

Towards robust identification of slow moving animals in deep-sea imagery by  
integrating shape and appearance cues

by

Marzieh Mehrnejad

B.Sc., Science and Research Branch of Azad University, 2010

A Thesis Submitted in Partial Fulfillment of the  
Requirements for the Degree of

MASTER OF APPLIED SCIENCE

in the Department of Electrical and Computer Engineering

© Marzieh Mehrnejad, 2015

University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by  
photocopying or other means, without the permission of the author.

Towards robust identification of slow moving animals in deep-sea imagery by  
integrating shape and appearance cues

by

Marzieh Mehrnejad

B.Sc., Science and Research Branch of Azad University, 2010

Supervisory Committee

---

Dr. Alexandra Branzan Albu, Co. Supervisor  
(Department of Computer and Electrical Engineering)

---

Dr. David Capson, Co. Supervisor  
(Department of Computer and Electrical Engineering)

## Supervisory Committee

---

Dr. Alexandra Branzan Albu, Co. Supervisor  
(Department of Computer and Electrical Engineering)

---

Dr. David Capson, Co. Supervisor  
(Department of Computer and Electrical Engineering)

### ABSTRACT

Underwater video data are a rich source of information for marine biologists. However, the large amount of recorded video creates a 'big data' problem, which emphasizes the need for automated detection techniques.

This work focuses on the detection of quasi-stationary crabs of various sizes in deep-sea images. Specific issues related to image quality such as low contrast and non-uniform lighting are addressed by the pre-processing step. The segmentation step is based on color, size and shape considerations. Segmentation identifies regions that potentially correspond to crabs. These regions are normalized to be invariant to scale and translation. Feature vectors are formed by the normalized regions, and they are further classified via supervised and non-supervised machine learning techniques.

The proposed approach is evaluated experimentally using a video dataset available from Ocean Networks Canada. The thesis provides an in-depth discussion about the performance of the proposed algorithms.

# Contents

<b>Supervisory Committee</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Table of Contents</b>	<b>iv</b>
<b>List of Tables</b>	<b>vi</b>
<b>List of Figures</b>	<b>vii</b>
<b>Acknowledgements</b>	<b>viii</b>
<b>Dedication</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The detection of deep-sea animals in cabled observatory video data . . . . .	1
1.2 Focus of Thesis . . . . .	3
1.3 Outline of Thesis . . . . .	3
<b>2 Related Work</b>	<b>5</b>
2.1 Underwater Image Analysis . . . . .	6
2.2 Sealife detection, tracking, measurement and behavior analysis . . . . .	9
2.3 Machine learning methods for underwater image analysis . . . . .	11
<b>3 Proposed Approach</b>	<b>13</b>
3.1 Module A - Preprocessing . . . . .	13
3.2 Module B - Segmentation . . . . .	19
3.3 Module C - Feature Extraction . . . . .	27
3.4 Classification . . . . .	29
3.4.1 Feed-forward Neural Network . . . . .	31

3.4.2 Self Organizing Maps . . . . .	33
<b>4 Experimental Results</b>	<b>35</b>
4.1 Performance Evaluation . . . . .	35
4.2 Database . . . . .	37
4.3 Discussion on Segmentation and Classification Results . . . . .	37
<b>5 Conclusions</b>	<b>44</b>
<b>Glossary</b>	<b>47</b>
<b>Bibliography</b>	<b>48</b>

# List of Tables

Table 3.1	Segmentation thresholds in the RGB color space. . . . .	20
Table 3.2	Segmentation thresholds in the HSV color space. . . . .	20
Table 3.3	Segmentation thresholds in the YIQ color space. . . . .	24
Table 3.4	Evaluation results of the three color spaces. . . . .	26
Table 4.1	Evaluation results of the Feed-Forward Neural Network. . . . .	39
Table 4.2	Evaluation results of the Self-Organizing Maps. . . . .	40
Table 4.3	Results of the cascade classification I . . . . .	42
Table 4.4	Results of the cascade classification II . . . . .	43

# List of Figures

Figure 1.1 Geographical distribution of ONC's cabled observatories . . . . .	2
Figure 1.2 Grooved tanner crab ( <i>Chionoecetes tanneri</i> ) . . . . .	4
Figure 2.1 Underwater image recorded by Remotely Operated Vehicle. . . . .	6
Figure 2.2 Underwater image recorded by stationary camera. . . . .	7
Figure 3.1 Main steps of the modular proposed approach . . . . .	14
Figure 3.2 Frames from the videos in ONC's [1] database. . . . .	15
Figure 3.3 Pre-processing steps in the HSV color-space. . . . .	16
Figure 3.4 Pre-processing steps in the RGB color-space. . . . .	17
Figure 3.5 Pre-processing steps in the YIQ color-space. . . . .	18
Figure 3.6 CLAHE Algorithm. . . . .	18
Figure 3.7 The generated histograms in the RGB color space. . . . .	21
Figure 3.8 The generated histograms in the HSV color space. . . . .	22
Figure 3.9 The generated histograms in the YIQ color space. . . . .	23
Figure 3.10 Eccentricity = $\frac{c}{a}$ . . . . .	24
Figure 3.11 Segmentation process in the RGB color space. . . . .	25
Figure 3.12 Segmentation process in the HSV color space. . . . .	25
Figure 3.13 Segmentation process in the YIQ color space. . . . .	26
Figure 3.14 Feature extraction process. . . . .	27
Figure 3.15 Image after performing segmentation and feature extraction. . . . .	28
Figure 3.16 Feature vector structure. . . . .	30
Figure 3.17 Feed-forward neural network. . . . .	31
Figure 3.18 Self Organizing Map. . . . .	34
Figure 4.1 Intermediate results obtained after the segmentation module. . . . .	36
Figure 4.2 The mean $ME$ and the neural network size. . . . .	38
Figure 4.3 Cascade Classification. . . . .	41

## ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors **Dr. Alexandra Branzan Albu** and **Dr. David Capson** for granting me the opportunity to do this degree, continuous guidance and support.

Thank you to my parents, **Zahra Zhaleh** and **Reza Mehrnejad** for their endless support and encouragement throughout this journey.

Special thanks to my labmates: Anissa Agahchen, Frederic Jean, Trevor Beugeling, Kawthar Moria and Jeremy Svendsen, for the friendship, advice and support.

## DEDICATION

I would like to dedicate this thesis to my loving family for their never ending encouragements throughout my life:

Mahdih Mehrnejad

Mohsen Mehrnejad

Reza Mehrnejad

Zahra Zhaleh

# Chapter 1

## Introduction

In recent years there has been a growing interest in gathering and processing various types of data such as text, images, audio files, etc. in order to retrieve meaningful information. The field of Computer Vision focuses on automatically making sense of real world imagery data in terms of making useful decisions about real physical objects based on sensed images. Our project explores the analysis of 'big imagery data' from the underwater environment. In this chapter we present contextual information about our research project as well as our motivation.

### 1.1 The detection of deep-sea animals in cabled observatory video data

The marine environment has an enormous importance for the health of our planet. Although the ocean covers over 70% of our planet's surface, very little of the volume of the ocean has been explored to date.

Traditional ocean exploration methods involve ship cruises, which gather data using submersibles equipped with a variety of sensors. Such methods offer high spatial resolution, as well as the flexibility to explore a variety of geographical sites; however, temporal sampling is severely limited by the frequency of cruises. Chave et al. [12] mention that the use of ships as primary observational platform has yielded a series of snapshot views of the oceans, which have limited resolution in time. However, high temporal resolution is critical for ocean exploration initiatives that are primarily interested in observing the impact of complex dynamic processes such as climate change on ecosystems.

Cabled observatories represent a relatively novel paradigm of ocean data collection, which enables researchers to understand, analyze, and model complex natural phenomena of high spatio-temporal variability. A cabled observatory consists of a network of nodes equipped with multiple sensing instruments; the nodes are positioned on the seafloor according to a predefined deployment plan. Underwater cables provide nodes with power, and allow for the remote control of the operating mode of various sensors; they also enable data streaming from each sensing unit. Chave et al. [12] provide a thorough description of a generic cabled observatory, and underline the main differences between the design processes behind cabled observatories and commercial submarine telecommunication systems. Their chosen case study is the NEPTUNE (North-East Pacific Time-Series Undersea Networked Experiments) Canada observatory as of 2006, which was officially launched in December 2009. Ocean Networks Canada (ONC) operates the world-leading NEPTUNE and VENUS (Victoria Experimental Network Under the Sea) cabled ocean observatories. Figure 1.1 shows the current node distribution in the NEPTUNE (deep-sea) and VENUS (coastal) cabled observatories.



Figure 1.1: Geographical distribution of nodes in the NEPTUNE and VENUS cabled observatories (2014) (Image was adapted from ONC [1])

Cabled observatories exhibit three distinctive characteristics with respect to the traditional ship-based exploration paradigm. First, they enable continuous, non-invasive presence in the environment. The temporal sampling frequency of the various sensors is set up according to the needs of biologists, as well as to the nature of the phenomena under observation. Second, the co-located sensors enable long-term correlational studies. For instance, video cameras are co-located with more basic sensors that measure salinity, water temperature etc. Thus, variations in species abundance and/or behavior can be correlated to physical changes in their habitat. Third, interactivity enables scientists to change the field of view of the cameras, or to modify the temporal sampling frequency of various instruments.

Seafloor cameras are essential for the non-invasive monitoring of animal life in deep-sea environments. High volumes of imagery data are typically collected and archived. It is extremely difficult, if not impossible, for biologists to inspect all the relevant data for the purpose of species identification and behavior analysis. This ‘big data’ problem emphasizes the critical need of computer vision techniques for automatic detection of certain species.

## 1.2 Focus of Thesis

This thesis focuses on the problem of species identification from still images, using only appearance cues. This is a less explored area, as most approaches for sea-life detection target highly mobile species (fish) and use motion as the primary detection cue. However, there are many deep-sea animals, which are stationary or quasi-stationary; examples include crabs, sea anemones, sea urchins etc.

The main goal of this work is to automate the detection of stationary grooved crabs (See figure 1.2 ) in the videos recorded by ONC Canada’s cabled observatory system. Once the automatic detection is performed, the results are verified and evaluated against detection performed by a manual human classifier.

## 1.3 Outline of Thesis

The thesis is structured as follows: Chapter 2 provides an overview of work related to our study. The proposed approach is presented in chapter 3. Experimental results and validation are presented in chapter 4. The thesis ends by outlining future research directions and by providing some concluding remarks.



Figure 1.2: Grooved tanner crab (*Chionoecetes tanneri*) recorded at 834 m depth, 9 August 2006.

## Chapter 2

# Related Work

During the last decades there has been a considerable increase of interest in the exploration of the underwater environment. One of the forms of data that is collected from the ocean and plays an important role in understanding the underwater environment, is visual data. In order to capture a natural, undisturbed view of the underwater habitats, it's desirable to avoid invasive observation methods such as having divers collect data. The presence of humans among underwater species can easily result in changes in the behavior of the animals. Another disadvantage of these invasive methods is that they are labour intensive. A minimally invasive approach to collecting data, which has the least effect on the ecosystem is the installation of sea-floor mounted cameras and sensors that operate with artificial light. These devices can communicate the natural state of the underwater environment to marine biologists [32][15][19][34]. These visual data contain valuable information about the variation of the species' behavior, abundance, etc. under certain environmental conditions[30][34][32]. However, this enormous amount of data introduces a "big data" problem and therefore the analysis of such data requires automation techniques[32][15][7].

The analysis of underwater images enforces a set of distinctive challenges due to the optical properties in the underwater medium. A considerable amount of research has been done to overcome these challenges; the next section is a detailed review of these issues and the proposed solutions.

## 2.1 Underwater Image Analysis

The attenuation of light as it travels through the water, results in a significant degradation in the quality of the images captured from the ocean. This attenuation is caused by absorption (i.e. light loses its energy) and scattering (i.e. the change in the direction of the individual photons) [27]. These processes are caused by the water's physical properties and also by dissolved organic matter [27]. The major issues that are present in underwater images are blurring of the image features, limited visibility range, low contrast and bluish appearance (see Figure 2.1) .

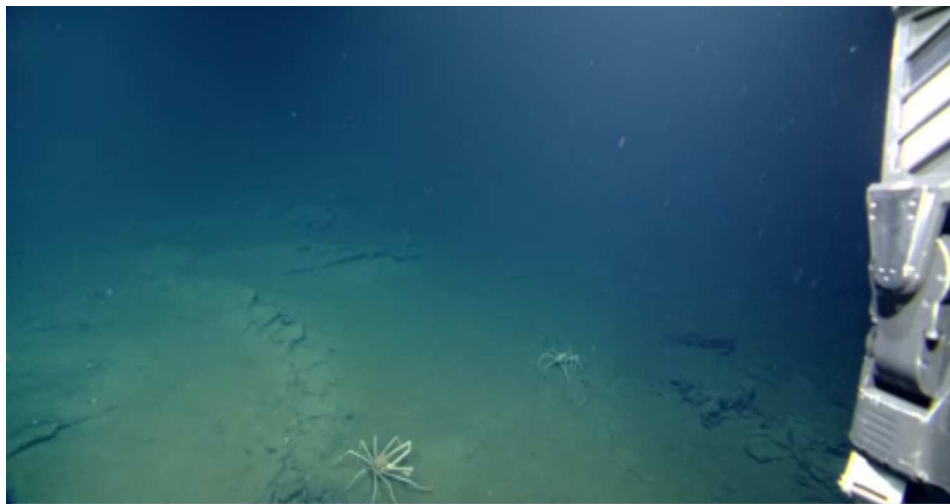


Figure 2.1: Underwater image recorded by Remotely Operated Vehicle (ROV). Image was adapted from NEPTUNE [1]

Natural light is usually not sufficient for imaging the sea floor and artificial lighting is used as a solution. In addition to the above mentioned problems, artificial lighting causes non-uniform illumination, which results in a bright spot in the center of the image and a poorly illuminated area surrounding it (Figure 2.2) [16].

Prior to detection of sea-life in underwater images, the above mentioned problems are to be addressed. Different solutions to these issues are generally regarded as pre-processing techniques. As it is shown in the survey conducted by Raimondo et al. [27], a large number of pre-processing techniques have been proposed for underwater images.

In general, pre-processing can be discussed from two view points:

1. Image restoration techniques



Figure 2.2: Underwater image recorded by stationary camera. Image was adapted from NEPTUNE [1]

These techniques apply algorithms that restore corrupted images by using depth estimation of a given object in the scene and many model parameters which are extremely variable (e.g. attenuation and diffusion coefficients) [27].

## 2. Image enhancement and color correction methods

These methods apply qualitative subjective criteria that don't rely on any physical model for the image formation process and therefore there is no need for information about the environment [27].

We will focus on image enhancement and color correction methods, because the goal is to address solutions which can be applied on images of locations where the physical characteristics of the environment are variable. Image enhancement is achieved by improving the contrast, which can be accomplished through various techniques. Bazeille et al [5] and Campos et al [16] recommend Homomorphic filtering and Contrast Limited Adaptive Histogram Equalization (CLAHE). Homomorphic filtering and CLAHE both reduces the effects of non-uniform lighting and results in a more balanced image. In addition, Bazeille et al [5] suggest anisotropic filtering

which reduce the noise and enhances the edges.

Regarding the color degradation of underwater videos, the goal is to process the image in order to remove the different undesired color casts (i.e. a tint of a particular color, which is usually unwanted and affects the whole image) that are present due to non-uniform illumination.

The videos taken in the aquatic environment present a strong and non uniform color cast [11]. This means that the cast has a different color and intensity in foreground, background, shadows and highlights. Several techniques are used to remove color casts, such as the ones discussed bellow.

The White Patch Algorithm estimates the cast from specular reflections or from an object that is originally white in the scene [10]. The Gray World Algorithm estimates the cast from the mean of the image [10].

The cast in highlights is best estimated by the White Patch method, while the cast in the mid-tones and shadows is more appropriately estimated by the Gray World algorithm. In images where highlights, mid-tones and shadows are present throughout the image, the union of these two methods enables the correction of the entire image. The Hybrid Gray World and White Patch Algorithm applies a different method depending on whether the zone in the image corresponds to shadows, highlights or mid-tones. In other words, a different method is used to correct a certain zone, depending on the intensity level of the pixels in that zone [10].

Typical color constancy methods such as Gray World and White Patch are not suitable pre-processing techniques for underwater images. These methods are global and can only deal with uniform color casts. Since the Gray World method is based on the mean of the image, it estimates the most noticeable cast in the image and therefore it only corrects the shadows of the image. Whereas, the White Patch method only estimates and corrects the cast in the highlights. ACE is the most desirable algorithm since it widely adapts to different non uniform color casts and it is unsupervised.

Due to non-uniform illumination, there are several casts throughout underwater images, and therefore global color correction algorithms are not sufficient in dealing with this issue. An appropriate solution for our problem is the Automatic Color Enhancement technique (ACE) which was first proposed by Rizzi et al [28]. The ACE algorithm is used for the unsupervised enhancement of digital images [28]. This method combines the Gray World and White Patch equalization mechanisms, while considering the spatial distribution of color information; therefore the parameters describing midtones, shadows and highlights do not need to be defined manually.

Chambah et al. [11] recommend this approach for underwater images. The distinctive feature of this algorithm is that it corrects the color in an image despite the variability in the lighting throughout the image.

Ancuti et al. [3] proposed a new strategy that enhances underwater imagery. This approach overcomes the color degradation and low contrast of underwater images by combining the following two input images:

- The color corrected version of the original image

The goal is to process the image in order to remove undesired color casts that are present due to variant illuminants. The color correction methods explained in the beginning of this section can be used for this task (preferably the ACE method which doesn't rely on manually set parameters).

- The contrast enhanced version of the image

The aim is to use an image that has a relatively good contrast and includes the details of the scene which have been degraded. An appropriate method for this task would reduce the noise while preserving the edges and therefore increasing the sharpness. Several techniques can be used for this purpose, including anisotropic filtering, bilateral filtering and median filtering. The technique that has been applied by Ancuti et al [3] is CLAHE, because it operates in a fully automated manner.

The weights of the fusion process determine which pixel should appear in the final image. These weights are determined depending on the desired output. The choice of the optimum combination of pre-processing techniques is an important step towards solving these problems. Ancuti et al [3] use a white balancing operation and CLAHE for color correction and contrast enhancement, respectively.

After improving the quality of the visual data, the main analysis is performed. Analysis usually includes tasks such as detection, tracking, identification and/or the study of the behaviour of certain species.

## **2.2 Sealife detection, tracking, measurement and behavior analysis**

Underwater videos are generally recorded by remotely operated vehicles, aquatic robots, seafloor mounted cameras, etc. These videos are used to study various types

of fish, bacterial mat coverage, squat lobsters, plankton and to assess megafaunal densities, etc.

The detection of sealife plays an important role in performing economic measurements and estimations. In addition, the analysis of fish behaviour aids marine biologists in their studies on pollution and climate change. Most of the conducted research on underwater animals focus on studying sealife that exhibit motion. This enables applying motion detection as the main or first step in the analysis of visual data. A recent review on detection and measurement of fish in underwater videos was performed by Shortis et al [31].

An approach to automatically measuring fish in videos was performed by Tillet et al. [35]. This method employs 3D Point Distribution Models (PDM), that include locations on the silhouette of the fish. The PDM of the species of interest does not depend on the scale and orientation of the fish. The PDM is fitted to the edge fish images by minimizing an energy function. In the minimization process, the edges are chosen based on their magnitude, direction and closeness to the PDM. An extension to this research was developed by Lines et al. [24] by implementing the above mentioned technique in low contrast areas and also by introducing a fish detection step.

The location of the animal of interest can be identified by background subtraction techniques which are discussed bellow.

An evaluation of algorithms that employ background subtraction techniques in order to detect stationary foreground objects, was performed by Bayona et al. [4]. In most cases where the camera's position is fixed and the changes in illumination are gradual, background-subtraction techniques are very popular [23] [6]. A significant number of these techniques have been proposed with the purpose of being employed in surveillance cameras and their aim is to develop an algorithm that is invariant in illumination changes and also suitable in complex background environments.

After locating the fish, in order to measure its abundance or understand its behaviour, tracking is usually performed by computing motion trajectories[32] [8] [15] [34] [33]. In Spampinato et al.'s work [34] [33] fish are tracked by combining two algorithms; shape feature matching and histogram matching. Boom et al. [8] and Spampinato et al. [32] track fish in various frames by comparing their features. The features include various color descriptors along with the position of the fish in the frame. In the algorithms presented by Fier et al. [15] fish are tracked and counted by comparing their shape outline and motion vectors.

## 2.3 Machine learning methods for underwater image analysis

A very popular tool for working with large volumes of data and attempting to extract semantically meaningful patterns from data, is machine learning. Machine learning enables computers to learn from available data and make predictions. Machine learning tasks are generally classified into two categories: supervised and unsupervised learning. Supervised learning is performed in two phases of training and testing. In the training phase, the algorithm is provided with sample inputs and their corresponding desired outputs (training data). The algorithm learns a general rule between the inputs and outputs. The testing phase consists of presenting the algorithm with testing data and retrieving the algorithm's output. The algorithm's performance is evaluated by comparing its output of the testing data with the desired output. In unsupervised learning, input data is presented to the algorithm without providing desired outputs. The algorithm learns the structure of the input data by itself and provides outputs. The algorithm's performance is evaluated by comparing its output with the desired output. Machine learning techniques are commonly applied in performing classification tasks.

In the studies conducted on the underwater environment, machine learning is applied to classify various species of fish [29] [19], determine the presence of certain sea-life [30] [18] or determine whether the behaviour of a number of individual animals from the same species is different from others [34]. Support Vector Machines (SVMs) are the most commonly used machine learning algorithms in underwater image analysis tasks. Other classification techniques such as k-nearest neighbours (kNN), Random Forest (RF), Self-organizing maps (SOMs), etc have been applied, but SVMs have proved to be more successful in recent research.

After the animal has been located, various features are extracted that represent the animal for the classification task. The choice of features is based on the nature and quality of the image. To distinguish between species, choosing features that are similar among classes is usually avoided, e.g. if two types of fish have very similar color, employing features that represent color properties will usually decrease or have no effect upon the classifier's performance. In general, features describing texture, shape and color are commonly used [30] [29] [19] [34] [8] [18] [32] to form feature vectors that are the inputs of the classifiers.

SVMs perform very well in classification problems involving two classes. When

more than two classes are present, SVMs are extended to multi-class classifiers where each class is compared against all other classes. Successful classification among various species of sealife has widely been performed by SVMs [18] [32] [8] [30] [29]. Hu et al. use SVMs in order to classify regions into various plankton categories. Co-occurrence matrices are used as features representing each class of plankton [18]. In Rova et al.'s work [29], SVMs are applied in classifying two fish species that are similar in shape but vary in texture. The textural properties form the feature vectors which are the inputs of the classifier SVM. Schoening et al. classify 9 regions (8 taxon and the background) using SVMs. Each SVM determines whether the region corresponds to a certain taxon or not (all others). Color and texture information is generally used to describe each pixel. Boom et al. apply hierarchical classifier SVMs to classify 15 fish species. The input features consist of color, contour and texture information [8].

In this thesis, the low quality of the videos has been addressed by employing several pre-processing steps that are suitable for our main task which is the detection of stationary grooved crabs. Since these crabs exhibit very little motion and in most cases no movement, motion detection techniques are not applicable to their detection process. The features that make these species distinguishable from their surroundings are color and shape which have been used to locate potential crab regions. In order to provide more reliable detection results, the potential crab regions are further classified by machine learning techniques. The next chapter contains a detailed description of the modules in the detection process.

## Chapter 3

# Proposed Approach

Our research is focused on the development of an algorithm that detects stationary grooved crabs in videos recorded by ONC's cabled observatory system. The intended application of this algorithm is to aid marine biologists in the analysis of underwater videos. Since the grooved crabs in our database exhibit very little or in most cases no motion throughout the videos, their detection does not consider motion cues. Due to the low quality and also the nature of underwater imagery, pre-processing techniques had to be applied prior to the detection process.

Our approach is designed as a sequence of four main modules. The pre-processing module performs pre-processing in order to improve contrast and color. The segmentation module segments the image and identifies potential crab regions of interest (ROI) using color, size and shape cues. The feature extraction module is based on edge detection applied to the ROIs, followed by the normalization of the ROIs and the generation of feature vectors. The classification module categorizes these feature vectors into crab and non-crab classes using supervised and unsupervised machine learning algorithms. Figure 3.1 presents the steps in all of these modules. The remainder of this chapter presents a detailed description of each module.

### 3.1 Module A - Preprocessing

In order to enable the efficient analysis of underwater images; a set of pre-processing algorithms must be performed. The choice of the pre-processing algorithms to be applied is based on the quality of the images. Sample images from the videos in the database from the Barkley Canyon site on the ONC cabled observatory [1] are

- Module A - Preprocessing
  1. Perform pre-processing steps to prepare the frames for the detection techniques.  
OUTPUT: Color image with improved color and contrast.
- Module B - Segmentation
  1. Perform color-based segmentation by thresholds derived from color histograms.  
OUTPUT: Binary image.
  2. Filter binary image according to shape and size criteria.  
OUTPUT: Binary image of potential crab regions.
- Module C - Feature extraction
  1. Perform local edge detection around potential crab regions.  
OUTPUT: Binary image of potential crab regions with distinguished appearance (visible legs).
  2. Normalize each region using its regular moments.  
OUTPUT: Potential crab regions invariant to scale and translation.
  3. Generate fixed-size crab and non-crab templates from their corresponding regions of interest.  
OUTPUT: Fixed-size square matrices reshaped into vectors as inputs of the classifier feed-forward neural network or self organizing map.
- Module D - Classification
  1. Classify the crabs and non-crabs by feed-forward neural network or self organizing map.  
OUTPUT Binary: 1 for crab and 0 for non-crab.

Figure 3.1: Main steps of the modular proposed approach

presented in Figure 3.2. As it is shown in the sample images; the dark areas in the top of the frames, limit the detection of crabs that are located in those areas. In addition, parts of the man-made instruments and also some rocks have a similar color to crabs.



Figure 3.2: Frames from the videos in ONC's [1] database.

This module consists of three steps which are described below and illustrated by example results shown in Figures 3.3, 3.4 and 3.5 for various color spaces.

### 1. Radial Brightening

This step is a simple and intuitive solution for correcting the gradual decrease of brightness due to the use of an artificial spotlight. The V value of each pixel  $(x, y)$  in the HSV space is increased based on the pixel's distance to the center point  $(x_0, y_0)$  at the bottom of the image:

$$V_{new}(x, y) = V_{old}(x, y) + KD(x, y, x_0, y_0) \quad (3.1)$$

where,  $K = 0.00025$  (chosen empirically) and  $D$  is the Euclidean distance of each point  $(x, y)$  from the center point  $(x_0, y_0)$ .

## 2. LASER point elimination

A pair of LASER spots are present in the center of the lower half of all frames, which result in many false detections. In this pre-processing step, these points are automatically detected by color-based segmentation and also shape (circular) and size descriptors. In order to avoid false detection, the detection algorithm is performed in a rectangular region located at the center of the lower half of each frame. In order to determine the thresholds for color-based segmentation, sample patches representing LASER spots are described in terms of their color histograms.

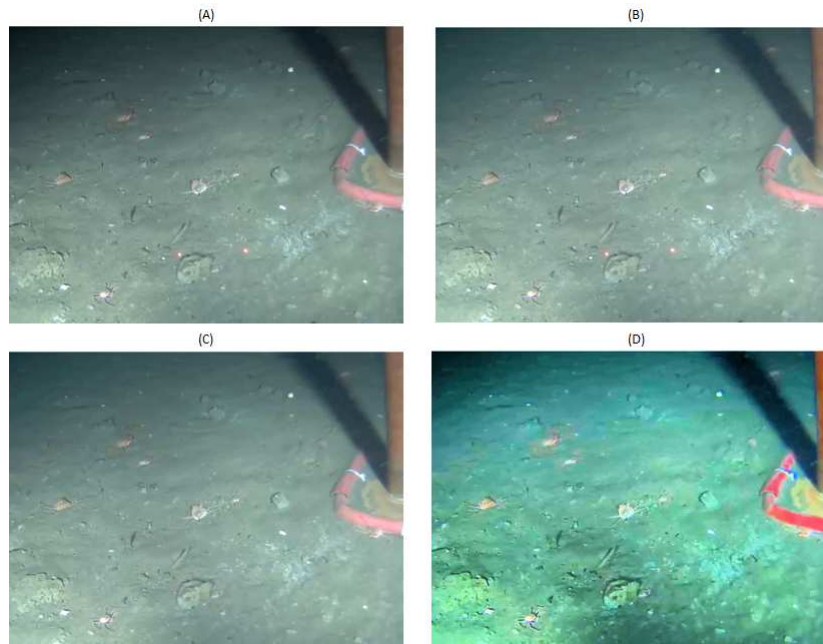


Figure 3.3: A: Original image. B: Image after radial brightening. C: Image B after LASER point elimination in the HSV color-space. D: Image C after performing the CLAHE algorithm.

## 3. Contrast Enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE)

In order to be able to distinguish shape and color features of the crabs, the contrast of the images needs to be increased. The image has to have relatively good contrast and include the details of the scene. An appropriate method for this task would reduce the noise while preserving the edges and therefore increasing the sharpness. Several techniques have been recommended for this

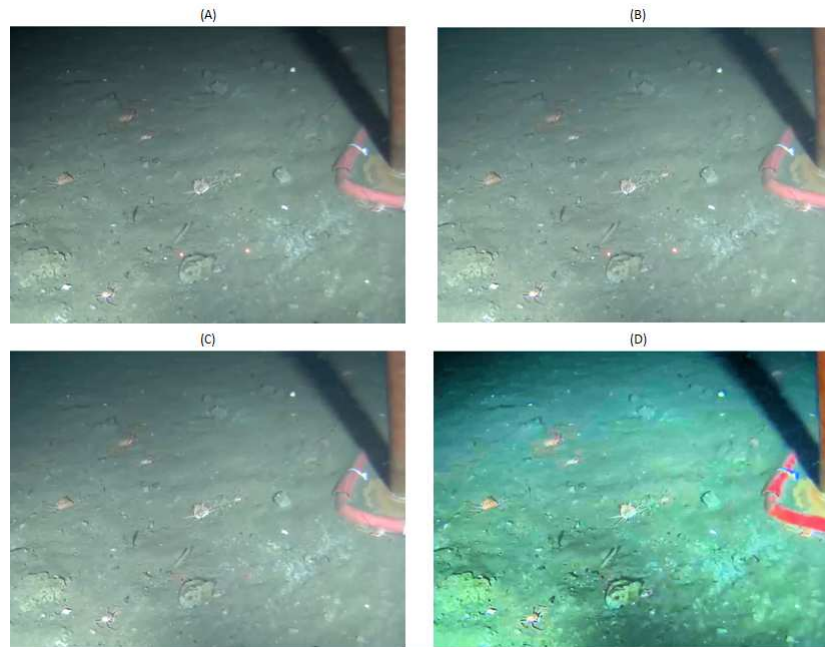


Figure 3.4: A: Original image. B: Image after radial brightening. C: Image B after LASER point elimination in the RGB color-space. D: Image C after performing the CLAHE algorithm.

purpose (for underwater images), including anisotropic filtering [5], homomorphic filtering [16] [5] and Contrast Limited Adaptive Histogram Equalization (CLAHE). CLAHE is a pre-processing step that has been discussed by Eustice [14] for underwater images and originated from Zuiderveld [37] who first used it for low-contrast medical imagery.

In this work we adopt CLAHE to improve the contrast of the images. Due to non-uniform lighting, global contrast enhancement techniques fail to augment the image quality. Local histogram equalization also performs poorly, since it is very time consuming and it amplifies noise in poorly contrasted areas. CLAHE is a suitable technique, because it operates in a fully automated manner and overcomes the two above mentioned shortcomings.

CLAHE operates on small regions in the image and applies histogram equalization on each pixel based on its surrounding pixels (its contextual region), rather than the entire image [26]. The contrast in each region is enhanced, so that the histogram on the output region approximately matches the histogram of a suitable gray level distribution (Rayleigh is optimal for underwater imagery

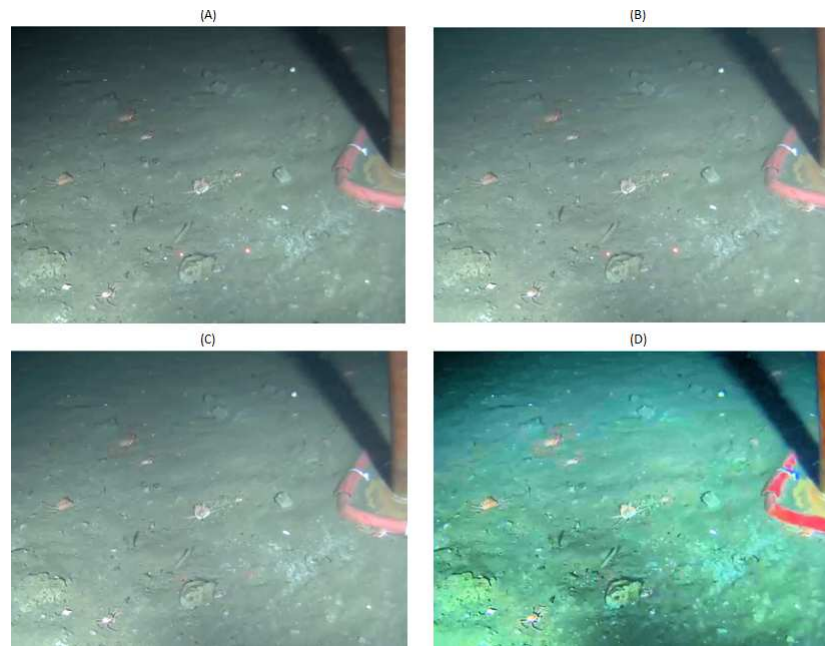


Figure 3.5: A: Original image. B: Image after radial brightening. C: Image B after LASER point elimination in the YIQ color-space. D: Image C after performing the CLAHE algorithm.

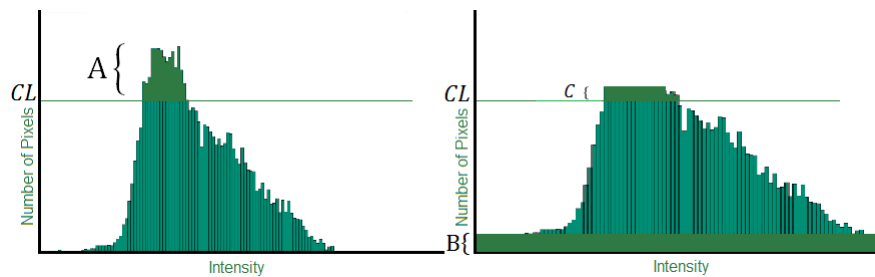


Figure 3.6: The area (A) above the clip limit (CL) is redistributed to B (The area of A is equal to B). The redistribution will push some bins over the clip limit again (region C in the figure), resulting in an effective CL that is larger than the prescribed limit and the exact value of which depends on the image. If this is undesirable, the redistribution procedure can be repeated recursively until the excess (C) is negligible.

[14]). The neighboring regions are then combined using bi-linear interpolation to avoid the formation of artificial boundaries. The contrast, especially in homogeneous areas, has to be limited in order to avoid noise amplification. To limit the amount of contrast enhancement, a clip limit (CL) is defined and used; CL is a multiple of the average histogram bin contents. Each bin can only contain

a maximum of CL, the remainder is uniformly distributed among the other bins (Figure 3.6). This limits the slope of the cumulative histogram which is used in the calculation of the gray level transform. The clipping factor is intended to decrease the noise amplification in low contrast regions. In addition to expanding the contrast, this technique occupies a larger portion of the intensity range than the original image. Therefore, the enhancement results in maximizing the contrast between adjacent structures.

In this work, the CLAHE algorithm was applied to improve the contrast of the images in our database. However, since we are working with low quality images, the CLAHE algorithm introduces noise that corrupts the color segmentation. Therefore the original images are used in the segmentation module and the contrast enhanced images are used in the feature extraction module.

## 3.2 Module B - Segmentation

Given that the color of grooved crabs is similar to human skin color, we have investigated related work on color-based human skin detection. As shown by Albiol et al [2], the choice of color space for describing color information is critical for the design of a successful skin detector. However, there is no clear consensus on which color space is optimal for skin detection applications. Albiol et al [2] recommend the use of the HSV space on the basis that it provides some robustness to variable illumination. Brand and Mason [9] work with RGB spaces and use the R/G ratio as the main color descriptor. Wand and Branstein [36] propose YIQ, a new color space derived from the standard NTSC. To investigate which color space works best for our task, we followed the recommendation of Elgammal et al [13] and collected a database of patches representing crab and non-crab regions. All patches were described in terms of their color histograms in each of the considered color spaces (RGB, HSV and YIQ). Figures 3.7, 3.8 and 3.9 show the histograms corresponding to each color space.

These histograms enabled the determination of specific threshold values for color-based thresholding, which was performed independently in each of the three color spaces:

- RGB

As shown in Figure 3.7, the R/G, B/G, B/R values for crab and non-crab patches are all confined to a limited range. The partial overlap in the B/G range

of the two classes was solved by setting thresholds based on all three ratios. A similar solution was used for the other two color spaces. The thresholds for the ratios were manually determined from their corresponding histograms in order to maximize the discriminability between crabs and non-crabs. The thresholds are presented in Table 3.1.

Table 3.1: Thresholds of crab and non-crab regions in the RGB color space.

Class	R/G Range	B/G Range	B/R Range
Crab	0.9106 - 1.2521	0.8448 - 1.0786	0.7002 - 1.1321
Non-crab	0.617 - 0.9117	0.8757 - 1.1563	0 - 2.2514

- HSV

According to the histograms in Figure 3.8, the hue (H) and saturation (S) values of crab and non-crab patches are confined to a fairly limited range. There is significant variation for the V value for both classes; this is why the V value is not considered in the thresholding process. The thresholds are presented in Table 3.2.

Table 3.2: Thresholds of crab and non-crab regions in the HSV color space.

Class	S Range	H Range
Crab	0 - 0.0020659	0 - 0.1694
	0.0103 - 0.3409	0.5281 - 0.6876
		0.8868 - 1
Non-crab	0.095 - 0.345	0.0499 - 0.1097
	0.365 - 0.395	0.3415 - 0.5785

- YIQ

As shown in Figure 3.9, the I and Q values of crab and non-crab patches are confined to a small range (with some overlaps in Q). The Y values are in a wide range of values, therefore the Y values were not considered for thresholding purposes. The thresholds are presented in Table 3.3.

The binary images resulting from color-based thresholding are postprocessed using size and shape considerations. The regions that do not belong to the crab area range are eliminated. We use an eccentricity measure in order to retain only compact and

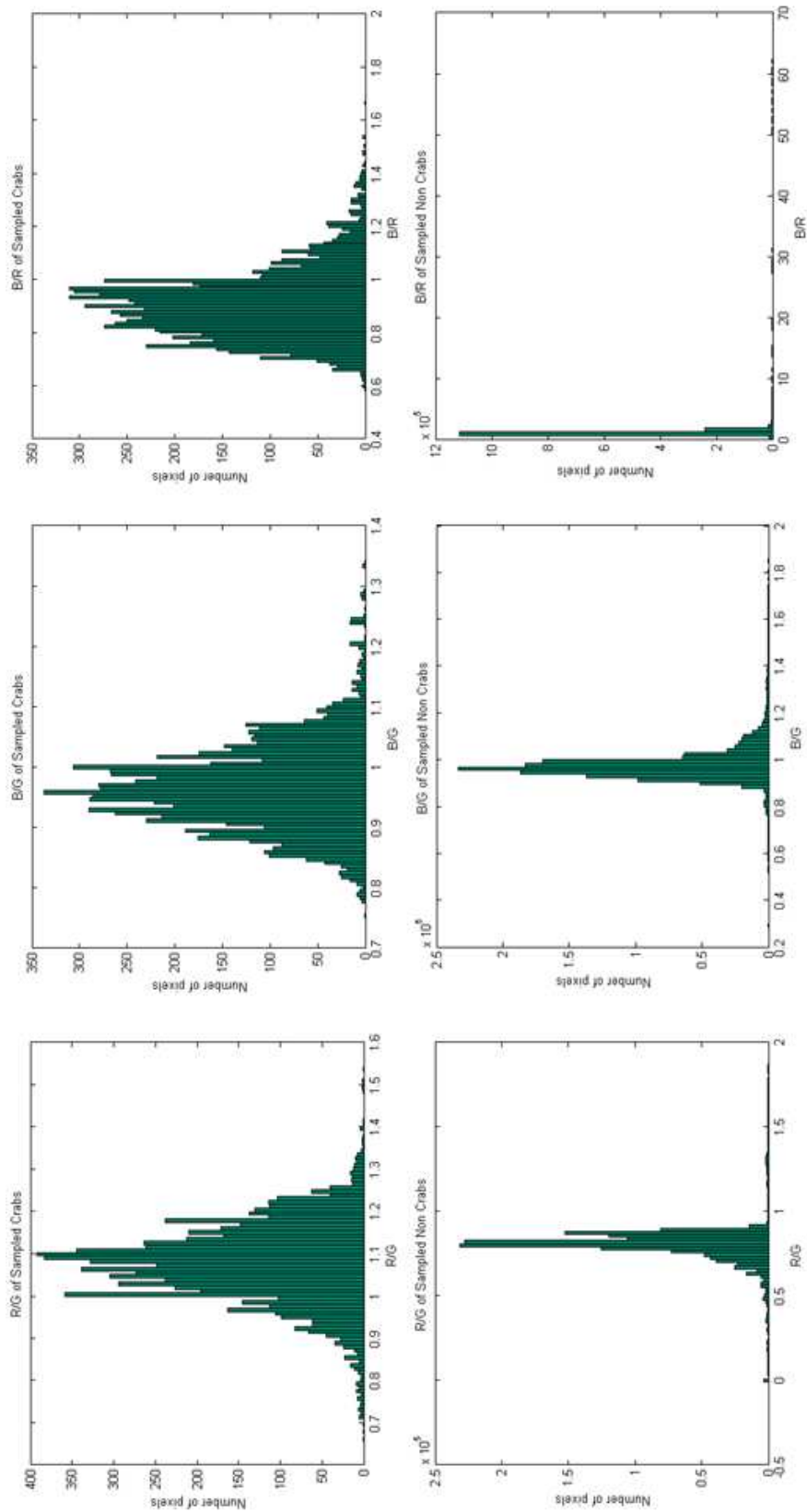


Figure 3.7: The generated histograms in the RGB color space.

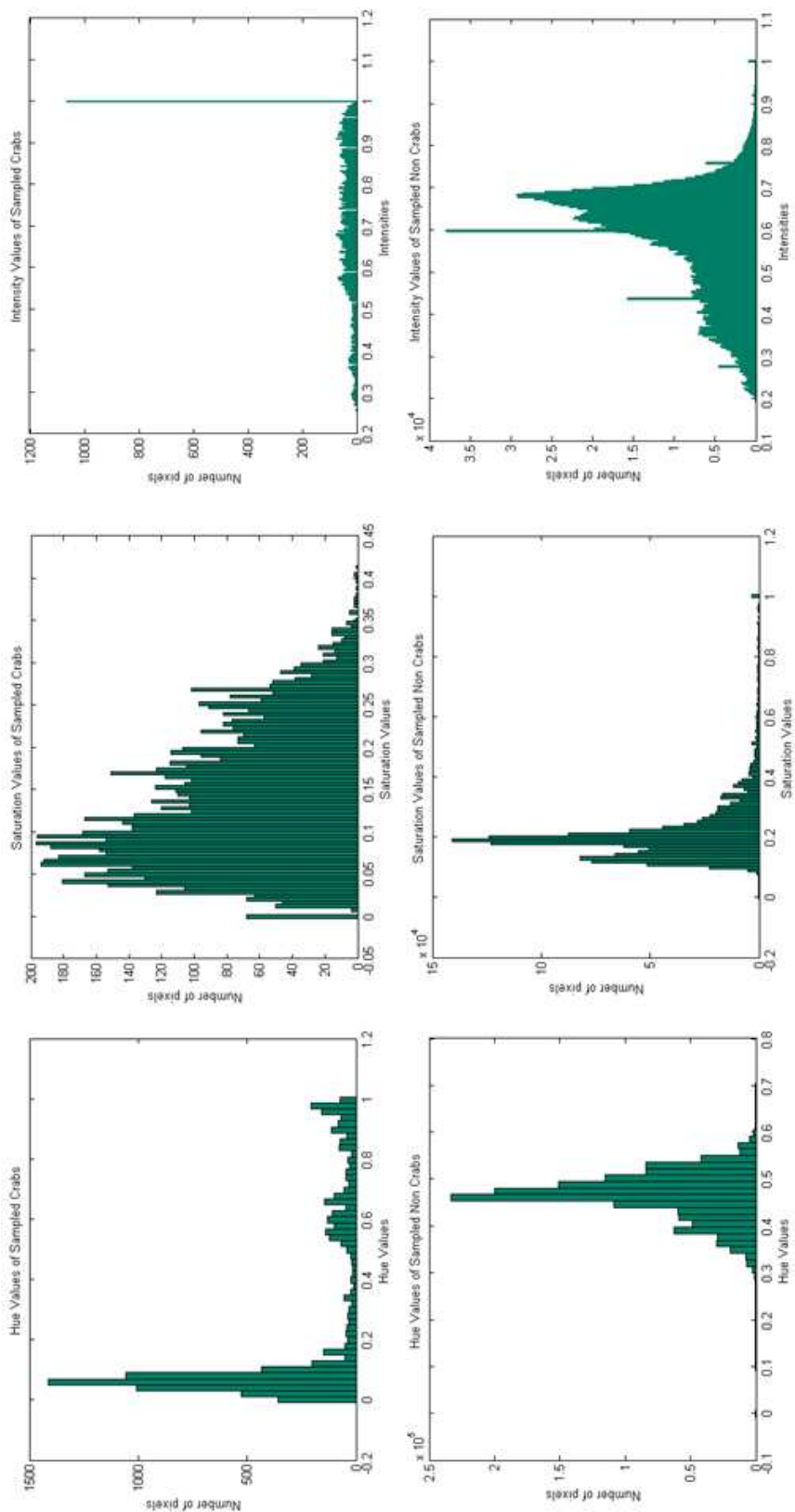


Figure 3.8: The generated histograms in the HSV color space.

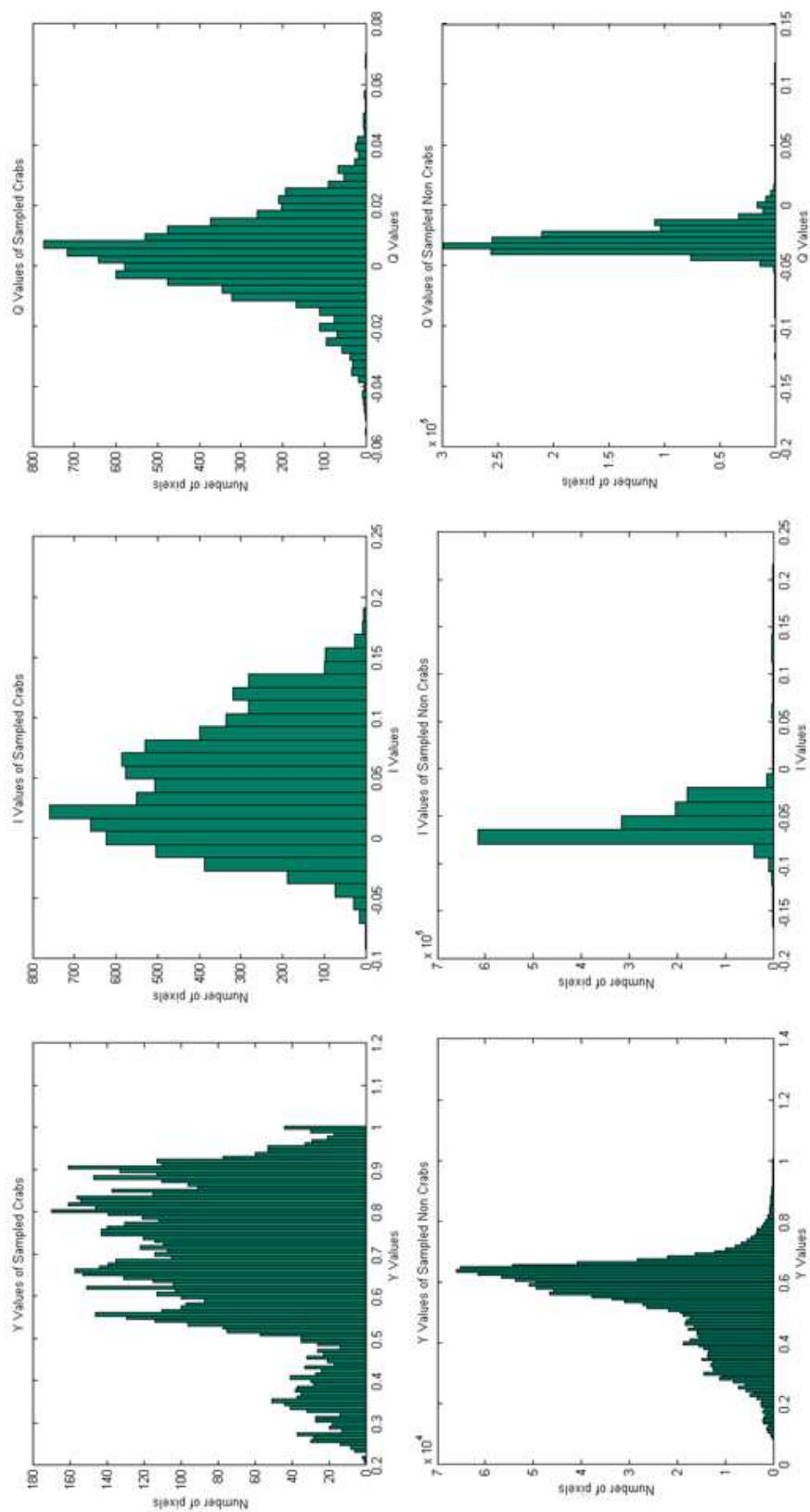


Figure 3.9: The generated histograms in the YIQ color space.

Table 3.3: Thresholds of crab and non-crab regions in the YIQ color space. (The values are coordinates;  $I \in [-0.5957, +0.5957]$   $Q \in [-0.5226, +0.5226]$ )

Class	I Range	Q Range
Crab	-0.038 - 0.1361	-0.0214 - 0.0257
Non-crab	-0.0937 - -0.02	-0.0501 - -0.007494

circular regions which are more likely to represent crabs. Figure 3.10 is a visual demonstration of how eccentricity is generally calculated. Only the regions with an eccentricity measure (0-0.98) are conserved as candidate regions for crab detection. One may note that the selected range is very inclusive, to account for incomplete detections of crabs due to partial occlusion or low contrast. These steps are presented in Figures 3.12, 3.11 and 3.13 for each color space.

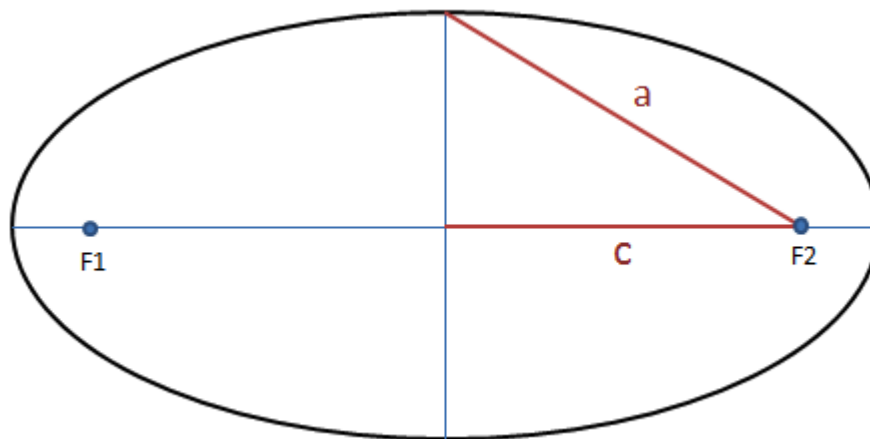


Figure 3.10: Eccentricity =  $\frac{c}{a}$ .

At this stage, the performance of detecting crab and non-crab regions by color-based thresholding was evaluated for every color space. This evaluation determined the optimal color space for locating potential crab regions. At this point recall is of most interest, because a high recall means fewer missed detections. Reducing the false detections is addressed in the classification module. As shown in Table 3.4, the HSV space produces the best results. Therefore, the following modules work with images resulting from HSV color space thresholding.

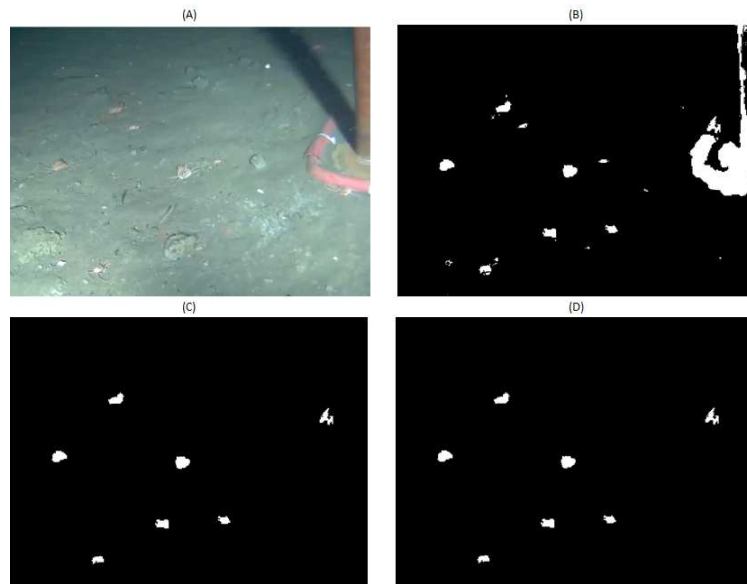


Figure 3.11: Images illustrating the segmentation process in the RGB color space. A) Original image. B) Binary image output by color-based thresholding. C) Image after the elimination of regions with undesirable area range. D) Image after the elimination of regions with undesirable eccentricity values.



Figure 3.12: Images illustrating the segmentation process in the HSV color space. A) Original image. B) Binary image output by color-based thresholding. C) Image after the elimination of regions with undesirable area range. D) Image after the elimination of regions with undesirable eccentricity values.

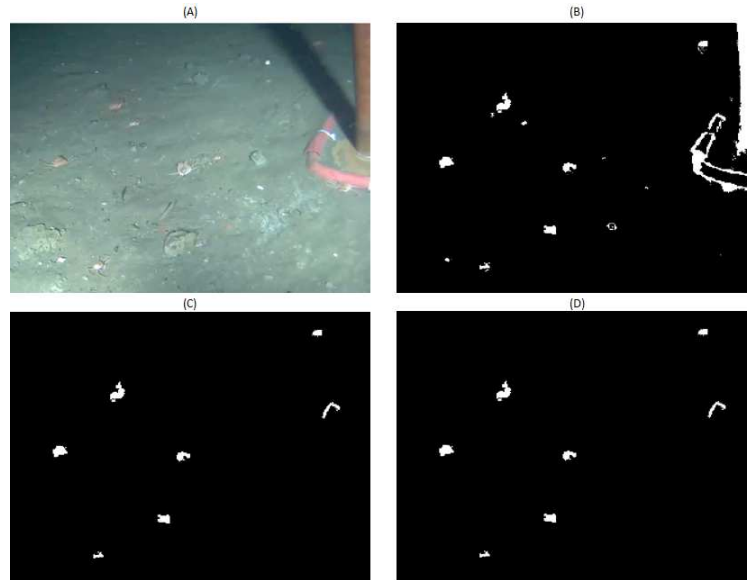


Figure 3.13: Images illustrating the segmentation process in the YIQ color space. A) Original image. B) Binary image output by color-based thresholding. C) Image after the elimination of regions with undesirable area range. D) Image after the elimination of regions with undesirable eccentricity values.

Table 3.4: Evaluation results of the three color spaces.

Color Space	Precision	Recall	F-score
HSV	0.205	0.705	0.317
RGB	0.113	0.645	0.192
YIQ	0.119	0.596	0.198

### 3.3 Module C - Feature Extraction

Since the candidate regions are mostly circular and the color-based segmentation module isn't able to preserve the legs of the crabs (due to the low quality of videos and also to the distance of the camera from the sea bed), local edge detection is performed on the contrast enhanced images (the output of the CLAHE method, e.g. Figure 3.3D) around the candidate regions resulting from the segmentation module. At this point, the image resulting from local edge detection and the segmented image from module B (Figure 3.13D) are overlaid (Figure 3.14C) and the connected components which do not belong to the crab size range are eliminated (Figure 3.14D). This process enables the recovery of the legs of the crabs.

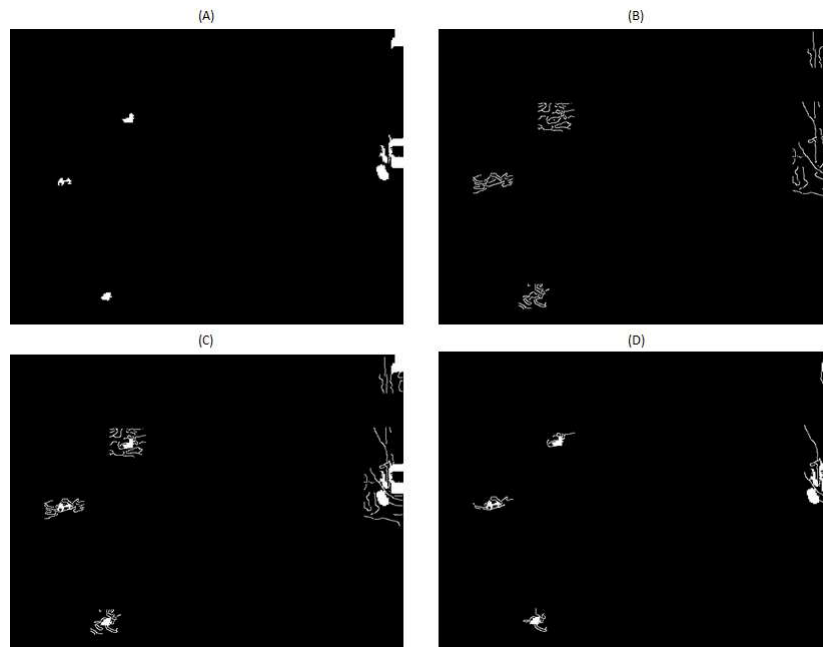


Figure 3.14: A) Binary image resulted from the segmentation module. B) Binary image output by local edge detection around the ROIs of (A) performed on contrast enhanced image (Figure 3.3). C) Image resulted from overlaying (A) and (B). D) Image after the elimination of regions with undesirable area values.

The crabs in our database appear in different sizes and orientations. Therefore, the candidate regions, need to be invariant to scale and translation. Here, we applied Khotanzad and Hong's approach [20] in order to form regions that are invariant to scale and translation. As recommended by Khotanzad and Hong [20], each region  $f(x, y)$  is initially normalized by its regular moments with the following process:

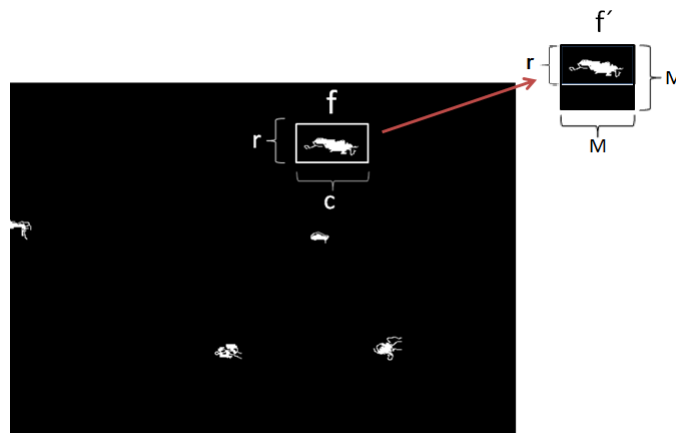


Figure 3.15: A binary image after performing segmentation and feature extraction process. Each one of the segmented binary regions  $f$  are transferred onto an  $M \times M$  square region  $f'(x, y)$ . In this case  $r \leq c$  and therefore  $M = c$ .

1. Each region  $f(x, y)$  is transferred onto an  $M \times M$  square region  $f'(x, y)$  where

$$M = \max[r, c] \quad (3.2)$$

$r$  and  $c$  are the height and width of  $f$ , respectively (Figure 3.15).

2. For consistency in various sizes  $(i, j) \in [1, M]$  are mapped onto  $(x, y) \in [-1, +1]$ .
3. The regular moments  $m_{pq}$  are calculated by:

$$m_{pq} = \sum_{x=-1}^{x=+1} \sum_{y=-1}^{y=+1} x^p y^q f(x, y), \quad p, q = 0, 1, 2, \dots \quad (3.3)$$

4. The region is normalized to form  $g(x, y)$ :

$$g(x, y) = f'\left(\bar{x} + \frac{x}{a}, \bar{y} + \frac{y}{a}\right) \quad (3.4)$$

where  $\bar{x}$  and  $\bar{y}$  are the centroids of region  $f'$ :

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (3.5)$$

and

$$a = \sqrt{\frac{B}{m_{00}}} \quad (3.6)$$

where  $B$  is the number of object pixels in the region.

The resulting region  $g(x, y)$  is invariant to scale and translation. Since the largest ROI was  $224 \times 224$ , the binary images corresponding to each region are all scaled to be stored as  $224 \times 224$  matrices. These matrices are reshaped as single column vectors (See Figure 3.16) in order to form input feature vectors for the classifiers described in the next module.

## 3.4 Classification

As presented in Table 3.4, there are few missed detections and many false detections after the segmentation process. In order to eliminate these false detections and further

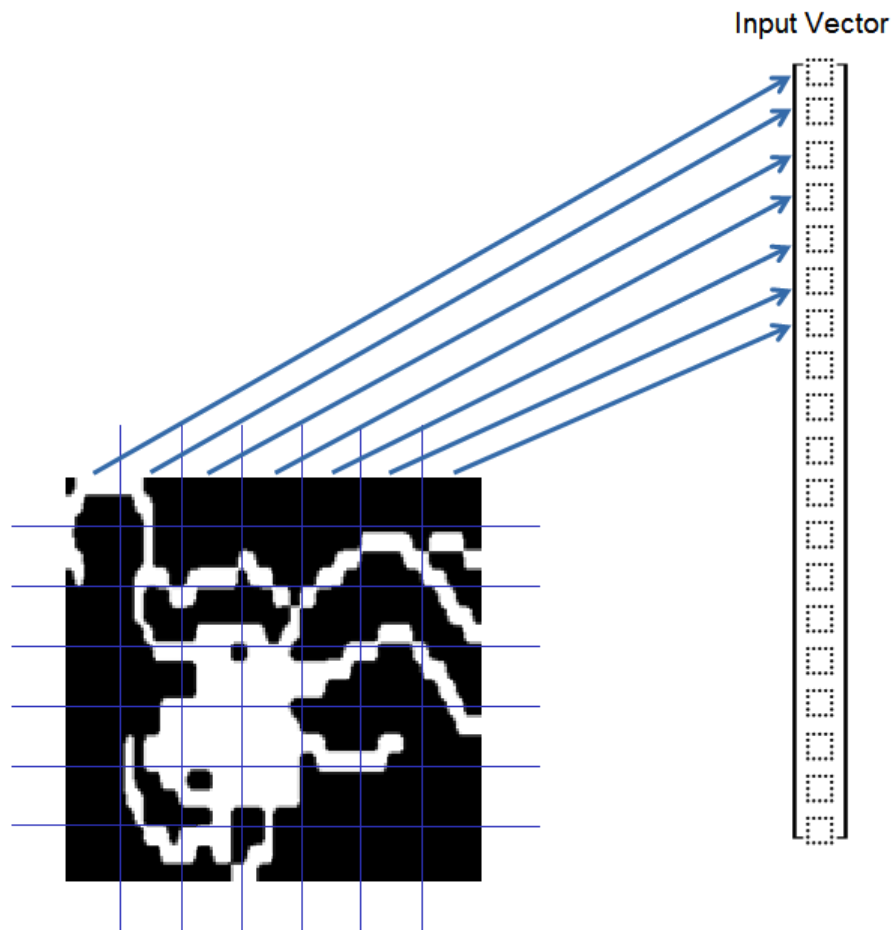


Figure 3.16: Feature vector structure. The process in which each ROI's matrix is reshaped to form a feature vector.

classify the ROIs, machine learning algorithms are applied. In order to apply machine learning algorithms, a database of crab and non-crab regions is formed by the manual classification of the candidate regions into crabs and non-crabs. The remainder of this chapter describes the process in which these classifiers are developed.

### 3.4.1 Feed-forward Neural Network

As stated by Hornik [17] feed-forward neural networks (Figure 3.17) may be considered as universal identifiers. In other words, these neural networks can be applied in many identification tasks. We will refer to the feed-forward neural networks as NNs in this thesis. One hidden layer is usually sufficient for classification tasks. However, the number of neurons in the hidden layer varies based on the data. This section explains the method applied to determine a suitable sized NN for the classification of the candidate regions and details the steps taken for training the classifier NNs.

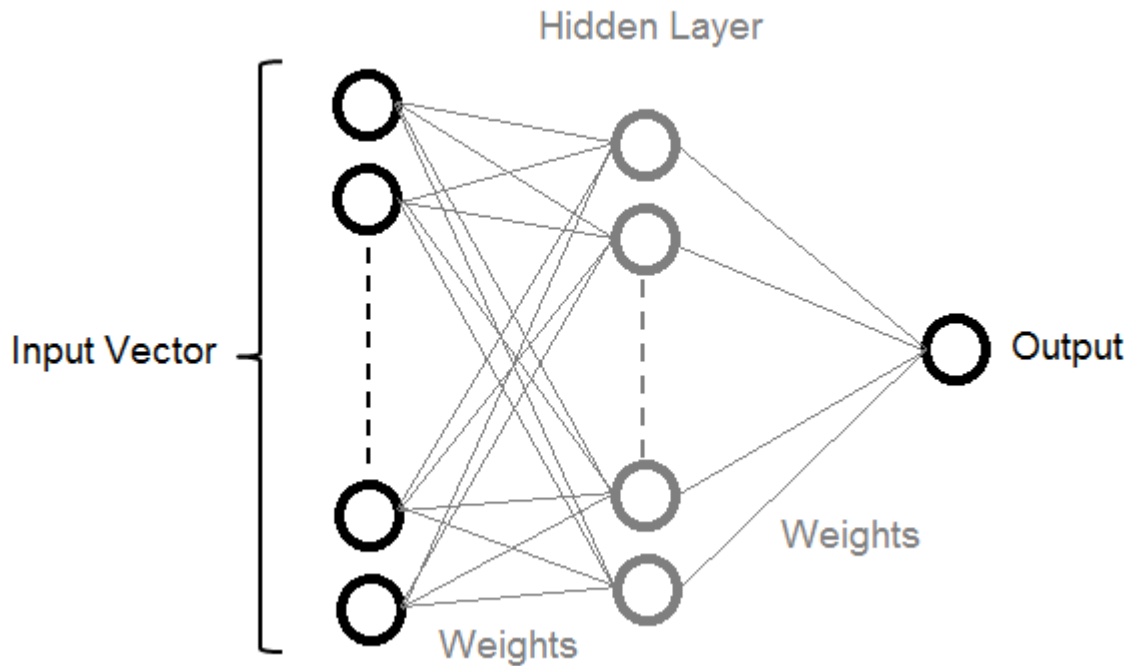


Figure 3.17: Feed-forward neural network.

### Determination of the optimal size of the neural network

In order to determine the suitable number of neurons in the hidden layer of the NN, the guided selection mechanism [22] was adopted:

1. The crab and non-crab vectors are used to form a training set of 380 and a testing sets of 150 images. The training set is used to train a set of 250 NNs of sizes 1 to 24. (e.g. 250 NNs with 1 neuron in the hidden layer, 250 NNs with 2 neurons in the hidden layer, etc.)
2. The mean error (ME) of each NN in classifying the test set is calculated as follows:

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_i - T_i) \quad (3.7)$$

where  $n$  is the number of images (150) and  $Y_i$  and  $T_i$  are the predicted and observed classes, respectively.

3. The mean ME is calculated for each set of 250 NNs (same size).
4. The mean ME is used to determine the suitable size of NN for our data. The NN size that has a generally low ME is optimal, but the fact that a large NN size will result in over fitting has to be taken into consideration.

### Training the neural network

Feature vectors corresponding to crab and non-crab regions are used to form various training and test sets in order to train and evaluate classifier NNs. More specifically these vectors are used to form 30 datasets that each comprise a training set and a test set. Training and testing of classifiers is generally performed such that a third of the data is used for testing and the remainder is applied in training the classifier [32]. For each dataset, the training set contains 372 samples, while the testing set contains 150 samples. The training sets are used to train 500 NNs. For each data set the following process is performed:

1. 500 NNs are trained using the training set. The initial weights that are assigned to each NN are chosen randomly, which makes each NN's classification performance different.

2. Each one of the NNs is evaluated by calculating its error in the classification of the corresponding test set.
3. The NN with the highest F-score in classifying the test set is chosen as the optimal classifier for its corresponding data set.

### 3.4.2 Self Organizing Maps

The Self-Organizing Map (SOM) proposed by Kohonen [21] is one of the most popular unsupervised learning algorithms among neural network models. The SOM is widely applied for the classification of large datasets with high dimensions. The training phase results in clustering the samples on the grid, such that the similar samples are close to each other, while the dissimilar ones are distant.

A dataset composed of vectors representing crab and non-crab regions is fed to 500 SOMs; which are all assigned with random initial weights (Figure 3.18). The variance in their initial weights, differentiates their performance. Through the training process of each SOM, the vectors are separated and similar vectors are placed in the same class. The SOM with the best classification ability among the 500 SOMs is chosen as the optimum classifier SOM.

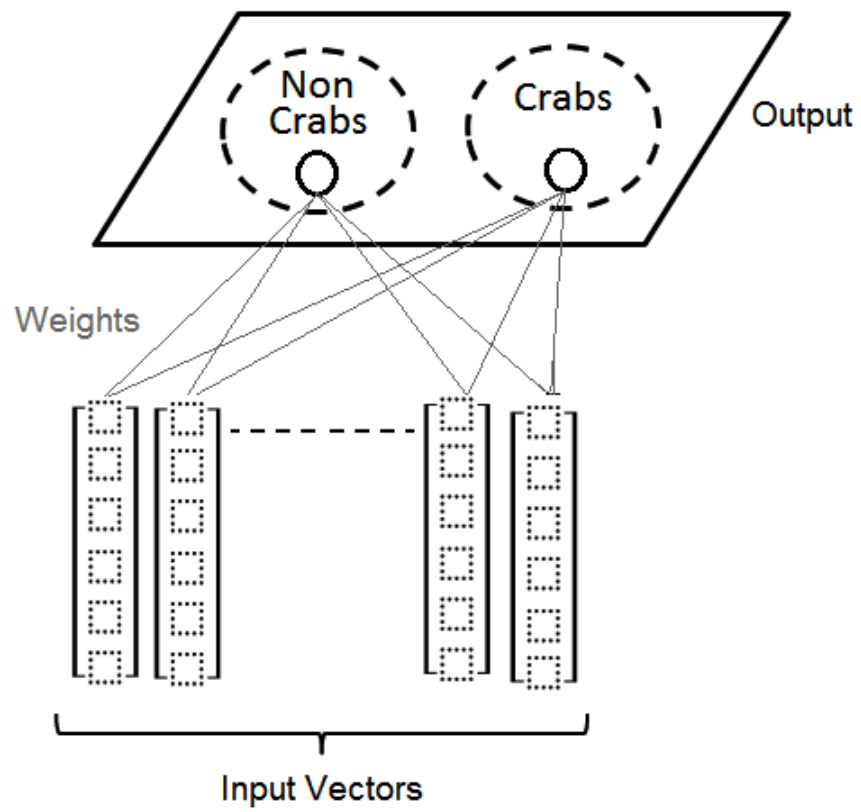


Figure 3.18: Self Organizing Map.

# Chapter 4

## Experimental Results

In the present chapter, the performance of the proposed approach is evaluated by calculation of the precision, recall and F-score of the classification module. The algorithms were implemented using MATLAB's image processing and neural network toolboxes [25].

### 4.1 Performance Evaluation

In order to assess the performance of the proposed approach's segmentation and classification modules, a human user manually evaluated the images and determined whether the detected regions in the proposed approach correspond to crabs or non-crabs. Sample output images of the segmentation module, that were used for validation are presented in Figure 4.1.

The evaluation criteria are as follows:

$$Precision = \frac{correct\ detections}{correct\ detections + false\ detections} \quad (4.1)$$

$$Recall = \frac{correct\ detections}{correct\ detections + missed\ detections} \quad (4.2)$$

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4.3)$$

where:

- **False detections** correspond to cases where the algorithm indicates that a region of interest (ROI) is a crab, but the human operator considers that ROI

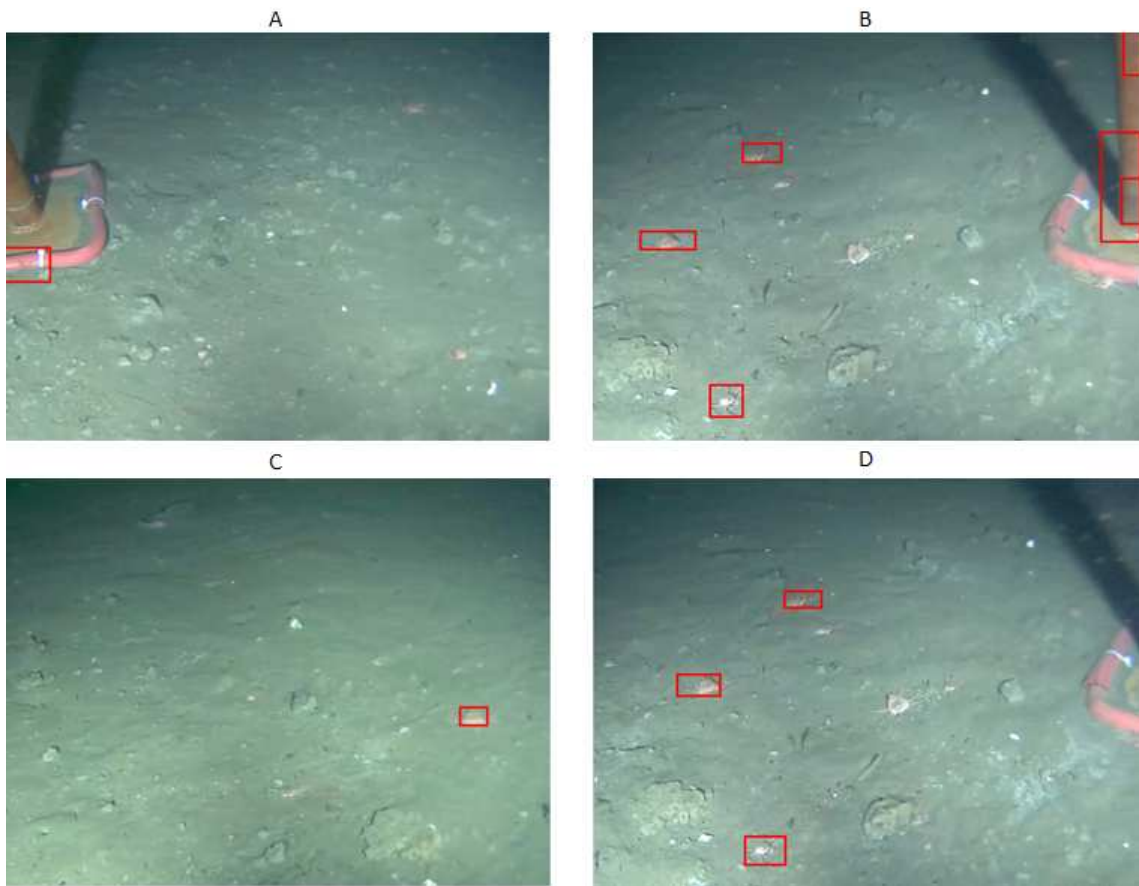


Figure 4.1: Intermediate results obtained after the segmentation module.

to be a non-crab.

- **Correct detections** are ROIs that are classified as crabs by both the algorithm and by the human observer.
- **Missed detections** are ROIs that the manual classification indicates to be crabs but the algorithm considers as non-crabs.

A high recall would indicate that a large portion of the crabs have been detected, whereas high precision means that most of the ROIs that have been classified as crabs, were indeed crabs.

The F-score is a measure of the classification's accuracy which considers both precision and recall.

## 4.2 Database

Our experimental database is composed of still images of the sea floor extracted from three 5 minute-long videos with  $480 \times 640$  spatial resolution and recorded at 30 fps with a stationary sea floor-mounted camera at the Barkley Canyon location on the NEPTUNE observatory (Ocean Networks Canada) [1]. The training image set consists of 100 images extracted from one video. The image patches which were used to create the colour histograms and determine the thresholds for the colour-based segmentation were acquired from the same training set. The test image set consists of 200 images extracted from the two remaining videos. The test image set was used for evaluating the overall performance of the classifier, as well as for measuring intermediate results such as the performance of the segmentation module.

## 4.3 Discussion on Segmentation and Classification Results

Segmentation-specific results are presented in Table 3.4. Missed detections in the segmentation step occur mostly when the crab merges with an adjacent non-crab region with similar colour; the size of the detected region is considered too large for a crab candidate (Figure 4.1B). A similar issue occurs when two crabs are located next to each other. False detections are more frequent than missed detections and they are due to detecting non-crab regions of similar color to crabs (Figure 4.1A). However, false detections are not of much concern, since the subsequent classification module is designed to separate the crab from the non-crab regions identified via segmentation.

The classification performance of a supervised NN and an unsupervised SOM was evaluated. In order to apply these classifiers, input feature vectors had to be formed from the binary ROIs (image patches) that resulted from the segmentation module. The binary ROIs formed in the feature extraction module (described in Figure 3.15) were manually classified as crab and non-crab patches. The largest ROI was  $224 \times 224$ , therefore all ROIs were scaled to  $224 \times 224$  square matrices composed of zeros and ones. These elements were reshaped to a  $1 \times 50176$  vector (Figure 3.16) and consequently normalized in the  $[-1,+1]$  range.

An NN's architecture is based on the number of neurons in the hidden layer/s. A suitable size for an NN depends on the data, therefore a technique was used to

determine this parameter for our data. Figure 4.2 summarizes our experiments for determining the optimal size of the network. The number of neurons in the hidden layer varied from 1 to 24. As it is shown in the plot, the error remains the same between 13 and 15 before increasing again. Considering the fact that larger NNs with lower errors will result in over-fitting the classifier; the optimal size number of neurons was found to be 14.

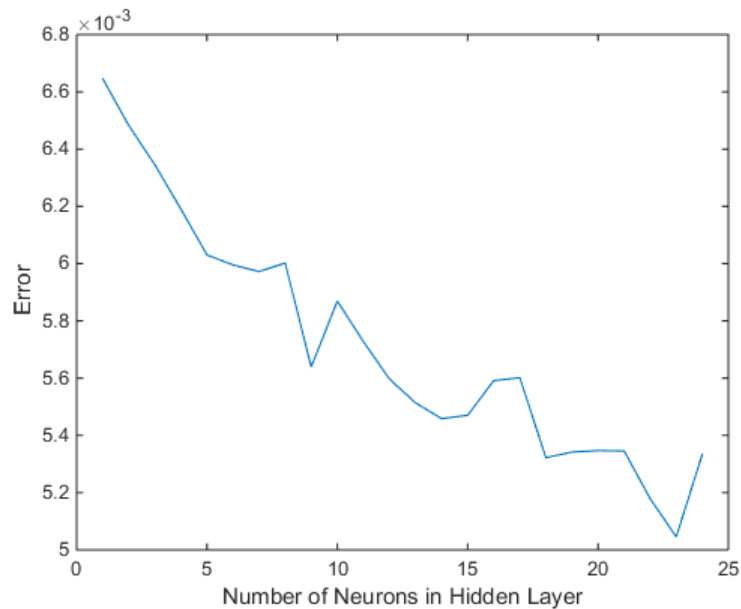


Figure 4.2: The mean  $ME$  and the neural network size.

Table 4.1 presents the classification performance of the NNs for each of the 30 randomly formed data sets described in the classification module of chapter 3. The NNs were trained by 372 training vectors and evaluated by 150 testing vectors. One may note that the performance of the classifiers trained and tested with different data sets varies significantly. This suggests that the samples used in training the NN play a critical role in its performance. The best classification has been performed on data set 17. The results also show that the feed-forward neural networks generally have a higher recall and lower precision rate. The high number of false detections correspond to non-crab regions that have a crab-like appearance (a circular blob with several lines attached to it).

The testing data in each one of the 30 data sets described earlier, were classified by Self-Organizing Maps (SOMs). Table 4.2 presents the classification results. These

Dataset	Precision	Recall	F-score
1	0.544118	0.755102	0.632479
2	0.618182	0.772727	0.686869
3	0.666667	0.711111	0.688172
4	0.543478	0.694444	0.609756
5	0.522388	0.795455	0.630631
6	0.557377	0.829268	0.666667
7	0.641509	0.666667	0.653846
8	0.584906	0.775000	0.666667
9	0.537313	0.837209	0.654545
10	0.651163	0.682927	0.666667
11	0.540541	0.869565	0.666667
12	0.576923	0.714286	0.638298
13	0.506667	0.926829	0.655172
14	0.634146	0.590909	0.611765
15	0.500000	0.857143	0.631579
16	0.568627	0.725000	0.637363
17	0.606557	0.840909	0.704762
18	0.562500	0.710526	0.627907
19	0.538462	0.795455	0.642202
20	0.566667	0.739130	0.641509
21	0.494624	0.867925	0.630137
22	0.532468	0.872340	0.661290
23	0.568966	0.868421	0.687500
24	0.591837	0.743590	0.659091
25	0.576271	0.739130	0.647619
26	0.609375	0.795918	0.690265
27	0.615385	0.744186	0.673684
28	0.645833	0.704545	0.673913
29	0.507463	0.850000	0.635514
30	0.483871	0.769231	0.594059
Average	0.569809	0.774832	0.652220

Table 4.1: Evaluation results of the Feed-Forward Neural Network.

classifiers have a similar classification behavior to the NNs in their results (i.e. high recall and lower precision). The best classification has been performed on data set 22.

The NNs show a better classification performance compared to the SOMs. In an attempt to improve the classification results, cascade classification was applied. Cascade classification consists of applying back to back classifiers in order to achieve

Dataset	Precision	Recall	F-score
1	0.285714	0.571429	0.380952
2	0.452381	0.863636	0.593750
3	0.296296	0.711111	0.418301
4	0.410959	0.833333	0.550459
5	0.447761	0.681818	0.540541
6	0.450000	0.878049	0.595041
7	0.453333	0.666667	0.539683
8	0.490909	0.675000	0.568421
9	0.250000	0.604651	0.353741
10	0.461538	0.731707	0.566038
11	0.439024	0.782609	0.562500
12	0.436364	0.571429	0.494845
13	0.440000	0.804878	0.568966
14	0.247525	0.568182	0.344828
15	0.492308	0.761905	0.598131
16	0.442857	0.775000	0.563636
17	0.452055	0.750000	0.564103
18	0.382716	0.815789	0.521008
19	0.477612	0.727273	0.576577
20	0.430556	0.673913	0.525424
21	0.500000	0.603774	0.547009
22	0.520548	0.808511	0.633333
23	0.437500	0.736842	0.549020
24	0.454545	0.769231	0.571429
25	0.485714	0.739130	0.586207
26	0.546875	0.714286	0.619469
27	0.507692	0.767442	0.611111
28	0.506849	0.840909	0.632479
29	0.460526	0.875000	0.603448
30	0.384615	0.897436	0.538462
Average	0.434826	0.740031	0.543964

Table 4.2: Evaluation results of the Self-Organizing Maps.

better classification results. Cascade classification was used in order to decrease the false detections. In other words, the second classifier was an attempt to further classify the regions that were assumed to be crabs by the first classifier (See Figure 4.3). In this phase, the supervised and non-supervised classifiers were cascaded in two ways, as follows:

- Feed-Forward Neural Networks followed by Self-Organizing Maps

- Self-Organizing Maps followed by Feed-Forward Neural Networks

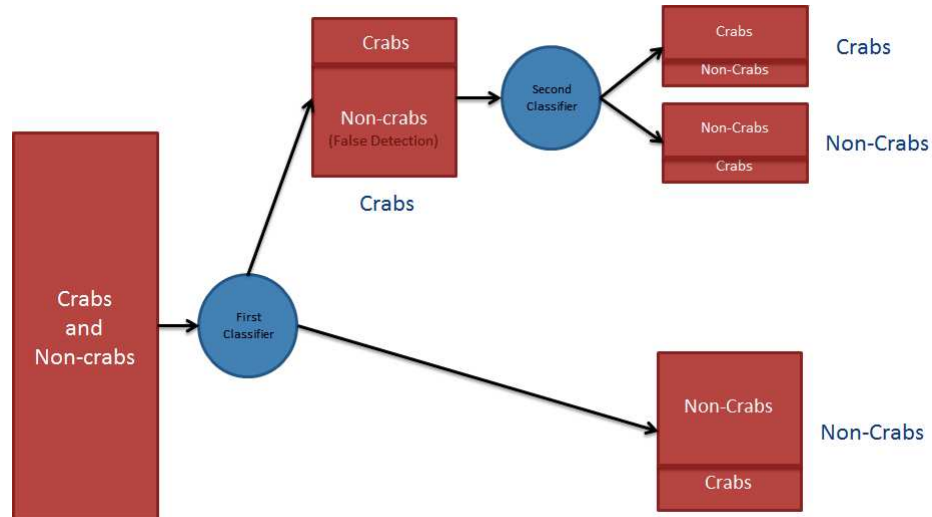


Figure 4.3: Cascade Classification. The first classifier recognizes a significant number of non-crabs as crabs (false detection) The second classifier attempts to identify these false detections and separate them from the crab class.

The ROIs that were classified as crabs by the first classifier were passed on to the second classifier for further classification. The ROIs that were determined as non-crabs by the first classifier and also the classified ROIs by the second classifier were combined to form the final classification results that are presented in tables 4.3 and 4.4. Based on the obtained results we may conclude that cascade classifiers do not improve the classification performance and the NNs prove to be the best classifiers for our task.

Dataset	Precision	Recall	F-score
1	0.553191	0.530612	0.541667
2	0.604651	0.590909	0.597701
3	0.800000	0.622222	0.700000
4	0.555556	0.694444	0.617284
5	0.562500	0.613636	0.586957
6	0.651163	0.682927	0.666667
7	0.674419	0.568627	0.617021
8	0.620000	0.775000	0.688889
9	0.586207	0.790698	0.673267
10	0.666667	0.682927	0.674699
11	0.508197	0.673913	0.579439
12	0.600000	0.500000	0.545455
13	0.500000	0.902439	0.643478
14	0.709677	0.500000	0.586667
15	0.519231	0.642857	0.574468
16	0.580000	0.725000	0.644444
17	0.682927	0.636364	0.658824
18	0.553191	0.684211	0.611765
19	0.538462	0.636364	0.583333
20	0.657143	0.500000	0.567901
21	0.519481	0.754717	0.615385
22	0.594595	0.468085	0.523810
23	0.600000	0.789474	0.681818
24	0.666667	0.717949	0.691358
25	0.659091	0.630435	0.644444
26	0.585366	0.489796	0.533333
27	0.651163	0.651163	0.651163
28	0.675676	0.568182	0.617284
29	0.550000	0.550000	0.550000
30	0.491228	0.717949	0.583333
Average	0.603882	0.643030	0.615062

Table 4.3: Evaluation results of the cascade classification of Feed-Forward Neural Networks followed by Self-Organizing Maps.

Dataset	Precision	Recall	F-score
1	0.633333	0.387755	0.481013
2	0.632653	0.704545	0.666667
3	0.781250	0.555556	0.649351
4	0.594595	0.611111	0.602740
5	0.520833	0.568182	0.543478
6	0.574074	0.756098	0.652632
7	0.642857	0.529412	0.580645
8	0.615385	0.600000	0.607595
9	0.621622	0.534884	0.575000
10	0.676471	0.560976	0.613333
11	0.566667	0.739130	0.641509
12	0.606061	0.476190	0.533333
13	0.533333	0.780488	0.633663
14	0.636364	0.318182	0.424242
15	0.551724	0.761905	0.640000
16	0.641026	0.625000	0.632911
17	0.617021	0.659091	0.637363
18	0.589744	0.605263	0.597403
19	0.555556	0.681818	0.612245
20	0.609756	0.543478	0.574713
21	0.538462	0.528302	0.533333
22	0.553846	0.765957	0.642857
23	0.600000	0.710526	0.650602
24	0.585366	0.615385	0.600000
25	0.576923	0.652174	0.612245
26	0.614035	0.714286	0.660377
27	0.604167	0.674419	0.637363
28	0.707317	0.659091	0.682353
29	0.543860	0.775000	0.639175
30	0.482759	0.717949	0.577320
Average	0.600235	0.627072	0.604515

Table 4.4: Evaluation results of the cascade classification of Self-Organizing Maps followed by Feed-Forward Neural Networks.

# Chapter 5

## Conclusions

This work aims at automating the detection of stationary grooved crabs of various sizes in underwater videos. High volumes of imagery data are recorded from sea-life, but it's almost impossible for researchers to manually inspect all these data.

One of the main challenges in analysing underwater imagery is low contrast and color issues which are caused by the physical properties of the underwater medium and artificial lighting. These problems were addressed by performing a sequence of pre-processing techniques. The pre-processing steps decreased the missed detections which were due to insufficient lighting and also improved the appearance of the legs of the crabs by contrast enhancement. The detection of crab regions considers color, shape and size properties. The segmentation performance in the YIQ, RGB and HSV color spaces was examined and the optimal color space for color-based segmentation was found to be the HSV color space. At this stage, a high portion of the crabs were correctly detected, while a large number of non-crabs were also detected as crabs. These false detections were addressed by supervised and unsupervised classifiers. In order to apply classification, the candidate regions resulting from segmentation were described by feature vectors. Local edge detection was performed on the contrast enhanced images, around the candidate regions resulting from segmentation. The resulting image has a better representation of the legs of the crabs. The candidate regions in the image (crabs and non-crabs) are in various sizes, therefore they were all normalized by their regular moments to be scale and translation invariant. The normalized regions were reshaped to form input feature vectors. Both classifiers have a higher recall than precision. This means that most of the crabs are identified; however there are some false detections. In addition, it's clear that the classifier's performance varies based on the training set. In an attempt to improve the precision

of the classifiers, they were cascaded, but cascade classification did not improve the overall classification performance. The optimal classifier was found to be feed-forward neural networks.

One of the most important properties of a trained classifier is its generalization capability; the classifier's ability to classify a wide range of samples that have not been used in the training process. Future work will focus on the following:

- Improving the generalization capability of the classification. This can be explored by applying Convolutional Neural Networks, which have proved to be highly successful in recent image classification tasks.
- Exploring computer vision techniques that improve the segmentation module and minimize the number of missed detections.
- Develop a more efficient method to determine the optimal samples to be included in the training set.



# Glossary

**ACE** Automatic Color Enhancement.

**CLAHE** Contrast Limited Adaptive Histogram Equalization.

**HSV** Hue, Saturation, Value color space.

**NEPTUNE** North-East Pacific Time-series Undersea Networked Experiments.

**NN** feed-forward Neural Network.

**ONC** Ocean Networks Canada.

**RGB** Red, Green, Blue color space.

**ROI** Region Of Interest.

**SOM** Self Organizing Maps.

**SVM** Support Vector Machines.

**VENUS** Victoria Experimental Network Under the Sea.

**YIQ** Luminance (Y) In-phase, Quadrature color space (NTSC color space).

# Bibliography

- [1] Ocean networks canada @ONLINE. <http://www.oceannetworks.ca/>, 2015.
- [2] Alberto Albiol, Luis Torres, and Edward J Delp. Optimum color spaces for skin detection. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 1, pages 122–124. IEEE, 2001.
- [3] Cosmin Ancuti, Codruta Orniana Ancuti, Tom Haber, and Philippe Bekaert. Enhancing underwater images and videos by fusion. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 81–88. IEEE, 2012.
- [4] Álvaro Bayona, Juan Carlos SanMiguel, and José M Martínez. Comparative evaluation of stationary foreground object detection algorithms based on background subtraction techniques. In *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*, pages 25–30. IEEE, 2009.
- [5] Stephane Bazeille, Isabelle Quidu, Luc Jaulin, Jean-Philippe Malkasse, et al. Automatic underwater image pre-processing. *Proceedings of CMM'06*, 2006.
- [6] Michael D Beynon, Daniel J Van Hook, Michael Seibert, Alen Peacock, and Dan Dudgeon. Detecting abandoned packages in a multi-camera video surveillance system. In *Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, 2003.*, pages 221–228. IEEE, 2003.
- [7] Bastiaan J Boom, Jiyin He, Simone Palazzo, Phoenix X Huang, Cigdem Beyan, Hsiu-Mei Chou, Fang-Pang Lin, Concetto Spampinato, and Robert B Fisher. A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage. *Ecological Informatics*, 2013.

- [8] Bastiaan J Boom, Phoenix X Huang, Cigdem Beyan, Concetto Spampinato, Simone Palazzo, Jiyin He, Emmanuelle Beauxis-Aussalet, Sun-In Lin, Hsiu-Mei Chou, Gayathri Nadarajan, et al. Long-term underwater camera surveillance for monitoring and analysis of fish populations. *VAIB12*, 2012.
- [9] Jason Brand and John S Mason. A comparative assessment of three approaches to pixel-level human skin-detection. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 1, pages 1056–1059. IEEE, 2000.
- [10] Majed Chambah, Bernard Besserer, and Pierre Courtellemont. Recent progress in automatic digital restoration of color motion pictures. In *Electronic Imaging 2002*, pages 98–109. International Society for Optics and Photonics, 2001.
- [11] Majed Chambah, Dahbia Semani, Arnaud Renouf, Pierre Courtellemont, and Alessandro Rizzi. Underwater color constancy: enhancement of automatic live fish recognition. In *Electronic Imaging 2004*, pages 157–168. International Society for Optics and Photonics, 2003.
- [12] Alan D Chave, Gene Massion, Hitoshi Mikada, et al. Science requirements and the design of cabled ocean observatories. 2006.
- [13] Ahmed Elgammal, Crystal Muang, and Dunxu Hu. Skin detection-a short tutorial. *Encyclopedia of Biometrics*, 2009.
- [14] Ryan Eustice, Oscar Pizarro, Hanumant Singh, and Jonathan Howland. Uwit: Underwater image toolbox for optical image processing and mosaicking in matlab. In *Underwater Technology, 2002. Proceedings of the 2002 International Symposium on*, pages 141–145. IEEE, 2002.
- [15] Ryan Fier, Alexandra Branzan Albu, and Maia Hoeberechts. Automatic fish counting system for noisy deep-sea videos. In *Oceans-St. John's, 2014*, pages 1–6. IEEE, 2014.
- [16] Rafael García Campos, Tudor Nicosevici, and Xavier Cufí i Solé. On the way to solve lighting problems in underwater imaging. © *OCEANS'02 MTS/IEEE, 2002, vol. 2, p. 1018-1024*, 2002.
- [17] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

- [18] Qiao Hu and Cabell Davis. Automatic plankton image recognition with co-occurrence matrices and support vector machine. *Marine Ecology Progress Series*, 295:21–31, 2005.
- [19] Phoenix X Huang, Bastiaan J Boom, and Robert B Fisher. Underwater live fish recognition using a balance-guaranteed optimized tree. In *Computer Vision–ACCV 2012*, pages 422–433. Springer, 2013.
- [20] Alireza Khotanzad and Yaw Hua Hong. Invariant image recognition by zernike moments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(5):489–497, 1990.
- [21] Teuvo Kohonen and Self-Organizing Maps. vol. 30 of springer series in information sciences, 1995.
- [22] Erik M Laxdal, Rafael Parra-Hernandez, and Nikitas J Dimopoulos. Guided construction of training data set for neural networks. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 6, pages 5905–5910. IEEE, 2004.
- [23] Liyuan Li, Weimin Huang, Irene Yu-Hua Gu, and Qi Tian. Statistical modeling of complex backgrounds for foreground object detection. *Image Processing, IEEE Transactions on*, 13(11):1459–1472, 2004.
- [24] JA Lines, RD Tillet, LG Ross, D Chan, S Hockaday, and NJB McFarlane. An automatic image-based system for estimating the mass of free-swimming fish. *Computers and Electronics in Agriculture*, 31(2):151–168, 2001.
- [25] MATLAB. *version 8.4.0 (R2014b)*. The MathWorks Inc., Natick, Massachusetts, 2014.
- [26] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.
- [27] Schettini Raimondo and Corchs Silvia. Underwater image processing: state of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing*, 2010, 2010.

- [28] Alessandro Rizzi, Carlo Gatta, and Daniele Marini. A new algorithm for unsupervised global and local color correction. *Pattern Recognition Letters*, 24(11):1663–1677, 2003.
- [29] Andrew Rova, Greg Mori, and Lawrence M Dill. One fish, two fish, butterflyfish, trumpeter: Recognizing fish in underwater video. In *MVA*, pages 404–407, 2007.
- [30] Timm Schoening, Melanie Bergmann, Jörg Ontrup, James Taylor, Jennifer Dannheim, Julian Gutt, Autun Purser, and Tim W Nattkemper. Semi-automated image analysis for the assessment of megafaunal densities at the arctic deep-sea observatory hausgarten. *PloS one*, 7(6):e38179, 2012.
- [31] Mark R Shortis, Mehdi Ravanbakskh, Faisal Shaifat, Euan S Harvey, Ajmal Mian, James W Seager, Philip F Culverhouse, Danelle E Cline, and Duane R Edgington. A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences. In *SPIE Optical Metrology 2013*, pages 87910G–87910G. International Society for Optics and Photonics, 2013.
- [32] Concetto Spampinato, Emmanuelle Beauxis-Aussalet, Simone Palazzo, Cigdem Beyan, Jacco van Ossenbruggen, Jiyin He, Bas Boom, and Xuan Huang. A rule-based event detection system for real-life underwater domain. *Machine vision and applications*, 25(1):99–117, 2014.
- [33] Concetto Spampinato, Yun-Heh Chen-Burger, Gayathri Nadarajan, and Robert B Fisher. Detecting, tracking and counting fish in low quality unconstrained underwater videos. In *VISAPP (2)*, pages 514–519. Citeseer, 2008.
- [34] Concetto Spampinato, Daniela Giordano, Roberto Di Salvo, Yun-Heh Jessica Chen-Burger, Robert Bob Fisher, and Gayathri Nadarajan. Automatic fish classification for underwater species behavior understanding. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams*, pages 45–50. ACM, 2010.
- [35] Robin Tillet, Nigel McFarlane, and Jeff Lines. Estimating dimensions of free-swimming fish using 3d point distribution models. *Computer Vision and Image Understanding*, 79(1):123–141, 2000.

- [36] Ce Wang and Michael S Brandstein. Multi-source face tracking with audio and visual data. In *Multimedia Signal Processing, 1999 IEEE 3rd Workshop on*, pages 169–174. IEEE, 1999.
- [37] Karel Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc., 1994.