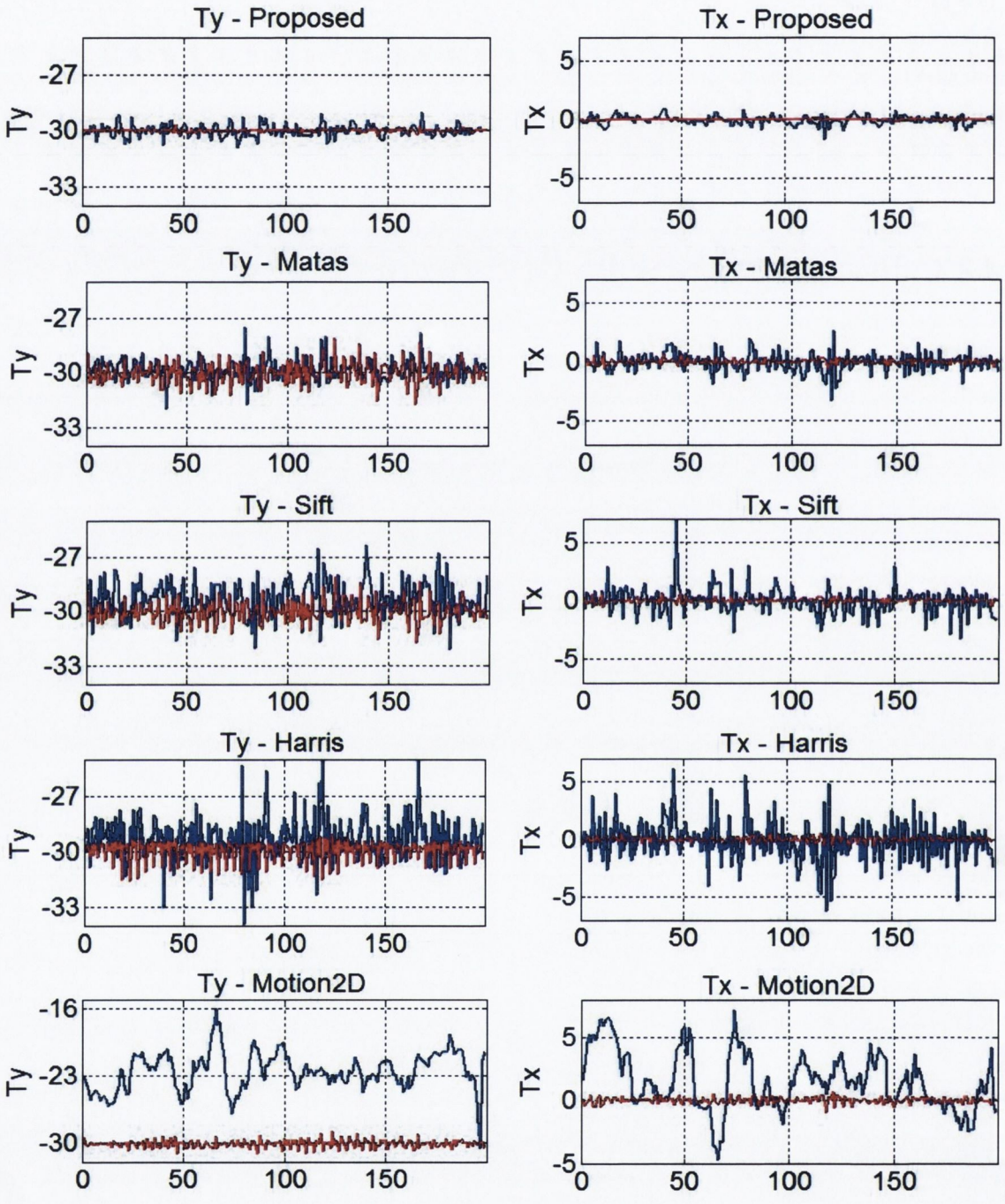


Figure 4.7: Estimated x and y translations obtained from the ground truth $\{T_x = 0, T_y = -30\}$ sequences without (red) and with (blue) vignetting, from the proposed, Matas et al. [53], Brown and Lowe [9] using SIFT, Li et al. [65] and Gracias and Santos-Victor [26] using corner features, and the hierarchical gradient based approach by Spindler and Bouthemy [79] in the 1st, 2nd, 3rd, 4th and 5th rows respectively.

in this sequence, giving errors which average below two pixels. These errors are relatively low compared to the hierarchical gradient based approach by Spindler and Bouthemy [79], which gave an average error of 9.3 pixels. This large increase in error is because this method uses the image intensities in an exhaustive searching scheme, which are altered significantly in the vignetting sequence. In the first sequence however, it is observed that this method gave the best results. This observation shows that if this technique is limited to the minimally degraded regions within the video, its performance will improve. The second key observation is the proposed algorithm performed the best overall, with Matas et al. [53] in second, Brown and Lowe [9] in third with sift features, Li et al. [65] and Gracias and Santos-Victor [26] in fourth with corner features, and Spindler and Bouthemy [79] is fifth with their hierarchical gradient based approach. This shows that, among these methods, blob tracking via the use of MSER features is best suited for this application, with sift and corner features, and the hierarchical gradient based approach being the second, third and fourth best choices respectively. Also, because the proposed method gave better results than Matas et al. [53], this shows the block matching step introduced in this work does improve the accuracy of the system.

4.2.3 Rendering Comparison

In the previous chapter it is shown from the experiments involving real seabed survey video, that residual degradations are still present after the vignetting correction is performed. Because of this, the ability of the proposed rendering scheme to capture the well lit image details will be examined in this experiment. This examination is performed by comparing the quality of the mosaic generated from ground truth video containing the vignetting degradation, to the corresponding section from the original mosaic that these sequences are derived from. To compare the quality, the mean absolute error (MAE) is used. These results are also compared to the rendering techniques from three previous authors. The first technique is from Brown and Lowe [9], which uses a circular shaped linear weighting function centered at the image center to perform image selection followed by a two stage multi band blending technique. The main reason for this comparison is to examine if selecting image details from the center of the degradation in these sequences has any advantage to using the image center, as this is the main difference between the proposed and this previous method. The second and third techniques by Li et al. [65], and Gracias and Santos-Victor [26] employ statistical techniques for rendering overlapping regions such as using the mean and median values respectively. As only the rendering techniques are being compared in this experiment, the ground truth motion values are used for generating the mosaic from each respective technique.

Figure 4.8 shows sections of the mosaics generated from various methods. Analysis of these results show the proposed algorithm obtaining the lowest MAE, with Gracias and Santos-Victor [26] using the median the next best method, Brown and Lowe [9] in third with their weighting function centered at the image center, and Li et al. [65] is fourth best using the mean value.

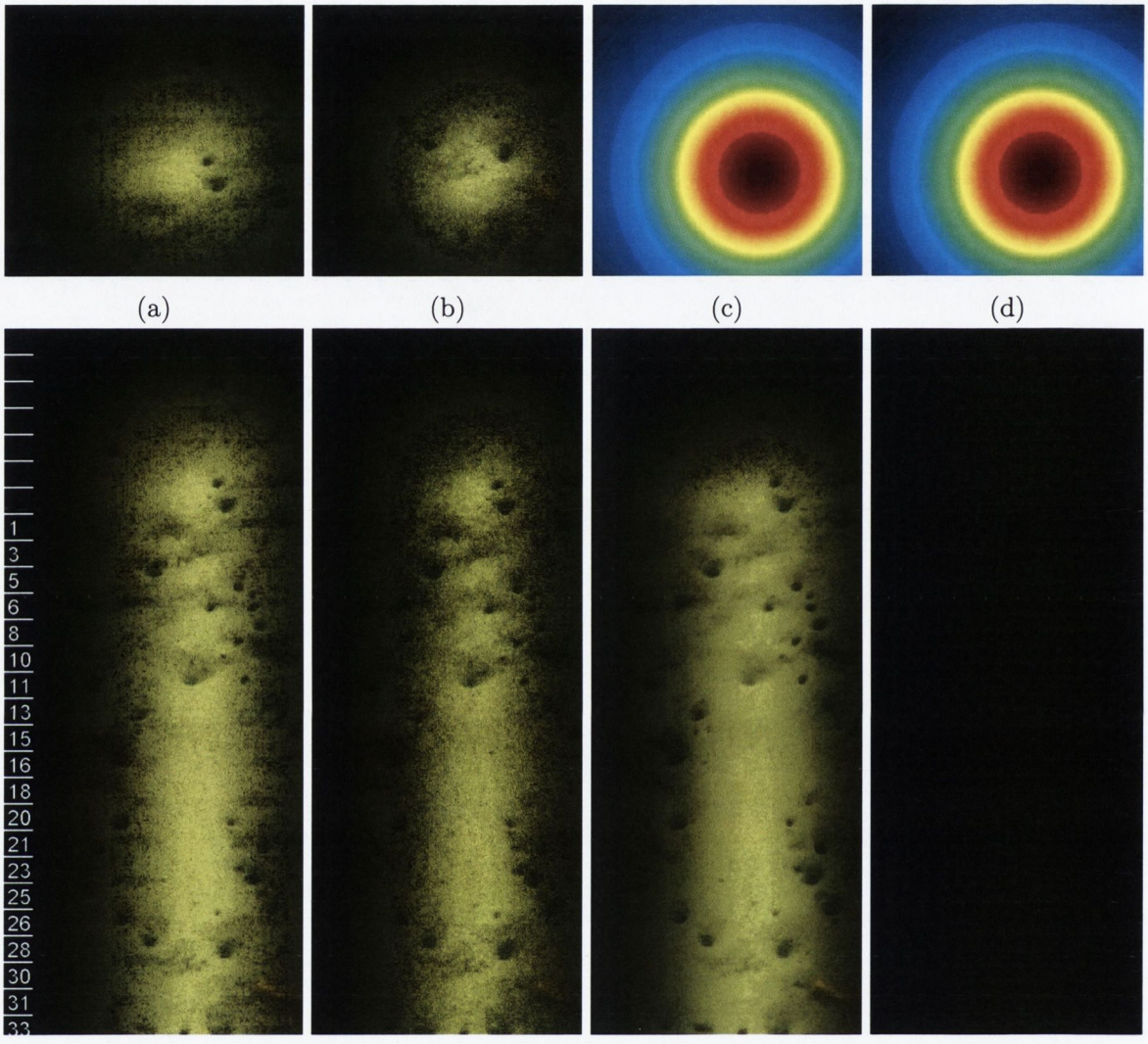


Figure 4.8: Rendering results. Sample ground truth frames (a) 1 and (b) 31, (c) estimated vignetting, and (d) weighting function used for image selection. Mosaics rendered from (e) proposed method, (f) Brown et al. [9], (g) Gracias et al. [26], and (h) Li et al. [65].

Method	Proposed	Brown et al. [9]	Gracias et al. [26]	Li et al. [65]
Rendering Technique	weighting fn. centered at vignetting center	weighting fn. centered at image center	median	mean
MAE	52.65	82.87	55.95	118.63

Table 4.2: Mean absolute error (MAE) between ground truth mosaic and mosaics generated from the ground truth video with vignetting, from each of the various methods examined.

Using the median value is a good alternative to the proposed method in this case as the motion is significantly large and constant, resulting in the well lit image details remaining at the middle position among the overlapping regions. Even some of the fuzzy noise that was added to this sequence is seen to be eliminated in this method. Using the mean value however is not a very good choice for these sequences as degraded values among the overlapping regions degrade the well lit ones to a very high extent, which results in an overall dark mosaic. Some image seams are also noticed in the results from this method. The results obtained from the third method by Brown and Lowe [9] in Figure 4.2, looks identical to the proposed results, but just a little darker. This darker results shows using the center of the vignetting degradation as the center of the weighting function to perform image selection, can improve the quality of the generated mosaic.

4.2.4 Results with Actual Underwater Videos

Having shown the proposed algorithm can align images and render overlapping regions effectively, it is now tested on 23 actual underwater survey videos of the seabed. Each of these videos is of PAL format, has approximately 1500 frames, and possess a wide variety of degradations. Twenty one of these videos are surveys from different Nephrops habitats, where the seabed type is mainly muddy, and last two are from different surveys where the seabed type mainly comprises of pebbles and shells. These different seabed types differ largely in colour and texture characteristics due to their different sediments. Sample images along with sections of the mosaics generated from the i) proposed, and previous methods from ii) Brown and Lowe [9], iii) Gracias and Santos-Victor [26], and iv) Li et al. [65] are given in Figures 4.9, 4.10, 4.11, 4.12, 4.13, with additional sets in Appendix B and D.

Analyzing the results in Figure 4.9, which is obtained from a typical Nephrops survey video given, two points are worth mentioning. First, the center for the weighting function is offset below the estimated vignetting center, because that location is usually well lit and has the least geometric distortion. This is why the details are smaller in the results obtained from Brown and Lowe [9], which uses the image center for their weighting function center. The next item worth mentioning is the bright and blurry results obtained by the previous method by Gracias and Santos-Victor [26] that uses the median value for rendering. This undesired effect is because of poor visibility of features at the top of each frame, which causes the features of the merged result to also be less clear. This characteristic is however averaged away in the method of Li et al. [65], as most of the overlapping regions are of good quality. The degradations at the top of the image did however cause the results from this method to be darker than the rest, with a light green line in the middle that corresponds to one of the laser spots used in this survey apparatus.

The usefulness of offsetting the weighting function center to the bottom of the Nephrops survey videos is also seen in the sequence in Figure 4.11. As seen this sequence involves intense

lighting, and only the proposed method captured the important image details in this case. The results in Figure 4.10, which are from the seabed sediment that is mainly pebbles, are also very interesting. As seen, the results obtained by the previous method by Brown and Lowe [9] is a little brighter than the proposed result. This extra brightness is due to the more intense lighting at the image center than at the center of the weighting function used by the proposed method. Another interesting observation from this sequence are the blurry results obtained from the previous methods by Gracias et al. [26] and Li et al. [65] respectively. These poor results can be attributed to misalignment errors together with the residual degradation in this case after the vignetting correction is performed.

4.3 Conclusion and Future Work

In this chapter three main contributions are made to the mosaicking literature. First, is the use of vignetting ideas to capture image details from well lit regions in underwater survey video, to generate high quality mosaics. Second, is the combination of feature matching and refinement approaches to improve homography estimation process. Lastly, is a method for indexing sections of the generated mosaic to its corresponding video frames. The simulated experiments performed verifies that the novel image capturing and homography estimation procedures introduced in this work does obtain improved results compared to four state of the art mosaicking algorithms from the literature.

Although most of these initial results show the proposed algorithm can generate high quality mosaics, two main problems are spotted. The first problem is the loss of the valuable high frequency details when motion blur occurs. This phenomenon occurred in the Shelly seabed sequence when the camera is moving too fast, as shown in Figure 4.12. In the future this problem may be rectified by exploring the use of deblurring algorithms, or excluding these blurry frames from the mosaic generation process. The next problem spotted in these tests is a trail of laser dots that occurs as a result of the movement of the lasers used these surveys. Figure 4.13 shows an instance of this phenomenon. In the future this problem may be solved by tracking these laser regions and masking them out of the rendering process.

A comparison of using mosaics and video for identifying Nephrops complexes is given in Appendix B. Sample results obtained from three scientists using mosaics in this comparison are given in Figure 4.14. As seen in Figure 4.14 and shown in Appendix B, there are inconsistencies in the counts obtained from different scientists. One of the key observations made is that the counts obtained from the mosaics are generally greater than those from the corresponding videos. This observation possibly implies that the enhanced visibility and field of view offered in mosaics can help to improve the accuracy of the Nephrops analysis procedure. Overall, the scientists agreed it was much easier to spot relationships among the Nephrops burrows with the mosaics as opposed to the original videos. The twenty mosaics generated in this chapter from the various Nephrops habitats will be used in the next chapter for performing burrow recognition.

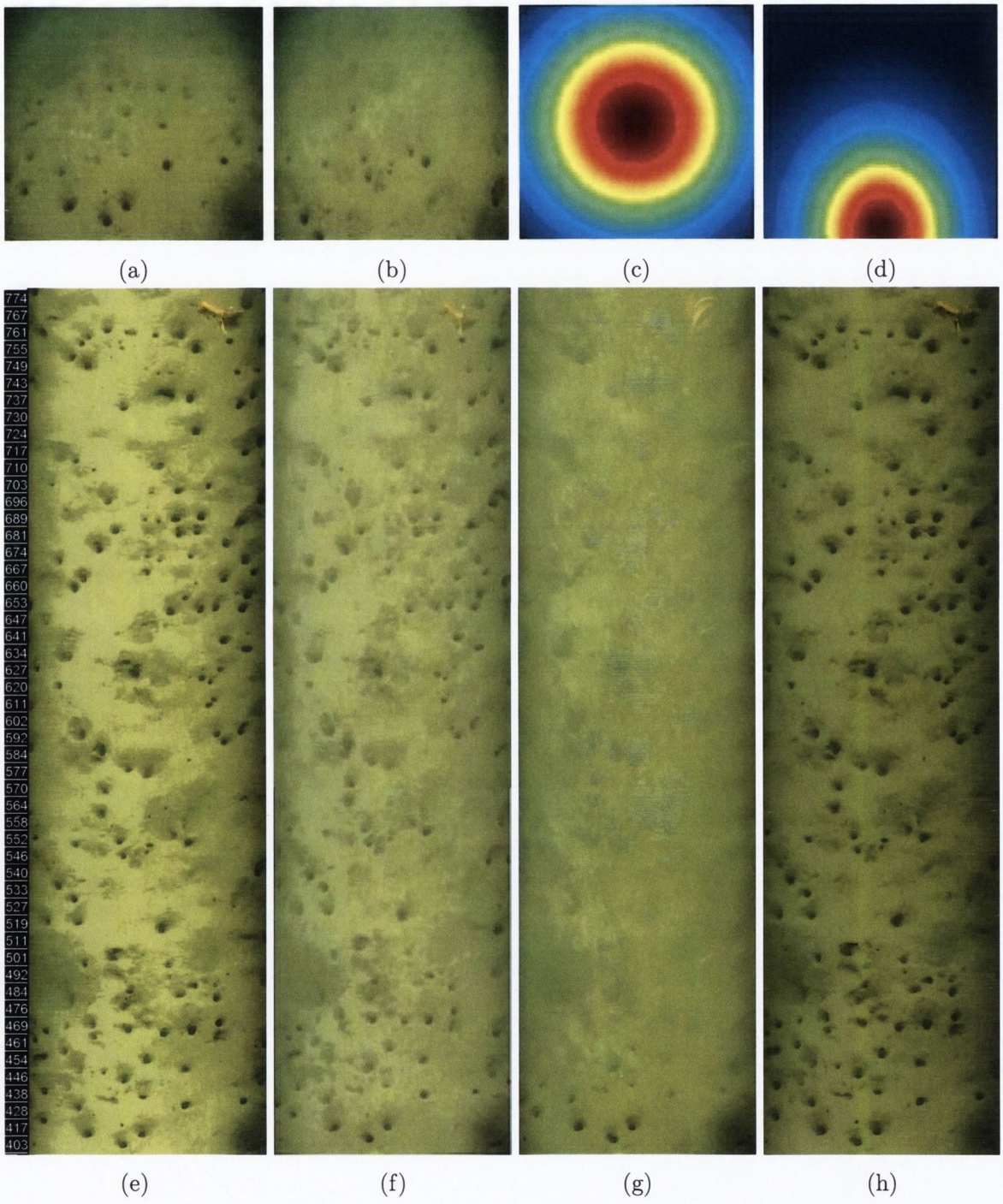


Figure 4.9: Original frames (a) 403 and (b) 527, (c) estimated vignetting, and (d) weighting function used for image selection. Mosaics rendered from (e) proposed technique, (f) Brown et al. [9], (g) Gracias et al. [26], and (h) Li et al. [65].

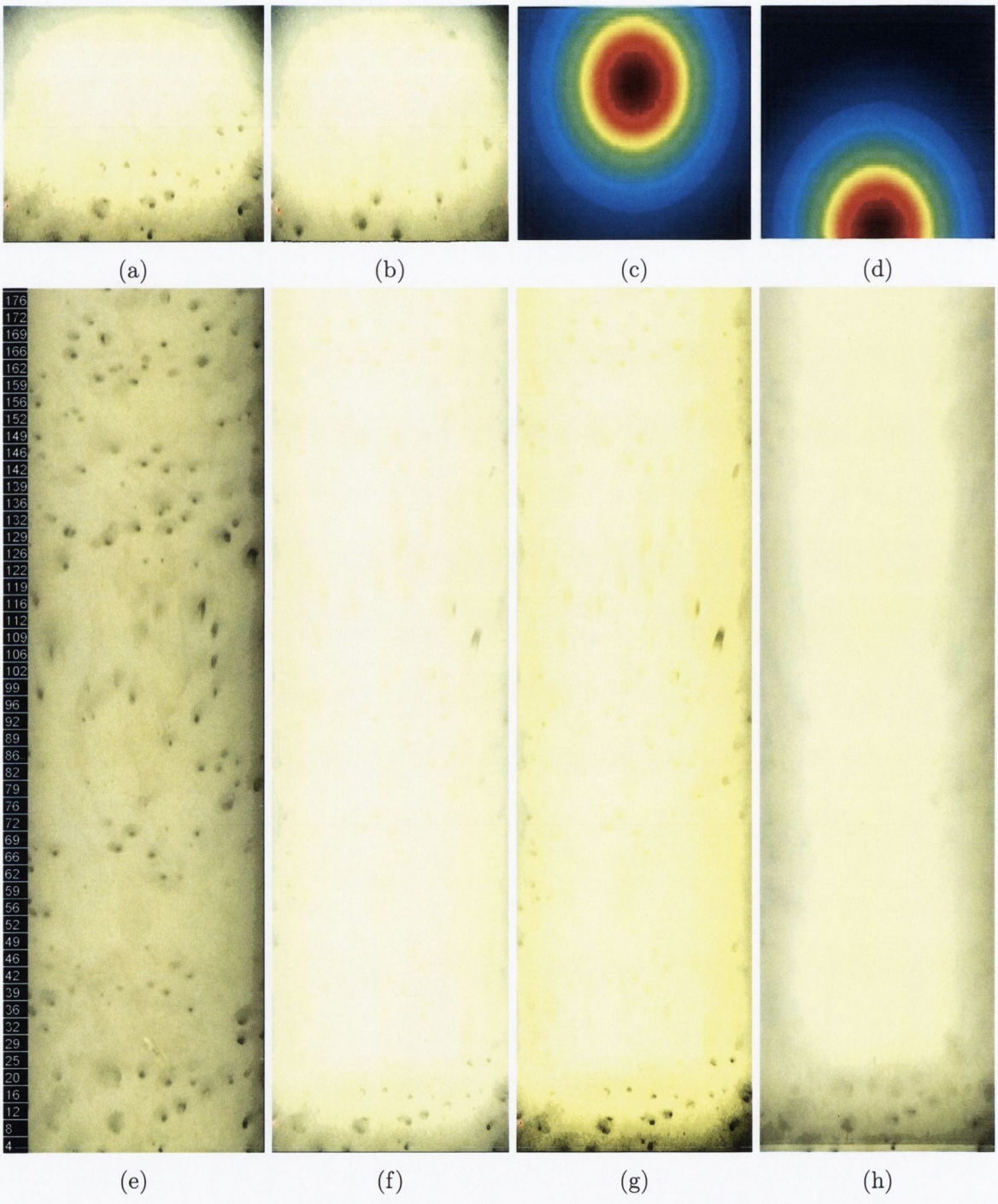


Figure 4.10: Original frames (a) 4 and (b) 64, (c) estimated vignetting, and (d) weighting function used for image selection. Mosaics rendered from (e) proposed technique, (f) Brown et al. [9], (g) Gracias et al. [26], and (h) Li et al. [65].

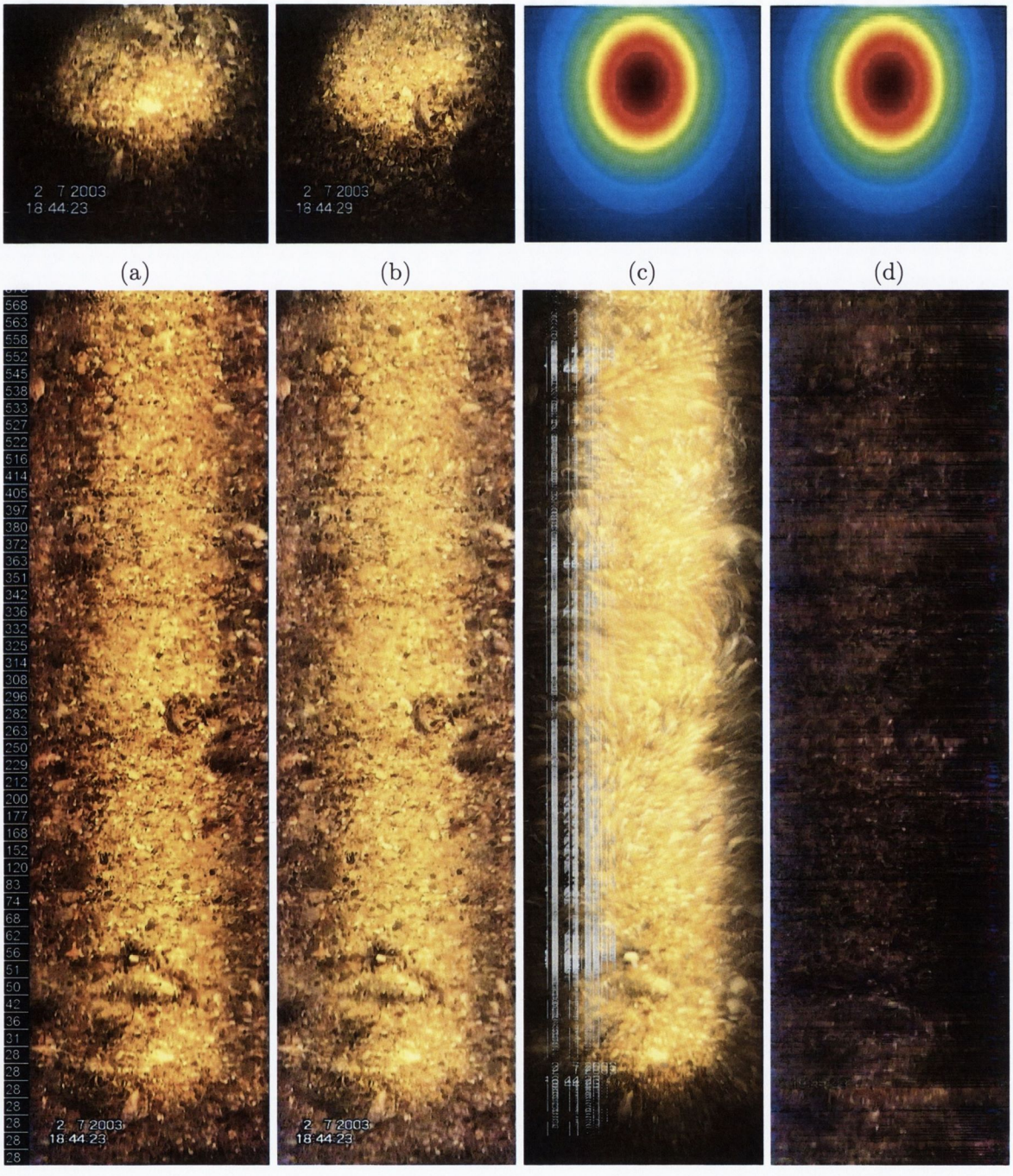


Figure 4.11: Original frames (a) 28 and (b) 282, (c) estimated vignetting, and (d) weighting function used for image selection. Mosaics rendered from (e) proposed technique, (f) Brown et al. [9], (g) Gracias et al. [26], and (h) Li et al. [65].

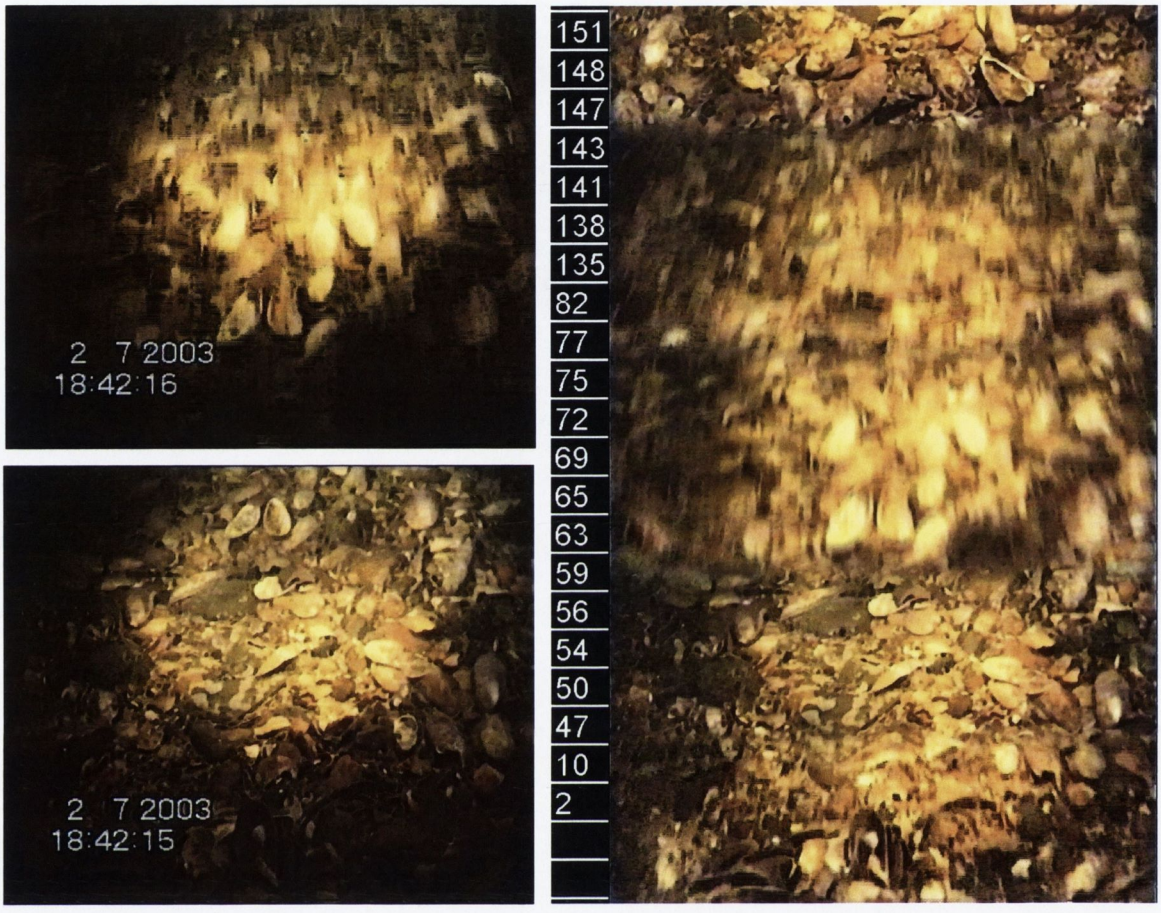


Figure 4.12: Loss of image details in mosaic on the right as a result of motion blur in captured images on the top right frame.

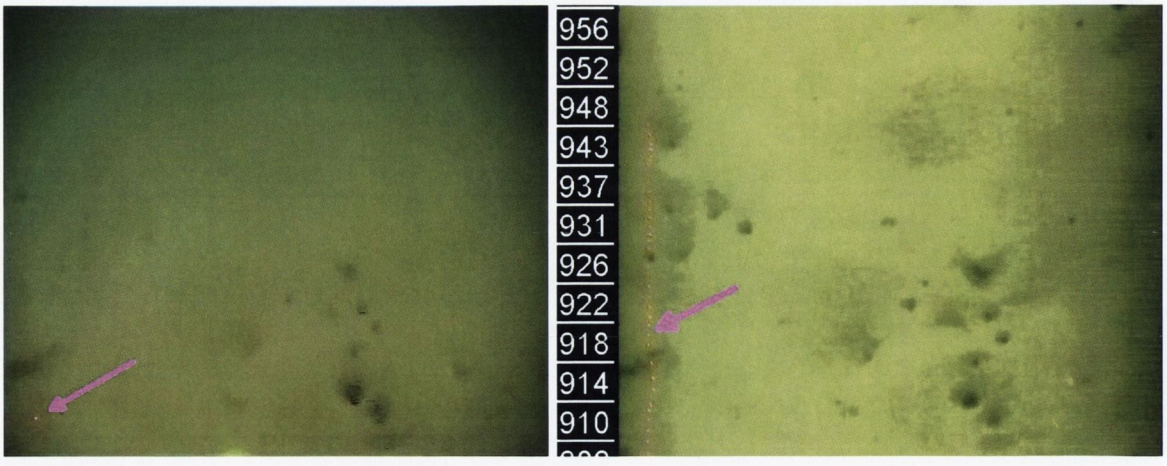


Figure 4.13: Moving laser dot in survey frames (left) sometimes result in a trail of laser dots in the generated mosaic (right), as indicated by the pink arrows.

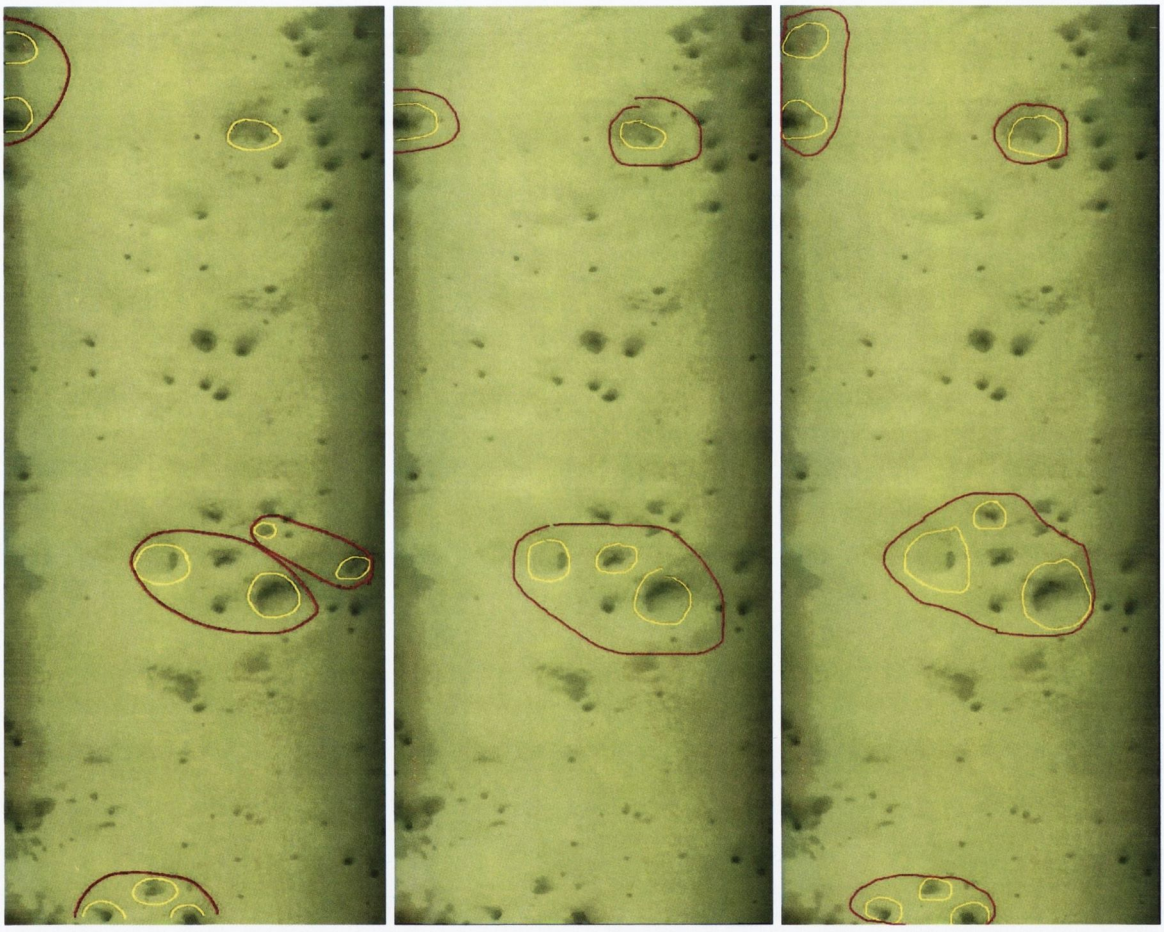


Figure 4.14: A section of test mosaic-3 showing the manual selections of Nephrops burrows (yellow) and their corresponding complexes (red), obtained from scientists: (Left) Adrian Weetman from the Marine Laboratory in Scotland, (Middle) Alessandro Ligas from the Biosciences Institute in Belfast, (Right) Jennifer Doyle from the Marine Institute in Galway

5

Burrow Recognition Using Mosaics

This chapter considers the detection and classification of burrows. Burrow analysis not only allows the estimation of the Nephrops population, but can also be used as an indicator of population densities of other species [52]. Some of these other species include shrimps, and a variety of crabs such as the four bearded rockling, Calocaris Macandreae, Geryon Tridens, and Goneplax. Scientists currently detect and count burrows manually from surveillance video. Given the very long duration of these recordings and the thousands of shapes that are candidate burrows, this is a tedious and error prone task. Automated analysis as presented in this chapter gives the marine scientist a much needed tool for improved reliability and increased throughput.

To identify burrows automatically from these survey videos is not an easy task due to three main challenges. First is the geometric distortion and narrow field of view present in these videos. Second, is the uneven lighting and colour degradations experienced in this type of environment due to the water medium. Lastly, is the large variety of burrows, which differ vastly in shape, size, texture, and colour characteristics. These challenges are illustrated in the sample surveillance video frames in Figure 5.1.

Previous work in automated Nephrops burrow identification was presented by Lau et al. [46], which is briefly discussed at the end of Chapter 2. Their work however suffers from three main shortcomings when tested on our real data set. First, because of the blurriness of these images, their object detection process which relies on edges alone, produces incomplete segmentations, as seen Figure 5.6 (f). Secondly, they use a strict set of rules, in a decision tree framework to perform classification, which may have worked well for their data set, but might not be applicable to other data sets. Lastly, to verify their video based results still require scientists to tediously

inspect of thousands of frames manually.

To improve on these short comings, four key contributions are introduced in this proposed work. First, mosaics are used for performing object recognition, which improves visibility, and simplifies the tedious video inspection process to the browsing of a single image. Secondly, a novel object detection method is developed that uses the difference of Gaussians image to target dark burrow-like regions. In this method, segmentation and shape modeling techniques are employed, which capture most of the object regions. The third contribution is a new feature set for this application that is motivated by a current scientific description of Nephrop burrows [36]. Some of these features include the dark entrance area of the burrow, its diameter, associating animal claw marks near the burrow entrance. Two key advantages of using these features is that marine scientists can easily relate to them, and they provide further statistical information such as the percentage of small burrow systems etc. Lastly, to get around the problem of using strict rules for classifying objects with a large diversity in size and shape features, the use of two well established supervised learning classification schemes, KNN and SVM, are explored. These schemes use training data from a large variety of burrow and non-burrow objects found in these videos to aid in the classification process, and can be updated with new data to adopt to most situations.

The design of this new burrow recognition application is accomplished in five stages involving: i) Data Collection, ii) Object Detection and Grouping, iii) Feature Choice and Extraction, iv) Classification Model Selection, v) Optimization and Training. Details on each of these design stages are now presented, followed by an overview of the proposed classification pipeline. Afterwards, a comparison with the state of the art technique introduced by Lau et al. [46] is given. Both of these classification schemes are also compared against random selection, along with an investigation into their robustness to operate at various levels of additive Gaussian white noise. After this rigorous evaluation, the most relevant and interesting results obtained are discussed.

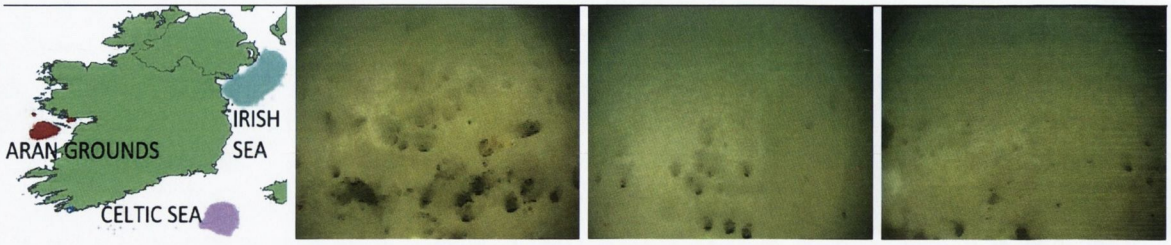


Figure 5.1: (a) Main Nephrops harvesting grounds in Ireland, Aran Grounds (red), Celtic Sea (purple) and Irish Sea (blue), with corresponding seabed images in (b), (c), and (d) respectively.

5.1 Data Collection

The training and testing data for these experiments are obtained from ten 2-minute sequences (PAL format) of actual underwater surveillance videos. These were supplied by marine scientist, Jennifer Doyle [1] (one of our collaborators in the Marine Institute), and represent real data used for Nephrops analysis by the Marine Institute. Figure 5.1 shows the locations from which these sequences are gathered around Ireland and the changing nature of the seabed. Using these video sequences, ground truth data is created in three steps. First, the corresponding video mosaics are generated using the algorithm described in the previous chapter. Then, an expert (Jennifer Doyle) manually selects the burrow regions, which are then fully segmented using the proposed object detection and grouping algorithm. The other objects that are detected from the proposed detection algorithm are then labeled as non-burrow items. Figure 5.2 illustrates these three steps. During the selection process the original frames are visually inspected to ensure the

[1] jennifer.doyle@marine.ie from the Marine Institute, Galway, Ireland

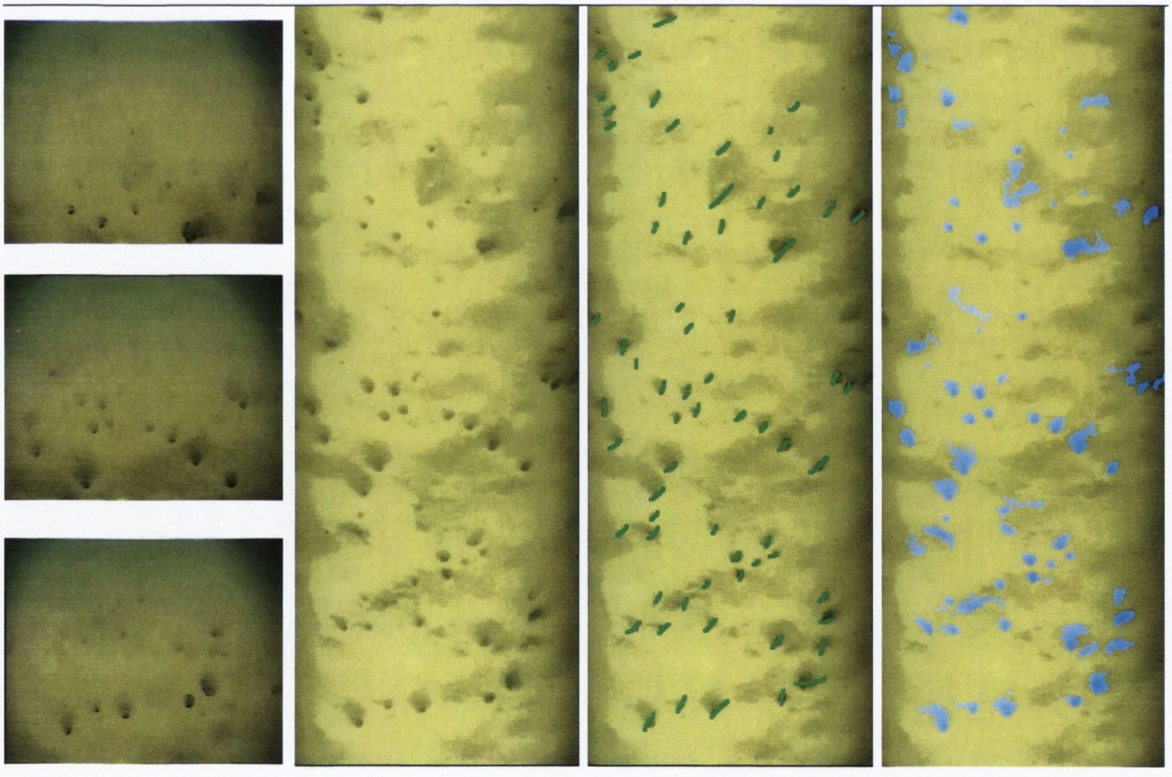


Figure 5.2: Ground truth creation steps. First the original video frames, shown in the 1st column are used, to generate a mosaic, as seen in the second column. Then an expert manually selects the burrow regions shown by the green markings in the 3rd column. Lastly the proposed segmentation algorithm is used to obtain the entire burrow (and non-burrow) regions.

Test Sequence	1	2	3	4	5	6	7	8	9	10
Burrows	527	654	596	737	785	546	968	1035	855	597
Non-Burrow	1430	4532	3400	4959	4262	3798	3573	3602	5200	2412
Location	AG	AG	IS	CS	CS	IS	IS	AG	AG	AG

Table 5.1: Ground truth burrow and non-burrow objects in each test sequence and their corresponding survey locations at the Aran Grounds (AR), Celtic Sea (CS) and Irish Sea (IS).

integrity of each mosaic, which proved to be accurate in all cases. Also, to ensure the accuracy of the manually selected burrows, they were verified with another expert. Table 5.1 illustrates the number of burrows labeled in each test sequence along with their respective location.

5.1.1 The nature of burrows

The ground truth data is now analyzed for clues that would be useful for choosing the most appropriate features and classifiers for this application. The main clue observed, and confirmed from discussions with scientists at the marine institute, is that the burrows generally appear as dark roughly ovoid regions in the video, as shown in Figure 5.1 (b). In addition to their characteristic dark color, scientists also look for additional features to recognize burrows such as their entrance region, its diameter, and associated animal claw marks that the creature creates while maneuvering in and out of the entrance. Figure 5.3 illustrates these features.

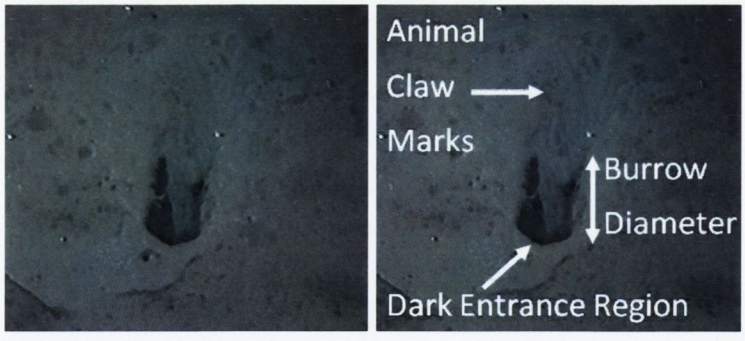


Figure 5.3: Sample image of a burrow (left), with some of its scientific features (right).

5.2 Object Detection and Grouping

The first stage of the burrow recognition system is to detect candidate burrow regions in the generated mosaic. A segmentation approach is used to accomplish this task because these images are usually very blurry, and detecting parts of the objects with techniques such as edges [46], might not be effective in some cases. Additionally, the scientifically important features such as

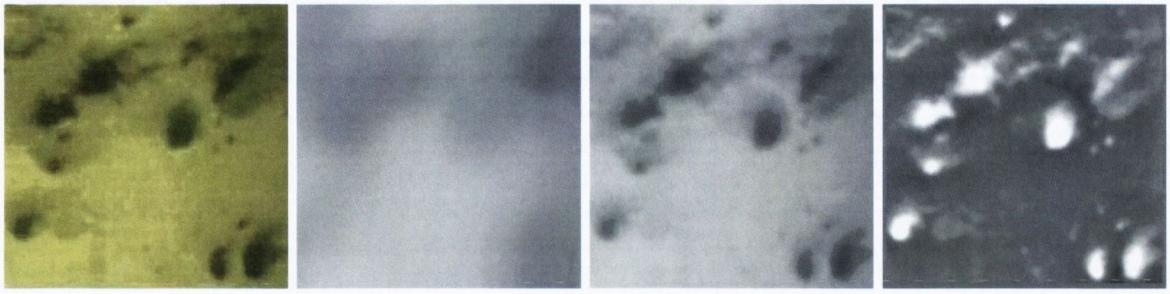


Figure 5.4: Locating the dark regions in the mosaic (a), by subtracting a heavily blurred gray scale version (b), from a lightly blurred version (c). This highlights the local dark (candidate burrow) regions as local maxima regions, as seen in (d).

its burrow diameter, animal claw mark region, dark entrance area etc., can only be extracted from the entire region. The segmentation approach is performed by targeting the characteristic dark appearance of burrows for their detection. To ensure the dark areas in these images due to uneven lighting are not targeted, detection is performed in the difference of Gaussians image, where the influence of absolute brightness has minimal effect. This overall procedure has three main steps involving: i) generating a dark region map, ii) segmentation, iii) labeling and splitting, which are now explained.

5.2.1 Dark Region Map Generation

The first stage in this detection algorithm is to locate dark regions in the mosaic, I. This is achieved by generating a dark region map as: $I_d = I * G_1 - I * G_2$. Where G_1 and G_2 are two dimensional Gaussian functions with 71 and 5 taps, and corresponding variances of 30 and 2 respectively. The large value of G_1 is chosen in relation to the average burrow diameter (71 pixels), so that most of these objects would be effectively blurred out. With the burrows blurred out, an equivalent homogeneous sandy background image is created, and when subtracted from the lightly blurred version ($I * G_2$), all of the locally dark (candidate burrow) regions are highlighted as local maxima regions. To obtain larger maxima values and hence improve detection, gamma correction is performed on the original image, $I = I^{\gamma}$, where $\gamma = 1.5$ is used, prior to the generation of I_d. Figure 5.4 illustrates the generation of this dark region map.

5.2.2 Segmentation

The candidate burrow regions are now obtained by performing segmentation on the dark region map. To obtain two scientifically significant burrow features [12], the dark entrance and lighter animal claw mark regions, a three layer segmentation map $L(\mathbf{x})$ is estimated in which the labels are defined as follows.

$$L(\mathbf{x}) = \begin{cases} 2 & \text{The dark entrance} \\ 1 & \text{The lighter intensity animal claw mark regions} \\ 0 & \text{Homogenous sandy background regions} \end{cases}$$

Following a Bayesian framework, the MAP estimate for $L(\mathbf{x})$ is generated by maximizing $p_o(L(\mathbf{x}) = \alpha | I_d(\mathbf{x}), \neg L(\mathbf{x}))$ where $\neg L(\mathbf{x})$ is the respective 3×3 neighborhood pixel labels of image position $\mathbf{x}$. Factorizing the posterior using Bayes Law [18], and dropping the notation $\mathbf{x}$ for clarity, gives:

$$p_o(L = \alpha | I_d, \neg L) \propto p_k(I_d | L = \alpha) p_r(L = \alpha | \neg L) \tag{5.1}$$

where p_k and p_r are the likelihood and prior terms. The likelihood is assumed to be Gaussian as follows.

$$p_k(I_d | L = \alpha) \propto \exp - \left[\frac{(I_d - I_\alpha)^2}{2\sigma_\alpha^2} \right] \tag{5.2}$$

where $\alpha = \{0, 1, 2\}$, and $\{I_0, I_1, I_2\}$ are the mean values of the background, claw mark and dark entrance regions respectively, and $\{\sigma_0^2, \sigma_1^2, \sigma_2^2\}$ are their corresponding variances. To enforce spatial smoothness within these segmentations, a Gibbs energy function [23], with a 3×3 pixel neighborhood, is used for the prior, $p_r(.)$, as:

$$p_r(L(\mathbf{x}) = \alpha | \neg L) \propto \exp - \left[\Lambda \sum_{k=0}^{7} \lambda_k |\alpha - L(\mathbf{x}_k)| \right] \tag{5.3}$$

where $\lambda_k = 1/||\mathbf{x} - \mathbf{x_k}||$, is a weight inversely proportional to the distance between the current site $\mathbf{x}$ and the respective neighbor $\mathbf{x_k}$ in a 3×3 neighborhood, and Λ is a global weighting factor, set as $\Lambda = 1$ in these experiments.

Figure 5.5: a) Original, b) Dark region map, and c) Segmented candidate burrow regions, with the dark entrances in pink and the animal claw mark regions in blue.

Good initial estimates for these parameters and labels are obtained using k-means clustering on I_d, with 4 clusters to represent the different levels of burrow shading and the background. The 4 clusters are ranked with respect to the intensity of the cluster centroid. The background ($L(\mathbf{x}) = 0$) and claw mark ($L(\mathbf{x}) = 1$) regions are labeled with cluster members associated with the first and second minimum intensity centroid values respectively, and the dark entrance ($L(\mathbf{x}) = 2$) regions with the other two cluster members. The parameters $\{I_0, I_1, I_2\}$ are set using the three smallest centroid values, $\{w_1, w_2, w_3\}$ as $\{w_1, (w_1 + w_2)/2, (w_2 + w_3)/2\}$. While $\{\sigma_0, \sigma_1, \sigma_2\}$ are set as $1.5\{(I_1 + I_0), (I_1 + I_0), (I_1 + I_2)\}$, so that the prior terms would have an equivalent weighting to the likelihood terms. Minimization of p_o is performed using the Iterated Conditional Modes [6] scheme, where a checkerboard scan is utilized until there are no further changes in labels or a maximum of 10 iterations is completed. Sample results obtained using this three layer segmentation procedure are shown in Figure 5.5.

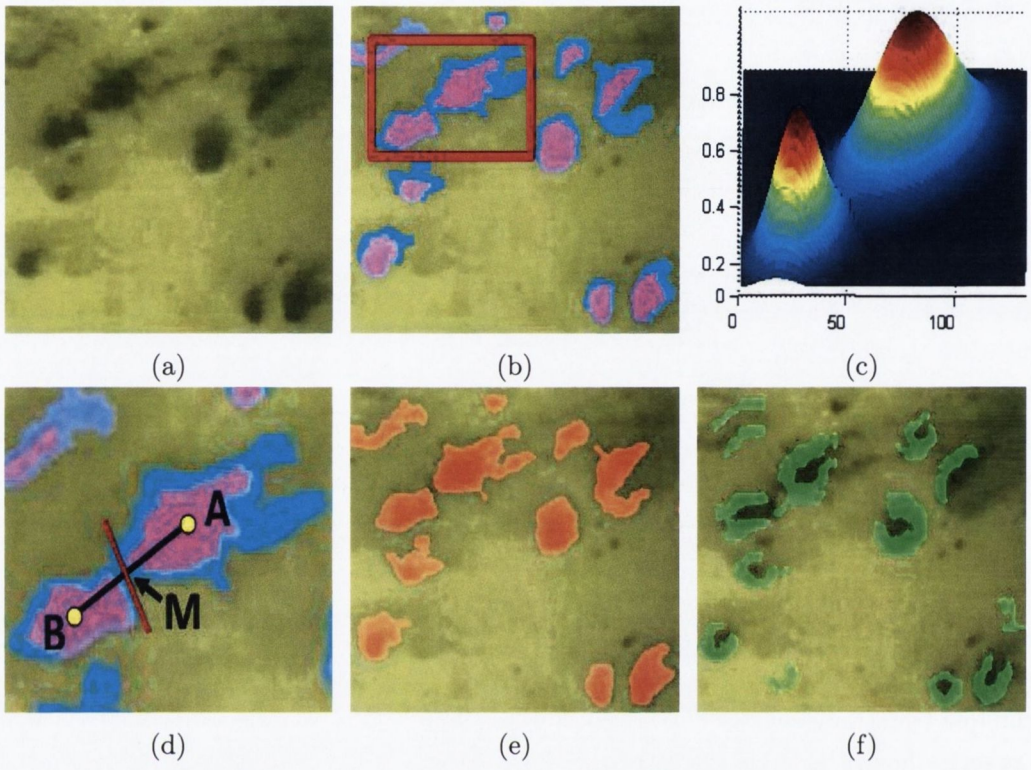
5.2.3 Labeling and Splitting

Locally connected claw mark and dark entrance regions, $L(\mathbf{x}) = \{1, 2\}$, of each candidate burrow are now labeled with unique identification numbers. The Connected Component Analysis technique by Sammet et al. [68], with a 3×3 neighborhood, is used to perform these labellings. In practice however, there are instances where multiple non-connected dark entrances are segmented together because of a common animal claw mark region, as illustrated in the red box in Figure 5.5 (b). As each dark entrance region corresponds to an individual burrow, these composite regions must be split. This splitting is accomplished in two steps. First, the shape of the particular composite region is modeled with a mixture of Gaussians equal to the number of dark entrance regions. Then, each component is separated from its neighbor, at the point where their local mixing weights are equal, along the line joining their respective means. Figure 5.6 illustrates this splitting process.

These mixture parameters are optimized using the Expectation Maximization algorithm [15]. In this algorithm, the mean, μ, and covariance matrix, Ω, of each component are initialized as $\mu = \sum \mathbf{x} I_n(\mathbf{x}) / \sum \mathbf{x}$, and $\Omega = \sum (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T$. Where $I_n(\mathbf{x})$ is the normalized intensity of the corresponding dark entrance area region component at image position $\mathbf{x}$, given by: $I_n(\mathbf{x}) = I_d(\mathbf{x}) / \sum I_d(\mathbf{x})$. This separation is performed on the dark region map, I_d as in this domain the intensity profile of the burrows are Gaussian-like i.e. it decreases from the center, whereas in the raw image it is the opposite.

5.3 Feature Choice and Extraction

In practice, although only dark regions are targeted in the object detection phase, a large percentage of the objects detected are not burrows. To eliminate these false alarms, Lau et al. [46] introduced 7 features. In this work a new set of features is developed, which are matched



(a) (b) (c)

(d) (e) (f)

Figure 5.6: (a) Original, (b) Initial segmentations with dark entrances (pink), claw marks (blue), and composite region (red box). This composite region is split by modeling the dark entrance regions with a (c) mixture of Gaussians, and then (d) separating them at point (M) where their components are equal. The final segmentation results are shown in (e), and (f) those from the previous method by Lau et al. [46].

more closely to the scientific observations. Brief descriptions of this existing feature set by Lau et al. [46] is now given (full details can be found in [46]), followed by details of the new feature set.

5.3.1 Existing Feature Set

In this set of features, Lau et al. [46] explored size, shading, shape and texture characteristics of burrows for their identification. There are seven features in total, which are described as follows:

Size

The size of objects are examined using the conspicuity map feature.

Conspicuity Map (c_o) . This feature is extracted as the total number of pixels in the segmented region. This value is then scaled by 2000, to range typically between 0 and 1.

Shading

As burrows are mainly dark, the grayscale shading of objects is useful for their identification. This characteristic is examined using the average intensity feature.

Average Intensity (a_i) . This feature is extracted as the percentage of grayscale pixels that are brighter (p_b) and darker (p_d) than the region average value i.e. $A_I = \{B_r, D_k\}$.

Shape

The overall shape of the object is examined using three features, the slant angle, run-length ratio, and shape orientation/ranking.

Slant Angle(s_a) . This feature is mainly developed for eliminating vertical trawl mark regions. It is computed as the percentage of times the distance between the line joining the minima of each vertical column and the line connecting the first and last pixels of the region, is greater than 10 (at 10 equally spaced positions). This value is then scaled by 10, to range typically between 0 and 1.

Run-length ratio (r_h) . Is the ratio of the vertical and horizontal run lengths [25] of the region image intensities. For this feature, burrow regions usually have values above 3, whereas lobster and other regions are around 1, as shown by Lau et al. [46].

Shape Orientation/ranking ($\mathbf{s_r}$) . Depicts the uniformity of the region shape based on a coded grid system. The system is created by subdividing the minimum bounding box of the region into a 4×4 grid. Each block in this grid is labeled (i.e coded) with a power of 2, in ascending order (0 to 3) from the top left corner, as shown in Figure 5.7. The codes corresponding to the middle pixel of each row and column of the region are then recorded. The feature is then extracted as the mean and variance of these codes, $\mathbf{s_r} = \{s_m, s_v\}$, which are then scaled by the quantity of codes to range between 0 and 1.

Texture

The grayscale texture of objects is examined using the Cross-over counting and homogeneity/co-occurrence matrices features.

Cross-over counting (c_c) . This feature measures the variation in the pixel intensities along the horizontal and vertical direction of the region. To count this variation a center baseline in each direction is established as the row and column with the minimum pixel intensity, while scanning horizontally and vertically respectively. Then the feature is calculated along the horizontal and vertical direction by counting the number of times the line joining the pixel with the minimum intensity between consecutive rows crosses the horizontal baseline.

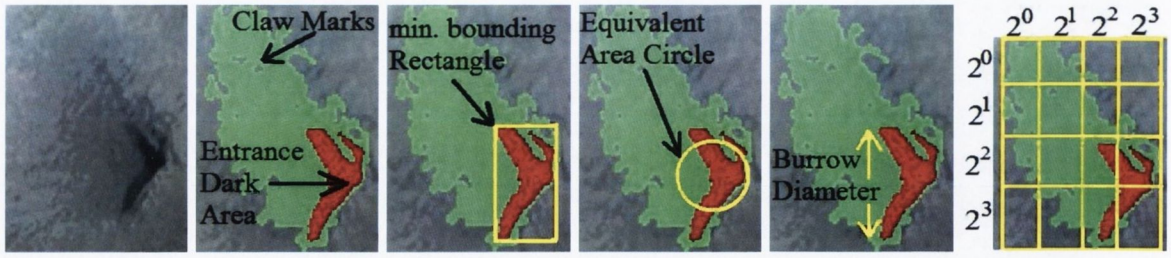


Figure 5.7: (a) Original image. Proposed feature extractions showing the: (b) segmentations of claw marks/scrapings (green) and entrance dark area (red), (c) the corresponding minimum bounding rectangle, (d) the equivalent area circle, and (e) the burrow diameter. (f) The existing shape ranking grid system used by Lau [46].

In a similar fashion, counts are performed along the vertical direction. Lastly the horizontal and vertical counts are summed. To have similar values as the rest of features, this feature is scaled by 20 to range between 0 and 1.

Homogeneity/co-occurrence matrices ($\mathbf{c_m}$) . Describes the texture of an object in terms of four pixel pair relationships obtained using gray-level co-occurrence matrices [28]. The first two relationships, x_l and x_h, are the sum of the horizontal and vertical pixel intensity pairs that are lower and higher than the region median intensity value respectively. While the last two pairs, y_h and y_v, are the horizontal and vertical pixel intensity pairs that are greater than a high threshold value (set as 100), to identify regions of high contrast, $\mathbf{c_m} = \{x_l, x_h, y_h, y_v\}$.

5.3.2 New Feature Set

The proposed set is based on the four main characteristics marine scientists search for in their current burrow analysis procedure [36]. As a result, marine scientists easily relate to these features and can use their values to verify decisions made during analysis. The first three characteristics are the condition, size and shape of the burrow entrances, and the fourth is the presence of claw marks or scrapings that creatures make while maneuvering in and out of burrows. These characteristics are important, as underwater studies by Marrs et al. [52] have shown that they are related to the occupancy, type and size of specific species. To capture these four vital characteristics, the following seven features were examined.

Burrow Entrance Condition

Burrows with caved-in entrances are deemed inactive and consequently not counted. Their key distinguishing characteristic is the absence of their dark entrance region. To identify these inactive burrows, one feature, the entrance dark area, a_d, was used.

Entrance Dark Area (a_d) . This feature is extracted as the number of pixels in the entrance dark region (obtained from the segmentation step). This value is then scaled by 1000, to range typically between 0 and 1.

Burrow Size

Underwater investigations conducted by Marrs et al. [52] using physical sampling and video surveillance techniques, have shown that burrows of various species can be identified based on their respective diameter. In their physical sampling analysis, a vernier caliper was used to take the burrow diameter meaurements by means of placing it into the entrance of the respective burrow. While for the video analysis, burrow diameters are measured as the longest diagonal along the burrow opening [36]. To capture this measurement, the burrow diameter feature, b_d, was used.

Burrow Diameter (b_d) . This particular measurement is extracted as the maximum distance between any two pixels in the dark entrance region. This value is then scaled by 100, to range typically between 0 and 1.

Presence of Claw marks

Displaced sediment due to species activity is commonly present around active burrows. It manifests as a brighter region surrounding the dark entrance region, as shown in Figures 5.7 (a)-(b). To examine the influence of this characteristic on burrow identification, the scrapings/claw mark feature, c_s, is used.

Scrapings/Claw Marks (c_s) . This feature was extracted as the percentage of the object region area outside the entrance dark region.

Core Dark Region Shape

The underwater studies carried out by Mars et al. [52] have also shown that the burrow entrances of various species have characteristic shapes. To capture this valuable information, four shape features: i) image moments, b_M, ii) Eccentricity, b_e, iii) Rectangularity, b_r, and iv) Circularity Fit, b_c, are extracted from the dark entrance regions, as follows:

Image Moments ($\mathbf{b_M}$) . These descriptors have been shown in the literature [2] [80] [35] to be effective at shape recognition tasks. They are basically specific weightings of the region pixel intensities with respect to the center of mass of the object [35]. For these experiments all seven moments are used, $\mathbf{b_M} = \{m_1, m_2, m_3, m_4, m_5, m_6, m_7\}$, as defined in [35].

Eccentricity (b_e) . This value describes how elliptical in shape the region is, with a value of 0 representing a perfect circle and 1 corresponding to a straight line segment respectively.

It is obtained by fitting an ellipse to the region. This is achieved using the relationship derived in [29] where the coefficients of an ellipse are equated to the first and second order moments of the respective region.

Rectangularity (b_r) . The rectangularity of a shape is the ratio of the region area to the area of the minimum bounding rectangle (see Figure 5.7 (c)). It has a maximum value of 1 for a perfect rectangle.

Circularity Fit (b_c) . This value describes how circular the region is, with a value of 1 representing a perfect circle. It is extracted as the percentage of the region area within a circle of radius, $r^2 = a_d/\pi$ positioned at the region center of mass, as shown in Figure 5.7 (d).

5.4 Classification Model Choice

The last stage of the recognition system is to classify the detected objects into burrow and non-burrow classes. To cater for the large diversity in burrow size and shape features, the use of two well established supervised learning classification schemes, a K-Nearest Neighbor (KNN), and a Support Vector Machine (SVM), are explored. The use of a non-parametric classifier (KNN), and one that uses linear discriminant functions (SVM) to perform classification, are explored because it is not known if the selected features would follow a particular model. The key advantage these two schemes offer, in comparison to the previously used Decision Tree scheme [46], is that they incorporate the use of training data into their classification process. The use of this data not only allow these systems to identify a large variety of burrows, but also facilitates easy adoption to new data sets. This adoption is performed by simply retraining the system with a new training set. An examination on combining these two classifiers with the hybrid scheme discussed in chapter 2, is also explored in the next section.

5.5 Optimization and Training Data Selection

These classification systems can become very complex depending on a number of factors such as the size of the feature space, and the quantity of training data etc. Apart from being computational expensive, the main drawback of overly complex systems is that they can classify the training data effectively, but may not perform well on other test sets. This situation is commonly referred to as overfitting [18]. To ensure this situation has not occurred, it is important to verify the classifier generalizes well with different training and testing data. To perform this verification three items have to be selected: i) features, ii) training data, and iii) the various model parameters for each classifier.

As a large data set (1000s of burrows) is being used for these experiments, it would be difficult to simultaneously select all three items, so selection is performed in a sequential procedure. In this procedure an appropriate training set and model parameters are chosen for selecting an

optimal feature subset. Then using this subset of features, optimal model parameters, and a training set that generalizes well with different test sets are selected. For feature selection, two approaches are examined. In the first approach, the redundancy among the entire feature set is examined directly using Principal Component Analysis [18]. While in the second approach, as there are not many features, the efficiency of all their various combinations are examined exhaustively. The steps undertaken to optimize the KNN, SVM and their hybrid combination, using this selection procedure are now presented.

5.5.1 Optimization of KNN Classifier

The performance of this classifier depends on three items, features used, the training data, and the neighborhood value. Optimization of these items is performed in three steps. First, a training set and a neighborhood value is selected. In this case, mosaic-5 is selected for training, and the data in the remaining mosaics are used for testing. This mosaic is initially selected for training as from visual inspection, it contains a large variety and quantity of burrow and non-burrow objects. While for the neighborhood, a value of $k = \sqrt{n_b} = 28$ is used (as recommended by Duda et al. [18], where $n_b = 785$, is the number of objects of interest (i.e. burrows) in the training set). Using this training set and neighborhood value, an optimal subset of features is selected. Then, in the second and third steps of this optimization, this subset of features, is used to select the best training set and neighborhood value respectively. Results obtained in each step of this sequential optimization approach are now presented.

5.5.1.1 Feature Selection for KNN

To select the best set of features, the performance of all 16,383 combinations of the fourteen new and existing features are examined. To spot any trends, the performances of the individual features are first examined, followed by an analysis of their combinations. From the individual feature analysis, two interesting points were noted. Firstly, as shown in Figure 5.8, the classification errors obtained from each of the new features are all below 5%, with the dark entrance area, a_d, being the best at 3.4% and the image moments, b_M, being the worst at 4.6%. These values were very good in comparison to the existing features, in which only the co-occurence matrices, c_m, obtained a low classification error of 4.6%, while the rest averaged 39.5%. The second interesting point noticed is that among the top 20% of combinations with the lowest classification error, each of the features, $\{\mathbf{c_m}, a_d, c_c\}$, are present in more than 75% of them. This level of consistency is very high compared to the rest of features, which are only present between 40-55% of these top combinations, as shown in Figure 5.8.

To examine the performance of the new, existing [46], and entire feature set combinations, plots of their classification error, recall and precision results are given in Figures 5.9 (a)-(c). As it is difficult to show all of these actual values, only the performance from the top ten and last five combinations from each set with the lowest classification error are given in Table C.1

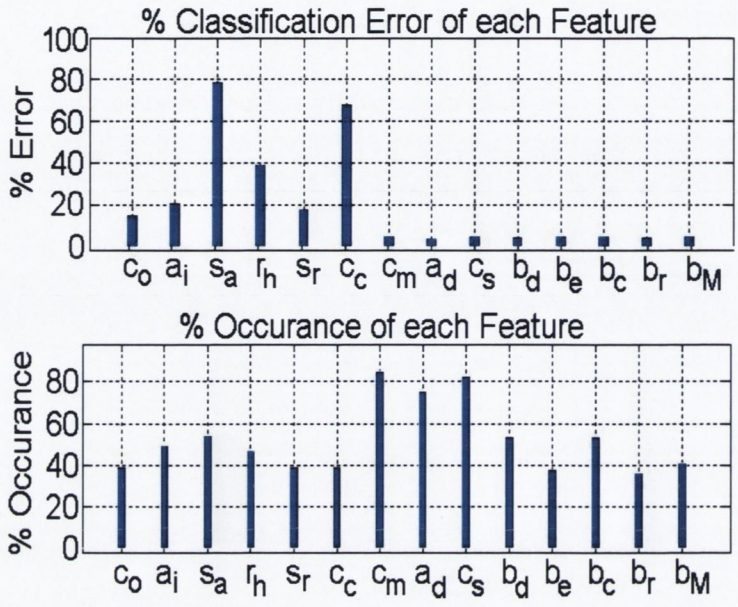
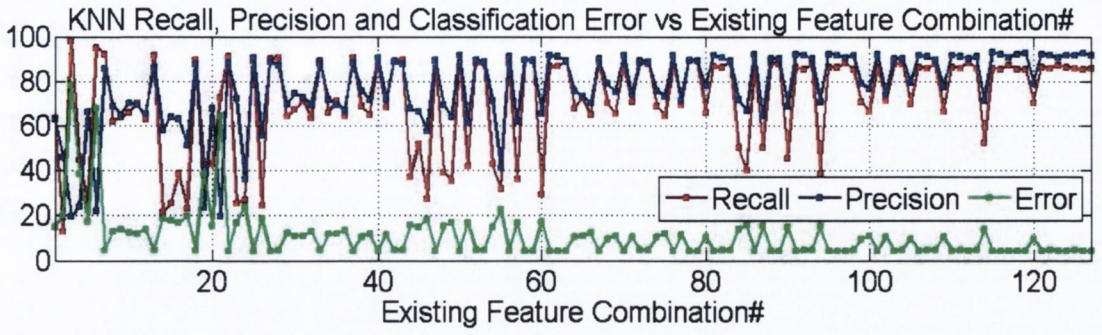


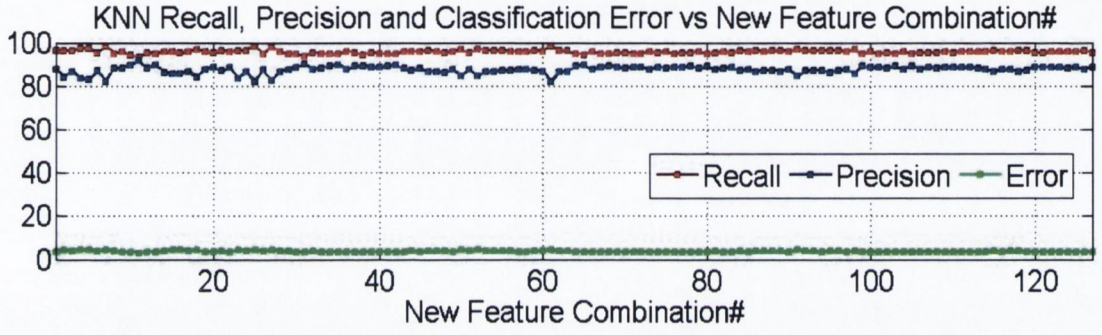
Figure 5.8: (Top) classification error obtained from each feature using the KNN, and (Bottom) their occurence among the top 20% of all combinations with the lowest classification error.

in Appendix C. From analyzing these results two interesting points are noted. Firstly, the classification errors from combinations involving the new features only, are all below 5%, with $\{a_d, b_c\}$ being the best at 3.1%, and $\{b_e, b_M\}$ the worst at 4.6%. These values are very good in comparison to the combinations involving the existing features only, where only those containing the $\mathbf{c_m}$ feature obtain classification errors below 5%, and those that do not average 17.4%. In this existing feature set, the $\{r_h, c_c\}$ combination perform the best, giving a classification error of 4%, while the s_a feature by itself is the worst at 78%. The second interesting point noticed is that 99% of the combinations from the entire set obtained classification errors below 5%, which is almost twice as much from the existing set by Lau et al. [46]. This implies that the addition of the new features generally improve the performance of the existing ones.

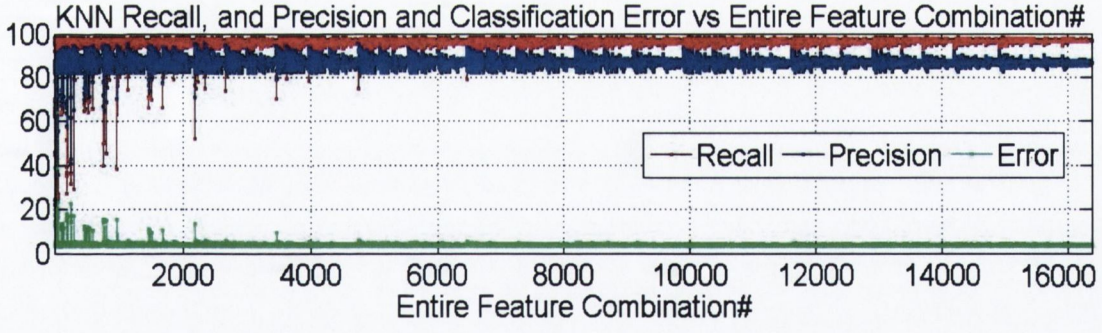
Among all these combinations, the $\{a_d, b_c\}$ from the new set remained the best with a classification error of 3.1%, and the s_a feature by itself from the existing set [46] remained the worst at 78%. The performance of the combination $\{a_d, b_c\}$ is even superior to the 3.6% classification error using PCA with the first 15 principal components, as detailed in Appendix C. Because of these superior reesults, with fewer features, PCA is not used in this classification scheme. Instead, the optimum combination of the dark entrance area and circularity fit features ($\{a_d, b_c\}$), is selected for use, as it gives the lowest classification error, and contains one of the most consistent features, a_d. Also, the use of just two features to model burrows improves user interpretability, and also lessens processing time.



(a)



(b)



(c)

Figure 5.9: The classification error (green), recall (red), and precision (blue) obtained from the KNN of the proposed system using feature combinations from the (Top) existing set by Lau et al. [46], (Middle) proposed set, and (Bottom) the proposed.

5.5.1.2 Training Data Selection

Using the feature subset of $\{a_d, b_c\}$ and the previous neighborhood value $k = 28$, a training set that generalizes well is now selected. This selection is performed by examining the performance of the system with different testing and training data. As each of the ground truth mosaics created contain a large amount of data, the usefulness of each one for training is examined. In these experiments, each of the mosaics is used for training, and tested against the sum of the

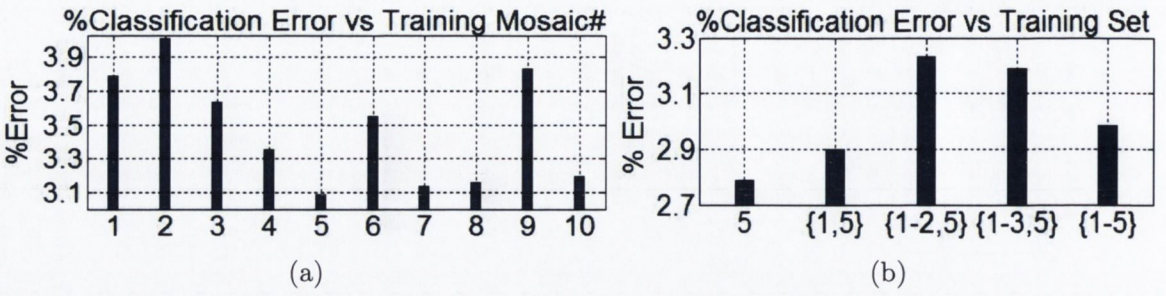


Figure 5.10: Classification error obtained from KNN when using training data as: (a) different mosaics, and (b) mosaic-5 combined with other mosaics.

objects in the remaining mosaics.

Figure 5.10 (a) shows the results from each individual mosaic when used for training. As seen, the classification errors obtained are low, ranging between 4-5%, with mosaic-5 performing the best at 3.1%, and mosaic-2 the worst at 4%. These low values imply the classifier generalizes well across the different training and test sets examined, and the data from any of the mosaics would be sufficient for training in these experiments. To examine if the performance of the system can be further improved, data from different combinations of mosaics are now examined for training. To keep this examination simple, different combinations with mosaic-5 are examined, as it achieved the lowest classification error. Using this limitation, mosaics 1-4 are cumulatively combined with mosaic-5 as training data, and tested against the sum of items from mosiacs 6-10.

The results obtained from these various combination experiments are given in Figure 5.10 (b). Examining these results show two interesting observations. First, the system continues to generalize well against this new test set, as the classification errors obtained are below 4%. Secondly, the errors obtained were all above the error obtained from using mosaic-5 for training by itself. This shows that randomly combining data from mosaics 1-4 with mosaic-5, for training, slightly degrades the performance of the system. These minor degradations are due to mosaics 1-4 containing a significant amount of small burrows that have a similar size to the non-burrow objects in the testing set. Based on these results, mosaic-5 is selected as the training set for use in these experiments, as it generalizes well against the data collected.

5.5.1.3 Neighborhood Selection

An optimum neighborhood value is now selected by analyzing the performance of the classifier with a range of k values from 1 to 100. These experiments are performed with the optimum feature set of $\{a_d, b_c\}$, the efficient training set of mosaic-5, and tested on the sum of items in the remaining nine mosaics. The classification errors obtained from these experiments are given in Figure 5.11, which show the system behaving in three states. In the first state, from $k = 1$ to $k = 4$, it is unstable with large fluctuations in classification errors, then it goes into an approximate steady state between $k = 5$ and $k = 30$, after which it generally degrades with

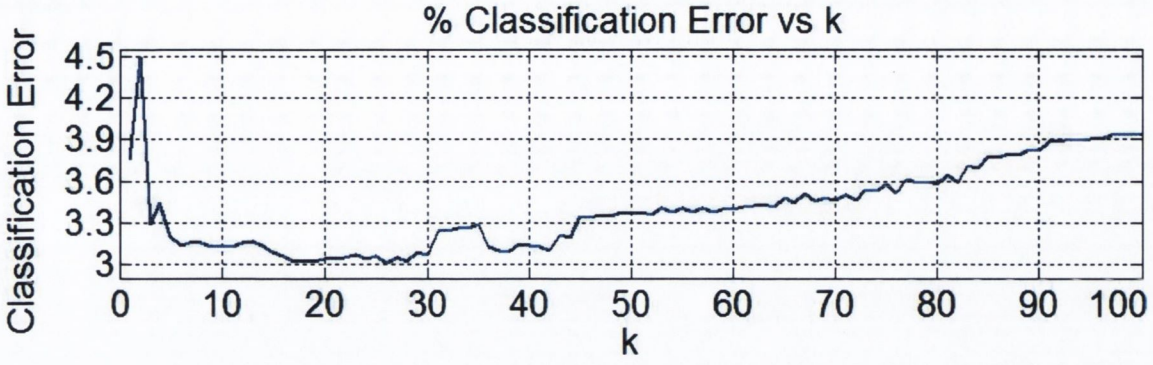


Figure 5.11: (a) Classification errors from KNN for different neighbourhood, k values.

increasing values of k. In the steady state, the system achieves the lowest classification errors between 3-3.1%, compared to the unstable state where the highest value of 4.5% at $k = 2$ is obtained. An interesting observation in the unsteady state is that a low classification error of 3.75% is achieved at $k = 1$. Using this value would significantly reduce system computations, as sorting the training data to obtain the nearest neighbours of the query object is no longer necessary. But, for these experiments, as speed is not the primary concern but accuracy, the value of $k = 28$ is selected as the optimal value, because it achieves the lowest classification error of 3.01%, and it lies in the stable region of the system. Although smaller values such as $k = 17$ obtain similar results, the larger value of $k = 28$ would help compensate more for imbalances among the classes, with additional training data in the future.

5.5.1.4 Removing Redundant Training Data

To reduce system processing, redundant training points are removed from the initially selected training set of mosaic-5. This is accomplished by using mosaic-5 as the training set to classify the sum of the items in the remaining nine mosaics, and then remove the training points that are not used in the procedure. In this procedure the optimal feature set of $\{a_d, b_c\}$, and neighborhood value of $k = 28$, are used. Using this reduction method, the original training set of 785 burrows and 3476 non-burrow items is reduced to 766 burrows and 176 non-burrow objects. This shows that there is a significant amount of redundant non-burrow items in the original training set.

5.5.2 Optimization of SVM Classifier

As mentioned in chapter 2, the SVM classifier from Matlab [54], is used to perform these experiments. Details of the specific settings used are also given in chapter 2. In addition to these settings, the performance of this classifier depends on two main items, the training data, and the features used. For the training data, the efficient and compact set obtained for the KNN from mosaic-5, as discussed previously, is utilized, and for feature selection, the PCA and

exhaustive combination approaches, as used for the KNN, are explored.

5.5.2.1 Feature Selection for SVM

Similar to the KNN, to select the best set of features for the SVM, the performance of all 16,383 combinations of the fourteen new and existing features are examined. To spot any trends, the performance of the individual features are first examined, followed by an analysis of their combinations. Figure 5.12 shows the classification errors obtained from each feature and their occurrence among the top 20% of combinations with the lowest classification error. Comparing these two graphs with the ones from the KNN in Figure 5.8, show the best performing features in the two systems are those from the new set (less than 5%), and the co-occurence matrices (6.2%) from the existing set. The six remaining existing features give a high average classification error of 32%, which is 17.3% greater than their average from the KNN. Another similar trend noticed is the dark entrance area feature is also present in more than 70% of the top combinations with the lowest classification error. In this scheme, the shape ranking feature (s_r) also has a high occurance of 73.2%, while the rest of features average 46.5%. This high level of consistency of the dark entrance area feature in both the KNN and SVM schemes, imply it is the most stable and suitable feature from the entire set for burrow identification.

The classification error, recall and precision results obtained from the combinations of the new, existing and entire feature sets are given in Figures 5.13 (a)-(c). Additionally, the performance of the top ten and five worst combinations with the lowest classification error are given in Table C.2, in Appendix C. Comparing these figures to the corresponding ones from the KNN in Figures 5.9 (a)-(f), three similarities are noted. Firstly, the classification errors from all of the combinations involving the new features only remained below 5%, with $\{c_s, b_c, b_r, b_M\}$ being the best at 3.6%, and $\{c_s, b_r\}$ the worst at 4.9%. Secondly, for the existing set, the co-occurence matrices remained the most valuable feature, as combinations with it give average classification error, recall and precision values of 5.5%, 95.5%, and 81.8%, while those without it average

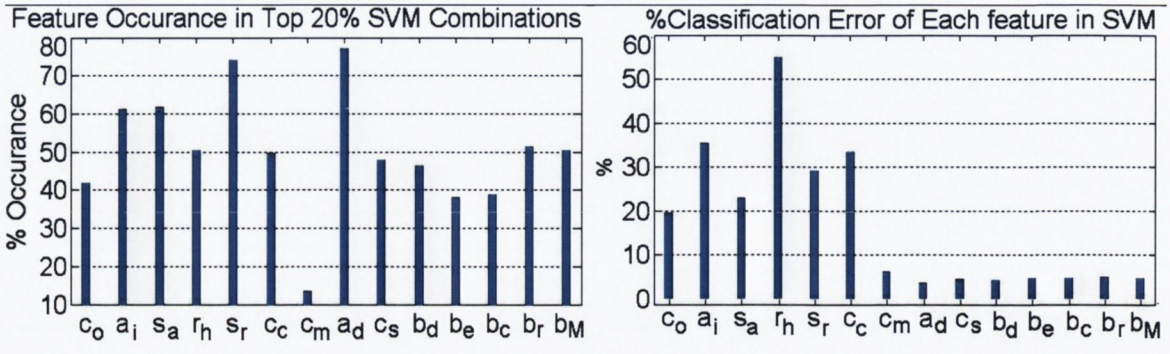


Figure 5.12: (Left) classification error obtained from each feature using the SVM, and (Right) their occurance among the top 20% of all combinations with the lowest classification error.

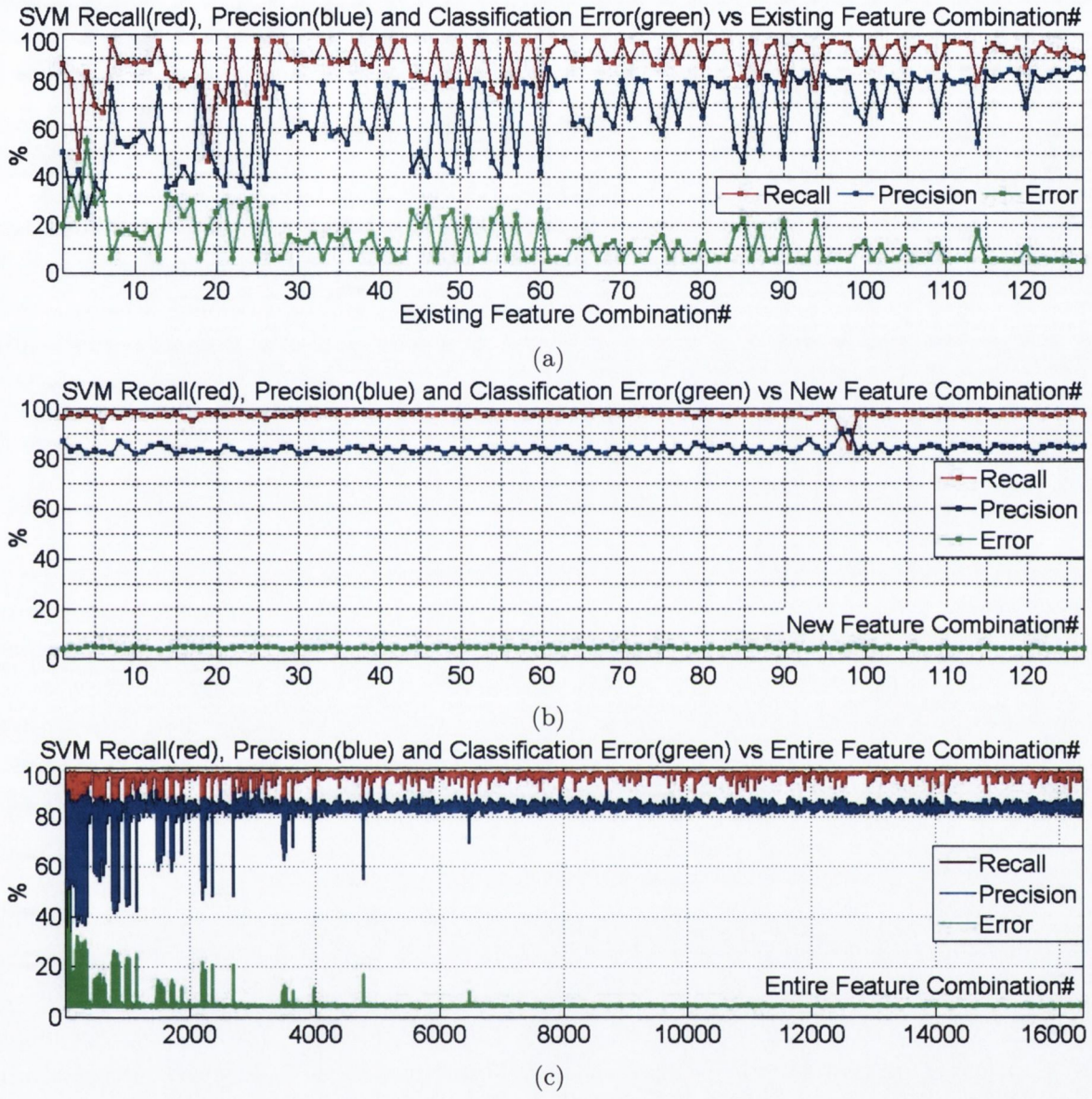


Figure 5.13: The percentage classification error (green), recall (red) and precision (blue) of feature combinations from the (a) existing, (b) new, and (c) entire sets.

19.9%, 81.8%, and 51.5% respectively. In this set, the $\{r_h, s_r, c_m\}$ combination performs the best with 4.6% classification error, 92.5% recall and 85.2% precision, while the r_h feature by itself performs the worst, with corresponding values of 54.9%, 84.0% and 24.2%. The last similarity spotted among these results and those from the KNN, is the general improvement in system performance with the addition of new features to the existing ones from Lau et al [46]. Evidence of this is seen as a large percentage (98.5%) of combinations from the entire feature set obtain classification errors below 5%, which is twice as much from combinations with the existing set

only. Among all combinations, the $\{r_h, a_d, b_M\}$ combination has the lowest classification error of 3.2%, with 95.9% in recall and 88.6% precision. While the run-length ratio by itself has the highest classification error of 54.9%, with 84.0% in recall and 24.2% in precision.

For this classification scheme, the use of PCA is also unnecessary, as superior results with less components are achievable with the original features. In particular the PCA best results of 3.7% in classification error, as shown in Appendix C, is achieved when all 25 components are used, compared to the original features, in which the combination of $\{r_h, a_d, b_M\}$ achieves 3.2%. This optimum combination of $\{r_h, a_d, b_M\}$, is selected for use in this classifier, as it gives the lowest classification error, and contains the most consistant feature, a_d. Also, as with the KNN, the use of a small subset of features (i.e. three) to model burrows improves user interpretability, and also lessens processing time.

5.5.3 Classifier Combination

To investigate if combining these two optimized classification schemes can improve results, the use of the hybrid approach, detailed in chapter 2, is examined. In this approach, as a pre-processing step, the KNN first prunes the SVM training set of objects whom k-nearest neighbors are not all of the same class. Then, the classification procedure is split amongst the two classifiers, where the KNN identify objects whom k-neighbors are all of the same class, and the SVM classifies the rest. To examine if this combination provides improved performance, it is compared to the individual results obtained from the optimized KNN and SVM schemes. Additionally, to examine if the pruning stage improves the performance of the system, its performance along with the SVM is evaluated with and without the pruned training data.

The results of these experiments, for testing on the sum of items from the nine remaining mosaics (excluding mosaic-5, as it is used for training), are given in Table 5.2. Analysis of these results show the hybrid scheme achieves a classification error of 3.23%, which is not superior to performance to the individual KNN (3.01%) and SVM schemes (3.23%). Also, pruning the training data degrades the performance of the SVM and the hybrid schemes, as their classification

Classifier	Classification Error	Recall	Precision
KNN	3.01	93.53	91.34
SVM	3.23	95.86	88.63
SVM_p	4.59	98.38	81.88
H_w	3.23	95.86	88.63
H_p	4.53	98.36	82.13

Table 5.2: Classification error, recall and precision results from the KNN, SVM and hybrid (H_w) classifiers without and with pruned training data (SVM_p, H_p).

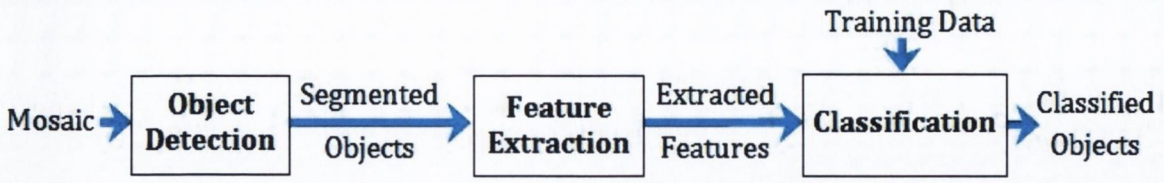


Figure 5.14: The processing pipeline of the proposed burrow recognition system involving: i) Object Detection, ii) Feature Extraction, and iii) Classification (which uses Training Data)

error without pruning (3.23%) is lower than with prunning (4.59% and 4.53%). In light of these results, the hybrid scheme examined here will not be considered further.

5.6 Proposed Classification Pipeline

With the features, training data, and classification models selected and optimized, burrow recognition can now be performed. The overall processing pipeline in this recognition system comprises of three major stages involving: i) object detection, ii) feature extraction and iii) classification, as shown in Figure 5.14. Candidate burrow objects are first detected by targeting dark contrasting regions in the generated video mosaics with the segmentation scheme described in Section 5.2. The relevant features are then extracted from each of these candidate objects and inputted into the respective supervised machine learning algorithm where they are finally classified. In this final stage, the machine learning algorithm incorporates information from the training data into the classification procedure.

5.7 Results

The optimized KNN and SVM classification frameworks are now evaluated. This evaluation is performed with four experiments using the nine remaining mosaics in Table 5.1 (i.e. excluding mosaic-5, as it is used for training) and their corresponding video sequences. In the first experiment, to examine which system is most suitable for this application, the performance of the KNN is compared to the SVM using all nine test mosaics. Then to examine how these systems perform to a previous state of the art system, they are compared to the video-based technique purposed by Lau et al. [46] using i) video and ii) mosaics. In the third experiment, the performance of all these classification schemes are compared to random guessing using Receiver Operating Characteristic plots. Lastly, because of the poor visibility conditions in some underwater environments, the robustness of these systems are examined in the presence of noise. The analysis of each of these experiments is as follows:

5.7.1 Comparison of Proposed method using KNN and SVM

To examine which of the classification schemes in the proposed method is more suited to this application, their performances are evaluated on the nine test mosaics. The classification error, recall and precision results obtained from these experiments are given in Figures 5.15 (a)-(c) and Table 5.3. Analysis of these results show the performance of the two classifiers are very similar. The average classification error from the KNN of 3.39% is however marginally lower than the SVM average of 3.60%. These classification errors suprisingly translated to the KNN achieving higher precision, but lower recall values than the SVM, in every test case. In particular, the average recall and precision values obtained from the KNN are 92.76% and 91.46%, while that from the SVM are 95.07% and 88.43% respectively. Based on these marginal differences, it is

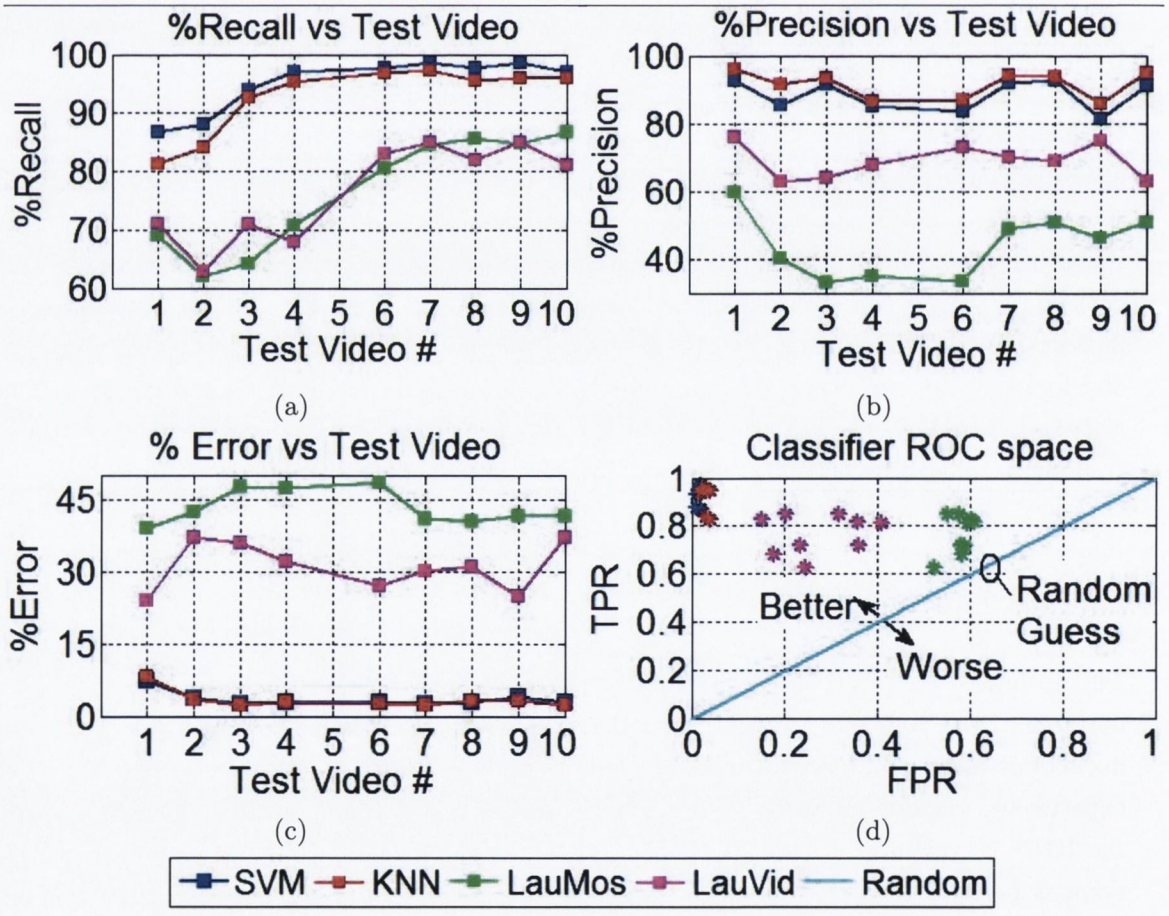


Figure 5.15: (a) Recall, (b) Precision, (c) Classification Error, and (d) ROC plots from the KNN (blue) SVM (red), Lau et al. [46] using video (purple), and mosaics (green) from each of the nine test mosaics, and (e) the legend used in all plots.

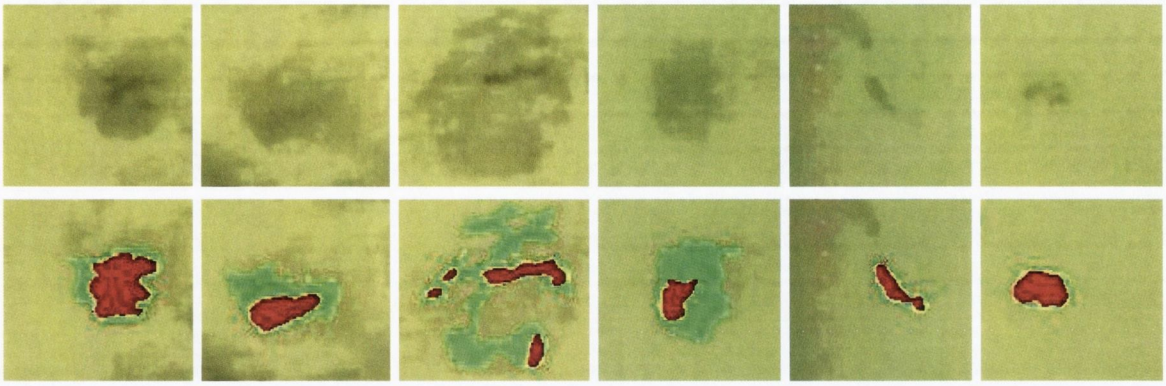


Figure 5.16: Falsely detected burrows by the proposed system because they have significant dark entrance regions. (Top) Original, and (Bottom) corresponding segmentations of dark entrance (red) and claw mark (green) regions.

difficult to make a decision as to which classification scheme is best suited for this application. The decision is even more difficult as all of the different parameters of the SVM such as kernel types etc. have not been exhaustively examined. So for this application either classifier would do, but the SVM would generally have a higher recall, whereas the KNN would generally have a higher precision and give marginally lower classification errors. Another key point drawn from the accuracy of the SVM is that burrow and non-burrow features can be separated with a decision boundary.

Further analysis of these results show two anomalies, the relatively high classification errors in mosaic-1 from both the KNN and SVM, and their low recall values in mosaics 1-2. The main source of these anomalies are due to the larger quantity of small and low-contrast burrows in these mosaics. Examples of these burrows along with other burrows that the system missed are shown in Figure 5.17. As seen, they exhibit very little dark entrance area, which make them difficult to identify with these classification schemes, as the non-burrow items in the training set have similar characteristics. As marine scientists are currently not interested in burrows of this small size and there are a relatively low quantity of low contrasting burrows in these test videos, further analysis to identify them is not pursued. Other sources of errors in the system are due to caved-in burrows that still exhibit a high level of contrast, as shown in example false alarms

Classifier	Classification Error	Recall	Precision
KNN	3.39	92.76	91.46
SVM	3.60	95.07	88.43

Table 5.3: Average classification error, recall and precision results from the KNN and SVM, calculated from the nine test mosaics.

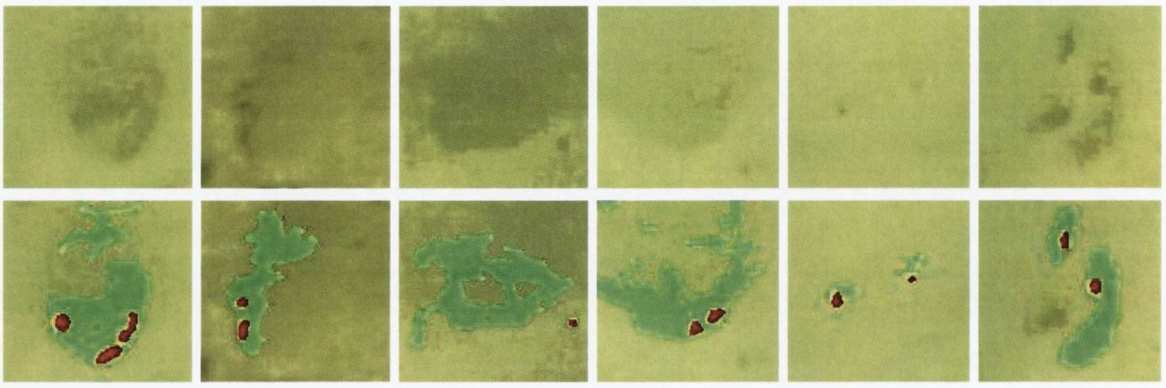


Figure 5.17: Burrows missed by the proposed system because they do not have sufficient dark entrance regions. The top row shows the original, and the bottom rows is the corresponding segmentations obtained, with dark entrance regions in red and the claw mark regions in green.

in Figure 5.16. Apart from these false alarms and missed burrows, the system does detect most of the burrows, as the average recall of KNN is 92.76%. Examples of these correctly burrows are given in Figure 5.18, which is a section of test mosaic-10.

5.7.2 Comparison with Previous Work

The performance of the KNN and SVM classifiers are now compared to the previous state of the art video-based technique proposed by Lau et al. [46] using: i) video and ii) mosaics. The video comparison is performed by manually cross referencing the classified objects obtained in each frame with the corresponding ground truth mosaic. The classification error, recall and precision results obtained from each mosaic, for this comparison, are given in Figures 5.15 (a)-(c). Analysis of these results show the KNN and SVM achieve improved results to the previous method. In detail, compared to the previous system using video, the average classification error, recall and precision values improved by 27.1%, 16.8% and 19.1%, using the proposed system. While in comparison to the previous method using mosaics, these corresponding values are improved by 39.1%, 16.8% and 44.7%. These results verify that it is not only possible to use mosaics to detect burrows, but improved results are achieved using this proposed technique compared to the previous method. The degradation in performance with the previous method using mosaics, compared to video, is due to the absence of the four-frame object consistency step in their algorithm. This step could not be performed with mosaics as they are only single images, and as a result, additional spurious objects due to noise are detected, hence explaining the degradation in the system precision. Examples of these spurious false alarms, along with missed and correctly detected burrows, are shown in Figure 5.18.

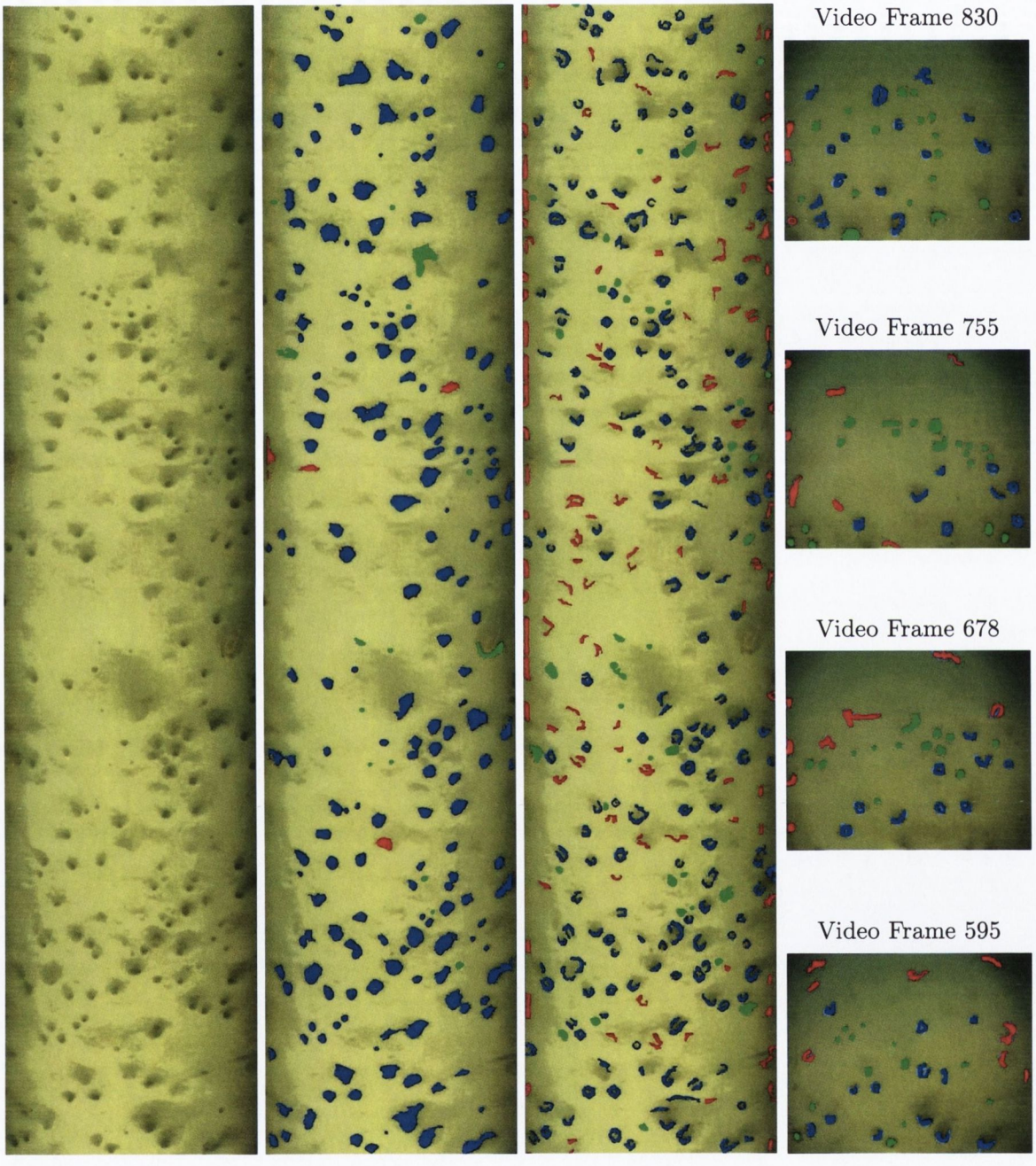


Figure 5.18: (a) Original, and correctly detected (blue), missed (green), and false alarm (red) burows, from (b) KNN, and previous method by Lau et al. [46] using (c) mosaics, and (d) video.

5.7.3 Comparison with Random Guess

The results obtained from random guessing are now used to benchmark the effectiveness of each classification scheme. To accomplish this task the line of no discrimination in the Receiver Operator Characteristic (ROC) space is used as the random guess benchmark, which is explained in Section 2.1. The results obtained from the test mosaics for each classifier in the ROC space are given in Figure 5.15 (d), which show all classifiers perform superior to random guessing. The average False Positive Rate (FPR) and True Positive Rate (TPR) values of 0.03 and 0.95 obtained by the SVM are closest to the perfect classification point of (0,1), the KNN is second with 0.02 and 0.93. In third place is the previous method by Lau et al. [46] technique using video with corresponding values of 0.28 and 0.76, and last is their technique using mosaics with 0.53 and 0.76. The increase in FPR using mosaics with the previous technique (in comparison to using video) is mainly due to the absence of the four-frame object consistency step in their algorithm, as explained earlier.

5.7.4 Robustness with Noise

Some of these video recordings suffer poor visibility due to floating sediments in the water medium. To simulate and examine the effects of this phenomenon on the performance of each classification scheme, they are now evaluated with different levels of additive white Gaussian noise. For these experiments as different images would have different initial levels of noise which would be difficult to estimate accurately, only one image is used, test mosaic-10. These simulated noisy images are created by adding the noise: $Z(\mathbf{x}) = \sigma N(0,1)$, to the original image. Where $Z(\mathbf{x})$ is a pseudorandom value at image location $\mathbf{x}$, that is drawn from the standard normal distribution $N(0,1)$, and σ is the respective noise level. Seventeen simulated noisy images are created for testing, with noise levels ranging from $\sigma = 0$ to $\sigma = 80$, in respective increments of $\sigma = 5$. Samples of some of these simulated noisy images, along with segmentations of the correctly detected, missed, and false alarm objects that are obtained from the purposed and previous methods are given in Figure 5.20. The corresponding performance of these methods in terms of their classification error, recall, precision, and ROC space plots, are given in Figure 5.19.

Analysis of these results are made with particular attention to the level of noise when experts can no longer perform analysis manually. This level is used to benchmark the robustness of a classification system, as beyond this limit results are not useful anymore, as they cannot be verified. To find this limit, four marine scientists were given the simulated noisy test images and asked which ones they could and could not analyze with confidence. From this simple visual experiment all four experts confirmed they could not analyze the data with confidence after noise levels of $\sigma = 20$, which is used as the limit in these experiments. In analyzing the performance of proposed system with the KNN and SVM at this level, the initial average recall and precision values only degrade by $\{5.9, 12.1\}$ to give values of $\{90.1, 81.7\}$. These

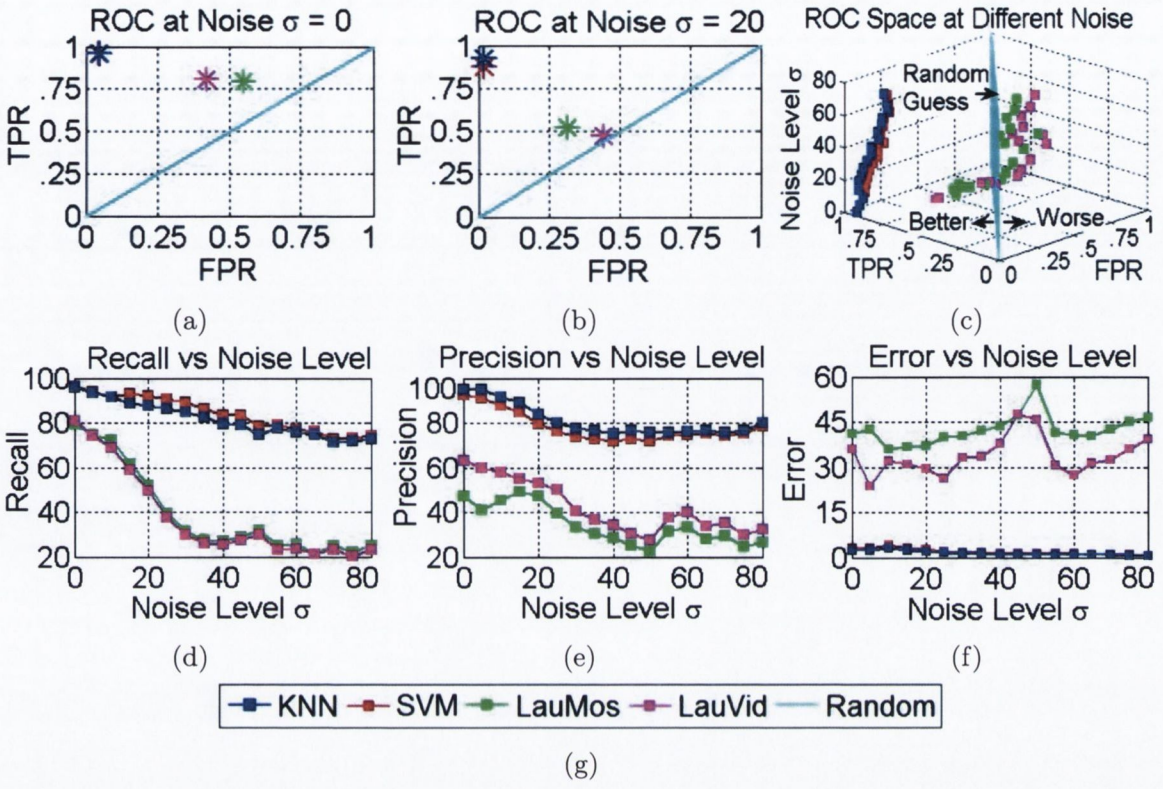


Figure 5.19: ROC plots of SVM, KNN, Random guess, Lau et al. [46] using video (LauVid) and mosaics (LauMos) at noise level (a) $\sigma = 0$, (b) $\sigma = 20$, (c) $\sigma = 0$ to $\sigma = 80$, and their corresponding (d) Recall, (e) Precision, and (f) Classification errors, and the (g) legend used.

values are very good in comparison to the previous method, which degraded by $\{31.5, 0.0\}$ using video and by $\{27.7, 10.0\}$ using mosaics, to give values of $\{49.5, 53.0\}$ and $\{52.1, 47.3\}$ respectively. Examination of the ROC plots show these degradations resulted in the previous method performing only marginally better than random guess, in contrast with the proposed method which maintain a superior performance.

Based on these initial results, the proposed method proves to be sufficiently robust to noise for this application, while the previous method is not. Even with extreme levels of noise ($\sigma = 80$), the purposed method with the KNN and SVM still performs good, obtaining average recall and precision values of $\{73.7, 79.2\}$, compared to $\{24.3, 29.5\}$ from the previous method using video and mosaics. Another interesting point noted is the area of the segmentations obtained from each method decreases with increasing levels of noise. In this analysis however, no deductions could be drawn from analysis of the classification errors obtained, because the number of objects detected increased significantly with the level of noise. This explains the decreasing trend in classification error obtained from the purposed method and fluctuating values in the previous method, as shown in Figure 5.19.

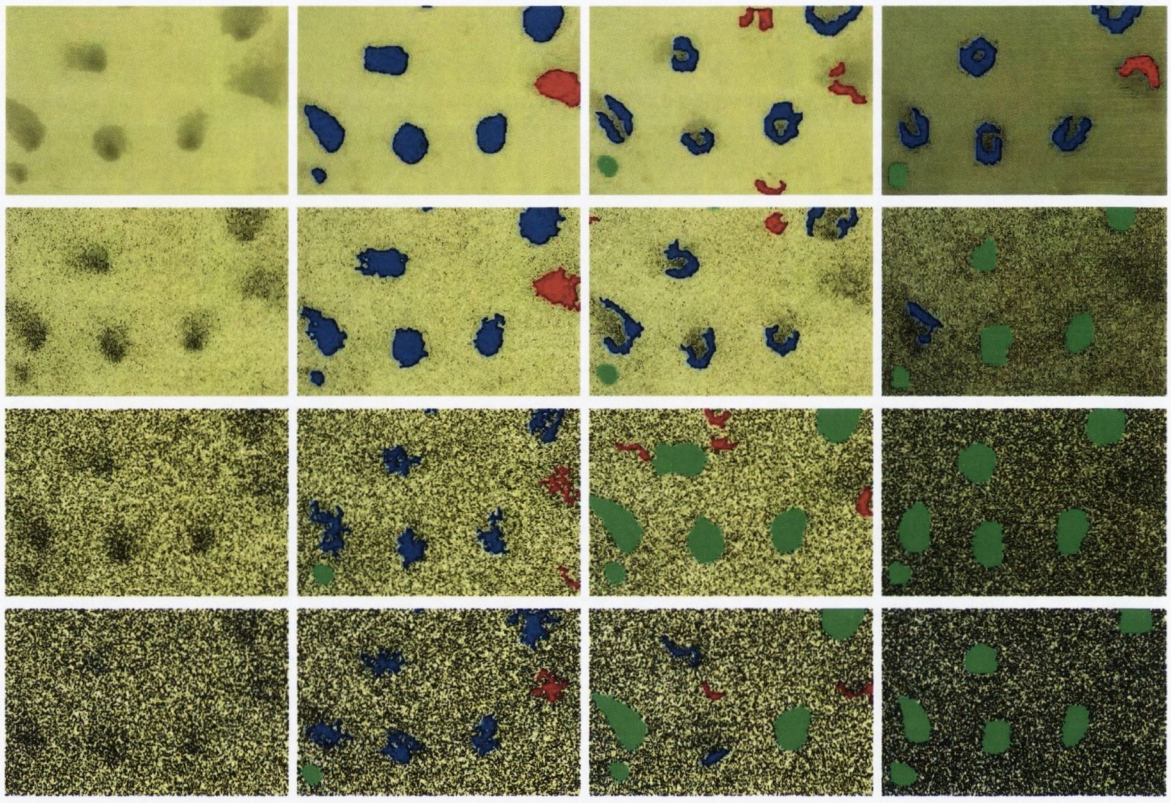


Figure 5.20: Simulated Noisy image (1^{st} Column), and correctly detected (blue), missed (green), and false alarm (red) burrows, using the KNN (2^{nd} Column), and previous method by Lau et al. [46] using mosaics (3^{rd} Column) and video (4^{th} Column), at noise levels (Top Row) $\sigma = 0$, (2^{nd} Row) $\sigma = 20$, (3^{rd} Row) $\sigma = 40$, and (4^{th} Row) $\sigma = 80$

5.8 Conclusion

In this chapter a novel technique for detecting burrows in marine surveillance videos is presented. This technique improves substantially on the previous state of the art method introduced by Lau et al. [46] using four key contributions. First, mosaics are used for performing object recognition, which improves visibility and reduces the tedious video inspection process currently performed by scientists to the browsing of a single image. Secondly, the use of classical segmentation techniques for performing object detection does capture most of the burrow regions, in contrast with the previously used edge detection method where only edges of the burrows are captured. Third, a new feature set is developed, which improves the performance of the existing set, and provides useful information as it is based on key characteristics scientists use in their current burrow analysis [36]. Although only a subset of these features are used for classification in these experiments, the rest can be used for providing useful information for scientists such as the diameter of burrows and the presence of animal claw marks. Lastly, to identify a large diversity

of burrows, the use of supervised learning classification schemes (KNN and SVM) are explored. These schemes use training data which can always be updated to adopt to any situation, as opposed to the strict set of rules used in the previous decision tree classification framework.

From the analysis of the various experiments performed throughout this chapter, five conclusions are made. First, the newly developed features are valuable for this application, as the KNN and SVM frameworks generally give superior results with them, when compared to using the existing features [46]. Second, because the KNN and SVM perform similarly, either of them can be used for this application, however the SVM generally gets a higher recall, whereas the KNN generally gets a higher precision and a marginally lower classification error. Third, because the recall and precision of the SVM is high, a linear relationship can be established to separate burrow and non-burrow features effectively, which in comparison to the KNN methodology, would reduce processing for this application. Additionally, these high performance values show that it is possible to use mosaics to detect burrows in underwater surveillance videos. Lastly, the results from the test sequences and noise analysis show the purposed system does obtain superior results, and is also more robust to additional noise, than the previous state of the art system by Lau et al. [46].

From the experimental analysis performed throughout this chapter six conclusions are made.

1. The newly developed features are valuable for this application, as the KNN and SVM frameworks generally give superior results with them, when compared to using the existing features [46].

2. Either the KNN or the SVM could be used for this application as they both performed similarly. However, it is observed that the SVM generally gets a higher recall, whereas the KNN generally gets a higher precision and a marginally lower classification error.

3. A linear relationship can be established to separate burrow and non-burrow features effectively as the SVM obtained high recall and precision values from the various experiments conducted. This relationship would reduce processing for this application in comparison to the KNN methodology.

4. It is possible to use mosaics to detect burrows in underwater surveillance videos as both the KNN and SVM recognition systems performed efficiently on the test sets examined.

5. The results from the test sequences and noise analysis show the purposed system does obtain superior results, and is also more robust to additional noise, than the previous state of the art system by Lau et al. [46].

6. The proposed system exhibited no signs of significant overfitting on the test sequences examined, as it performed far superior to that of random guessing in the ROC space. Also, the high recall and precision values obtained by the system indicate the classifica-

tion models are describing burrows effectively and not random noise (if overfitting was occurring).

For overfitting concerns in the future, it is unknown how well the system will generalize on different data sets, as it was only examined on data obtained from the Marine Institute in Galway, Ireland. In this institute data is collected by pulling sleds along the seafloor attached with lights and cameras. In other parts of the world however, the method of data collection employed may be different, such as in Portugal where underwater surveys are performed by pulling trawl nets attached with lights and cameras. These various collection methods result in data sets with dissimilar visual characteristics such as camera geometric distortions, lighting, motion, and even the burrow features to a certain extent if the geographical location is significantly different [52]. In order for the proposed system to generalize well on these different data sets it may need recalibrating. Another possible option for dealing with these different data sets is to develop and optimize separate recognition systems that are specific to the particular type of data. This is one aspect of the system that should be carefully examined in the future.

As a practical point of interest, when the scientists from the Marine Institute Galway are shown these results they agreed that this algorithm has the potential to significantly improve their current manual analysis procedure. Preliminary work with regards to detecting Nephrops themselves in mosaics is given in Appendix D. In this work, a similar approach is adopted from this chapter where segmentation is used for object detection and supervised learning classification schemes (KNN and SVM) are explored for recognition. From the experiments performed, the system obtained high recall and precision values (87.5% from the KNN). These results show that it is also possible to use mosaics to detect Nephrops in underwater surveillance videos.

6

Conclusions

This thesis presents a system for the analysis of seabed surveys that are used in the maintenance of Nephrops stocks off the coast of Ireland. The problem is challenging because of the heavy illumination degradation in the observed video as well as the difficulty in identifying the key features used by scientists in their manual analysis. Three main algorithmic components have been presented: enhancement, summarization and recognition. The following sections give a brief review of the key contributions made in each of these algorithms, and propose ideas for future work.

6.1 Image Enhancement

The image enhancement aspect of this work, detailed in chapter 3, involves correcting the radial degradations (vignetting) associated with the illumination distribution of the light source and the absorption from water in these images. To perform this correction three key contributions are made that improve on the state of the art vignetting correction technique by Kim & Pollefeys [45], as follows.

1. The introduction of a new spatial degradation model that:

   (a) Combines ideas from the vignetting [45, 91, 92] and underwater colour correction literatures [8, 69].

   (b) Does not restrict the shape and central location of the deteriorations to being circular and centered at the image center, but instead allows them to be elliptical and have a

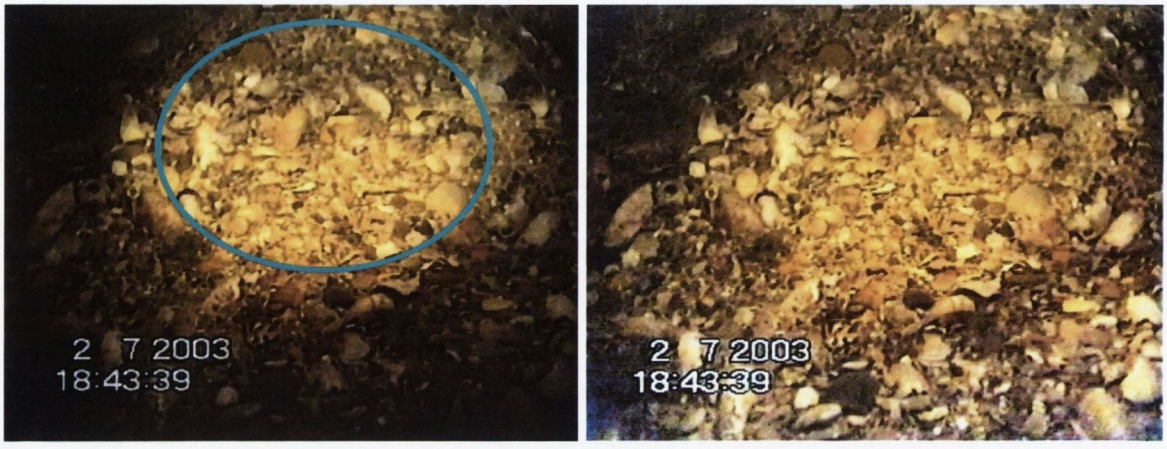


Figure 6.1: Original degraded image on the left with footprint estimate superimposed in blue, and the correction obtained using the proposed method is shown on the right.

central location at the center of the light beam footprint on the sea floor.

(c) Parameterizes the differing levels of degradation in each of the colour channels due to absorption from water.

2. A Bayesian approach for estimating the various parameters for this model, which uses point correspondences from consecutive frames to provide information on the illumination changes on the sea floor.

3. A novel correction procedure that can account for instances where radial degradations in illumination only occur outside of the footprint of the light beam on the sea floor, as seen in Figure 6.1. This procedure involves estimating the extent of the footprint region and incorporating it into a gain field where only the pixels outside of this region are amplified. As a result of this procedure, over-amplification can be prevented in the corrected images.

Apart from testing on a large set of underwater videos, experiments were performed that analyzed the impact degradations with different i) camera response functions, ii) shape, iii) central location, iv) footprint size, and v) colour deteriorations, would have on the performance of the proposed method. From the results obtained, four conclusions were made.

1. The proposed algorithm can remove a substantial amount of the degradations present in these videos, and hence improve their visibility tremendously. Sample results obtained are shown in Figure 6.1.

2. The camera response function does affect the final appearance of the corrected image, and hence incorporating it into the proposed method may improve its performance.

3. The high degradation rate that is present when transiting out of the footprint region affects the shape estimates, and hence overall performance of the algorithm.

4. For degradations that are not circular in shape and centered at the image center, the proposed algorithm does achieve superior results compared to the state of the vignetting method by Kim & Pollefeys [45]. However, when the shape and center estimates are incorporated into their algorithm improved results are obtained. This improvement verifies the need for good shape and center estimates for correcting these types of degradations.

Although initial experiments show the proposed method can enhance visibility in actual underwater sequences, it has one main drawback in that it needs motion for good parameter estimates. For videos where there is little or no motion, this drawback can be overcome by allowing the user to initialize the system. This should not be a difficult task as there are not many parameters. To facilitate this initialization, the use of a Graphical User Interface that shows the corrections obtained when different parameters are altered, would be very useful.

6.2 Content Summarization

The next major area of research in this thesis is content summarization. This involves creating a large area view or mosaic of the surveyed sea floor by aligning and rendering overlapping video frames together. To generate high quality mosaics from these underwater survey videos three key contributions are made.

1. Improving the existing blob-based image alignment technique of Matas et al. [53], for these sequences by:

   (a) Performing blob detection in the Difference of Gaussians image, where the absolute brightness due to the uneven illumination in these images has minimal impact on feature extraction.

   (b) Refining the feature matching phase with a pixel matching technique to get increased precision in the location of features.

   (c) Using a Bayesian framework for registration which takes advantage of the smooth continuous motion in these videos to robustly align images when there is little or no matching features available.

2. For rendering overlapping regions, the estimates for the center of the light beam footprint on the sea floor is used to capture well lit image details in the generated mosaic.

3. The introduction of a system to cross reference sections of the generated mosaic to its corresponding video frames.

In addition to testing on a wide corpus of underwater survey videos, experiments were performed using videos created with known motion and simulated degradations. The results from these tests show the new alignment and rendering techniques obtains improved results compared

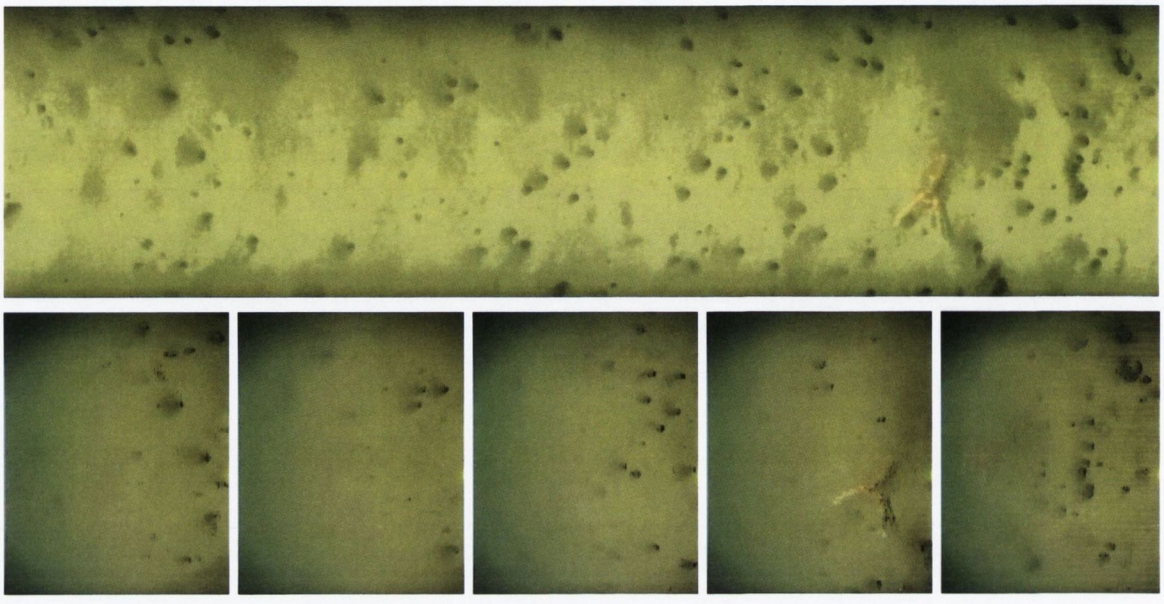


Figure 6.2: (top) Mosaic generated from 200 frames, samples of which are shown in bottom row.

to three state of the art mosaicking algorithms from the literature [9,26,65]. Additionally, when tested on actual Nephrops survey videos, the mosaics generated improved the visual context, making it much easier to quickly view the burrows and their inter-relationships, as seen in Figure 6.2.

Although most of these initial results show the proposed algorithm can generate high quality mosaics, two general cases are observed that can degrade its performance. First, is the loss of the valuable high frequency details when motion blur occurs. In the future this problem may be rectified by exploring the use of deblurring algorithms, or excluding these blurry frames from the mosaic generation process. The next problem spotted in these tests is a trail of laser dots that occurs as a result of the movement of the lasers sometimes attached to the survey apparatus. This problem can be solved by tracking these laser regions and masking them out of the rendering process.

6.3 Content Analysis

The last major area of research is identifying burrows automatically from the generated video mosaics. This item involves first detecting candidate regions and then classifying them as burrow or non-burrow objects. To perform this task, four key contributions are made that improve on the video-based state of the art system developed by Lau et al. [46], as follows.

1. Recognition is performed using mosaics, which summarizes the results in a single image for scientists to inspect.

2. A novel object detection technique is developed that:

   (a) Performs detection in the difference of Gaussians image to robustly detect candidate burrow regions in unevenly lit areas.

   (b) Uses state of the art segmentation and shape modeling techniques to capture two key scientific burrow components, which are their dark entrance and the surrounding animal claw mark regions.

3. A new feature set is introduced for distinguishing between burrow and non-burrow objects. This set is motivated by the current scientific description of Nephrops burrows [36], such as the burrow diameter, dark entrance area, the presence of animal claw markings etc., allowing experts to relate easily to them.

4. To identify a large diversity of burrows, the use of supervised learning classification schemes (KNN and SVM) are explored. These schemes use training data which can always be updated to adopt for different situations, as opposed to the strict set of rules used in the previous decision tree classification framework.

In addition to evaluating the KNN and SVM, experiments were performed comparing the performance of the proposed system with the state of the art by Lau et al. [46], and against random guessing. From the results obtained, five conclusions are made.

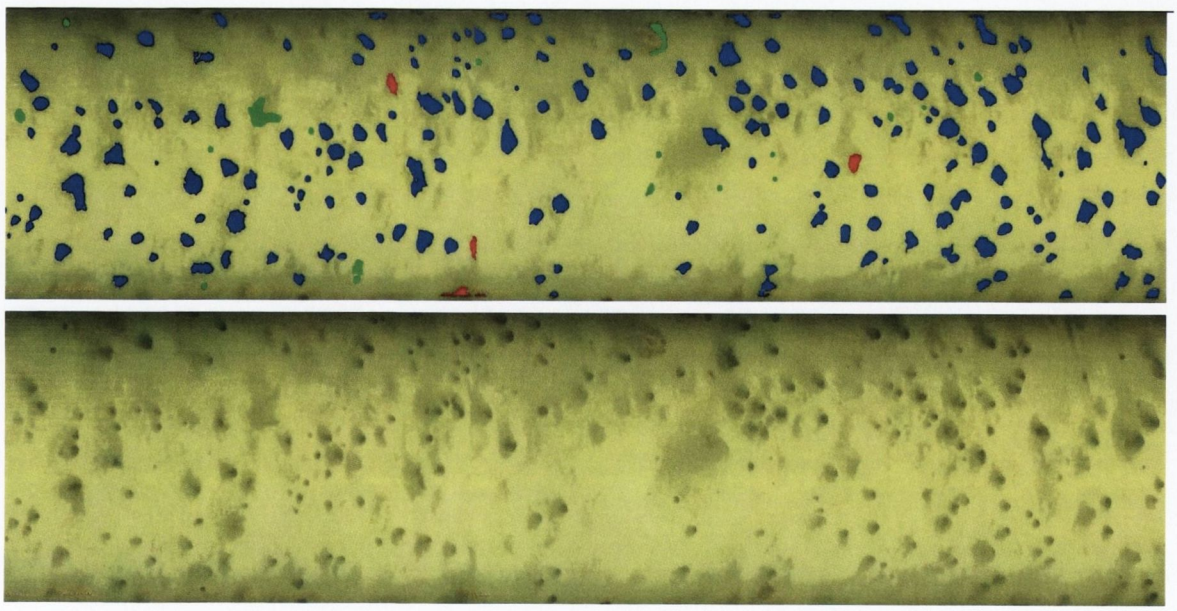


Figure 6.3: (top) Original Mosaic generated from 200 frames, correctly detected (blue), missed (green), and false alarm (red) burrows, using the KNN from the proposed system.

1. the KNN and SVM perform similarly, however with the training set used, the SVM generally gets a higher recall, whereas the KNN generally gets a higher precision and a marginally lower classification error.

2. The proposed system obtains superior results, and is also more robust to additional noise, than the previous state of the art system by Lau et al. [46].

3. The newly developed features are valuable for this application as the KNN and SVM frameworks generally give superior results with them, when compared to using the existing features [46].

4. The high recall and precision values (95.1% and 88.4%) obtained from the system implies.

   (a) For the SVM classifier, a decision boundary can be used to separate burrow and non-burrow objects, which in comparison to the KNN methodology, would reduce processing for this application.

   (b) It is possible to use mosaics to detect burrows in underwater surveillance videos.

Sample results obtained using the KNN in this system are shown in Figure 6.3. From a cursory glance, the burrow recognition algorithms developed give good results (recall and precision over 90% on the test sets examined). Preliminary work with regards to detecting Nephrops themselves in mosaics is given in Appendix C. In this work, a similar approach is adopted from this chapter where segmentation is used for object detection and supervised learning classification schemes (KNN and SVM) are explored for recognition. From the experiments performed, the system obtained high recall and precision values (87.5% from the KNN). These results show that it is also possible to use mosaics to detect Nephrops in underwater surveillance videos.

Having established the robustness of the burrow detection work, Nephrops complex identification can now proceed with some confidence. This is the main future work that has to be done next. Similar to the image enhancement technique, a Graphical User Interface would also be useful for this application. In this interface the video and corresponding mosaic with and without detected objects can be viewed side by side to facilitate quick referencing and verification of the automated results. Also, editing capabilities of this GUI can allow scientists to select and store the burrows and associated clusters they identify, for comparison with other experts.

6.4 Final Thoughts

As a practical point of interest, marine scientists are quite pleased with these initial results, and commented as follows. For the image enhancement technique, they found correcting the illumination degradations in these images did improve visibility which helped them in their manual inspection. For the content summarization, they said it is much easier to spot spatial relationships among burrows with the mosaics than with the original videos. Lastly, for the

burrow recognition, they said the accuracy of the system is sufficient enough to benchmark the total number of burrows within the surveyed area. This benchmark value provides further statistical information for them such as the variability of different species found in that particular area [52].

A comparison of using mosaics and video for identifying Nephrops complexes is given in Appendix B. In this comparison it is shown there are inconsistencies in the counts obtained from different scientists. One of the key observations made is that the counts obtained from the mosaics are generally greater than the corresponding video, this possibly implies the improved visibility and field of view does help scientists. Form this comparison, the main advantage pointed out by scientists, is the use of mosaics allow them to easily resolve discrepancies in Nephrops burrow counts among different experts. An example of this discrepancy is shown in Figure 6.4, which is the selected Nephrop burrows (yellow) and their associated complexes (red), obtained from three separate scientists. These scientists are Adrian Weetman from the Marine Laboratory in Scotland, Alessandro Ligas from the Biosciences Institute in Belfast, and Jennifer Doyle from the Marine Institute in Galway. As seen, there is a large variation in the selected burrows and complexes among these three professionals. Currently, with the use of only video, scientists only use the burrow count number to compare results, as it is too tedious to annotate thousands of video frames. As a result, the actual burrows identified by different users is not known. Using just a number, discrepancies are resolved by either: i) repeating the entire manual inspection until the counts become similar, ii) discarding counts with large deviations, or iii) taking the average. But with the use of mosaics scientists can simply scan the selected results from other experts, identify the actual discrepancies, and resolve them quickly.

It is possible that the proposed burrow recognition application can further help to resolve these discrepancies in Nephrop complex counts. First, all of the detected burrows can be indexed. Then when users manually select their respective Nephrops burrows and their corresponding complexes, the indexing system can be used to automatically identify discrepancies among the different users.

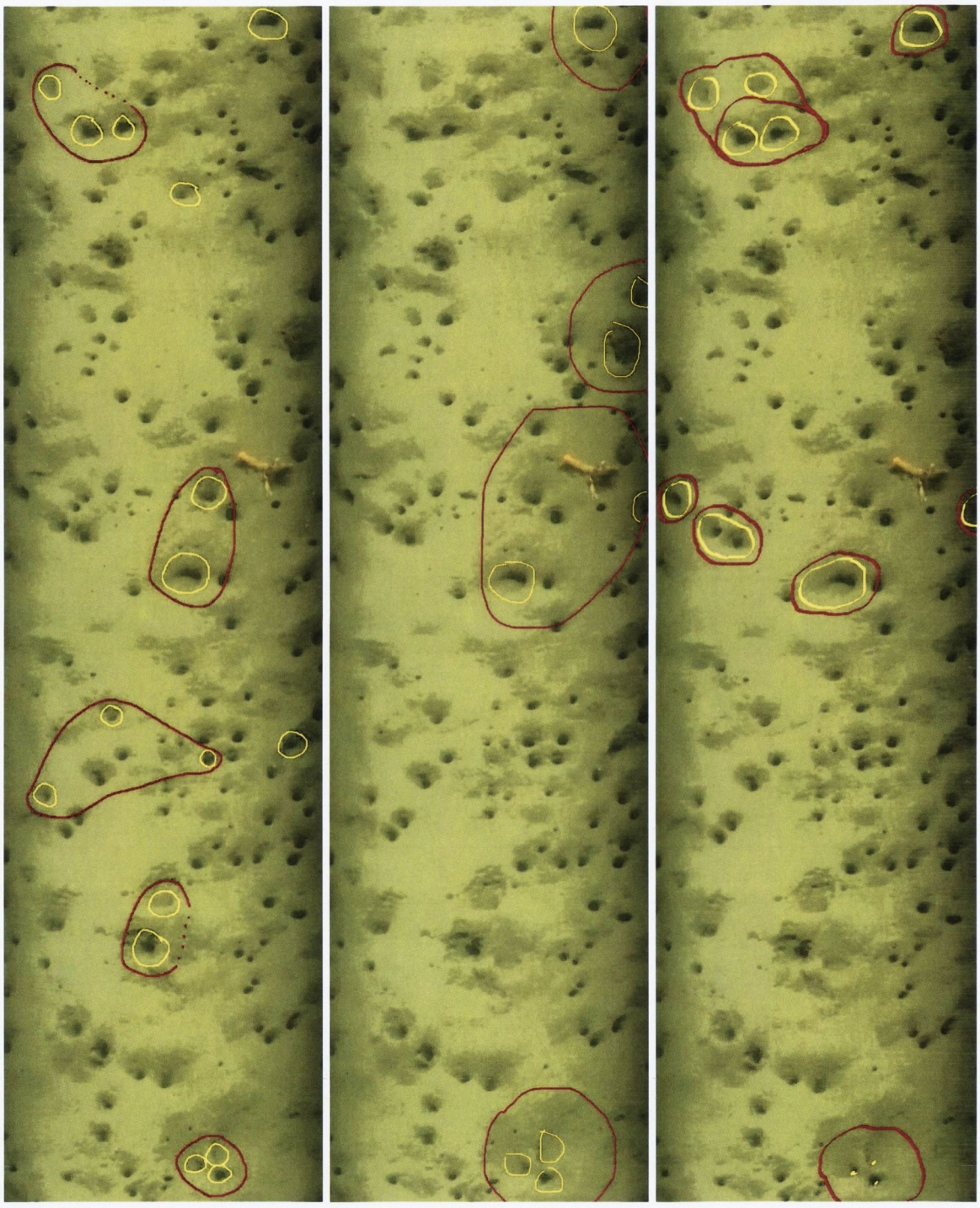


Figure 6.4: Manual selections of Nephrops burrows (yellow) and their corresponding complexes (red), obtained form scientists: (Left) Adrian Weetman from the Marine Laboratory in Scotland, (Middle) Alessandro Ligas from the Biosciences Institute in Belfast, (Right) Jennifer Doyle from the Marine Institute in Galway

A

Supplementary Results for Chapter 3

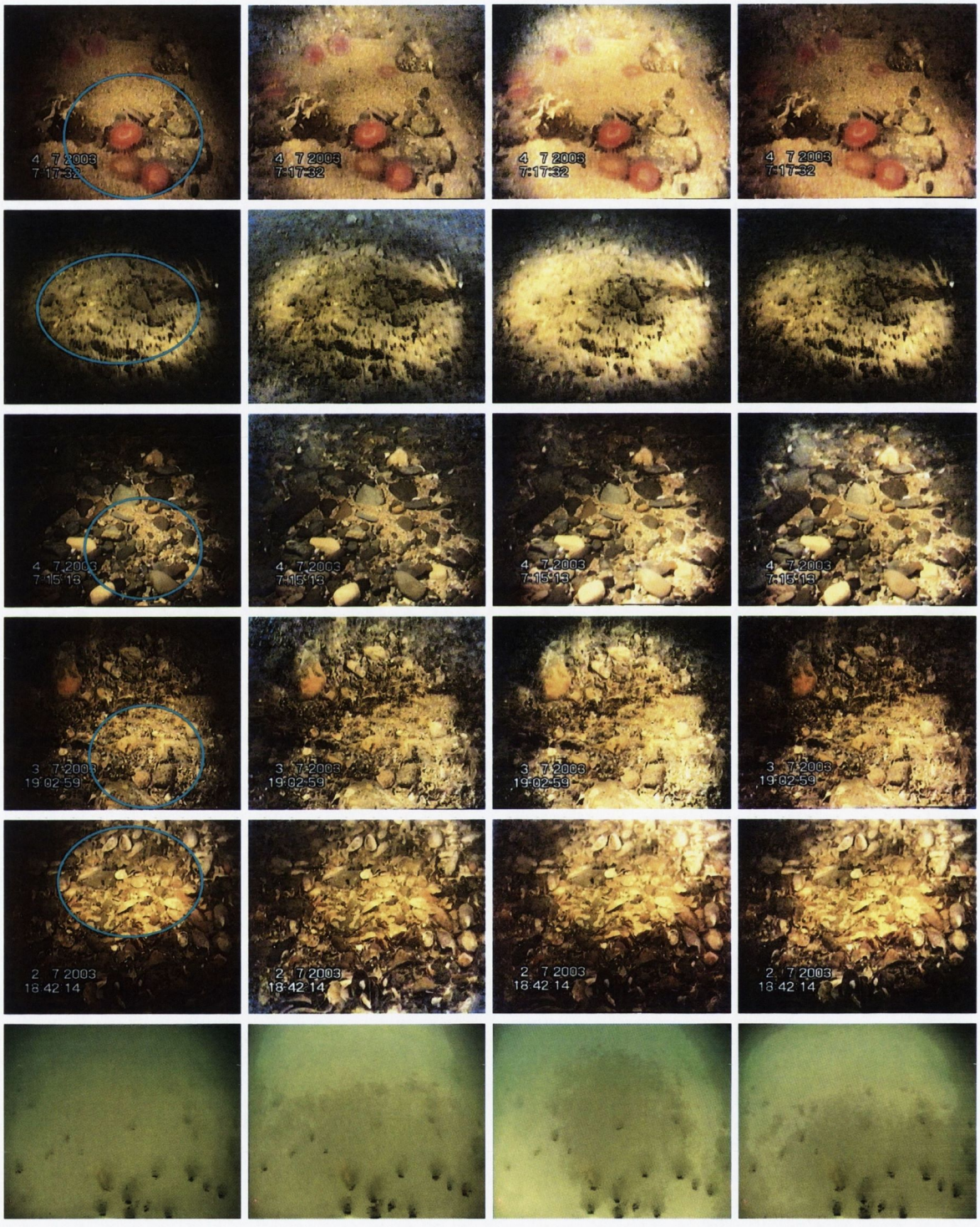


Figure A.1: (1^{st} Column) Original degraded images with footprint superimposed in blue. (2^{nd} Column) Proposed correction. (3^{rd} Column) Kim & Pollefeys [45] original correction. (4^{th} Column) Correction after incorporating $\mathbf{c}$ and $\mathbf{V}$ estimates into Kim & Pollefeys [45] method.

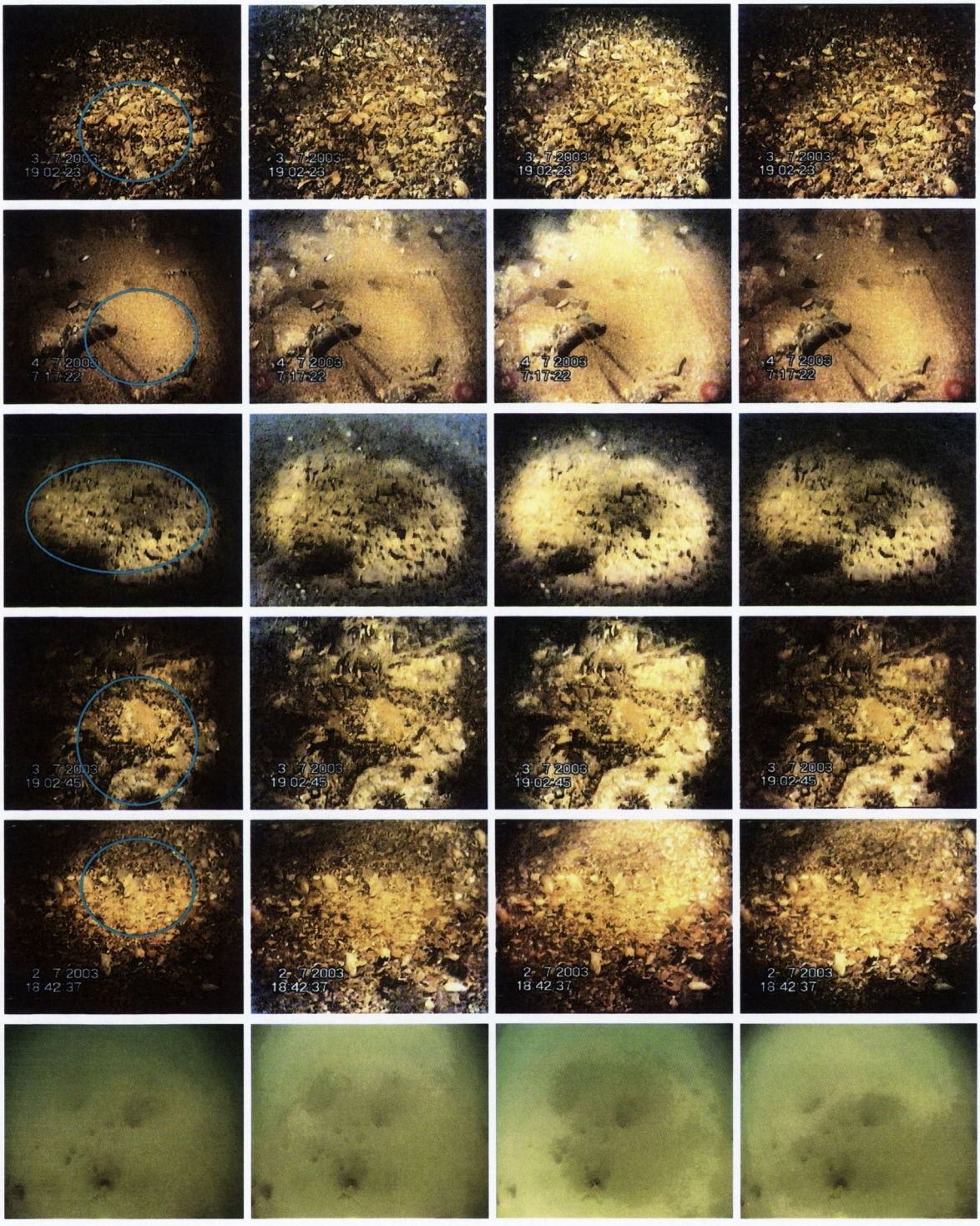


Figure A.2: (1^{st} Column) Original degraded images with footprint superimposed in blue. (2^{nd} Column) Proposed correction. (3^{rd} Column) Kim & Pollefeys [45] original correction. (4^{th} Column) Correction after incorporating $\mathbf{c}$ and $\mathbf{V}$ estimates into Kim & Pollefeys [45] method.

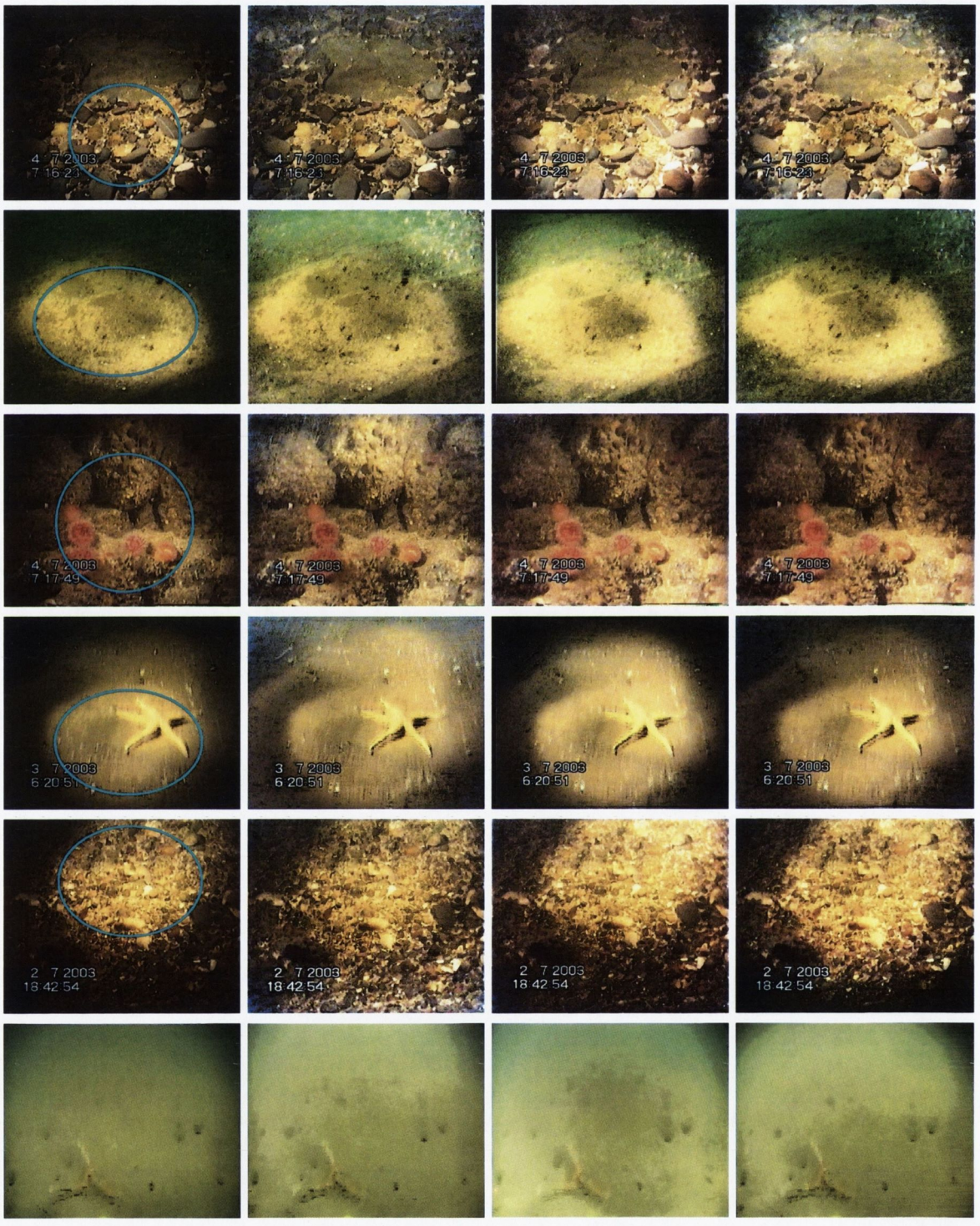


Figure A.3: (1^{st} Column) Original degraded images with footprint superimposed in blue. (2^{nd} Column) Proposed correction. (3^{rd} Column) Kim & Pollefeys [45] original correction. (4^{th} Column) Correction after incorporating $\mathbf{c}$ and $\mathbf{V}$ estimates into Kim & Pollefeys [45] method.

B

Analyzing the Selection of Nephrops Burrow Complexes from Different Scientists

This section presents an analysis of Nephrops complexes selected by three marine scientists: Jennifer Doyle [1], Alessandro Ligas [2], and Adrian Weetman [3]. The data for this analysis is made into two sets: videos and their corresponding mosaics. The videos are obtained from twenty 2-minute sequences (PAL format) of actual underwater surveillance videos. These were supplied by marine scientist, Jennifer Doyle, and represent real data used for Nephrops analysis by the Marine Institute. Using these videos, the corresponding mosaics are generated using the algorithm described in chapter 3. The test videos and mosaics were given (emailed) to each scientist with instructions to perform the Nephrops burrow complex counts as follows:

1. Analysis should be performed on all of the videos first, followed by a short break (about an hour) before beginning analysis on the mosaics.

2. Counting of the Nephrops complexes in the videos should be performed in the usual manner of clicking mechanical tally counters while inspecting the captured video playing at its recorded speed of 25 frames per second, on an 18 inch television screen.

3. Perform the mosaic analysis using the common commercial software, Microsoft Paint, to label the Nephrop complexes (red) and their corresponding burrows (yellow).

[1] jennifer.doyle@marine.ie from the Marine Institute, Galway, Ireland

[2] Alessandro.Ligas@afbini.gov.uk from the Fisheries and Aquatic Ecosystems Branch, AFBI - Agri-Food and Biosciences Institute, Belfast, Northern Ireland

[3] A.Weetman@marlab.ac.uk from the Scottish Government Marine Laboratory, Aberdeen, UK

From the three scientists, Adrian and Alessandro returned results from all 20 videos and mosaics, and Jennifer returned from the first 10. Their individual results are first analyzed, followed by a group examination as follows.

B.1 Individual Analysis

Plots of the video and mosaic complex counts obtained from each scientist are given Figure B.1, and sample labels from test mosaic-8 are shown in Figure B.2. Analyzing the results six key observations are made.

1. The counts from the video and mosaics are different in almost all cases. The standard deviation of the difference between the video and mosaic complex counts obtained by

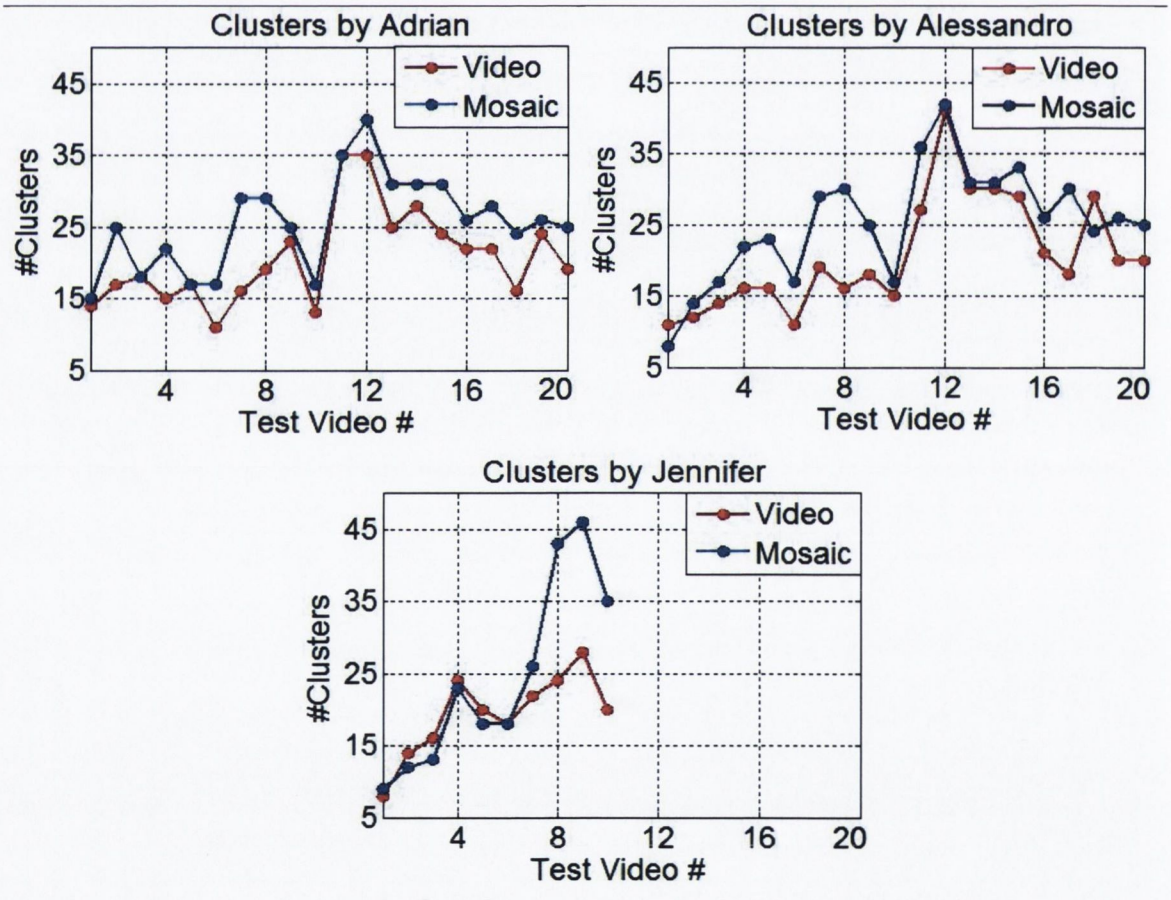


Figure B.1: Clusters counted in video (red) and corresponding mosaics (blue) from three marine scientists: (top-left) Adrian Weetman from the Marine Laboratory in Scotland, (top-right) Alessandro Ligas from the Biosciences Institute in Belfast, and (bottom) Jennifer Doyle from the Marine Institute in Galway.

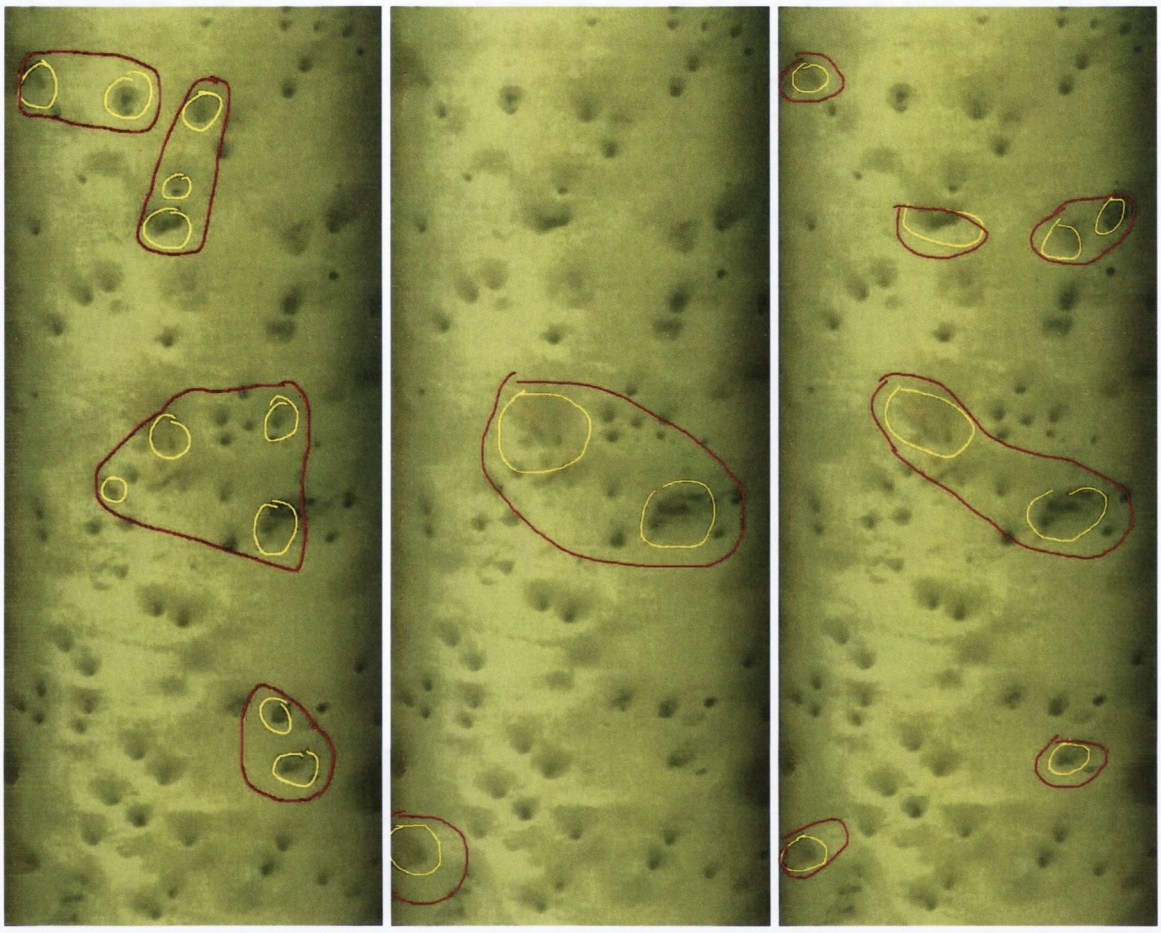


Figure B.2: Manual selections from test mosaic-8, of Nephrops burrows (yellow) and their corresponding complexes (red), obtained from scientists: (Left) Adrian Weetman from the Marine Laboratory in Scotland, (Middle) Alessandro Ligas from the Biosciences Institute in Belfast, (Right) Jennifer Doyle from the Marine Institute in Galway

Adrian, Alessandro and Jennifer are 3.5, 4.6 and 8.9.

2. Generally more complexes are identified in the mosaics than with video. This possibly means the improved visibility and field of view offered in the mosaics improves the inspection.

3. A Large discrepancy (average of 73.1%) among the video and mosaics counts is observed in test sequence 8. The reason for this anomaly is because of the large burrow density in this mosaic, as seen in Figure B.2. Under these circumstances scientists are very skeptical to select burrows as Nephrops are usually territorial creatures in nature and prefer to make their complexes in isolation to other creatures [52].

4. Upon examining the selections made in the mosaics, some instances are observed where the burrows selected do not fit the four characteristic Nephrop features (listed in Chapter-2) that scientists usually use as guidelines. Some of these ambiguities include:

   (a) Non-crescent shaped burrows were selected.

   (b) Burrows that did not have any sediment ejecta were selected.

   (c) Neighboring burrows that are not pointing towards a common center were grouped as a complex.

   (d) Burrows with Nephrops were not selected.

   (e) Burrows with very little contrast and dark entrance regions are selected.

5. The distance between some of the burrows grouped into a complex is very large, spanning over the frame width. Most of these cases occur when the diameter of one of the burrows is large. This highlights the advantage of using mosaics, as it is almost impossible to spot these relationships from the original video.

6. The direct path connecting burrows in complexes sometimes contain other burrows that are not associated with the complex, as seen in the third cluster from the top of the leftmost mosaic in Figure B.1.

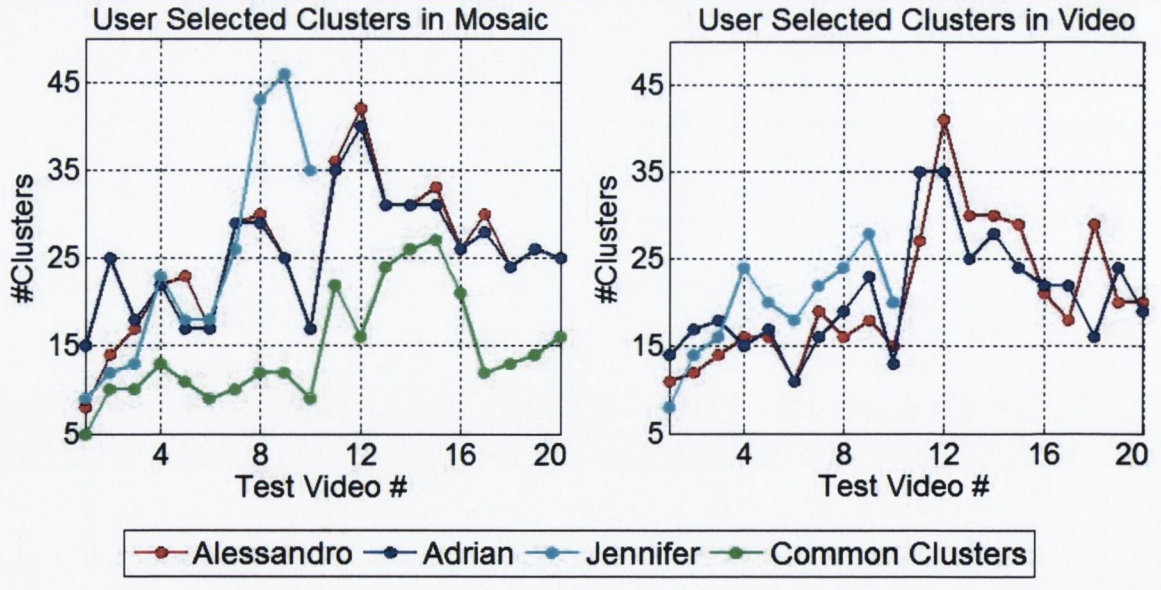


Figure B.3: Collected data form scientists (Left) Adrian Weetman from the Marine Laboratory in Scotland, (Middle) Alessandro Ligas from the Biosciences Institute in Belfast, (Right) Jennifer Doyle from the Marine Institute in Galway

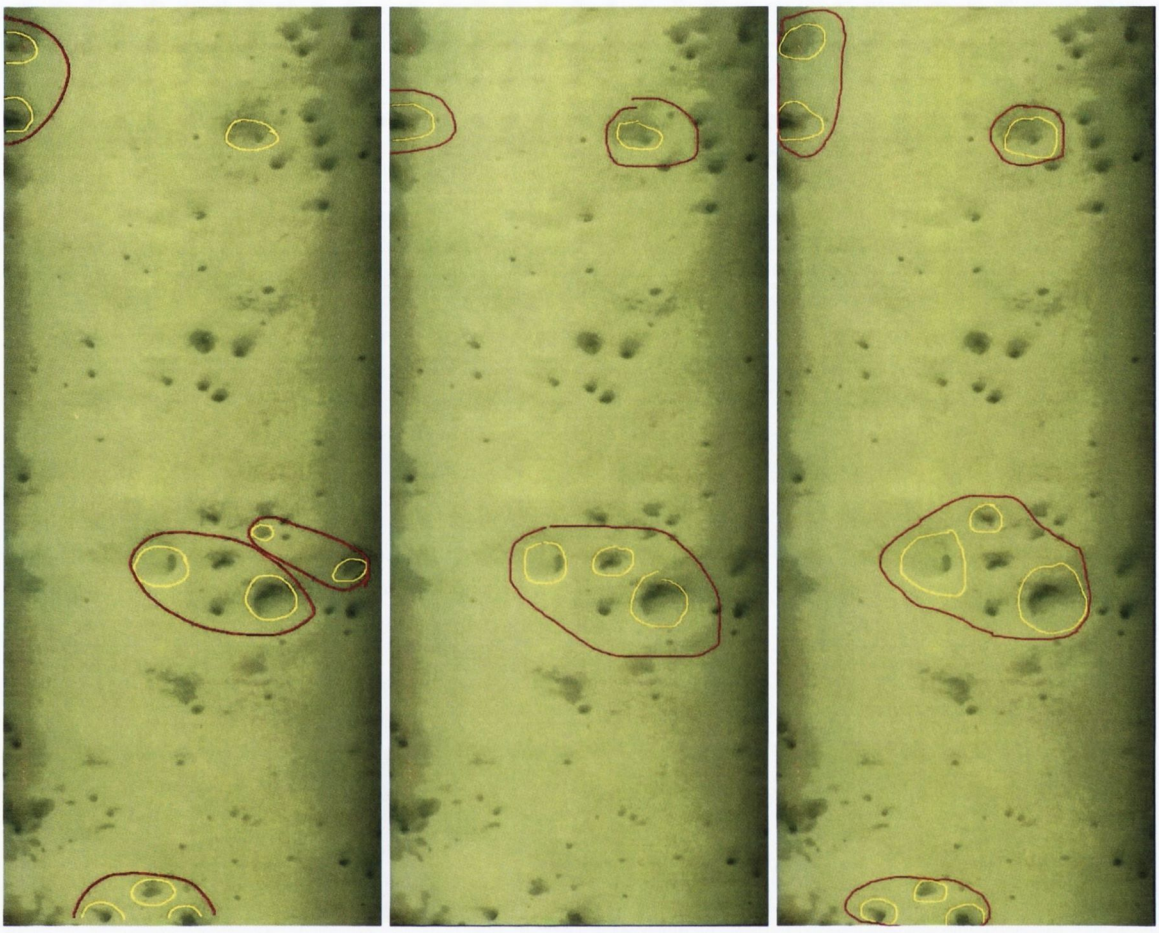


Figure B.4: Manual selections from test mosaic-3, of Nephrops burrows (yellow) and their corresponding complexes (red), obtained from scientists: (Left) Adrian Weetman from the Marine Laboratory in Scotland, (Middle) Alessandro Ligas from the Biosciences Institute in Belfast, (Right) Jennifer Doyle from the Marine Institute in Galway

B.2 Group Analysis

Plots of the complex counts obtained from each scientist using videos and mosaics are given in Figure B.3, and sample labels from test mosaic-3 are shown in Figure B.4. Analyzing the results five key observations are made.

1. Generally more complexes are identified in the mosaics than with video.
2. Video counts among the different users are generally consistent.
3. The counts obtained from the mosaics are also generally consistent. An exception to this generalization is seen in the counts from Jennifer in mosaics 8-10, which are largely different from the other scientists.

4. Although mosaic counts are consistent, visual inspection reveals that on average only 50% of the clusters selected by a given user are common to the rest of experts. These inconsistencies highlight the procedure is difficult and error prone.

5. There is larger percentage of common clusters in mosaics 1-3, and 13-16. The reason for this consistency is because in these mosaics the burrow density are not large, and burrows are well spaced out. This condition is ideal for Nephrops [52], hence scientists select burrows with more confidence.

B.3 Summary

This section presented an analysis of Nephrops complexes selected by three marine scientists using video and mosaics. The inconsistencies of the counts obtained highlight the selection of Nephrop complexes is difficult and error prone. One of the key observations made is that the counts obtained from the mosaics are generally greater than the corresponding video, this possibly implies the improved visibility and field of view does help scientists. Overall, the scientists agreed it was much easier to spot relationships among the Nephrops burrows with the mosaics as opposed to the original videos.

C

Supplementary Analysis for Chapter 5

C.1 Principal Component Analysis (PCA) for KNN

PCA [18] is performed for the KNN using the 25 component feature vector: $\{c_o, a_i, s_a, r_h, s_m, s_v, c_c, x_l, x_h, y_h, y_v, a_d, c_s, b_d, b_e, b_c, b_r, m_1, m_2, m_3, m_4, m_5, m_6, m_7\}$, of all objects in the training set. Figure C.1 (a) shows a plot of the eigenvalues in descending order. With the corresponding classification error, recall and precision values obtained from using the top number of principal components, in Figures C.1 (b) and (c) respectively. Analysis of these plots show the performance of the system ranged from 12.9 – 4.0% in classification error, 68.2 – 91.2% in recall and 67.1 – 88.1% in precision, from using the top and all 25 components respectively. With approximately 81% of the improvement within these ranges occuring with the use of only the top two components. Another interesting point observed is that the lowest classification error of 3.9%, along with high recall and precision values of 91.2% and 88.1%, is obtained with the use of only the top 15 components. Using this optimal number of components, the performance of the classifier is further examined with various neighbourhood values, k, ranging from 1-100, the results of which are given in Figures C.1 (d) and (e). Examination of these plots show optimal classification error, recall and precision values of 3.6%, 93.2% and 88.5% is achieved with a neighbourhood value of $k = 12$. From these observations it can be concluded that there is redundancy in the entire feature set, and its 25 component complexity can be reduced to 15 components, without any major loss in performance.

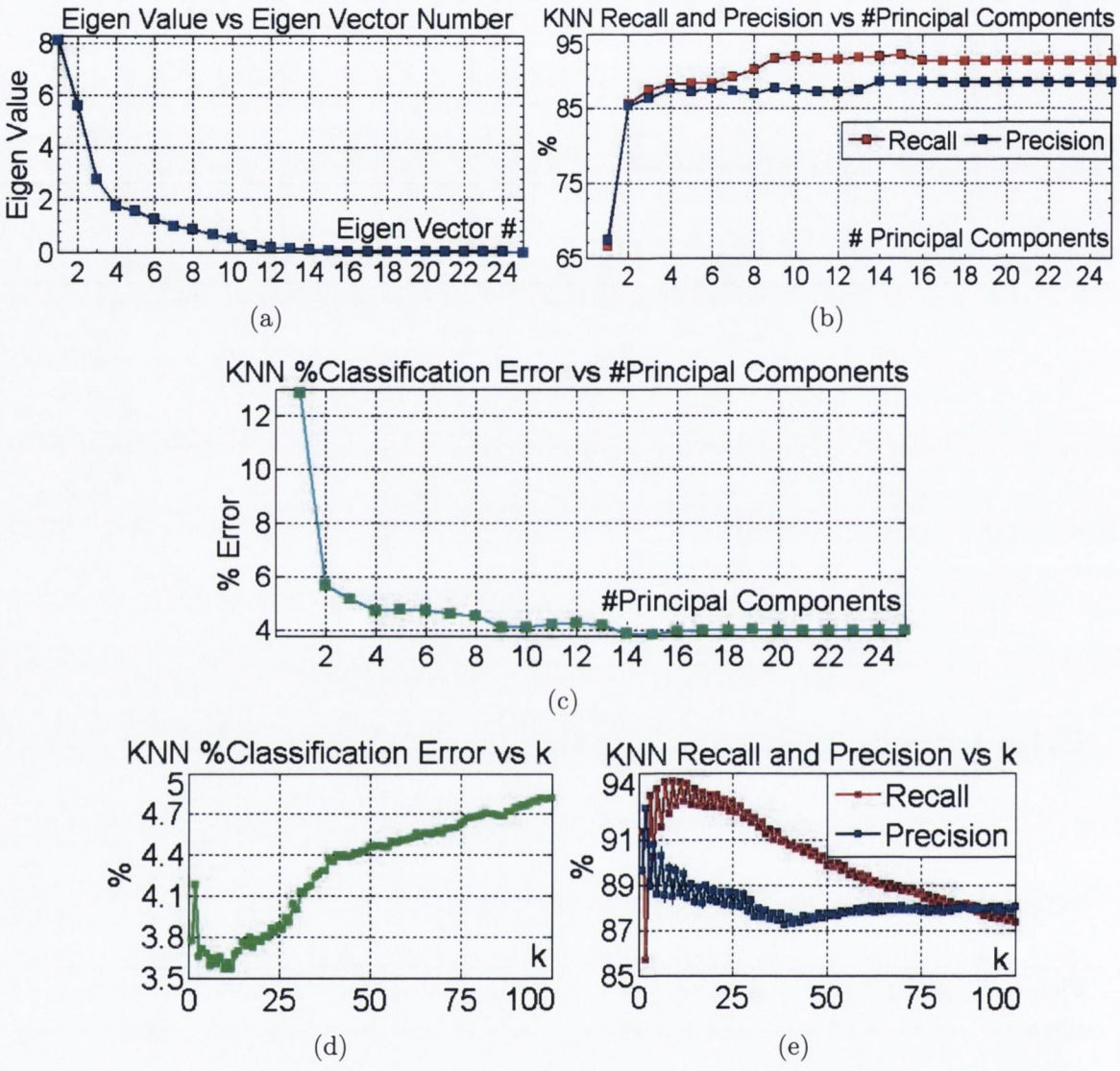


Figure C.1: (a) Eigenvalues of the Eigenvectors obtained from PCA, with (b) classification error, (c) recall and precision results from KNN, and (d) the classification error, with (e) recall and precision values from using the top 15 principal components with various k values.

C.2 PCA analysis for SVM

PCA is performed for the SVM with the same 25 component feature vector used in the KNN analysis, extracted from all objects in the training set. As the same training set is used, identical eigenvalues to the KNN in Figure C.1 (a), are obtained. The classification error, recall and precision values obtained from this system, from using the top number of principal components, are shown in Figure C.2. Comparing these results to the KNN in Figure C.1, the performance of the SVM with less than eight components is observed to be much worse than the KNN,

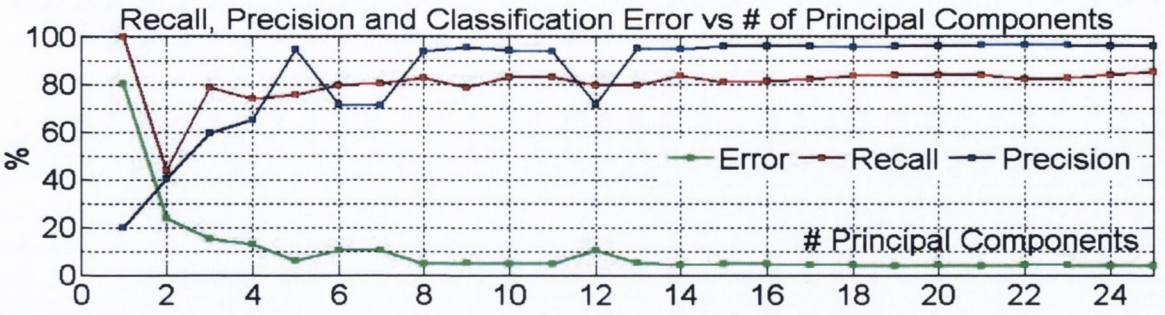


Figure C.2: Percentage classification error (green), recall (red) and precision (blue) results obtained from the SVM classifier vs the number of principal components used respectively.

giving an average classification error of 23.1%, compared to the KNN average of 6.05%. But with more eight components, the system stablizes to give an average classification error of 5.1%, which is only marginally worse than the KNN average of 3.9%. Another interesting observation is that beyond the use of eight components the recall values of the SVM are generally higher than the KNN, but the precision values are generally lower. Overall, the worse result of 80.4% in classification error, 19.6% in recall and 99.8% in precision, is obtained when only the top principal component is used. While the best result of 3.7% in classification error, 95.5% in recall and 84.8% in precision, is achieved when all 25 components are used. With the use of just eight components however these optimum values in classification error, recall and precision are only degraded by 1.4%, 3.2%, 2.9%. This observation shows that the 25 component complexity of the system can be reduced to eight, without any major loss in performance.

C.3 Results from KNN Exhaustive Feature Selection

Existing Feature Set					New Feature Set				Entire Feature Set			
Rank	Comb	R	P	E	Comb	R	P	E	Comb	R	P	E
1	r_h, c_m	88.9	90.6	4.02	a_d, b_c	93.3	91.3	3.05	a_d, b_c	93.3	91.4	3.05
2	s_a, r_h, c_m	88.1	91.1	4.04	a_d, b_M	94.8	89.7	3.17	a_i, s_a, r_h, c_c c_m, a_d, c_s, b_c	95.5	89.6	3.08
3	a_i, s_a, r_h, s_r c_m	85.8	92.8	4.11	a_d, b_d, b_c	94.6	89.8	3.17	s_a, c_m, a_d	93.3	91.2	3.08
4	c_o, s_a, r_h, c_m	87.4	91.3	4.12	a_d, b_d, b_c, b_M	95.2	89.0	3.26	c_o, a_i, s_a, r_h c_m, a_d, c_s, b_c	95.3	89.7	3.08
5	a_i, s_a, c_m	89.1	89.8	4.14	a_d, c_s, b_c	93.9	90.0	3.27	a_i, s_a, r_h, c_m a_d, c_s, b_d, b_c	95.4	89.7	3.08
6	s_a, c_m	91.2	88.1	4.15	a_d, b_d, b_M	95.2	88.9	3.28	s_a, r_h, c_m, a_d b_d	93.4	91.3	3.08
7	s_a, r_h, s_r, c_m	86.4	92.0	4.15	a_d, b_c, b_M	94.5	89.4	3.29	s_a, r_h, c_m, a_d c_s, b_d, b_c	95.4	89.6	3.08
8	s_a, s_r, c_m	89.4	89.4	4.15	a_d, c_s, b_d, b_c	94.3	89.5	3.30	c_o, a_i, s_a, r_h c_m, a_d, c_s, b_M	95.4	89.6	3.09
9	c_o, s_a, c_m	90.8	88.3	4.19	a_d, c_s, b_e, b_c b_r	94.8	89.1	3.30	c_o, a_i, s_a, r_h c_m, a_d, b_d	92.3	92.0	3.09
10	c_o, r_h, c_m	87.8	90.6	4.19	a_d, c_s, b_d, b_e b_c, b_r, b_M	95.3	88.6	3.32	a_i, s_a, r_h, c_m a_d, c_s, b_c	95.0	90.0	3.09
5^{th}L	r_h	44.7	24.1	38.6	b_e, b_c	96.3	83.3	4.54	r_h	44.7	24.1	38.6
4^{th}L	s_a, r_h	42.3	23.3	38.7	b_c	97.2	82.7	4.56	s_a, r_h	42.3	23.3	38.7
3^{rd}L	s_a, c_c	72.7	19.5	64.4	b_e, b_c, b_M	98.3	82.0	4.57	s_a, c_c	72.7	19.5	64.4
2^{nd}L	c_c	95.1	21.9	67.9	b_M	98.4	82.0	4.58	c_c	95.1	21.9	67.9
L	s_a	97.8	19.8	78.3	b_e, b_M	98.4	82.0	4.58	s_a	97.8	19.8	78.3

Table C.1: Recall (R), Precision (P), and Classification error (E) from KNN for top 10 and last (L) 5 feature combinations (ranked in lowest E) from existing, new, and entire feature sets.

C.4 Results from SVM Exhaustive Feature Selection

Existing Features					New Features				All Features			
Rank	Comb	R	P	E	Comb	R	P	E	Comb	R	P	E
1	r_h, s_r, c_m	92.5	85.2	4.62	c_s, b_c, b_r, b_M	95.9	87.3	3.55	r_h, a_d, b_M	95.8	88.6	3.23
2	a_i, s_a, r_h, s_r c_c, c_m	90.6	85.3	4.90	a_d	96.2	86.8	3.64	s_a, c_c, a_d, b_M	95.6	88.5	3.29
3	c_o, a_i, s_a, r_h s_r, c_c, c_m	90.4	85.4	4.93	a_d, c_s	96.2	86.8	3.64	c_c, a_d, b_M	96.1	88.0	3.34
4	a_i, s_a, r_h, s_r c_m	92.5	84.0	4.93	b_d, b_c, b_r, b_M	91.4	90.1	3.69	r_h, c_c, a_d, b_M	96.4	87.6	3.38
5	a_i, r_h, s_r c_c, c_m	91.2	84.4	4.95	a_d, b_M	97.3	85.7	3.73	r_h, c_c, a_d, b_d	95.9	88.0	3.39
6	c_o, a_i, s_a, r_h s_r, c_m	92.3	84.0	4.97	a_d, b_d, b_r, b_M	96.4	85.8	3.84	c_c, a_d, b_d, b_M	96.9	87.0	3.47
7	a_i, r_h, s_r, c_m	93.4	83.3	4.99	a_d, b_d, b_e b_c, b_r, b_M	97.9	84.9	3.84	s_a, r_h, c_c, a_d b_M	96.3	87.4	3.47
8	s_a, r_h, s_r, c_m	94.7	82.5	4.99	a_d, b_e, b_c, b_r	97.9	84.7	3.90	s_a, r_h, a_d, b_M	96.2	87.4	3.48
9	c_o, a_i, r_h s_r, c_c, c_m	91.5	84.4	5.00	a_d, b_c, b_r b_M	96.9	85.3	3.91	s_a, c_c, a_d, b_d b_M	96.6	87.1	3.49
10	c_o, s_a, r_h, s_r c_m	95.0	82.2	5.01	a_d, c_s, b_c b_r, b_M	96.9	85.3	3.91	c_o, s_a, c_m, a_d b_e, b_c, b_r	95.6	87.6	3.52
123	r_h, c_c	70.9	35.9	30.6	c_s, b_e	97.6	82.0	4.68	c_c, b_M	67.4	32.8	33.5
124	c_o, a_i	83.3	35.9	32.6	b_c, b_r	95.4	83.0	4.75	a_i, b_M	81.1	33.8	35.0
125	c_c	67.0	32.8	33.5	b_e, b_c, b_r, b_M	84.1	90.8	4.81	a_i	80.7	33.3	35.5
126	a_i	80.7	33.4	35.5	b_r	94.8	82.8	4.90	r_h, b_M	85.4	26.0	50.6
127	r_h	84.0	24.2	54.9	c_s, b_r	94.8	82.8	4.9	r_h	84.0	24.2	54.9

Table C.2: Recall (R), Precision (P), and Classification error (E) from SVM for top 10 and last (L) 5 feature combinations (ranked in lowest E) from existing, new, and entire feature sets.

D

Detecting Nephrops Using Mosaics

Although the population estimate of Nephrops is solely based on the quantity of their respective burrow complexes, their presence still provides useful information. Firstly, as they are highly territorial in nature [52], their occurrence in particular burrows strongly indicate that is their respective place of dwelling. This is one of the key features scientists use when identifying their complex systems. The quantity and size of these creatures also provide scientists with useful information with regards to the maturity of the respective population [52]. However, to recognize these creatures automatically is not easy because of: i) the large diversity in their shapes and sizes, and ii) the visibility challenges in these videos.

Previous work in automated Nephrops identification was presented by Lau et al [46], which is briefly discussed at the end of Chapter 2. Although good results are obtained from the test sets they used, their algorithm has four drawbacks:

1. The object detection process which uses edges produces incomplete segmentations, and may not be effective on the blurry images used in this work.

2. By processing in the gray scale space only, many other objects on the sea floor are detected, which can decrease the overall efficiency of the system.

3. Using a strict set of rules, via their decision tree classification framework might restrict the system from generalizing well with other data sets.

4. To verify the automated results, scientists still have to inspect the video.

To improve on these drawbacks, four key contributions are introduced in this work:

1. Mosaics are used for detecting and summarizing the automated results, which reduces the time spent tediously inspecting thousands of frames to the scanning of a single image.

2. Segmentation is for object detection, which obtain more complete object regions than using edges. The segmentation technique developed targets the bright pink-orange colour characteristics of the creature.

3. A new feature set that is motivated by a current scientific description of Nephrop burrows, which would be easy for marine scientists to relate to. Some of these features include the diameter of the creature, which provides further statistical information relating to the size and population age of the species [52].

4. Supervised learning schemes are used for classification, which can improve how the system generalizes on different data sets.

The design of the proposed Nephrop recognition system is accomplished in five stages involving: i) Data Collection, ii) Object Detection and Grouping, iii) Feature Choice and Extraction, iv) Classification Model Selection, v) Optimization and Training. Details on each of these stages is now presented, followed by a comparison with the state of the art technique introduced by Lau et al. [46].

D.1 Data Collection

The training and testing data for these experiments are obtained from twenty mosaics used in the previous section. Using these mosaics, ground truth data is created in two steps. First an expert, Jennifer Doyle [1], manually selects the Nephrop regions, which are then fully segmented using the proposed object detection algorithm. The other objects that are detected from this algorithm are labeled as non-Nephrop items. During the selection process the original frames are visually inspected to ensure the integrity of each mosaic, which proved to be accurate in all cases. Table D.1 illustrates the number of Nephrops labeled in each test mosaic.

[1] jennifer.doyle@marine.ie from the Marine Institute, Galway, Ireland

Mosaic	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Nephrops	3	0	1	1	0	1	0	4	3	1	5	1	4	1	3	1	2	2	1	2
Non-Nephrop	57	87	98	122	78	88	78	67	99	79	86	90	79	95	89	123	67	83	94	87

Table D.1: Ground truth Nephrop and non-Nephrop objects in each test mosaic.

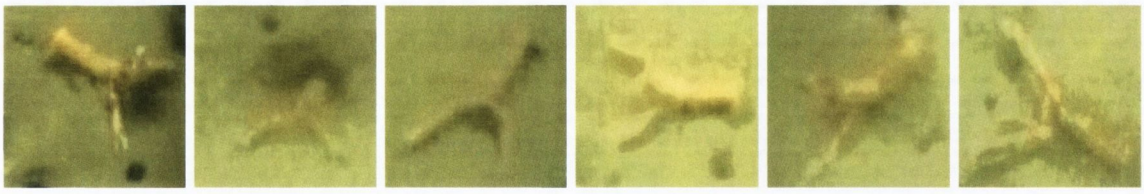


Figure D.1: Examples of Nephrops from actual underwater survey videos.

D.1.1 The Nature of Nephrops

The ground truth data is now analyzed for clues that would be useful for choosing the most appropriate features and classifiers for this application. The main clue observed, and confirmed from discussions with scientists at the marine institute, is that Nephrops appear as bright pink-orange regions in the video, as shown in Figure D.1. In addition to their characteristic appearance, scientists are also interested in the length of these creatures.

D.2 Nephrops Object Detection and Grouping

The first stage of the Nephrops recognition system is to detect candidate Nephrop regions in the generated mosaic. A segmentation approach is used to accomplish this task because these images are usually very blurry, and detecting parts of the objects with techniques such as edges [46], might not be effective in some cases. Additionally, the scientifically important features such as the diameter of the creature is most effectively extracted from the entire region. The segmentation approach is performed by targeting the characteristic bright contrasting appearance and pink-orange colour of Nephrops for their detection. To cope with the uneven lighting in these images, the bright contrasting characteristic is targeted in the difference of Gaussians image, where the influence of absolute brightness has minimal effect. The pink-orange colour characteristics are targeted in the Hue and Saturation colour channels. This overall procedure has three main steps involving: i) generating a bright region map, ii) segmentation, and iii) labeling, which are now explained.

D.2.1 Bright Region Map Generation

The first stage in this detection algorithm is to locate bright contrasting regions in the mosaic, I. This is achieved by generating a bright region map as: $I_b = I * G_2 - I * G_1$. Where G_1 and G_2 are two dimensional Gaussian functions with 71 and 5 taps, and corresponding variances of 30 and 2 respectively. Because of the large variance of G_1, a homogeneous sandy background image is created, which when subtracted from a lightly blurred version $(I * G_2)$, all of the local bright (candidate lobster) contrasting regions are highlighted as local maxima regions. To obtain larger maxima values and hence improve detection, gamma correction is performed on the original image, $I = I^{\gamma}$, where $\gamma = 1.5$ is used, prior to the generation of I_b. Figure D.2

illustrates the generation of this dark region map.

D.2.2 Segmentation

The candidate Nephrop regions are now obtained by performing segmentation on the i) bright region map, $I_b(\mathbf{x})$, and the ii) Hue, $I_h(\mathbf{x})$, and iii) Saturation, $I_s(\mathbf{x})$, colour channels. A two layer segmentation map $L(\mathbf{x})$ is estimated in which the labels are defined as follows.

$$L(\mathbf{x}) = \begin{cases} 1 & \text{Nephrop regions} \\ 0 & \text{Homogenous sandy background regions} \end{cases}$$

Following a Bayesian framework, the MAP estimate for $L(\mathbf{x})$ is generated by maximizing $p_o(L(\mathbf{x}) = \alpha | I_b(\mathbf{x}), I_h(\mathbf{x}), I_s(\mathbf{x}), \neg L(\mathbf{x}))$ where $\neg L(\mathbf{x})$ is the respective 3×3 neighborhood pixel labels of image position $\mathbf{x}$. Factorizing the posterior using Bayes Law [18], and dropping the notation $\mathbf{x}$ for clarity, gives:

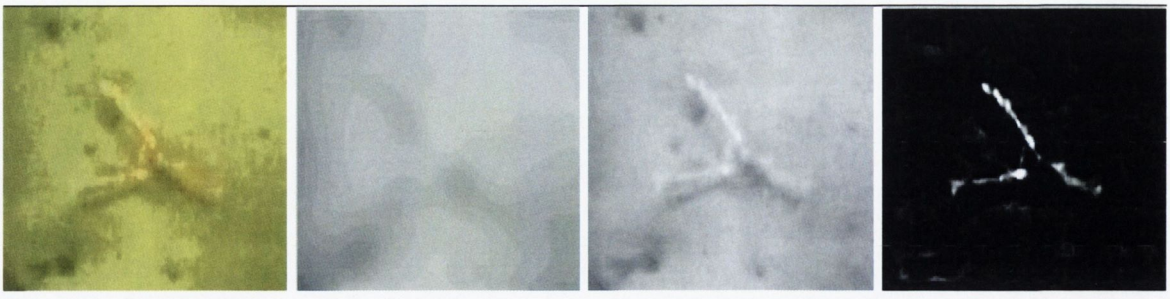
$$p_o(L = \alpha | I_b, I_h, I_s, \neg L) \propto p_b(I_b, | L = \alpha) p_h(I_h, | L = \alpha) p_s(I_s, | L = \alpha) p_r(L = \alpha | \neg L) \tag{D.1}$$

where $\{p_b, p_h, p_s\}$ and p_r are the likelihood and prior terms. The likelihoods are assumed to be Gaussian as follows.

$$p_b(I_b | L = \alpha) \propto \exp - \Big[\frac{(I_b - \hat{I}_{b,\alpha})^2}{2\sigma^2_{b,\alpha}} \Big]$$

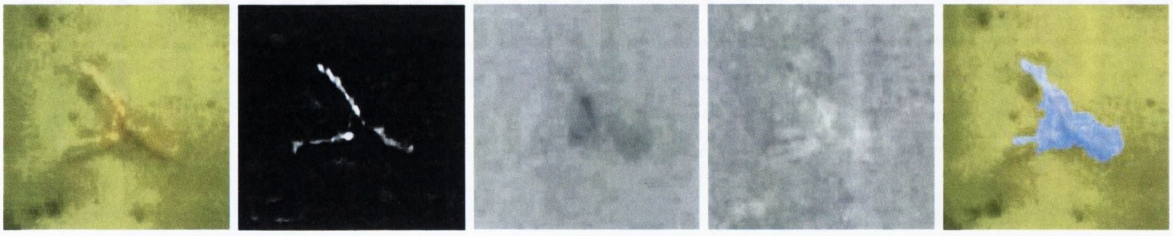
$$p_h(I_b | L = \alpha) \propto \exp - \Big[\frac{(I_h - \hat{I}_{h,\alpha})^2}{2\sigma^2_{h,\alpha}} \Big]$$

$$p_b(I_s | L = \alpha) \propto \exp - \Big[\frac{(I_s - \hat{I}_{s,\alpha})^2}{2\sigma^2_{s,\alpha}} \Big]$$



(a) (b) (c) (d)

Figure D.2: Generating the bright region map of the original image (a), by subtracting a heavily blurred gray scale version (b), from a lightly blurred version (c). This highlights the local bright (candidate Nephrops) regions as local maxima regions, as seen in the bright map in (d).



(a) (b) (c) (d) (e)

Figure D.3: Segmenting Nephrop in original image (a), by combining the (b) bright region map with the (c) Hue, and (d) Saturation channels, using the proposed segmentation algorithm. As seen most of the object region is segmented (e).

where $\alpha = \{0, 1\}$, and $\{\hat{I}_{b,0}, \hat{I}_{h,0}, \hat{I}_{s,0}\}$ and $\{\hat{I}_{b,1}, \hat{I}_{h,1}, \hat{I}_{s,1}\}$ are the characteristic mean brightness, hue and saturation values of background and Nephrops regions respectively, and $\{\sigma^2_{b,\alpha}, \sigma^2_{h,\alpha}, \sigma^2_{s,\alpha}\}$ are their corresponding variances. To enforce spatial smoothness within these segmentations, a Gibbs energy function [23], with a 3×3 pixel neighborhood, is used for the prior, $p_r(.)$, as:

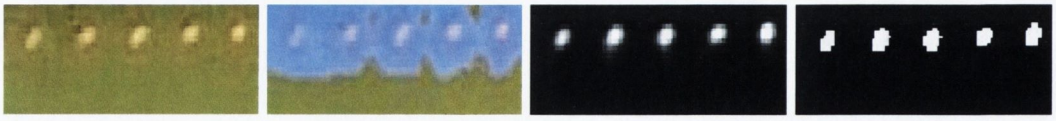
$$p_r(L(\mathbf{x}) = \alpha | \neg L) \propto \exp - \Big[\Lambda \sum_{k=0}^{7} \lambda_k |\alpha - L(\mathbf{x}_k)| \Big] \tag{D.2}$$

where $\lambda_k = 1/||\mathbf{x} - \mathbf{x_k}||$, is a weight inversely proportional to the distance between the current site $\mathbf{x}$ and the respective neighbor $\mathbf{x_k}$ in a 3×3 neighborhood, and Λ is a global weighting factor, set as $\Lambda = 1$ in these experiments.

Good initial estimates for the various parameters are obtained after analyzing the mean brightness, and colour characteristics of several of the ground truth Nephrop regions. The respective values of $\{\hat{I}_{b,0}, \hat{I}_{h,0}, \hat{I}_{s,0}\}$ and $\{\hat{I}_{b,1}, \hat{I}_{h,1}, \hat{I}_{s,1}\}$ were set to $\{100, 0.16, 0.65\}$ and $\{0, 0.17, 0.60\}$ respectively. While the corresponding variances are set as $\{\sigma^2_{b,0}, \sigma^2_{h,0}, \sigma^2_{s,0}\} = \{\sigma^2_{b,1}, \sigma^2_{h,1}, \sigma^2_{s,1}\}$ $= \{1/|I_{b,0} - I_{b,1}|^2, 1/|I_{h,0} - I_{h,1}|^2, 1/|I_{s,0} - I_{s,1}|^2\}$. Using these settings, minimization of p_o is then performed using the Iterated Conditional Modes [6] scheme, where a checkerboard scan is utilized until there are no further changes in labels or a maximum of 10 iterations is completed. Sample results obtained using this segmentation procedure are shown in Figure D.3.

D.2.3 Labeling

Locally connected Nephrop regions, $L(\mathbf{x}) = 1$, are now labeled with unique identification numbers. The Connected Component Analysis technique by Sammet et al. [68], with a 3×3 neighborhood, is used to perform these labellings.



(a) (b) (c) (d)

Figure D.4: Extraction of laser dots. (a) Original, (b) segmentation obtained from proposed system, (c) bright region map, and (d) detected laser regions.

D.3 Feature Choice and Extraction

In practice, although only bright and pink-orange regions, characteristic of Nephrops are targeted, a large percentage of the objects detected are not Nephrops. Most of these other objects are due to the calibration laser dots that were created in the mosaicking process, which are also bright and orange-pink in appearance, as seen in Figure D.4. To eliminate these false alarms, four features are examined. From inspecting the ground truth data it is noticed that some Nephrops regions have similar sizes and shapes. To use this knowledge for their identification two size features, and one shape feature are used. These features are the: i) dark entrance area (a_d), ii) burrow diameter (b_d), and iii) eccentricity (b_e), which are defined in chapter 5. To eliminate the false alarms due to the laser regions, a fourth feature is created, Laser Dots (l_d), which is defined as follows.

Laser Dots (d). To distinguish the false alarm regions containing lasers dots from the Nephrops regions, each region is searched for the presence of these round intense dots, using two steps. First k-means clustering using 3 clusters is performed on the bright region map, I_b, of the particular object using centroid locations $\{0, 50, 120\}$. Then, regions from the cluster with the largest intensity value that have eccentricity and burrow diameter features within $\{b_e = \pm 0.2, b_d = \pm 5 pixels\}$ of the characteristic values of $\{b_e = 0.4, b_d = 12 pixels\}$, are classified as laser dots. This feature is extracted as the quantity of these laser dots. Figure D.4 illustrates this procedure.

D.4 Classification Model Selection

The last stage of the recognition system is to classify the detected objects into burrow and non-burrow classes. To cater for the large diversity in burrow size and shape features, the use of two well established supervised learning classification schemes, a K-Nearest Neighbor (KNN), and a Support Vector Machine (SVM), are explored. The use of a non-parametric classifier (KNN), and one that uses linear discriminant functions (SVM) to perform classification, are explored because it is not known if the selected features would follow a particular model. The key advantage these two schemes offer, in comparison to the previously used Decision Tree scheme [46], is that they incorporate the use of training data into their classification process.

The use of this data not only allow these systems to identify a large variety of burrows, but also facilitates easy adoption to new data sets. This adoption is performed by simply retraining the system with a new training set.

D.5 Training and Optimization

These classification systems can get very complex depending on a number of factors such as the size of the feature space, and the quantity of training data etc. Apart from being computational expensive, the main drawback of overly complex systems is that they can classify the training data effectively, but may not perform well on other test sets. This situation is commonly referred to as overfitting [18]. To ensure this situation has not occurred, it is important to verify the classifier generalizes well with different test data. To perform this verification three items have to be selected: i) features, ii) training data, and iii) the various model parameters for each classifier.

D.5.1 Feature Selection

For feature selection, the four features described earlier are used, which are the area (a_d), diameter (b_d), eccentricity (b_e), and laser dots (d).

D.5.2 Training Data Selection

For training and testing data, as there are only 36 Nephrops in the entire data set collected, mosaics 1-10 are used for training and mosaics 11-20 are tested on. From the data in table D.1, this means the training set has 14 Nephrops and 853 other objects, and the testing set comprises of 22 Nephrops and 893 other objects. This mosaics are selected for training as from visual inspection, the Nephrops and other objects they contain have a large variety in shape in size characteristics.

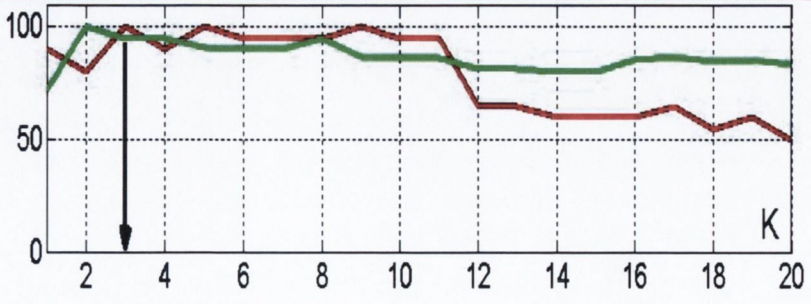


Figure D.5: Recall (red) and Precision (green) from KNN for different neighborhood, k values. Black arrow indicates highest recall and precision average

D.5.3 Parameter Selection

As mentioned in chapter 2, the SVM classifier from Matlab [54], is used to perform these experiments. Details of the respective parameter settings used in this system are also given in chapter 2. For the KNN, a optimum neighborhood value, k, is selected by analyzing the performance of the classifier on the test set with a range of k values from 1 to 20.

The recall and precision values obtained from these experiments are given in Figure D.5. Analysis of these results show the system behaves steadily from $k = 1$ to $k = 11$, with recall and precision averaging approximately 80%. But after $k = 11$ the performance of the system generally degrades with the precision averaging approximately 75%, and the recall 55%. This degradation can possibley be attributied to the small quantity of Nephrops in the training set compared to the number of other objects. Another interesting observation is with a value of $k = 1$ 75% precision and 80% recall is obtained. Using this value would significantly reduce system computations, as sorting the training data to obtain the nearest neighbours of the query object is no longer necessary. But, for these experiments, as speed is not the primary concern but accuracy, the value of $k = 3$ is selected as the optimal value, because it achieves the largest recall and precesion value average of 90%, and it lies in the stable region of the system.

D.6 Results

The KNN and SVM classification frameworks are now evaluated. This evaluation is performed with two experiments using mosaics 11-20 in Table D.1 and their corresponding video sequences. In the first experiment, the performance of the KNN is compared to the SVM using all ten test mosaics. Then to examine how these systems perform to a previous state of the art system, they are compared to the video-based technique purposed by Lau et al. [46] using i) video and ii) mosaics.

D.6.1 Comparison of Proposed method using KNN and SVM

The recall and precision results obtained from testing these two classification schemes on mosaics 11-20 are given in Figure D.6 and Table D.1. Analyzing these results show the KNN performs the best with average recall and precision values of 87.5%. The SVM maintains this level of recall, but its precision drops to 62.5%. This drop in precision is probably due to SVM not being able to establish an effective decision boundary from the limited amount of Nephrops in the training set. Another key observation made is the recall and precision values from the both systems are zero in test mosaic-4. This scenario occurred because there is only one Nephrop in this test case whose appearance is too faint to be detected by the proposed system. A sample image of this Nephrop is shown in Figure D.9.

From these high recall and precision values obtained, two conclusions are drawn. First, both systems generalize well across the majority of these test mosaics. Secondly, this is potentially

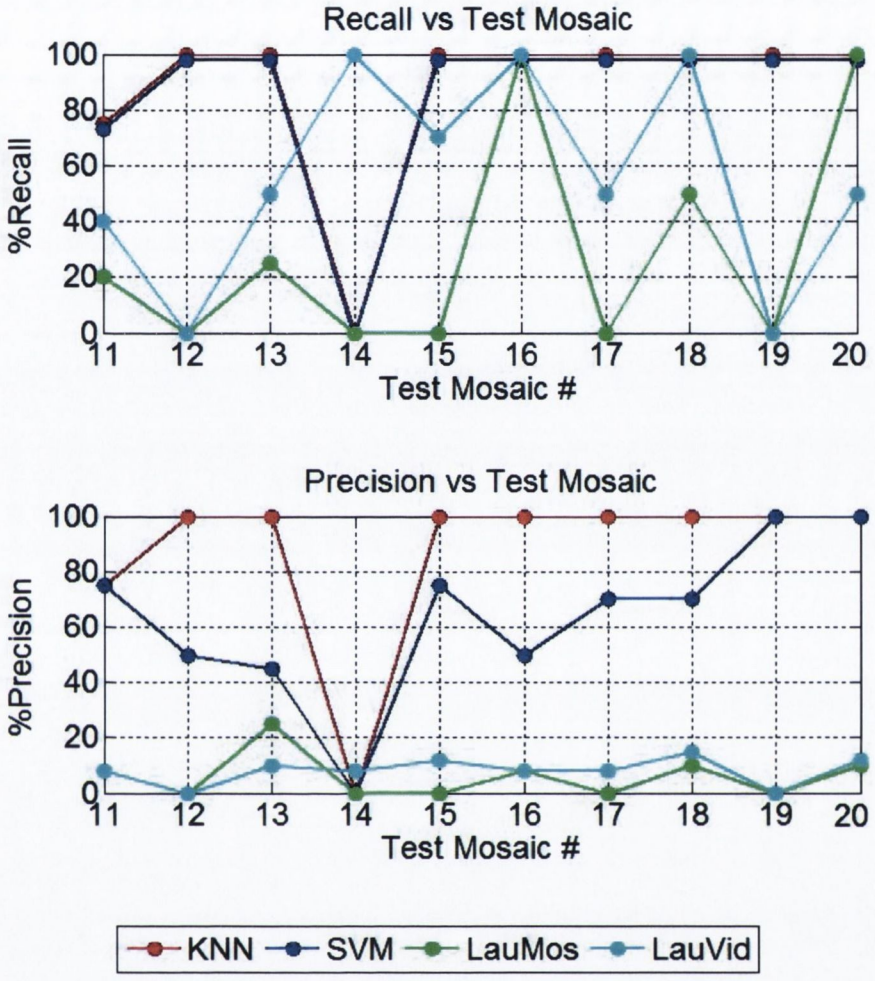


Figure D.6: (Top) Recall and (Bottom) Precision plots from the KNN (red), SVM (blue), Lau et al. [46] using video (cayan), and mosaics (green), from test mosaics 11-20.

a very good approach for this application as both systems are able to identify the majority of the 22 test Nephrops fairly accurately from mosaics containing thousands of different objects. Although the KNN performed superior in these test cases, it is difficult to conclude if it is better suited for this application than the SVM, as the number of Nephrops used for training and testing in these experiments are small.

D.6.2 Comparison with Previous Work

The performance of the KNN and SVM classifiers are now compared to the previous state of the art video-based technique proposed by Lau et al. [46] using: i) video and ii) mosaics. The video comparison is performed by manually cross referencing the classified objects obtained in each frame with the corresponding ground truth mosaic. The recall and precision results obtained

from each mosaic, for this comparison, are given in Figure D.6.

Analysis of these results show the proposed system achieves superior results to the previous method. In detail, compared to the previous system using video, the average recall and precision values improved by 31.5% and 79.4%, using the KNN from the proposed system. While in comparison to the previous method using mosaics, these corresponding values are improved by 58% and 81.2%. These results verify that it is not only possible to use mosaics to detect Nephrops, but improved results are achieved using this proposed technique compared to the previous method. The degradation in performance with the previous method using mosaics, compared to video, is due to the absence of the four-frame object consistency step in their algorithm. This step could not be performed with mosaics as they are only single images, and as a result, additional spurious objects due to noise are detected, hence explaining the degradation in the system precision. Examples of correctly detected, missed and false alarms obtained from these experiments are shown in Figures D.7, D.8, D.9.

D.7 Conclusion

In this chapter a novel technique for detecting Nephrops in marine surveillance videos is presented. This technique improves substantially on the previous state of the art method introduced by Lau et al. [46] using three key contributions. First, mosaics are used for performing object recognition, which improves visibility and reduces the tedious video inspection process currently performed by scientists to the browsing of a single image. Secondly, the use of classical segmentation techniques for performing object detection does capture most of the burrow regions, in contrast with the previously used edge detection method where only edges of the burrows are captured. Lastly, to identify a large diversity of burrows, the use of supervised learning classification schemes (KNN and SVM) are explored. These schemes use training data which can always be updated to adopt to any situation, as opposed to the strict set of rules used in the previous decision tree classification framework.

The high recall and precision values obtained from the system (87.5% from the KNN) show it is possible to use mosaics to detect Nephrops in underwater surveillance videos. As a practical point of interest, when the scientists from the Marine Institute Galway are shown these results they agreed that this algorithm has the potential to assist with their current manual analysis procedure.

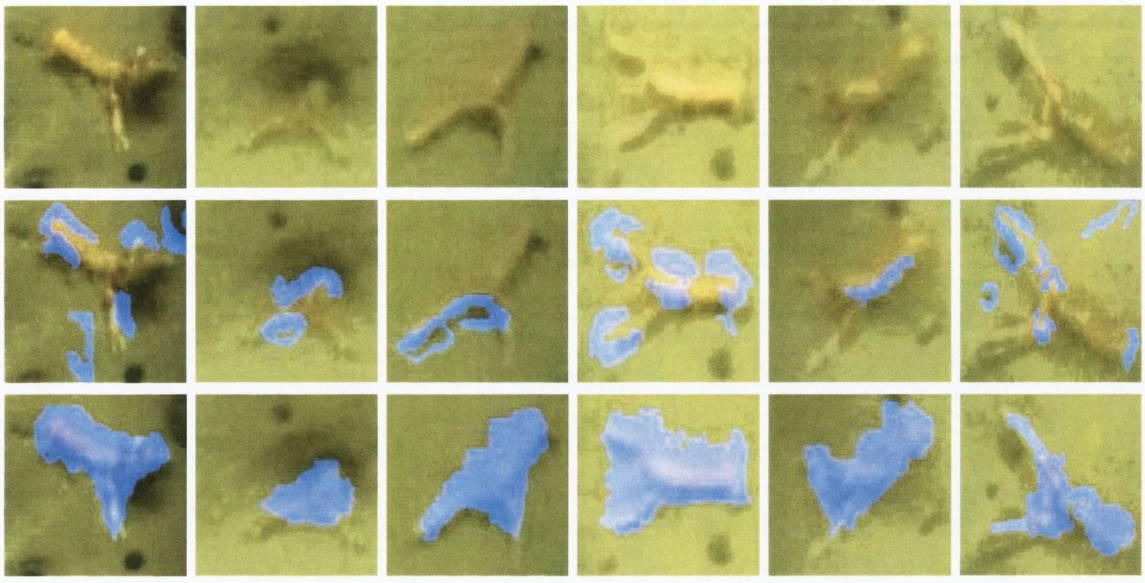


Figure D.7: Examples of correctly detected Nephrops. The top row are the original images, the middle row are the segmentations obtained by Lau [46], and last row are the segmentations obtained using the proposed method. As seen the proposed method obtains most of the object regions.

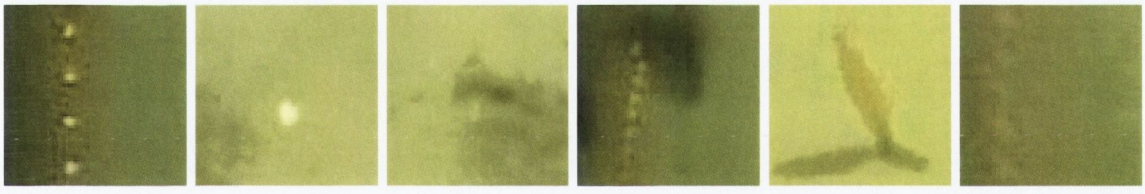


Figure D.8: The first three images are false alarms obtained using the previous method of Lau et al. [46]. The last three images are false alarms obtained from both the previous method of Lau et al. [46], and the proposed method.

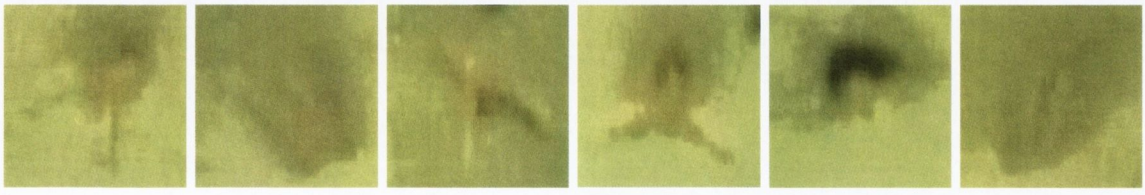


Figure D.9: Examples of missed Nephrops using previous method by Lau et al. [46] because of low contrast but detected with the proposed method (first four images), and samples missed by both methods (last two images).

Bibliography

[1] J. Ahlen, D. Sundgren, and E. Bengtsson. Application of underwater hyperspectral data for color correction purposes. *Journal on Pattern Recognition and Image Analysis*, 17(1):170–173, July 2007.

[2] A. Amanatiadis, V. Kaburlasos, A. Gasteratos, and E. Papadakis. Evaluation of shape descriptors for shape-based image retrieval. *IET Journal of Image Processing*, 5(1):493–499, April 2011.

[3] N. Asada, A. Amano, and M. Baba. Photometric calibration of zoom lens systems. In *Proceedings from the International Conference on Pattern Recognition*, volume 1, pages 186–190, Vienna, July 1996.

[4] S. Bazeille, I. Quidu, and L. Jaulin. Color-based underwater object recognition using water light attenuation. *Journal on Intelligent Service Robotics*, 5(2):109–118, July 2012.

[5] S. Bazeille, I. Quidu, L. Jaulin, and P. Malkasse. Automatic underwater image pre-processing. *Sea Tech Week Journal*, 2(1):8–13, July 2006.

[6] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–30, May 1986.

[7] M. Bond, E. Babcock, E. Pikitch, D. Abercrombie, N. Lamb, and D. Chapman. Reef sharks exhibit site-fidelity and higher relative abundance in marine reserves on the mesoamerican barrier reef. *Pone Journal*, 7(3):1–8, May 2012.

[8] D. Bongiorno, M. Bryson, and W. S. Dynamic spectral-based underwater colour correction. In *Proceedings of the IEEE International Conference on Oceans (OCEANS 2013)*, volume 1, pages 1–9, Bergen, May 2013.

[9] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, September 2007.

[10] P. Burt and E. Adelson. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics*, 2(4):217–236, September 1983.

[11] M. Caimi, M. Kocak, and L. Asper. Developments in laser-line scanned undersea surface mapping and image analysis systems for scientific applications. In *Proceedings of the IEEE International Conference on Oceans (OCEANS 1996)*, volume 1, pages 81–85, Biloxi, May 1996.

[12] N. Campbell, H. Dobby, and N. Bailey. Investigating and mitigating uncertainties in the assessment of scottish nephrops norvegicus populations using simulated underwater television data. *ICES Journal of Marine Science*, 66(1):646–655, May 2009.

[13] A. Castano, C. Anderson, R. Castano, T. Estlin, and M. Judd. *Intensity-based Rock Detection for Acquiring Onboard Rover Science.* Jet Propulsion Laboratory, National Aeronautics and Space Administration, CA, USA, 1st edition, pages:1-7, 2003. http://trs-new.jpl.nasa.gov/dspace/handle/2014/38819.

[14] P. Cloud and A. Gibor. The oxygen cycle. *Journal of Scientific American*, 5(10):110–123, September 1970.

[15] P. Dempster, M. Laird, and B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, September 1977.

[16] Z. Dongxiang, Z. Hong, and N. Ray. Texture based background subtraction. In *Proceedings of the IEEE International Conference on Information and Automation (ICIA 2008)*, volume 1, pages 601–605, Changsha, China, June 2008.

[17] A. Douglas and E. Kerr. Derivation of the cosine fourth law for falloff of illuminance across a camera image, May 2007. http://dougkerr.net/Pumpkin/articles/.

[18] R. Duda, P. Hart, and D. Stork. *Pattern Classification.* Wiley-Interscience, NY, 2nd edition, pages:182-184, 2001.

[19] H. Dunlop, R. Thompson, and D. Wettergreen. Multi-scale features for detection and segmentation of rocks in mars images. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2007).*, volume 1, pages 1–7, Minneapolis, MN, USA, June 2007.

[20] J. Fox, R. Castano, and C. Anderson. Onboard autonomous rock shape analysis for mars rovers. In *Proceedings of the IEEE Conference on Aerospace*, volume 5, pages 5–2052, Pasadena, CA, USA, June 2002.

[21] R. Garcia, T. Nicosevici, and X. Cufi. On the way to solve lighting problems in underwater imaging. In *Proceedings from the IEEE International Conference on Oceans (OCEANS 2002)*, volume 2, pages 1018–1024, Mississippi, USA, October 2002.

[22] A. Gebali, B. Albu, and M. Hoeberechts. Detection of salient events in large datasets of underwater video. In *Proceedings of the IEEE International Conference on Oceans (OCEANS 2012)*, volume 2, pages 1–10, Hampton Road, VA, USA, October 2012.

[23] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, May 1984.

[24] D. Goldman and J. Chen. Vignette and exposure calibration and compensation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2276–2288, December 2010.

[25] S. Golomb. Run-length encodings (corresp.). *IEEE Transactions on Information Theory*, 12(3):399–401, June 1966.

[26] N. Gracias and J. Santos-Victor. Automatic mosaic creation of the ocean floor. In *Proceedings from the IEEE International Conference on Oceans (OCEANS 1998)*, volume 1, pages 257–262, Biloxi, MS, USA, May 1998.

[27] D. Grossberg and K. Nayar. Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1272–1282, January 2004.

[28] M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(1):610–621, June 1973.

[29] R. Haralick and L. Shapiro. *Computer and Robot Vision.* Addison-Wesley, Boston, 1st edition, pages:620-660, 1992.

[30] C. Harris. Determination of ego-motion from matched points. In *Proceedings of the third Alvey Vision Conference*, volume 1, pages 189–192, Cambridge, UK, June 1987.

[31] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, Cambridge, UK, 2nd edition, 2003.

[32] W. Hassan, P. Birch, B. Mitra, N. Bangalore, R. Young, and C. Chatwin. Illumination invariant stationary object detection. *IET Journal of Computer Vision*, 7(1):1–8, May 2013.

[33] R. Haywood. Acquisition of a micro scale photographic survey using an autonomous submersible. In *Proceedings from the IEEE International Conference on Oceans (OCEANS 1986)*, volume 2, pages 338–343, New York, USA, September 1986.

[34] H. Holden and E. LeDrew. Hyperspectral discrimination of healthy versus stressed corals using in situ reflectance. *Journal of Coastal Research*, 17(4):850–858, July 2001.

[35] K. Hu. Pattern recognition by moment invariants. *Journal on Information Theory*, 49(1):179–187, January 1961.

[36] ICES. *Report of the Workshop and training course on Nephrops burrow identification.* ICES CM, H. C. Andersens Boulevard 44-46, DK-1553 Copenhagen V., Denmark, lrc:03 edition.

[37] IESNA. *The IESNA Lighting Handbook.* Illuminating Engineering Society of North America, New York, 9th edition, pages:46-47, 2000.

[38] K. Iqbal, M. Odetayo, A. James, R. A. Salam, and H. Talib. Enhancing the low quality images using unsupervised colour correction method. In *Proceedings from the IEEE International Conference on Systems Man and Cybernetics (SMC)*, volume 2, pages 1703–1709, Istanbul, Turkey, June 2010.

[39] M. Irani and P. Anandan. Video indexing based on mosaic representations. *Proceedings of the IEEE*, 86(5):905–921, May 1998.

[40] J. Jain and A. Jain. Displacement measurement and its application in interframe image coding. *IEEE Transactions on Communication*, 29(12):1799–1808, September 1981.

[41] K. Jain. *Fundamentals of Digital Image Processing.* Englewood Cliffs, NJ: Prentice Hall, 1st edition, 1989.

[42] R. Jain, R. Kasturi, and G. Schunck. *Machine Vision.* McGraw-Hill, NJ: Prentice Hall, international edition, 1995.

[43] S. Kang and R. Weiss. Can we calibrate a camera using an image of a flat textureless lambertian surface? In *Proceedings from the European Conference on Computer Vision*, volume 2, pages 640–653, Dublin, Ireland, July 2000.

[44] F. Kelly, A. Drygajlo, and N. Harte. Speaker verification in score-ageing-quality classification space. *Journal of Computer Speech and Language*, 27(5):1068–1084, November 2013.

[45] S. Kim and M. Pollefeys. Robust radiometric calibration and vignetting correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4):562–576, April 2008.

[46] Y. Lau, L. Correia, P. Fonseca, and A. Campos. Estimating norway lobster abundance from deep-water videos: an automatic approach. *IET Journal of Image Processing*, 6(1):22–30, January 2012.

[47] K. Lebart, E. Trucco, and M. Lane. Real-time automatic sea-floor change detection from video. In *Proceedings of the IEEE International Conference on Oceans (OCEANS 2000)*, volume 2, pages 1337–1343, Providence, USA, September 2000.

[48] T. Liu, J. Zhang, and F. Qi. A novel video keyframe extraction algorithm. In *Proceedings from the IEEE International Symposium on Circuits and Systems (ISCAS 2002)*, volume 4, pages 149–152, Arizona, USA, May 2002.

[49] X. Liu and J. Tang. Mass classification in mammograms using selected geometry and texture features, and a new svm-based feature selection method. *IEEE Systems Journal*, 1(99):1–11, November 2013.

[50] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, June 2004.

[51] L. Maired, D. Gwenael, A. Hodgsonc, and M. Frederic. Automated marine mammal detection from aerial imagery. In *Proceedings of the IEEE International Conference on Oceans (OCEANS 2013)*, volume 1, pages 1–4, San Diego, California, USA, September 2013.

[52] J. Marrs, A. Atkinson, J. Smith, and M. Hills. Calibration of the towed underwater tv technique for use in stock assessment of nephrops norvegicus. final report to the european commission contract 94/069. study project in support of the common fisheries policy (xiv/1810/c1/94). Technical report, January pages:1-155, 1996.

[53] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of the British Machine Vision Conference*, volume 1, pages 761–767, Cardiff, UK, September 2002.

[54] MATLAB. *(R2012a)*. The MathWorks Inc., Natick, Massachusetts.

[55] M. Mehrnejad, A. Albu, D. Capson, and M. Hoeberechts. Detection of stationary animals in deep-sea video. In *Proceedings of the IEEE International Conference on Oceans (OCEANS 2013)*, volume 1, pages 1–4, San Diego, California, USA, September 2013.

[56] C. Mobley. *Light and Water Radiative Transfer in Natural Waters*. Academic Press, New York, 1st edition, May 27 1994.

[57] J. Mundy. *Object recognition in the geometric era: a retrospective*. Springer-Verlag Berlin Heidelberg, New York, 1st edition, pages:3-29, 2006.

[58] R. Muralidharan and C. Chandrasckar. Object recognition using svm-knn based on geometric moment invariant. *International Journal of Computer Trends and Technology*, 12(3):215–220, June 2011.

[59] S. Omachi and M. Omachi. Fast template matching with polynomials. *IEEE Transactions on Image Processing*, 16(8):2139–2149, September 2007.

[60] W. Pan, L. Yangke, L. Ying, and W. Shuhang. Dehazing model based on multiple scattering. In *Proceedings from the 3rd International Congress on Image and Signal Processing (CISP)*, volume 1, pages 249–252, Yantai, China, July 2010.

[61] A. Panagopoulos, W. Chaohui, D. Samaras, and N. Paragios. Illumination estimation and cast shadow detection through a higher-order graphical model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*, volume 1, pages 673–680, Providence, RI, USA, June 2011.

[62] S. Peleg and J. Herman. Panoramic mosaics by manifold projection. In *Proceedings from the IEEE Conference on Computer Vision and Pattern Recognition (CVPR97)*, volume 2, pages 338–343, San Juan, Puerto Rico, October 1997.

[63] C. Peng, H. Mei, and G. Yihong. Extract highlights from baseball game video with hidden markov models. In *Proceedings of IEEE International Conference on Image Processing*, volume 1, pages 609–612, Rochester, New York, 2002.

[64] S. Pons, J. Piera, and J. Aguzzi. Video-image processing applied to the analysis of the behaviour of deep-water lobsters (nephrops norvegicus). In *Proceedings of the IEEE International Conference on Oceans (OCEANS 2010)*, volume 1, pages 1–4, Sidney, Australia, May 2010.

[65] L. Qing-Zhong, G. Xiao-Ling, and W. Wen-Jin. Fast video mosaic construction for observation of large static scenes. In *Proceedings from the WRI World Congress on Computer Science and Information Engineering*, volume 1, pages 349–354, Los Angeles, California, USA, April 2009.

[66] M. Raman and A. Himanshu. A comprehensive review of image enhancement techniques. *Journal of Computing*, 2(12):8–13, December 2010.

[67] N. Rea, R. Dahyot, and A. Kokaram. Modeling high level structure in sports with motion driven hmms. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '04)*, volume 3, pages 609–612, Montreal, Quebec, Canada, May 2004.

[68] H. Samet and M. Tamminen. Efficient component labeling of images of arbitrary dimension represented by linear bintrees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):579–586, May 1988.

[69] Y. Schechner and N. Karpel. Recovery of underwater visibility and structure by polarization analysis. *IEEE Journal of Oceanic Engineering*, 30(3):570–587, September 2005.

[70] A. Sedlazeck, K. Koser, and R. Koch. 3d reconstruction based on underwater video from rov kiel 6000 considering underwater imaging conditions. In *Proceedings of the IEEE International Conference on Oceans (OCEANS 2009)*, volume 1, pages 1–10, Bremen, May 2009.

[71] SFIA. *Seafood Industry Value Chain Analysis of Cod and Haddock and Nephrops.* KPMG AS, Centre for Aquaculture and Fisheries, Fjordgaten 68, 7010 Trondheim, Finland, 2nd edition, pages:400-592, March, 2004.

[72] C. Shifeng, C. Liangliang, L. Jianzhuang, and T. Xiaoou. Iterative map and ml estimations for image segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, volume 1, pages 1–6, Minneapolis, USA, June 2007.

[73] H. Shum and R. Szeliski. Construction of panoramic mosaics with global and local alignment. *International Journal of Computer Vision*, 2(1):101–130, July 2000.

[74] M. Smith and T. Kanade. Video skimming and characterisation through the combination of image and language understanding techniques. In *IEEE Transactions on Computer Vision and Pattern Recognition (CVPR 97)*, volume 4, pages 775–781, Arizona, USA, May 1997.

[75] P. Somol, J. Novovicova, and P. Pudil. *Pattern Recognition Recent Advances on Efficient Feature Subset Selection and Subset Size Optimization.* InTech, New York, 1st edition, 2010.

[76] K. Sooknanan, A. Kokaram, D. Corrigan, G. Baugh, J. Wilson, and N. Harte. Improving underwater visibility using vignetting correction. In *Proceedings of the IS&T/SPIE Electronic Imaging Conference on Visual Information Processing and Communication III (SPIE 2012).*, volume 1, pages 1–8, Burlingame, California, USA, February 2012.

[77] K. Sooknanan, A. Kokaram, J. Doyle, J. Wilson, N. Harte, and D. Corrigan. Mosaics for burrow detection in underwater surveillance video. In *Proceedings of the IEEE International Conference on Oceans (OCEANS 2013)*, volume 1, pages 1–7, San Diego, California, September 2013.

[78] K. Sooknanan, A. Kokaram, N. Harte, G. Baugh, J. Wilson, and D. Corrigan. Indexing and selection of well-lit details in underwater video mosaics using vignetting estimation. In *Proceedings of the IEEE International Conference on Oceans (OCEANS 2012)*, volume 1, pages 1–7, Yeosu, Republic of Korea, May 2012.

[79] F. Spindler and P. Bouthemy. Real-time estimation of dominant motion in underwater video images for dynamic positioning. In *Proceedings from the IEEE International Conference on Robotics and Automation*, volume 2, pages 1063–1068, Leuven, Belgium, May 1998.

[80] J. Suk and F. Tomas. Pattern recognition by affine moment invariants. *IEEE Transactions on Pattern Recognition*, 26(1):167–174, June 1993.

[81] G. Sulzberger, J. Bono, J. Manley, T. Clem, L. Vaizer, and R. Holtzapple. Hunting sea mines with uuv-based magnetic and electro-optic sensors. In *Proceedings from the IEEE International Conference on Oceans (OCEANS 2009)*, volume 1, pages 1–7, Biloxi, MS, USA, September 2009.

[82] G. Suraj and S. Guru. Secondary diagonal fld for fingerspelling recognition. In *Proceedings of the IEEE International Conference on Computing: Theory and Applications (ICCTA '07)*, volume 1, pages 693–697, Kolkata, March 2007.

[83] M. Tian, Y. Zhuang, and S. Chen. Improving support vector machine classifier by combining it with k nearest neighbor principle based on the best distance measurement. *Journal on Intelligent Transportation Systems*, 1(1):373–378, June 2003.

[84] L. Torres-Mendez and G. Dudek. Color correction of underwater images for aquatic robot inspection. *Journal on Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer*, 2(1):60–73, July 2005.

[85] M. Turk and A. Pentland. Eigenfaces for recognition. In *IEEE Transactions on Computer Vision and Pattern Recognition (CVPR 91)*, volume 1, pages 586–591, Maui, Hl, June 1991.

[86] M. Wall, A. Rechtsteiner, and R. Luis. *Singular value decomposition and principal component analysis: A Practical Approach to Microarray Data Analysis.* Kluwer: Norwell, MA, USA, 1st edition, pages:91-109, 2003.

[87] L. Wei, X. Weidong, and L. Lihua. An exploring study of multi-scale complexity texture descriptors for medical image retrieval. In *Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE 2008)*, volume 1, pages 2635–2638, Shanghai, China, May 2008.

[88] S. XiangJun, W. HaoXiang, and Z. Qian. Training support vector machine through redundant data reduction. In *Proceedings of the 4th International Conference on Internet Multimedia Computing and Service (ICIMCS '12)*, volume 1, pages 25–28, Wuhan, China, March 2012.

[89] S. Xiaoqiao and L. Yaping. Gene expression data classification using svm-knn classifier. In *Proceedings of the IEEE International Symposium on Intelligent Multimedia, Video and Speech Processing.*, volume 1, pages 149–152, Changsha, China, October 2004.

[90] G. Xin, Z. Zhi-Hua, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2234–2240, November 2007.

[91] Z. Yuanjie, Y. Jingyi, K. Sing Bing, S. Lin, and C. Kambhamettu. Single-image vignetting correction using radial gradient symmetry. In *Proceedings of the IEEE Conference on*

Computer Vision and Pattern Recognition (CVPR 2008), volume 1, pages 1–8, Alaska, USA, May 2008.

[92] Z. Yuanjie, S. Lin, C. Kambhamettu, Y. Jingyi, and S. Kang. Single-image vignetting correction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, volume 1, pages 1063–6919, Anchorage, AK, USA, June 2008.

[93] J. Zhang, H. Wu, D. Zhong, and W. Smoliar. An integrated system for content based video retrieval and browsing. *Journal of Pattern Recognition*, 30(4):643–658, May 1997.

[94] M. Zhen, S. Qi, and Y. Baoxian. Hybridized knn and svm for gene expression data classification. *Life Science Journal*, 6(1):61–66, June 2009.

[95] Y. Zhuang, Y. Rui, S. Huang, and S. Mehrotra. Adaptive key frame extraction using unsupervised clustering. In *Proceedings from the IEEE International Conference on Image Processing (ICIP 1998)*, volume 1, pages 866–870, Chicago, Illinois, USA, September 1998.