Automatic Detection of benthos & birds

Microphytobenthos cover and bird number detection on the Galgeplaat mudflat using terrestrial imagery

Penelope Rammos
September 2012
Cover illustration: image captured by one of the Argus cameras from the Argus Bio Platform positioned on the Galgeplaat mudflat in the Oosterschelde.
Automatic Detection of benthos & birds
Microphytobenthos cover and bird number detection on the Galgeplaat mudflat using terrestrial imagery

MASTER OF SCIENCE THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Geomatics at Delft University of Technology

Penelope Rammos

September 13, 2012

Faculty of Civil Engineering & Geosciences - Delft University of Technology
The thesis work was achieved with support from Deltares. Their cooperation is hereby gratefully acknowledged.
Delft University of Technology
Department of
Geoscience & Remote Sensing

The undersigned hereby certify that they have read and recommend to the Faculty of
Civil Engineering & Geosciences for acceptance a thesis entitled

**Automatic Detection of benthos & birds**

by

Penelope Rammos

in partial fulfillment of the requirements for the degree of

**Master of Science Geomatics**

Dated: September 13, 2012

Supervisor(s):

Prof.dr. M. Menenti (TU Delft)

Dr. R.C. Lindenbergh (TU Delft)

Dr.ir. J. Dijkstra (Deltares)

Reader(s):

Dr. E.A. Hendriks (TU Delft)
Ecological monitoring, i.e. the process of assessing the quality and health of natural habitats, is required to assess the impact of anthropogenic influence. This process is often inhibited by the expenses involved and a limited accessibility to the study site. The use of remotely sensed data then logically comes to mind as a potential solution.

In this thesis the focus is on the ecological monitoring of an intertidal mudflat located in the Oosterschelde known as the Galgeplaat. It possesses a monitoring platform standing 15 meters tall producing imagery that was previously used for monitoring the morphology of the mudflat. The goal is to examine the potential of using this available terrestrial imagery for ecological monitoring of the mudflat and whether it is possible to do this automatically.

Two separate case studies were formulated to investigate this potential: 1. automatic detection of microphytobenthos and 2. automatic detection of bird numbers. The primary focus in this thesis was put on the microphytobenthos case study which elicited the most interest from involved parties.

For the detection of microphytobenthos cover the multispectral camera from the platform was used. The main inhibiting factor was the presence of macroalgae in the images, which possesses similar spectral properties to that of microphytobenthos. Two methods were used to detect microphytobenthos: I. maximum likelihood classification combined with the masking of the macroalgae (the undesired target) and II. Kohonen’s self organising maps (SOM). The results of this case study indicated that distinguishment between microphytobenthos and macroalgae was best achieved with the Self organizing map (SOM) approach. The SOM approach was best suited to the identification of microphytobenthos (the desired target) but its potential use for only macroalgae identification may be considered excessive as the maximum likelihood approach (smaller complexity) produced similar results for this target.

For the detection of bird numbers the Bio pan/tilt/zoom camera on the platform was used. Consecutive snapshot images of the camera were used such that the motion of birds could be
taken advantage of. Background subtraction using a weighted mean background image and a standard deviation image was the most promising of the methods used to count birds in the 20 frame video sequences. The video sequences with the least zoom (far scale) produced the most erroneous detections. This probably as a result of the lack of movement visible in the video and the small size of the birds (more interference from noise).

The results suggest ecological monitoring, in this case of microphytobenthos and birds on the Galgeplaat, is indeed possible by using the available terrestrial imagery from the platform. In the current state of development however, the process cannot be claimed fully automatic yet as some a priori knowledge from the user’s part is still required. Also, it is advised unsuitable images (images with rain drops, too much glare or just too blurry) be automatically detected prior to detection as it is inhibiting the detection process. Regardless, the use of remotely sensed imagery for ecological monitoring proves promising in comparison with current ecological monitoring for several reasons. Provided that unsuitable images are filtered out, the problem of limited accessibility to the mudflat is ruled out (thus making it possible to produce high time resolution data) and no costs are required for lab work or transport to the mudflat for bird counting and the collection of specimen.
I am pleased to present you with my thesis, the last milestone in this 3 year journey through my Geomatics master. Choosing the Geomatics master’s track, instead of a master that matched my bachelor’s track, was certainly a decision that came with doubts. However, having reached the end of it now, I can say with confidence that everything along the way has made this choice all the worthwhile.

Choosing a thesis topic was troubling at first, but in the end I decided to choose the field that interested me most; image processing. I was also convinced that adding practical experience to my personal development would help me along the way after my master. By chance Roderik Lindenbergh, which soon after became my supervisor, knew of a graduation topic placed by the institute Deltares concerning automatic bird detection. This is how I met my second supervisor Jasper Dijkstra. There at Deltares I came to realise the importance of ecological monitoring; how it can help us understand an ecosystem’s dynamics and sustain or help create a balanced coexistence between anthropogenic activities and nature’s ecosystems. Consequentially, I also came across the difficulties that come with ecological monitoring, such as reaching the study site for in situ sampling and the costs involved in the process. This gave reason to investigate the potential of remote sensing for ecological monitoring purposes.

This thesis in particular focuses on the use of remote terrestrial imagery as a tool for monitoring ecological features on an intertidal mudflat located in the Oosterschelde, known as the Galgeplaat. Two case studies were selected; the identification of microphytobenthos cover and the automatic detection of bird numbers. The report is organised into general background knowledge on the ecological environment of the study site, digital camera essentials, and existing detection methods, and on the strategies and results of the two case studies. With this report I would like to present the results of my investigation on the potential of the available terrestrial imagery for ecological monitoring of the Galgeplaat.

For those who are unfamiliar with digital camera basics please go through Chapter 2 which possesses a brief description of camera terminology as well as the specifics of the cameras on the remote sensing platform. For those unfamiliar to ecological terminology and the species that reside in the Oosterschelde basin you can refer to Chapter 3. Chapter 4 presents a general introduction of the existing detection methods and the basics of the methods that were
used. The next chapters represent the methodology, results and conclusions of the two case studies. The thesis is then closed by coming back at the research question.

I would like to express my gratitude to the people without whom this work would not have been possible. Special thanks goes to both of my supervisors Roderik Lindenbergh and Jasper Dijkstra that always made the effort to read through all my (and yes, they were many) draft reports to give me detailed feedback. I thank both of them for giving me the opportunity to experience the working atmosphere at Deltares and consequentially to meet the very nice colleagues I came to know there.

I would also like to extend my appreciation to Dr. Emile Hendriks of the pattern recognition department for offering to be my co-reader for my thesis and to my graduation professor Massimo Menenti for giving me good advise. Great thanks also goes to Tom Ysebaert, Daphne van der Wal and Jeroen van Dalen from the IMARES and NIOZ institutes who shared their expertise on ecological matters of the Oosterschelde with me. Next, I would personally like to thank Roderik Koenders who introduced me to the concept of self organising maps and for helping me remember the purpose of the thesis when I lost track of it. I would like to thank all my friends (in particular Mey and Jiebo) for continuously rejuvenating my motivation and my dear parents for their full support during my academic years. Last but not least, my deepest thanks to my siblings Lyssandre and Irina and my boyfriend Yang for being there when I needed them the most.

I wish you a pleasant read,

Penelope Rammos
Delft, the Netherlands
September 2012
Dedicated to my grandparents
# Table of Contents

**Preface** iii

1 **Introduction** 1
   1.1 Current Status Monitoring Biological Features 3
       1.1.1 Current monitoring of birds 3
       1.1.2 Current monitoring of benthos 4
   1.2 Problem statement and suggested solution 6
   1.3 Research Question 7
   1.4 Thesis Structure 8

2 **The Argus BIO Platform** 11
   2.1 The Company: Deltares 11
   2.2 The Study site: Galgeplaat 12
   2.3 The Remote Sensing platform: Argus Bio 13
   2.4 Camera Basics 16
       2.4.1 Relevance of technical characteristics 18
   2.5 Input data specifics: Cameras and Images 19
       2.5.1 The Argus Cameras 19
       2.5.2 The Multispectral Camera 21
       2.5.3 The Axis Bio pan/tilt/zoom camera 23

3 **Oosterschelde Ecological Features** 25
   3.1 The Oosterschelde Ecosystem 25
   3.2 Birds and Benthos of the Rhine-meuse-scheldt delta 28
       3.2.1 Birds 28
       3.2.2 Benthic fauna and flora 30
   3.3 Ecological features of Interest 32
# 4 Existing Image Detection Methods

4.1 Impeding factors to detection .............................................. 35
4.2 Inventory of Detection methods ........................................... 38
  4.2.1 Morphological & Thresholding techniques ......................... 38
  4.2.2 Template Matching .................................................... 42
  4.2.3 Machine Learning Algorithms ....................................... 43
4.3 Pros and Cons of existing detection methods ......................... 45
4.4 Case studies ................................................................. 47
  4.4.1 Case Study I: Microphytobenthos cover .......................... 47
  4.4.2 Case Study II: Bird Number determination ....................... 51

# 5 Diatoms

5.1 Raw image Specifications .................................................. 55
5.2 The in situ sampled data ................................................... 56
5.3 Test Image set ............................................................... 60
5.4 Image Preprocessing ........................................................ 63
  5.4.1 Removing image text bands ......................................... 63
  5.4.2 Colour space Conversion ............................................. 64
  5.4.3 Resizing data size with Symmetrical extension .................. 64
  5.4.4 Homomorphic filtering .............................................. 65
  5.4.5 Contrast Stretching and Equalization ............................. 68
5.5 Image Processing ............................................................ 70
  5.5.1 Macroalgae Masking with Maximum Likelihood Classification .... 70
  5.5.2 Self Organising Maps ................................................ 76
5.6 Validation ................................................................. 93
  5.6.1 In situ macroalgae cover and computed cover comparison ........ 94
  5.6.2 Computed diatom cover and pigment ratio seasonal behaviour ... 100
  5.6.3 Computed diatom NDVI versus in situ pigment data ............. 104
  5.6.4 Grid representativeness ............................................ 106
5.7 Discussion ................................................................. 107
  5.7.1 Distinguishing between macroalgae and diatoms ................ 108
  5.7.2 Method Suitability .................................................. 110
  5.7.3 In situ data & Validation .......................................... 111
  5.7.4 General recommendations ......................................... 112
5.8 Conclusions ............................................................... 113
# Table of Contents

6 Birds

6.1 Raw Image Specifications .................................................. 115
6.2 Image Preprocessing ......................................................... 119
6.3 Bird Detection ................................................................. 119
   6.3.1 Output requirements ................................................... 120
   6.3.2 The Matlab Computer Vision system toolbox ......................... 120
   6.3.3 Frame Differencing .................................................... 120
   6.3.4 Background subtraction with a median filtered background image .... 121
   6.3.5 Background subtraction with a weighted mean image ................ 122
6.4 Results .............................................................................. 126
6.5 Method Suitability ............................................................. 129
6.6 Proposed post processing of results ......................................... 130
6.7 Discussion .......................................................................... 132
6.8 Conclusions ........................................................................ 134

7 Conclusion ........................................................................... 137

A Table of monitoring costs Galgeplaat ...................................... 139

B Inventory of Birds .................................................................. 141

C Absorbance spectra Fucoxanthin, Chlorophyll-a and Diadinoxanthin 143

D Introduction to SOMPAK ......................................................... 145
   D.1 SOMPAK parameters ....................................................... 145
   D.2 SOMPAK example ........................................................... 147

E Clustered Test Images SOM .................................................... 149

F Specifics of images used for daily variance assessment of Macroalgae 153

G Matlab codes bird detection .................................................... 155

H Error types per frame .............................................................. 167

Bibliography ............................................................................ 171
## List of Figures

1.1 Zeeland Province of the Netherlands ........................................ 2
1.2 Box corer; [NIOZ.nl] .......................................................... 5
1.3 Gridded frame used on the Galgeplaat, image courtesy of NIOZ .......... 5
1.4 Thesis Structure .............................................................. 9

2.1 The Oosterschelde Tidal Basin ............................................. 12
2.2 The storm surge barrier and the Galgeplaat mudflat .................... 13
2.3 The Argus Bio station ........................................................ 14
2.4 The Argus Bio location on the Galgeplaat ................................. 15
2.5 The location of the five permanent quadrants around the Argus Bio station ... 15
2.6 The Argus Bio station cameras .............................................. 19
2.7 Example images Argus cameras .............................................. 20
2.8 Pixel Resolution Map of the 4 Argus cameras at the Galgeplaat .......... 21
2.9 Example images of the Multispectral camera ............................. 22
2.10 Spectral response curves of the Multispectral RGB bands .............. 23
2.11 Example images of the Bio pan/tilt/zoom camera ........................ 24

3.1 Simplified food web Oosterschelde ...................................... 27
3.2 Eurasian Curlew and Oystercatcher ....................................... 29
3.3 Mussels and benthic copepoda ............................................. 31
3.4 Sea Lettuce and Bladderwrack ............................................. 31

4.1 Problem factors for detection; unfit images .............................. 36
4.2 Problem factors for detection; presence of non target features .......... 36
4.3 RGB and corresponding false colour composite ........................ 41
4.4 Normalised Differential Vegetation Index image ........................ 42
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5</td>
<td>Example of a SOM training</td>
<td>45</td>
</tr>
<tr>
<td>4.6</td>
<td>A visual of individual diatoms and diatom biofilm cover</td>
<td>48</td>
</tr>
<tr>
<td>5.1</td>
<td>General flowchart of the Diatom detection procedure</td>
<td>55</td>
</tr>
<tr>
<td>5.2</td>
<td>Location of the six sampling grids in the plot below the Argus tower</td>
<td>57</td>
</tr>
<tr>
<td>5.3</td>
<td>In situ measured Pigment concentrations over time</td>
<td>58</td>
</tr>
<tr>
<td>5.4</td>
<td>Fucoxanthin to chlorophyll-a pigment ratio over time</td>
<td>59</td>
</tr>
<tr>
<td>5.5</td>
<td>Average in situ percentage coverage estimations of macroalgae</td>
<td>60</td>
</tr>
<tr>
<td>5.6</td>
<td>Detailed Preprocessing task block</td>
<td>63</td>
</tr>
<tr>
<td>5.7</td>
<td>The homomorphic filtering process</td>
<td>66</td>
</tr>
<tr>
<td>5.8</td>
<td>Fast Fourier transform of the natural logarithm of the luminance band</td>
<td>66</td>
</tr>
<tr>
<td>5.9</td>
<td>The modified Butterworth filter frequency domain response with different starting parameters</td>
<td>67</td>
</tr>
<tr>
<td>5.10</td>
<td>Detailed processing task block for the Maximum Likelihood macroalgae masking approach</td>
<td>71</td>
</tr>
<tr>
<td>5.11</td>
<td>The defined ROI’s in the plot below the Argus tower</td>
<td>72</td>
</tr>
<tr>
<td>5.12</td>
<td>The original False colour composite image and the masked macroalgae image</td>
<td>73</td>
</tr>
<tr>
<td>5.13</td>
<td>Display of the Macroalgae masking approach output</td>
<td>75</td>
</tr>
<tr>
<td>5.14</td>
<td>Detailed Processing task block for the SOM approach</td>
<td>78</td>
</tr>
<tr>
<td>5.15</td>
<td>The average RGB and NIR values of each test image over time</td>
<td>80</td>
</tr>
<tr>
<td>5.16</td>
<td>U-matrix and Sammon visualisation</td>
<td>85</td>
</tr>
<tr>
<td>5.17</td>
<td>Dendrogram of Trained network</td>
<td>86</td>
</tr>
<tr>
<td>5.18</td>
<td>Dendrogram with the clusters appointed to targets</td>
<td>91</td>
</tr>
<tr>
<td>5.19</td>
<td>Sections of the clustered image outputs</td>
<td>92</td>
</tr>
<tr>
<td>5.20</td>
<td>Tabular output of the SOM approach</td>
<td>93</td>
</tr>
<tr>
<td>5.21</td>
<td>Time series of the in situ average macroalgae cover and the computed cover</td>
<td>94</td>
</tr>
<tr>
<td>5.22</td>
<td>The RGB image of November</td>
<td>95</td>
</tr>
<tr>
<td>5.23</td>
<td>Sampling grid November</td>
<td>96</td>
</tr>
<tr>
<td>5.24</td>
<td>The in situ and computed average macroalgae cover and the water cover in percentages over time</td>
<td>97</td>
</tr>
<tr>
<td>5.25</td>
<td>Daily variance macroalgae cover</td>
<td>98</td>
</tr>
<tr>
<td>5.26</td>
<td>Computed diatom percentage cover over time of the SOM approach and the macroalgae masking approach</td>
<td>101</td>
</tr>
<tr>
<td>5.27</td>
<td>Diatom cover behaviour at 20 minute intervals</td>
<td>103</td>
</tr>
<tr>
<td>5.28</td>
<td>NDVI versus Chlorophyll-a for the two approaches</td>
<td>105</td>
</tr>
<tr>
<td>5.29</td>
<td>Diatom percentage cover versus fucoxanthin to chlorophyll-a ratio</td>
<td>106</td>
</tr>
<tr>
<td>5.30</td>
<td>Grid representativeness</td>
<td>107</td>
</tr>
<tr>
<td>5.31</td>
<td>Comparison NDVI and NDVI that uses the blue band instead of the red band</td>
<td>109</td>
</tr>
<tr>
<td>6.1</td>
<td>Birds as seen from one of the Argus cameras</td>
<td>116</td>
</tr>
<tr>
<td>6.2</td>
<td>First frames of the three selected video sequences; VID1, VID2 and VID3</td>
<td>118</td>
</tr>
</tbody>
</table>
List of Figures

6.3 First frame of VID4 ................................................................. 119
6.4 Example of a frame difference image of VID2 ............................. 121
6.5 Processing task block for the frame differencing approach .......... 121
6.6 Processing task block for background subtraction with a median filtered background image .................................................. 122
6.7 Example of a weighted mean image .......................................... 123
6.8 Example of a standard deviation image ...................................... 124
6.9 Processing task block for background subtraction with a weighted mean background image .................................................. 125
6.10 Visual example of bird detection output VID1, VID2 and VID3 ....... 128
6.11 Visual example of the bird detection output VID4 ...................... 129

A.1 Table of Costs of the experimental Sand Nourishment on the Galgeplaat 2007 to 2011 [1] ................................................................. 139

C.1 Absorbance spectra for the pigments Fucoxanthin and Chlorophyll-a [2] ...... 143
C.2 Absorbance spectra diadinoxanthin [3] ....................................... 144
List of Tables

2.1 Optical formats and the corresponding sensor sizes ........................................... 17
2.2 Argus camera orientation angles and Camera ID’s ............................................. 20
2.3 Multispectral camera boundary coordinates ...................................................... 22
3.1 Common bird classes per consumer types ......................................................... 29
4.1 General Advantages and Disadvantages of the detection classes ......................... 46
5.1 Test set Case Study I: dates, times and corresponding remarks ........................... 61
5.2 Image test set for Case Study I ............................................................................ 62
5.3 A visual of the image preprocessing procedure steps ........................................ 69
5.4 Image Training set details for training the SOM ................................................. 80
5.5 Correlation matrix of the eight input variables .................................................. 84
5.6 Specifics and performance of the different SOM runs ........................................ 88
5.7 Specifics and performance of the different SOM runs ........................................ 89
5.8 Specifics of the selected final settings for the SOM training ............................... 89
5.9 Cluster Averages of the ten classes ...................................................................... 90
5.10 Performance statistics of the two approaches .................................................... 100
5.11 Average and maximum deviation between the entire plot value and the mean of the six grids value .............................................................. 107
6.1 Test set Case Study II: Number of birds, dates, times and Zoom details .......... 117
6.2 The parameter settings for each of the bird detection methods .......................... 125
6.3 Bird count per frame for each of the three methods for all three video sequences 127
6.4 Average true positive (TP), false positive (FP) and false negative (FN) for each method for the three video sequences ..................................................... 130
B.1 Inventory of Birds of the Middle region of the Oosterschelde ............................ 141
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.1</td>
<td>Clustered images of the SOM approach</td>
<td>150</td>
</tr>
<tr>
<td>E.2</td>
<td>Clustered images of the SOM approach</td>
<td>151</td>
</tr>
<tr>
<td>E.3</td>
<td>Clustered images of the SOM approach</td>
<td>152</td>
</tr>
<tr>
<td>F.1</td>
<td>Dates, times and corresponding remarks of the images used to assess the daily variance of macroalgae cover</td>
<td>153</td>
</tr>
<tr>
<td>H.1</td>
<td>True positives (TP), false positives (FP) and false negatives (FN) for each frame of the 3 video sequences</td>
<td>168</td>
</tr>
<tr>
<td>H.2</td>
<td>True positives (TP), false positives (FP) and false negatives (FN) for each frame of the 3 video sequences</td>
<td>169</td>
</tr>
<tr>
<td>H.3</td>
<td>True positives (TP), false positives (FP) and false negatives (FN) for each frame of the 3 video sequences</td>
<td>170</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The Netherlands are globally renowned for their expertise in the field of water management and water defence systems. Civil works projects such as the Delta Works and the Zuiderzee Works (man made infrastructures comprised of a combination of coastal dams, sluices, dikes, land reclamation and water drainage works) make sure that the lands remain protected from imminent flood threats. As these works per definition alter the natural dynamics of the area it is clear they do not come without adverse effects on the ecology and morphology of the surrounding areas. Examples of adverse effects are increased erosion of mudflats and sand banks (also referred to as sand hunger) and decreased relief of mudflats. This is the case for the Galgeplaat mudflat located in the Oosterschelde (a tidal basin located in the southern province Zeeland of the Netherlands, see figure 1.1) as a consequence of the construction of the Oosterschelde Storm Surge Barrier [4]. Other examples of adverse effects are the appearance of brackish water species and the disappearance of habitat bounded species as was the case in lake Grevelingen also located in the province of Zeeland [5].
It is not so surprising then that monitoring the ecological and morphological dynamics in coastal and delta regions is necessary to assess the degree of impact of man made structures or other anthropogenic factors (such as water pollution, climate change). Particularly for the monitoring of the ecology this requires extensive field work (for instance to collect samples or conduct censuses of wildlife) and lab work (to prepare the collected samples and identify the different species). The expertise, facilities and time required for all this sampling, censusing, preparation and identification of species make the whole ecology monitoring procedure a pricey yet necessary endeavour.

In this thesis the focus lies on the Galgeplaat, an intertidal mudflat, which is currently suffering from sand hunger as a result of the Storm Surge Barrier construction. In trying to deal with this problem the Rijkswaterstaat (the Dutch department of Waterways and Public Works) performed an experimental sand nourishment in 2008 on a section of the mud flat. A remote sensing platform (equipped with multi spectral cameras) was placed on the Galgeplaat to monitor the effects of the experimental nourishment. The platform was named the Argus Bio platform and the images were initially used to monitor the morphological dynamics of the mud flat. Since these images also contain a lot of ecological information, their use for ecological monitoring of birds, and benthic organisms has been considered. The health of bird and benthic organism populations are considered indicators of an ecosystem’s health. Bird populations dwell if there is not enough food or breeding grounds present (caused by sand hunger for example). Benthic organisms (in particular the location dependent ones) can rapidly increase or decrease in their numbers with a small change in nutrient fluxes, salinity of the water or temperature of the water. In extreme cases species can go extinct as was the case for some benthic species in the Grevelingenmeer due to the construction of the dam (the water became sweet).
The project will be conducted in collaboration with Deltares, a renowned Dutch-based research institute and specialist consultancy specialized in matters related to delta and coastal zones, river basins, the soil and the subsurface of the land. The objective is to see whether the ecological monitoring procedure can be made more time and cost efficient by using imagery obtained from the remote sensing platform on the Galgeplaat.

The focus of this chapter is the delineation of the thesis’ research field. It initiates with general knowledge on the current status of biological feature monitoring and will be followed by a problem statement and suggested solution. The chapter will be closed with the formal definition of the research question and its sub questions.

1.1 Current Status Monitoring Biological Features

Delta areas, coastal regions and river basins possess multiple species from both the animal and plant kingdoms. In particular delta areas and estuaries, which form the main transition zones between the land and ocean environments, possess great wealth when it comes to species. The material flux (sediment, nutrients and organic material) that takes place between this land-ocean interface makes them the most productive natural habitats in the world [7]. Yet, it is estimated that by 2025 about 75% of the world’s population will be living in these coastal and delta regions [8]. One can only ponder whether the ecosystems’ current complex dynamics are bound to yield under the pressure of further anthropogenic involvement.

It is comprehensible then that the European union adopted the EU Water Framework Directive in 2000 (Directive 2000/60/EC, OJ L327). The directive’s environmental objectives requires member states to prevent the deterioration of the status of all the surface and groundwater bodies and to achieve a good status by 2015 [9]. In fact monitoring of algae and microphytobenthos such as diatoms are now a mandatory component for European member states. Other protocols for monitoring and assessment of benthos (in particular algae) are for example the algal sample analysis protocol that exists under the US Geological Survey National Water-Quality Assessment program (USGS-NAWQA) and the Stream Periphyton Monitoring Protocol from the New Zealand Ministry of the Environment [10, 11]. For the monitoring of wildlife of coastal and delta regions associations are also present. An example being the EBCC (the European Bird-Census Council) with its Pan European Common Bird Monitoring Scheme project [12]. In the following two sections a brief description of how birds and benthos are currently monitored is shown.

1.1.1 Current monitoring of birds

In the Netherlands bird populations are mostly registered by the Dutch National institute for Sea and Coast known as the RIKZ (this has been officially incorporated into the Dutch Rijkswaterstaat Waterdienst since 2007) which often collaborates with other corporations such as the SBB (StaatsBosBeheer) and amateur volunteers [13].

For the Galgeplaat the consultancy Agency Habitat Advies (specialized in ornithology, aquatic
and estuarine ecology) was employed in 2007 up till now to count the birds [1]. Birds are counted manually every 15 minutes during the periods of low tide over two days using binoculars or telescopes from aboard a ship [14]. Only birds present in eleven laid out counting compartments on the Galgeplaat of approximately 100 by 100 meters (bounded by poles) are counted. All active birds were counted as foraging birds and all inactive birds as resting birds. Gulls and terns were not included in the bird counting process [14]. The output of this monitoring is the maximum number of foraging birds per compartment (obtained by averaging the maximum counts observed during two days of counting), the average number of birds per compartment (average of the bird averages observed per compartment during two days of counting) and the foraging time in minutes (the sum of the total number of foraging birds times 15, the ‘sampling’ interval) [14]. Disturbances to the bird counting procedure are low flying motor powered gliders or the presence of predatory birds which scatter the foraging birds. The uncertainty of bird counts on the Galgeplaat in [14] is stated as unknown. Errors can arise from the previous mentioned disturbances, as a result of the perspective viewed through the telescope, as a result of glare and/or the large distance of the observer to the counting compartment (maximum was 900 meters). An individual compartment cannot be overseen completely with one telescope and the borders of a compartment are not visible (only the corners are marked with poles) [14]. As such birds at the edges of the compartments may not be counted accurately.

1.1.2 Current monitoring of benthos

The monitoring of benthic organisms refers to the presence (density and biomass), resettling rates and species diversity of benthos. Benthic organisms or benthos refer to the organisms that live in the benthic zone, defined as the lowest part of a water body such as the ocean or lake floor (it is comprised of the sediment surface and some sub surfaces of the sediment). Benthos can be divided into macrobenthos and microbenthos. Micro benthos refers to benthic organisms smaller than 0.1 mm and macro benthos the benthos larger than this. Within these two classes one can also distinguish between zoobenthos and phytobenthos (animal and plantlike organisms respectively)

A variety of ways exist to monitor and analyse micro and macro zoobenthos biomass numbers and their species diversity and the benthic flora. In most cases samples of the sediment are taken through the use of sediment traps or sediment cores (also known as sediment sample tubes or "steekbuizen"). For soft sediments in shallow waters box corers (see figure 1.2) are used. The tool was used in 2004 in the BIOMON monitoring programme commissioned by the National Institute for Coastal and Marine Management (or RIKZ, now part of the Waterdienst of the Rijkswaterstaat) in cooperation with the North Sea Directorate (DNZ-RWS) and the Royal Netherlands Institute for Sea Research (NIOZ) for the Dutch sector of the North sea [15].

In the case of the Galgeplaat gridded frames of 50 cm by 50 cm are used (see figure 1.3). In each individual grid square then sampling is performed for the sediment (3 cm deep) and the pigment (the upper 1 cm) and the casts of Arenicola (lug worms) and Lanice (bristle worms) are counted (the latter two being macro zoobenthos). Also the position and percentage cover
of Macroalgae is noted with respect to the grid and their corresponding state (alive, decaying).

![Box corer](NIOZ.nl)

**Figure 1.2:** Box corer; [NIOZ.nl]

![Gridded frame](NIOZ.nl)

**Figure 1.3:** Gridded frame used on the Galgeplaat, image courtesy of NIOZ

After extraction from the targeted sites the samples are cleaned in order to remove clay, organic matter and detritus (dead organic matter that is suspended in the water). Samples are often also sieved with a predetermined mesh size to facilitate the sorting (the Galgeplaat samples are sieved through a 1 mm mesh sized sieve). The cleaned specimens are then analysed through the use of microscopes. The species are identified and counted by skilled professionals or by estimation of several chemical components that are typical of a species [16]. One typical technique to perform quantification, identification and purification of micro zoobenthos is the HPLC technique (High-performance liquid chromatography). Common outputs of the analysis are values of density, biomass and diversity (using indexes such as BENTIX, BQI=benthic quality index and AMBI=AZTI Marine Biotic Index) [15, 17].

In some of the more recent research it is not uncommon to make use of remotely sensed data to monitor benthos as is the case for monitoring phytobenthos. Their coverage can be extracted by using the Normalized Differential Vegetation Index (NDVI) which is based on the reflectance properties of the targeted benthos in the red and near infra red regions of the electromagnetic spectrum. This index is often used to detect vegetation cover from satellite imagery such as can be acquired from the Aqua MODIS instrument (MODerate resolution
6 Introduction

Imaging Spectro radiometer). However, this does not allow for distinction between benthic macro algae and benthic micro algae [18]. Note that when remote sensing data is used, it is common to use actual sampling from the most representative sites of the region for ground truthing (as a means of data verification).

Current monitoring of birds and benthic organisms requires expert personnel, fieldwork and lab work. It is apprehensible that the procedure becomes an expensive investment, an example being the ecological monitoring costs of the Galgeplaat mud flat. Sediment analysis required an annual amount of 18,000 Euro from 2008 to 2011 according to the monitoring program of the Dutch department of Waterways and Public Works. For bird counting in 2007 the costs involved reached an annual sum of 17,826 Euro [1]. See Appendix A for the table of costs for monitoring the Galgeplaat. Aside from expenses the often limited accessibility to the study site produces another disadvantage; a low time resolution of sample data. On the Galgeplaat pigment sampling is done once a month.

1.2 Problem statement and suggested solution

The use of satellite imagery for monitoring algal blooms has shown the potential of using remote sensing for ecological monitoring. It is not uncommon to make use of the spectral reflectance properties of the target being investigated. However, this is not always possible. Birds (and therefore bird numbers) cannot always be detected solely on spectral signature, as their characteristics vary greatly depending on the species. In the case of benthos spectral reflectance could be taken advantage of but this doesn’t grant that a distinction can be made between the different flora and fauna species.

Automatic detection, using images or video data, is not an uncommon field of study, and has been used for a variety of applications such as face detection in commercial cameras, automated vehicle parking, automatic extraction of lesions from medical images, automatic counting of cyclists and pedestrians and much more. It is interesting for this case because in the absence of distinctive spectral reflectance properties it can make use of properties such as texture, edge detection, shape recognition and more. Because the different application fields have different parameters and different types of input data and output data there is a variety of existing automatic detection or pattern recognition techniques out there. Some examples of the different approaches to specific problems are background subtraction, algorithms that make use of feature parameters (such as the Hough transform) and algorithms that use data from reference images of the objects (such as the SIFT method).

Taking the previous into consideration a solution to the time and labour consuming behaviour of current monitoring of wildlife and benthic fauna and flora would be the use of an automatic detection algorithm which uses the data from the remote sensing platform already present on the Galgeplaat as input. For this to be achieved the targets should have detectable features. Features that could be used to detect wildlife and benthic organisms next to spectral characteristics are texture properties of the target (with or without respect to the background
1.3 Research Question

The foremost aim of this thesis is to assess the potential of the Argus Bio platform for ecological monitoring of the Galgeplaat. To do this the main research question was formulated as follows:

*Is it possible to automatically detect the main biological features (Birds and Benthic organisms) on the Galgeplaat mudflat from the available set of terrestrial images from the Argus-Bio platform and validate it with the available in situ data?*

The answer to this question depends on a variety of factors (i.e. fitness of the input data for detection, characteristics of targeted biological features, details of the available in situ data, required output analysis, etc.). As such a series of sub questions have been set up to assist in answering the main research question of the thesis:

**Sub questions input data:**
• Which benthic organism should be detected? (i.e. which species are of interest to Deltares and potential clients?)
• What characteristics of the benthic species and bird/s are detectable?
• What are the camera technical characteristics (image formats, bands captured, camera position & height, etc.)

Sub questions Detection method and software:
• What existing pattern recognition methods are most suitable for this application (i.e. the detection of biological features)?
• What are the impeding factors to detection?
• What software package(s) is/are most suitable for performing automatic detection for this particular application? Or is new technology necessary?
• What is the practicability of the selected detection method and software? (i.e. the ease of use of the method and software)

Sub questions Validation of algorithm:
• How will the algorithm deal with images that have no birds or benthic organisms in them?
• How will the algorithm deal with images in which birds partially occlude each other?
• How will the algorithm deal with the ambient conditions of the images? (ambient conditions such as the different weather conditions in the images and the variable direction of light at different times of the day)
• How can one validate the output of the algorithm?

Sub questions Argus Bio station:
• How can the current Argus platform be improved? (i.e. what type of sensors could contribute to a better performance of the automatic extraction/detection?)

1.4 Thesis Structure

The thesis is structured according to the outline shown in figure 1.4. Chapter 1 and Chapter 2 describe the scope and context of the thesis. Chapter 2 additionally possesses a section with camera basics to ease the transition of the reader into the specifics of the Argus Bio Platform cameras.

Background information is gathered in Chapter 3 and 4 regarding the ecological features of the study site and existing image detection methods. Chapter 3 closes with the reasoning behind the selection of the two targets that are to be detected based on interest of related institutes. Chapter 4 closes with the formal definition of the two case studies.

Chapter 5 and Chapter 6 revolve around the implementation of selected methods, the results and the validation of results for the two selected targets respectively. Both chapters are closed with a discussion and conclusions. The thesis is closed in the end with Chapter 7 by coming back at the research question.
1.4 Thesis Structure

Figure 1.4: Thesis Structure

**Research Scope & Context**
- Chapter 1: Introduction
- Chapter 2: The Deltas Argus Bio Platform

**Literature Research**
- Chapter 3: Oosterschelde ecological features
- Chapter 4: Existing image detection methods

**Methodology & Implementation**
- Chapter 5: Diatoms
- Chapter 6: Birds

**Conclusions**
Chapter 2

The Argus BIO Platform

This chapter gives a detailed description of the project context. The project context refers to the company which was collaborated with during the entire thesis process, the study site under consideration, the remote sensing platform on the study site and the specifics of the data obtained from the remote sensing platform.

2.1 The Company: Deltares

Deltares is a Dutch-based research and consultancy institute specialized in national and international matters related to delta areas, coastal zones, river basins, the soil and the subsurface of the land. Deltares was formed in 2008 and is in fact the product of the joined forces of WL| Delft Hydraulics, GeoDelft, the subsurface and groundwater unit of TNO (Dutch Organisation of Applied Scientific Research) and some sections of the Dutch Department of Public works and Waterways (Rijkswaterstaat). Their expertise is employed to achieve a safe and sustainable living environment for people in delta areas, coastal zones and river basins. This is done through for instance coastal zone management, flood risk management, hydraulic engineering, water resource management, monitoring of the water and soil quality, analysing ecosystems of delta regions and much more.

This thesis will be performed under the section of Water Quality & Ecology of the Marine & Coastal systems unit of Deltares. Assistance will also be provided from involved individuals from IMARES (Institute for Marine Resources and Ecosystem Studies) and NIOO-CEME (Netherlands Institute for Ecological Research-Centre for Estuarine and Marine Ecology, or NIOZ, from January 2012 onwards).

Monitoring of the Galgeplaat’s morphology and ecology using data obtained from the remote sensing platform placed on it is a project of Ecoshape-Building with Nature with which Deltares is involved. Monitoring the Galgeplaat (a mudflat) is important due to its significance to wildlife and benthic organisms. Monitoring helps to assess the degree of impact the
Galgeplaat receives from the storm surge barrier and the applied experimental sand nourishment.

2.2 The Study site: Galgeplaat

The Galgeplaat is a mudflat located in the Oosterschelde of the province of Zeeland, in the south of the Netherlands (See figure 2.1). A mudflat or tidal flat is a wetland formed due to tidal or river depositing of silts, clay and marine animal detritus (dead organic material). Mud flats are covered at high tides and exposed at low tides. They are the natural habitat of several location-dependent benthic species and the breeding and foraging grounds for both migratory and non migratory birds.

![Figure 2.1: The Oosterschelde Tidal Basin (highlighted blue) in the province of Zeeland (dark grey)](image)

In 1986 the Oosterschelde Storm surge barrier (Oosterscheldekering), located at the mouth of the Oosterschelde tidal basin, held its grand opening (See figure 2.2). This surge barrier was one amongst many of the dams constructed for safekeeping the inland of the Netherlands. The collection of these dam constructions is more familiarly known as the Delta Works, which were built in response to the 1953 North Sea flood. The Oosterschelde was initially designed to be a closed dam, this however received some unexpected resistance as people feared that the ceasing of the salt water environment and the tidal movement would have too many adverse affects on the ecology of the region. Also it was feared the barrier would cause the fishing industry of the region to suffer [19]. As a result the storm surge barrier was designed to still allow salt water to enter the Oosterschelde, yet this didn’t relieve it from adverse effects towards the tidal basin. Today, the storm surge barrier still causes, and will for the next couple of decades, perturbations in the vulnerable morphological dynamics of the region.

One particularly concerning adversity caused by the barrier is the sand hunger process. Sand hunger refers to the erosion of mudflats, sand banks and marshes. It is a concern for several reasons. Primarily, it results in loss of foraging grounds for wading birds. It also means the dikes along the Oosterschelde boundary will have to put up with stronger wave energies as they are losing their natural wave defences in front of them (i.e. if the mud flats and sand-banks are sufficiently present they act as wave dampers prior to the collision). Sand hunger thus also leads to an increased investment amount into dike reinforcement, that is, unless the sand hunger issue can somehow be dealt with.
One measure to deal with the problem of sand hunger is by conducting a sand nourishment (the supplementing of the sand bank with more sand). An experimental such nourishment was performed by the Dutch Department Waterways and Public Works in 2008 on the Galgeplaat. The Galgeplaat was chosen for this experimental sand nourishment due to its favourable position with respect to dredging locations (dredging refers to the removal of sand and silt from the base of a water body to facilitate shipping traffic) and due to its high erosion rates.

To assess the impact of the nourishment on the morphology, the Galgeplaat was subjected to a great deal of monitoring of the sand bank. This resulted in the installation of the Argus Bio station on the Galgeplaat by Ecoshape-Building with Nature.

2.3 The Remote Sensing platform: Argus Bio

The station was named the Argus Bio station as a consequence of it possessing four Argus cameras (See figure 2.3). Currently, Argus cameras are used globally for coastal monitoring. The Argus monitoring concept itself was initiated by the Coastal Imaging Lab of Oregon State University in 1980 and developed further during the 1990s with the discovery of the practicability of the Timex technique (refers to the use of time exposure images) for coastal morphology monitoring. The Argus concept does not only focus on the Timex technique, in fact it gathers three types of images: a snapshot image, a time exposure image and a variance image. Snapshot images don’t offer much information on the morphology of the coastal regions but do give an indication of the ambiance conditions. Time exposure images average out natural modulations (such as waves breaking on the shore) and depict these smoothened out. Moving objects are also removed from time exposure images (this refers to objects such as ships, people and vehicles on the beach). Lastly, the variance images indicate the regions that are changing in time (such as the sea surface due to tidal effects).

The data is temporarily stored on a computer on site which is connected to the outside with an internet connection. Because the process of collecting and retrieving data is automated the
operation costs for an Argus monitoring station are very low. As such it is understandable why the Argus monitoring system has been widely used globally to monitor coastal morphology since its completed development in 1992 [Cohen, A., 2009].

The Argus Bio platform on the Galgeplaat is 15 meters high and aside from possessing four Argus cameras also possesses a Multi spectral camera and a Bio pan/tilt/zoom Camera. The site’s coordinates are 55602.53, 397714.1214, 0 in the (EPSG) 28992 Amersfoort / RD coordinate reference system (the Dutch double stereographic RD 'RijksDriehoekstelsel" projection). See figure 2.4 for the location of the platform on the mudflat.

The multi spectral camera possesses an RGB channel, and two Near Infra red channels at bands 775 nm and 860 nm, and is pointed downwards (i.e. it makes near vertical images of the sediment right below the platform). The Bio pan/tilt/zoom camera is directed to five Permanent Quadrants around the tower (see figure 2.5). These images are oblique (images taken at an angle from the vertical). The Argus cameras are pointed in the direction of the experimental sand nourishment and produce highly oblique imagery (highly oblique means the horizon is visible in the images). Aside from visual sensors the tower also possesses a water level measurement sensor positioned at the base of the tower which regulates the multi spectral camera to start taking images when there is no water below the tower (although a small layer of water is occasionally detected in the multi spectral images). Finally the Argus Database system also possesses tide and wave data measured at the site of the station.
2.3 The Remote Sensing platform: Argus Bio

Figure 2.4: The Argus Bio location on the Galgeplaat; [Google Earth, satellite imagery from 2005]

Figure 2.5: The location of the five permanent quadrants around the Argus Bio station; [Google Earth, satellite imagery from 2005]
2.4 Camera Basics

This chapter briefly introduces camera terminology. Its purpose is to serve as an aid in understanding the more in depth details of the Argus Bio cameras that will be described later on. The main components of a digital camera are the sensor, the lens, the shutter and the aperture.

The sensor, also known as the digital sensor chip or sensor element, is responsible for turning the captured light into digital information and storing it on the memory card [21]. The most used digital sensor types today that are able to capture light and convert it to electrical signals are CCD sensors (Charge coupled device) and CMOS sensors (complementary metal oxide semiconductor). All the cameras on the Argus BIO platform possess CCD sensors. A CCD in essence is a matrix of individual photosensitive elements. Each individual element captures 70% of the incident light (to compare, a photographic film only makes use of 2% of the incident light). There are also three CCD sensors which basically consists of three separate charged coupled devices. Each of these takes separate measurements of different bands in the electromagnetic spectrum (such as red, green, blue) by the use of a dichroic prism which splits the light accordingly for the respective wavelengths and directs them to each of the respective CCD’s. This is done with a variety of optical coatings on the prism which can alter the way the light is reflected and transmitted after it enters the camera.

A camera sensor’s spatial resolution is assessed by the number of pixels per unit length (of the sensor). This number of pixels is commonly denoted in Mega Pixels (MP) which in essence describes the number of pixels present in an image (i.e. a 600 by 500 pixel image=300,000 pixels or 0.3 MP resolution) [22]. The larger the amount of Mega Pixels the bigger the image can be enlarged without losing its resolution [23]. Or in simpler terms, a 4 MP image has a better spatial resolution than a 2 MP pixel image of the same size.

Take note that the term ‘spatial resolution’ in remote sensing sectors has a different meaning, it refers to the size of the smallest object that can be resolved on the ground (i.e. the area captured by a single pixel) and not to the amount of pixels in an image [24]. From here on the term spatial resolution refers to the number of pixels in an image. The area size captured by a single pixel will be termed the pixel footprint or ground sampling distance (GSD).

Spatial resolution, together with the sensor size (which is comparable to the size of the negative film in analogue cameras) tells a lot about the quality of the image (as one can derive the pixel element size with these) [25]. The sensor size is often described with a fraction known as the optical format. This type of measurement of the sensor size originates from the 1950s and was originally used for TV tubes. This fraction says something about the diameter of the glass envelope of the tube but is in reality not mathematically related to the actual sensor size [25]. Despite this, the designation was stuck and is now commonly used. Common optical formats are 1/1.8", 1/2", 1/2.7" and 2/3". See Table 2.1 for typical optical formats and their corresponding sensor sizes.

The next main components of a digital camera are the lens and the aperture which in digital cameras are incorporated into one piece. The lens focuses the incident light onto the digital sensor. Lenses are often described with a variety of letters and numbers. For instance a
Table 2.1: Optical formats and the corresponding sensor sizes [25]

<table>
<thead>
<tr>
<th>Optical Formats</th>
<th>Sensor size [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>Tube Diameter [mm]</td>
</tr>
<tr>
<td>1/1.8&quot;</td>
<td>14.111</td>
</tr>
<tr>
<td>1/2&quot;</td>
<td>12.700</td>
</tr>
<tr>
<td>1/2.7&quot;</td>
<td>9.407</td>
</tr>
<tr>
<td>1/3&quot;</td>
<td>8.467</td>
</tr>
<tr>
<td>1/3.2&quot;</td>
<td>7.938</td>
</tr>
<tr>
<td>2/3&quot;</td>
<td>16.933</td>
</tr>
<tr>
<td>3/4&quot;</td>
<td>33.867</td>
</tr>
</tbody>
</table>

planar T* 1.4/85 ZF-IR lens describes the lens the multispectral camera possesses. At first sight this might seem confusing but a lens in general is described by its focal length and its aperture. The aperture controls the amount of light falling on the sensor (the aperture is the lens's diaphragm opening). It works like the iris of the eye; when too much light falls on the eye the pupil becomes small, when too little light falls on the eye the pupil dilates. The first number 1.4 describes the aperture and is known as the F/Stop or the F number. It is a ratio of the aperture diameter over the lens’s focal length [26] and as such is unit less. The smaller the F number the bigger the maximum aperture of the lens. The second number 85 describes the focal length of the camera in mm. So in this particular case an F number of 1.4 for an 85 mm lens means that the aperture has a diameter of 60 mm (F number= focal length/aperture diameter).

The focal length together with the sensor size describes the angle of view of the lens, the smaller the focal length the bigger the angle of view that is obtained. So in the earlier example there is the number 85, this corresponds to an 85 mm fixed focal length lens. If it was a zoom lens a range of focal lengths would be given instead of a single number [25]. Sometimes the angle of view is also referred to as the field of view (FOV) and it is directly related to the focal length as is indicated in equation 2.1.

\[
FOV = 2 \cdot \arctan \left( \frac{0.5 \cdot W}{f} \right) \tag{2.1}
\]

where W is the width of the sensor width and f is the focal length. Finally, the letters in a lens description usually refer to the producer and the mounting type (i.e. on which camera’s the lens can be mounted on).

The shutter of a camera controls the duration the incident light is allowed to fall onto the sensor. Together with the aperture it can be adjusted accordingly to the scene being captured to obtain good quality images. Shutter speed is measured in fractions of a second. i.e. a shutter speed of 1/5000 means that the shutter will be open for the duration of 1/5000th of a second [27].

Using the FOV or the focal length, the distance to and size of the object being captured and the spatial resolution (pixel dimensions) one can determine the pixel footprint or GSD
The pixel footprint indicates how much area in the real world a single image pixel captures as mentioned earlier.

2.4.1 Relevance of technical characteristics

Aside from the previously mentioned camera components one also has to consider their performance. This is important because to assess the quality of the spectral estimation (i.e. the quality of the produced image) it is necessary to look at how the internal and external parameters affect the imaging process [28]. This in turn is important to assess the influence of the camera parameters on the detection output such that the camera can be tuned optimally to the detection problem at hand. The quality of an image primarily depends on the sensors spectral sensitivity, the illumination and the scene properties [29].

When referring to illumination we are referring to the lighting conditions of the scene, such as the light incidence angle or the brightness. When referring to scene properties we are referring to the surface reflectance properties (i.e. diffuse surface, specular surface, absorbance of the surface) and the atmospheric conditions at the scene (humidity, temperature, etc.). The strength of the returned signal is affected by these scene properties, analysing their influence is thus important to assess the performance of the received signal.

The sensor spectral sensitivity (relative efficiency of light detection as a function of wavelength or frequency) is dependent on several factors. It is dependent on the performance of the sensor properties, such as the transmittance/reflectance of the filters in the camera (i.e. such as the performance of the dichroic prism in the camera), the performance of the lens and the performance of the CCD itself. If the lens is not performing well it can produce optical aberrations (distortions in the output image). Examples of optical aberrations are for instance chromatic aberration, comatic aberration, spherical aberration, distortion and astigmatism [30]. On top of this the overall performance is also affected by the quantization noise (experienced during analogue to digital conversion of data) and the random sensor noise [31]. For multispectral cameras the number of channels recorded and the channel bandwidths also have direct influence on the spectral resolution. This is particularly important when comparing band values obtained from different sensors. An example in practice is the comparison of NDVI (normalised difference vegetation index) values from different cameras. The NDVI is a measure for the amount of vegetation observed in the scene and uses the Near Infra Red band and the Red band in its computation (see chapter 4 for more detail). The NDVI can yield different results when acquired with different instruments inhibiting proper comparison between the two.

That the technical characteristics have influence on the produced result (the image) is clear, but their influence on the target detection to be performed should not be underrated either. Choosing the proper camera intrinsic parameters is essential to the detection process (i.e. improper choice of shutter speed can result in blurry images and thus inhibit the process of detection). In cases where the intrinsic parameters are already set it is important to have a measure of the image output quality to determine how the eventual error propagates through the detection process.

Penelope Rammos

Master of Science Thesis
2.5 Input data specifics: Cameras and Images

The Argus Bio Platform has 3 different types of cameras, which capture different parts of the Galgeplaat mudflat. See figure 2.6 for a visual of the camera types. This section will cover the specifics of the cameras and their respective images.

![Image of cameras](image)

**Figure 2.6:** The Argus Bio station cameras: Multi spectral camera, Bio pan/tilt/zoom camera and the 4 Argus cameras

2.5.1 The Argus Cameras

The 4 Argus cameras (See figure 2.6 c) used come from the Grasshopper Digital Camera line of Point Grey Research, known for their compactness, high resolution and high frame rates. The sensor type is a Sony progressive scan interline charge coupled device (CCD) with square pixels and a global shutter[ptgrey.com]. The progressive scan interline transfer refers to the scanning method of the CCD. The sensor model of the mounted cameras is a Sony ICX285 CCD 2/3”. As mentioned before the fraction 2/3”refers to the sensor size (see section 2.4). The Argus cameras have a spatial resolution (the pixel dimensions of the image) of 1.45 MP (1040 by 1392 pixels).

The mounted Argus cameras are directed towards the sand nourishment. Detailed information on their tilt, azimuth, horizontal fov (field of view) and roll can be seen in Table 2.2. The tilt of a camera describes its rotation with respect to the vertical z-axis. The azimuth describes the orientation of the camera in the xy-plane. The roll describes the rotation of the focal plane with respect to the horizon (a roll of 0 degrees being a landscape image and a roll of 90 degrees being a portrait image).

The images of the 4 Argus cameras have 10% overlap default but this can be changed manually if required. Some images of the 4 Argus cameras from 2011 can be seen in figure 2.7. The horizontal field of view as aforementioned in section 2.4 is directly related to the focal length of the camera. The focal length of the camera can be deduced by rearranging equation 2.1) by substituting the FOV values of Table 2.2 in the equation and using a sensor width of 8.800 mm. We then get focal lengths ranging from 11 mm (for the largest FOV) to 16 mm (for the smallest FOV) for the 4 Argus cameras.
Table 2.2: Argus camera orientation angles and Camera ID’s

<table>
<thead>
<tr>
<th>Angles [rad]</th>
<th>GPX01C</th>
<th>GPX02C</th>
<th>GPX03C</th>
<th>GPX04C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tilt</td>
<td>1.4050</td>
<td>1.3250</td>
<td>1.3268</td>
<td>1.3837</td>
</tr>
<tr>
<td>Azimuth</td>
<td>2.5023</td>
<td>1.9058</td>
<td>1.2275</td>
<td>0.6050</td>
</tr>
<tr>
<td>fov</td>
<td>0.5293</td>
<td>0.7430</td>
<td>0.7438</td>
<td>0.5323</td>
</tr>
<tr>
<td>Roll</td>
<td>0.0064</td>
<td>0.0034</td>
<td>0.0245</td>
<td>-0.0026</td>
</tr>
</tbody>
</table>

Because the cameras are tilted with a large angle the images become highly oblique. As a result the pixel footprints are spatially varying. The larger the distance between the location of interest and the camera, the bigger the pixel footprint. This spatially varying pixel resolution can be mapped in a so called pixel resolution map [32]. See figure 2.8 (Note that the resolution referred to in the map refers to the pixel footprint and not the spatial resolution and that the axes are logarithmic). Cross shore is perpendicular to the shoreline whereas alongshore is parallel. The Galgeplaat in this case does not have straight shorelines as most Argus sites have, so one can compare the cross shore direction to the vertical FOV pixel footprints of the cameras and the along shore direction to the horizontal FOV pixel footprints. The larger the distance from the camera (position x,y=0) the larger the pixel footprint becomes, ranging from 10 cm up to 30 m.

Figure 2.7: Images of the 4 Argus cameras taken on April 2nd at 11:06 PM in 2011.
2.5.2 The Multispectral Camera

The Multispectral camera (figure 2.6 a) also comes from the Grasshopper Digital Camera line of Point Grey Research. It is a custom 3CCD camera with the same sensor model as the Argus cameras and has a spatial resolution of 1.4 MP (the exact spatial resolution is 1384 (horizontal) by 1036 (vertical) with the pixel size being 6.45\(\mu\)m). The difference with the Argus cameras is that the sensor captures not only the RGB channel (400-700 nm) but also the NIR (775 nm) and NIR (860 nm) channels. It makes use of a prism with RGB/NIR Dichroic coating to direct the light into the 3 channels (RGB and the two NIR channels). The lens used on the sensor is a Zeiss lens planar T* 1.4/85 ZF. Z stands for Zeiss and the F stands for the mounting type (in this case Nikon’s F mount).

The multispectral camera is pointed to a patch of ground below the tower. Its images are close to vertical images (the camera is slightly tilted away from the tower to avoid capturing the tower base). Some example images of the Multispectral camera can be seen in figure 2.9. Because of the use of the prism to split the bands and direct them to their respective CCD’s there is a small time delay (from 1 to 4 seconds) between the RGB, NIR 1 and NIR 2 images (note how the bird in the RGB and NIR images has changed position). The patch that is captured is bounded by 4 wooden poles. The coordinates of these poles are given in Table 2.3 by the uneven numbers (1,3,5,7). The even numbers represent the coordinates right below the poles on the ground. The patch of ground is 4.59 metres by 4.65 metres. The approximate pixel footprint is 5 mm. It is an approximate value because the multispectral images are not entirely vertical, i.e. pixels whose corresponding ground points...
Figure 2.9: Images of the Multispectral camera taken September 1st at 12:30 PM in 2011. From left to right we have the RGB image, the NIR 775 nm band and the NIR 860 nm band.

are farther away from the camera have a larger pixel footprint then the ones that are closer to the camera).

Table 2.3: Multispectral camera boundary coordinates

<table>
<thead>
<tr>
<th>Boundary points</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS0001</td>
<td>55587.477</td>
<td>397720.616</td>
<td>8.711</td>
</tr>
<tr>
<td>GPS0002</td>
<td>55587.499</td>
<td>397720.673</td>
<td>-0.691</td>
</tr>
<tr>
<td>GPS0003</td>
<td>55583.623</td>
<td>397723.071</td>
<td>-0.770</td>
</tr>
<tr>
<td>GPS0004</td>
<td>55583.638</td>
<td>397723.111</td>
<td>-0.776</td>
</tr>
<tr>
<td>GPS0005</td>
<td>55586.341</td>
<td>397726.839</td>
<td>-0.693</td>
</tr>
<tr>
<td>GPS0006</td>
<td>55586.377</td>
<td>397726.864</td>
<td>-0.749</td>
</tr>
<tr>
<td>GPS0007</td>
<td>55589.729</td>
<td>397723.789</td>
<td>-0.687</td>
</tr>
<tr>
<td>GPS0008</td>
<td>55589.765</td>
<td>397723.801</td>
<td>-0.752</td>
</tr>
</tbody>
</table>

The spectral responses of the RGB bands can be seen in figure 2.10. Note though that the effect of the lens is not taken into account in this. The data originates from the Grasshopper Technical Reference manual obtained from the Point Grey official website [33].
2.5.3 The Axis Bio pan/tilt/zoom camera

The Bio pan/tilt/zoom camera is an Axis Network Dome Camera meant for professional video surveillance applications. On the platform however the camera does not take videos. Instead it captures RGB snapshot images of 5 Permanent Quadrants (or PQ’) around the Argus Bio platform (see figure 2.4). It does not take one snapshot image of each quadrant but rather a set of snapshot images scanned over the quadrant (with a shutter time of 1/1000 s to 1 s). A quadrant is approximately 50 by 50 metres. From this quadrant a number snapshot images are taken which are slightly overlapping. In PQ1 the number of snapshot images taken are 58, in PQ2 it is 30 images, in PQ3 33 images, in PQ4 and PQ5 18 snapshot images.

The spatial resolution of the images is 0.45 MP (576 by 768 pixels). The lens has a focal length of 4.1 mm to 73.8 mm tele (tele refers to teleconverter, a secondary lens mounted between the camera and the photographic lens). It is capable of 18 times optical zoom and a 12 times digital zoom. The camera itself is mounted on top of a pole on the platform itself. As such it is positioned higher than the other cameras but is also more subject to wind (which results into blurry images). Its height is at 16.895 metres above the mudflat. Example images of the Bio pan/tilt/zoom camera can be seen in figure 2.11. Unlike the Argus cameras a pixel resolution map (footprint) for the Bio pan/tilt zoom camera is not available. However, one can take advantage of the height of the camera and the tilt settings to determine the distance of the platform to an object in the centre of the image. The tilt angle for the Bio pan/tilt/zoom camera is given in degrees and indicates the tilt downwards from the horizontal plane.
Figure 2.11: Images of the Bio pan/tilt/zoom camera taken October 14th in PQ1 and PQ3 at 10:01 PM in 2011.
Chapter 3

Oosterschelde Ecological Features

This chapter’s objective is to give a general introduction to the Oosterschelde tidal basin ecosystem and thus indicate the importance of bird and benthic monitoring to assess the impact of the storm surge barrier and the sand nourishment. This will be followed by the details of a discussion held with people from the NIOZ and IMARES institutes concerning the ecological features of interest for ecological monitoring (Section 1.3, sub question input data).

3.1 The Oosterschelde Ecosystem

The material fluxes in the Oosterschelde tidal basin (such as nutrients, organic material, dead organic material, and silts) have altered since the construction of the storm surge barrier. As a result the ecosystem of the Oosterschelde has been affected. To determine how the ecosystem has been influenced one must first determine how the status of an ecosystem can be described. Even as we speak, estuarine and tidal basin ecosystem health models and indicators are constantly under development and improvement. There are a wide variety of such ecosystem health models and ecosystem indicators existing, examples being the Shannon-Weaver Index for biodiversity and the Estuarine Health Index known as EPI [34, 35]. Deckere and Meire [36] described the health of an ecosystem in terms of its resilience, vigour and organization. Resilience referring to the ability of the ecosystem to return to its original vigour and organisational structure after a certain amount of stress. The vigour of an ecosystem refers to the nutrient cycle of the ecosystem and the primary and secondary production processes. The organisation of an ecosystem refers to the structure and complexity of the food web [36]. The great variety of existing ecosystem models existing make it clear that an ecosystem cannot be assessed with simple rating systems. The intricate relationships between morphological, hydrodynamical, and ecological processes are clearly very complex. However, because these complex relationships exist, monitoring either one process indirectly tells something about the other process. Monitoring ecological processes, such as the biodiversity of a tidal basin, thus indirectly says something about the other processes.
The Oosterschelde has a rich biodiversity, making the region of national and global significance for marine and coastal processes. As such the Oosterschelde has become a protected area under the Natura 2000 Network, which is a European established network of protected areas following the regulations of the Habitats Directive of 1992 [37]. For a simple version of the Oosterschelde food web see figure 3.1 (this food web is based on information from [38] and [39]).

The focus in this thesis is on birds and benthic organisms. Taking a look at the food web diagram one can see that a variation in their population dynamics have influence on the other residing species. Also, morphology and chemical processes involved are affected (an example being some species of benthic micro algae which stabilize the sediment by producing extracellular polymeric substances, also known as EPS). It is thus of importance to monitor the population dynamics and resilience of bird and benthic species to external stresses (by monitoring for instance their recovery rates) to determine the ecological health status of the area. Note that the storm surge barrier is the primary external stress applied to the Oosterschelde but that the sand nourishment itself, although performed in consequence of the barrier, is also an external stress.
3.1 The Oosterschelde Ecosystem

Figure 3.1: Simplified food web Oosterschelde
3.2 Birds and Benthos of the Rhine-meuse-scheldt delta

The Rhine-Meuse-Scheldt Delta which contains our study area consists of an estuary (Wester schelde), a tidal basin (Oosterschelde) and some non-tidal basins (Grevelingen, Haringvliet, Veerse Meer) and can be characterised by their salt, brackish and sweet water zones. These zones make the co occurrence of different food webs possible. A quick introduction on the birds and benthic species of these zones will be given in the following.

3.2.1 Birds

After the Wadden islands of the Netherlands the Rhine-Meuse-Scheldt Delta estuary (the Westerschelde) and the tidal basin (the Oosterschelde) possess the most bird species in the Dutch lowlands. Birds feed on the variety of mussels, worms, shrimps, small fish and seaweed of the area. Because of the beneficial conditions the region is popular to migratory birds as well. The species present on the Galgeplaat consist mostly of waders, gulls, ducks and geese. These birds can be classified into four categories of food preferences; benthivores (birds that feed on benthos), herbivores (birds that feed on plants), omnivores (birds that feed both on plants and animals) and piscivores (birds that feed primarily on fish)[38]. The salt water zone is mostly characterized by waders. Towards the sweeter waters the number of waders become less. Brackish waters, next to the few waders, are also characterised by ducks such as the Eurasian Widgeon and the Common Shelduck. The sweet water zone is mostly characterized by the presence of omnivorous ducks such as the Winged teal. Table 3.1 contains a list of the birds classified according to the four consumer types and their habitats. The polyhaline, mesohaline and oligohaline zones in the table refer to the salinity of the respective waters, with polyhaline possessing the most salt content and oligohaline possessing the least [38].

See table B.1 in the Appendix B for an inventory of birds in the Scheldt estuary (images included). The number and species of birds are influenced by the presence of food but also by external factors such as climate conditions (such as severe winters) and a loss of breeding and foraging grounds (such as the loss of mudflats due to increased erosion rates). Figure 3.2 shows two bird species that can be seen on the Galgeplaat.
Table 3.1: Common bird classes per consumer types [38]

<table>
<thead>
<tr>
<th>Consumer type</th>
<th>Bird classes</th>
<th>Food source</th>
<th>Habitat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benthivore</td>
<td>Waders (Oyster catcher, Dunlin, Sanderling)</td>
<td>Macrobenthos</td>
<td>Mudflats (polyhaline zone)</td>
</tr>
<tr>
<td>Herbivore</td>
<td>Geese (Greylag goose), Ducks (Widgeon)</td>
<td>Salt marsh vegetation (Sea Clu-brush) Eelgrass</td>
<td>Salt-marshes (mesohaline zone)</td>
</tr>
<tr>
<td>Omnivore</td>
<td>Ducks (Mallard, Winged Teal)</td>
<td>Oligochaeta (aquatic worms), Organic Material</td>
<td>Oligohaline zone to sweet water zone</td>
</tr>
<tr>
<td>Piscivore</td>
<td>Cormorant, Tern</td>
<td>Clupeiforms (Herring-like fish)</td>
<td>Polyhaline zone</td>
</tr>
</tbody>
</table>

Figure 3.2: Eurasian Curlew (left) and Oystercatcher (right)
3.2.2 Benthic fauna and flora

As mentioned earlier in chapter 1, benthic organisms or benthos refer to the organisms that live in the benthic zone, defined as the lowest part of a water body such as the ocean or lake floor. Benthos can be divided into macrobenthos and microbenthos. Microbenthos refers to benthic organisms smaller than 0.1 mm. Benthic organisms are important because they comprise the bottom of the food chain and because some of them are responsible for stabilizing the sediment (and thus have direct influence on the morphology of the mudflat). Thus, if the Galgeplaat nourishment affects their population adversely than the regional food chain will be compromised and the morphology altered. As such the resettling rate of benthos is an important measure of how much impact the nourishment has on the ecology and morphology of the mudflat.

In the benthos communities one can also distinguish between the benthic fauna (or zoobenthos) and benthic flora (or phytobenthos). Common zoobenthos of the Oosterschelde are [39]:

- Lug worm (Arenicola Marina)
- Bristle worm (Spiio martinensis)
- Bristle worm (Scoloplos armiger)
- Bristle worm (Lanice)
- Laver spire shell (Hydrobia ulvae)
- Amphipod (Bathyporcais sp)
- Cockles (Cardiidae)
- Mussels (Mytilidae)
- Mud shrimp (Corophium volutator)
- Baltic clam (Macoma balthica)
- Nematodes or round worms (phylum Nematoda)
- Small crustaceans (benthic copepoda)

Not all of these zoobenthos are visible above the sediment. Lug worms and bristle worms for instance are always below the sediment but their position can be easily found by locating their coiled casts (defecated soil heaps) and blow holes. See figure 3.3 for two zoobenthos commonly found in the Oosterschelde.
3.2 Birds and Benthos of the Rhine-meuse-scheldt delta

Figure 3.3: Mussels (left) and benthic copepoda or ‘roeipootkreeftje’ in Dutch (right)

Macrophytobenthos (bigger than 1 mm) of the Oosterschelde species are mostly comprised of Algae (like red-brown and green algae) and sea grasses [40]. Common species in the Oosterschelde are:

- Bladderwrack (Fucus vesiculosus)
- Sea Lettuce (Ulva lactusa)
- Gut Weed (Ulva intestinalis)
- Eelgrass (Zostera marina)

See figure 3.4 for a visual of two macroalgae. Finally there are also the microphytobenthos of the Oosterschelde. Examples are for instance unicellular algae such as diatoms and the phytoplankton in the water. In science both the living and fossil assemblage of diatoms are useful indicators for changes in the ecology of an area. Also, they play an important role in the stabilization of the sediment as they secrete EPS (extracellular polymeric substance also known as biofilm) [41]. Phytoplankton are mostly important because they serve as the main food source of filter feeders such as cockles and mussels which are important for the quality of the water.

Figure 3.4: Sea Lettuce (left) and Bladderwrack (right)
3.3 Ecological features of Interest

The Argus Bio station images have already shown their use for morphological analysis of the Galgeplaat. The use of the images for biological features on the other hand is quite a new concept and several proposals have been made by the NIOZ institute, Deltares and IMARES. These included the extraction of features like macro algae cover, the determination of the recovery rates of macro algae and of Lanice (bristle worm) and Arenicola (lug worm) casts after storms, the detection of the number of birds and Arenicola casts present in an image and the detection of microphytobenthos (more specifically diatoms). To clarify, worm casts are coils of sand that have been egested (defecated) by the burrowing worms underneath.

Several aspects came forth during the discussion with people from NIOZ and IMARES. The use of NDVI (Normalised Difference Vegetation Index) to detect macroalgae cover using classification methods is already beyond the proposition phase, and is now under the process of being analysed and compared with Chlorophyll-a concentration values obtained from field sampling. Alongside this, the effect of Lanice cast numbers on the presence of macroalgae was to be determined (algae attach to Lanice casts) as well as the use of the NDVI values to say something about the state of the macroalgae (healthy, decaying).

Particular interest was shown in the identification of microphytobenthos. As of yet no clear distinction could be made between microphytobenthos and macroalgae by using solely NDVI. If microphytobenthos are to be detected using the Multispectral images it was advised the detection method should preferably use images and field samples (for validation) of months when there is little macro algae present (the macro algae cover the sediment and the presence of microphytobenthos underneath).

The detection of birds posed some questions, particularly in how the detection outputs would be validated. It was agreed that the Bio pan/tilt/zoom camera images followed by the Argus camera images would be the most suitable for the task as opposed to the Multispectral camera in which rarely a bird is present in the image. The problem that arises using Bio pan/tilt/zoom images is that the images present only a section of the overall PQ (permanent Quadrant). It would be very hard to count the birds in each section of the PQ’s in the field (as there are no lines in the PQ’s). Counting the birds in the entire PQ as a measure for validation would be much easier as the area is bounded by wooden poles. Suggestions were to use images of the entire PQ rather than use PQ section images. What has to be considered here is whether the resolution of those images will be sufficient enough to detect birds. Taking advantage of the movement of birds in an image would be beneficial (for instance by limiting the search space to where movement is detected), particularly if the resolution of the images is low. Currently the Bio pan/tilt/zoom camera scans over the 25 sections of the quadrant but it could be set such that it would take several subsequent snapshots of the same area. This requires some prior analysis in bird movement and speed to determine the time resolution necessary to detect the movement of a bird.

After discussion the following targets were set as the focus in this thesis:

Penelope Rammos  Master of Science Thesis
3.3 Ecological features of Interest

1. Case Study I: Microphytobenthos (more specifically Diatoms)

2. Case Study II: Birds

The main focus of the thesis is the first case study as it elicited more interest by the involved parties rather than the counting of birds. Case study II is elaborated as far as time permitted after finishing case study I.
Chapter 4

Existing Image Detection Methods

The objective of this chapter is to clarify the impeding factors to detection and with this determine which existing detection methods are most suitable for the detection of the selected targets (Section 1.3, sub questions on Detection method and software).

This chapter will start with a brief overview of the problems that should be dealt with during the detection process and is followed by a general description of existing image detection methods. The presented information serves as a mere basis for the selection of potential detection methods for the chosen case studies. The chapter is closed with the two case studies’ details and with this an analysis of the detectability of the characteristics of the targets will be laid out (Section 1.3, sub question on input data).

4.1 Impeding factors to detection

The detection procedure consists of several sub tasks (i.e. image preprocessing, processing and validation). They are dependent on the type of images used, the targets to be detected and the application of the detection. In the case of the Argus Bio platform there are several problems that need to be dealt with during the detection process. Examples being unfit images for detection (i.e. blurry images or images were raindrops on the lens are visible) or the presence of non-target features in the image. Figure 4.1 shows examples of unfit images, figure 4.2 shows examples of images with the presence of non target features.
Figure 4.1: Problem factors for the detection procedure; blurry images (left column) and raindrops on the lens (right column). These are images (clockwise) from the Multispectral camera, the Bio pan/tilt/zoom camera, and two of the Argus camera.

Figure 4.2: Problem factors for the detection procedure. First column shows images with bird reflections, second column shows the presence of a pole and in the third image the shadow of the platform is visible.
The main problems to be dealt with during the detection process are listed below:

- Varying illumination across an image (such as the brightness being darker at the edges than at the centre of the image)
- Varying illumination between the images (due to the different lighting conditions each day)
- Presence of other non-target features (poles, shadow of the platform)
- Images in which the target is occluded (for instance by raindrops on the camera lens)
- Varying backgrounds (such as before and after a storm, varying water levels, varying seasons)
- Blurry or low contrast images
- Camera movement (due to wind) in video data
- The size variation of the target in oblique and high oblique images
- Bird detection in images were the bird’s image is reflected on the water (which can lead to double detection) or birds that are in flight (landing or just taking off)
4.2 Inventory of Detection methods

Object detection from images is a well researched domain and as such a great variety of image processing methods have already been developed for applications such as face detection, vehicle tracking, vegetation detection and more. In a very broad sense these variety of existing techniques can be divided into three classes: The combination of morphological & thresholding techniques, template matching techniques and machine learning algorithms.

4.2.1 Morphological & Thresholding techniques

Morphological (referring to shapes) image processing is often used for image analysis, reconstruction and recognition applications, often in combination with thresholding techniques. Since they are often combined, the two were placed in the same class. Morphological and thresholding techniques take advantage of useful image components such as pixel intensity and colour, gradients between pixel values, textural properties of a group of pixels and shapes. A general description of these techniques is presented in the following.

Morphological techniques

Morphological operations work by applying a structuring element over an input image. A structuring element is a matrix made up of zeros and ones defining the neighbourhood of pixels that will be used in the operation. It slides over the image and processes the respective pixels, the shape and size of the structuring element defining the operation that will be performed. The most well known basic morphological operations are dilation, erosion, closing and opening of an image [42]. These can add and remove pixels to shape boundaries (dilation and erosion), remove peaks (erosion), close holes (closing) or open holes (opening) of a shape in an image.

Similar to structuring elements there are also filters (also known as kernels) not necessarily consisting of ones and zeros, which take morphological processing to the next level. With these filters for example edges can be detected by detecting sharp discontinuities in the image (examples of filters being the Canny, Sobel or Prewitt filters [43]), noise can be removed (by using median or averaging filters [44]) or certain illumination effects can be made negligible (by using a top hat filter, as was done in [45]). There are also filtering methods that correct for non uniform illumination across an image (for instance when one side of an image is brighter than the other side). Homomorphic filtering is one such method and aside from correcting for non uniform illumination it is also used for image enhancement and dynamic range compression [46, 47].

Thresholding techniques

Thresholding refers to the segmentation of an image by setting one or more thresholds to pixel values. Depending on whether the pixel complies to the threshold(s) it is classified to
the respective segment of the image. Determining the compliance of a pixel to a particular threshold is dependent on the classification method used.

Common classification/segmentation techniques are for instance the minimum distance classification method, the parallelepiped classification method and the Maximum Likelihood classification method [48]. These are all supervised classification techniques. This means the user is required to have foreknowledge of the classes to be classified (the user is asked to provide the algorithm with pixels representative of a class). Out of the three mentioned methods the maximum likelihood classification method is the most commonly applied for image segmentation [49]. It is based on the assumption that pixels in each class sample in the multidimensional space are normally distributed and the method uses Bayes theorem of decision making to determine the compliance of a pixel to a certain class [49, 50]. In essence what the method does is building a discriminant function for each class which determines the probability that a pixel belongs to each class. This is done by extracting the variance and covariance of the trained classes and using these to model a normal distribution for each class (with a mean vector and a covariance matrix). The advantage of the method is that it takes the variability of the classes into account. On the other hand it does assume the user selected signatures of a class are well selected (which is a subjective procedure).

Thresholding is also common in background subtraction techniques used in motion detection. In background subtraction techniques each image of an image sequence is compared to a background image. This produces an image in which each pixel represents the magnitude of change. The pixels whose change is over a particular threshold is classified as foreground (i.e the moving object) and otherwise as background. The simplest form of background subtraction is frame differencing which uses the previous frame as the background image. This is a method commonly used for its computational simplicity but is prone to the aperture problem [51]. The aperture problem refers to the failure to detect the motion of a homogeneous surface of an object [52]. Another problem is that objects must be continuously moving to be able to detect any variation in pixel intensity [53].

Other techniques used to estimate the background image are median filtering, approximate median and weighted mean determination of the image sequence [51, 54, 55]. The approximate median method compares each pixel in the current frame to an initial background model. If the pixel of the current frame is larger than the background pixel then the background pixel is incremented by one, otherwise it is decremented by one. It should then converge to a value where half of the pixels are larger and the other half smaller than its value (hence the name approximate median). This method does require a sufficient amount of frames to fully converge to the approximate median.

The weighted mean of an image sequence is computed with an adaptive filter (see equation 4.1).

\[ \mu_t = \alpha \ast P_t + (1 - \alpha) \ast \mu_{t-1} \]  

(4.1)

Where \( \mu_t \) is the mean computed up to frame \( t \), \( \alpha \) is the learning rate (values 0 to 1), and \( P_t \) the pixel value at time \( t \). This is also known as an exponential moving average (EMA) filter, named so because the applied weights decrease exponentially. In simpler terms it takes a particular percentage of the background pixel and a particular percentage of the current...
Existing Image Detection Methods

40 Existing Image Detection Methods

pixel to update the background [51, 55]. The higher the learning rate the quicker older observations are discounted. More complex but effective background subtraction techniques are the Mixture of Gaussian (MoG) approach and the Multimodal mean approach, which are particularly effective for images with dynamic backgrounds [56, 51]. The MoG approach is said to be computationally expensive and requires parameter optimization [53]. The Multimodal mean approach produces results comparable to MoG with a significant lower execution time but it too requires parameter optimization [51]. Despite their complexity, both the MoG and multimodal mean approaches can be used to produce background models in real time [56, 51, 57]. However, both need a sufficient amount of frames to produce a good background image.

Similar to the background subtraction principle, is the creation of optical flow images. With this method the motion of each pixel or a group of pixels is recorded in a matrix (its velocity magnitude and direction) using the difference in intensity of the same pixels in the different time frames [58]. The pixels that are “moving” together (same speed and direction) can then be identified as individual objects. Optical flow images are also prone to the aperture problem.

Importance of contrast for thresholding and morphological operations

The efficiency of thresholding and morphological processing is largely based on the degree of contrast of image components such as intensity, hue, saturation (first order level statistics, i.e. histogram properties) and texture (second order level statistics, i.e. co-occurrence of grey level frequencies between neighbouring pixels) [59]. That is why thresholding and morphological operations often come with the selection of proper colour models. Common example colour space models are RGB (red green blue), GREY, HSV (Hue Saturation Value), YCbCr (Y being the luminance and Cb and Cr are the blue and red differences colour components) or the L*a*b colour space (L for lightness and a and b for the colour dimensions) [60].

The efficiency of the HSV colour space for detecting predatory birds in aquacultural settings has already been analysed and compared against other methods by Louisiana’s State University [61]. For their application the HSV colour space proved quite successful compared to RGB and grey level images. However, the birds present in their images all possessed white coats.

Second order level statistics (texture) segment the image according to the variable textures present in the image. An example is image thresholding based on graph cut techniques (also known as min-cut) as was done in [62, 63]. Graph cut techniques are energy minimization methods (in this case to find a globally optimal segmentation solution) based on graph theory [64, 65, 66].

Alongside the perceptible colour bands also bands out of the human perception field can be used for object recognition (examples are the infrared and near infrared bands) or bands which can be artificially created (by using Principal Component Analysis for instance). Often, to discriminate components more effectively from one another different bands can combined, an image of this type is known as a colour composite image. The bands of the Landsat Thematic Mapper have often been combined in a variety of ways to produce such false colour composites.
A common false colour composite used to detect vegetation is replacing the red band with the near infrared band, the green with the red band and the blue with the green band. An example of this false colour composite using an image from the multispectral camera of this study can be seen in figure 4.3.

![Figure 4.3: RGB (left) and false colour composite (right) using image data acquired October 31 at 12:30 PM from the Multispectral camera. The red in the right image indicates the presence of algae](image)

Other methods which are used to detect vegetation by using the red and near infrared bands are the Normalised Differential Vegetation Index (see figure 4.4 for an example), the Transformed Normalised Differential Index (TNDVI) and the Ratio Vegetation Index (RVI). From these methods the NDVI is by far the most commonly used vegetation index, its values ranging between -1 and 1. The NDVI is defined in equation 4.2:

$$NDVI(x, y) = \frac{NIR(x, y) - Red(x, y)}{NIR(x, y) + Red(x, y)}$$ (4.2)

Where NIR(x,y) refers to the NIR band value at pixel location (x,y) and Red(x,y) refers to the red band value at pixel location (x,y) in an image. The rationale is that the chlorophyll in plants (chlorophyll-a) highly absorbs in the visible spectrum and strongly reflects in the infrared region. When plants are in the field of view of the spectral camera, the received red radiation is low and the received NIR radiation is high producing a high NDVI. Values between 0.3 to 0.8 generally indicate dense vegetation, values between 0 to 0.2 generally indicate soils or mud and values below zero indicate the presence of water, clouds or snow in the image. The NDVI, TNDVI and RVI methods do show some susceptibility to background soil interference [68]. SAVI (Soil adjusted Vegetation Index) and MSAVI (Modified SAVI) are vegetation indices developed in an effort to eliminate the interference of the soil. Other common indices are the IRGVI (Infrared Green Ratio Vegetation Index), the LIRGVI (the logarithm of IRGVI) and the specific index used in [69, 70] for microphytobenthos detection. The specific index used in [69, 70] is twice the 586 nm spectral band over the sum of the 496 nm and 675 nm band. However, the multispectral camera does not specifically record the 586 nm band and thus this index will not be used.
4.2.2 Template Matching

Template Matching, makes use of template images (also known as training sample images) which are compared with the input image. Template images contain the target feature often in different orientations, illumination conditions and sizes. In essence this technique compares two images (by determining their correlation values) to see whether the desired target was depicted in the input image. Different properties can be compared, for instance texture properties, edge properties, spectral reflectance properties, shape properties or properties in relation to one another (i.e. shape edges distance from the shape’s centre point). Common methods used in template matching operations are SIFT (Scale Invariant Feature Transform), KLT (Kanade-Lucas-Tomasi), the use of Moment Invariants, Hough transforms and Fourier Descriptors.

Template matching techniques are often combined with morphological and thresholding techniques for object recognition, an example being the detection of benthic crustaceans in [45] where SIFT and Fourier descriptors were used in combination with median and top hat filtering to remove noise and illumination effects respectively. Some template matching techniques such as KLT, are also used for following the motion of objects in a sequence of image snapshots or video data. This is also commonly referred as the ‘tracking’ of an object. Hough transforms can also be used for motion detection (such as the velocity Hough Transform or VHT). A common tracking algorithm is the Kalman filter. Tracking of an object is useful when target occlusion in the video sequence is high.

Although motion detection techniques were earlier classified under the morphological and thresholding class because they use image components (intensity, grey values, etc.) this technique can fall under template matching as well because invariant features of targets can also be used to detect motion (feature based approach versus gradient based approach).
4.2.3 Machine Learning Algorithms

Machine learning algorithms refer to machine systems that learn by adjusting their internal structure or data such that their expected performance is improved [71]. There are a variety of machine learning algorithms types but two in particular have been used before for object recognition and classification in imagery, they are the supervised and unsupervised learning algorithms. Under these two categories examples are Artificial Neural Networks (ANN’s) commonly used to find patterns in data (for instance to recognise birds as was examined in [61]), Support Vector Machines (SVM’s) which are commonly used for data classification [72], Naive Bayes classifier and clustering (such as k-means clustering, Ward clustering, density based clustering and distribution based clustering) used for image segmentation and object detection. Clustering is often confused with classification but this is incorrect. With classification the classes are predefined (it involves labelling of the classes beforehand) whereas clustering has no labelling involved [73]. It is the division of data in groups or so called clusters of similar objects [74]. Clustering techniques are seen as unsupervised learning algorithms of which the most commonly used types are the hierarchical clustering methods (such as ward clustering) and the data partitioning clustering methods (such as k-means clustering). Unsupervised means that no "teacher" is required to supervise the performance of the algorithm. Hierarchical clustering establishes a cluster taxonomy whereas partitioning clustering establishes a set of ‘flat’ partitions (i.e. non-overlapping clusters) [74].

Aside from the clustering algorithms which are one of most commonly used from the machine learning algorithms also the ANN’s are popular for image processing applications. ANN’s are learning algorithms that make use of networks of interconnected nodes or so called ‘neurons’. Each neuron is a computational unit which receives an input and processes that input to produce an output. The processing is either simple (for instance by summng the inputs) or very complex (each neuron could contain a network in itself) [75]. In general the ANN working principle consists of the learning phase (also commonly referred to as the training of the neural network) and the phase were the trained neural network is used (for instance for clustering). By modifying the weights of the synapses (connections between neurons) an ANN adapts its network structure according to the flow of data given to the network during the learning phase.

The Self Organising Map

A popular ANN model is the SOM (self organising map) which was developed by Teuvo Kohonen in the early 1980’s. SOM’s are unsupervised competitive learning networks. Unsupervised in this case means that no supervision of the performance of the algorithm during the learning phase is required. A Competitive learning network system refers to the technique in which the neurons in the artificial network compete with each other for the exclusive right to respond to a particular input pattern [76]. With response in the case of a SOM it is meant that the winning neuron is allowed the right to modify the weight of itself and its surrounding neighbours so that the network ‘fits’ the input data flown into the network better. Self Organizing maps for image segmentation are particularly useful for clustering multiple targets. For segmenting a single target in an image another approach is advisable due to the complexity of the method.
To further elaborate on the SOM method, in the learning phase each neuron is given an initial parametric reference vector with the same dimensionality as the input data. This is either done randomly (the algorithm initializes the reference vectors to random values) or linearly (the algorithm initializes the reference vectors in an orderly fashion along a two-dimensional subspace spanned by the two principal eigenvectors of the input data vectors) [77]. After this 'example' input data vectors are consecutively flown into the neural network. For each input vector the neurons compete to be the Best matching unit (BMU), the neuron whose parametric reference vector is closest to the input data vector. This is determined with the smallest Euclidean distance between the reference vector and the input data vector in N-dimensional space (the dimensionality of the input data). Once the best matching unit is selected it gets the exclusive right to modify the weight (or rather its parametric vector) of itself and its surrounding neighbours so that the next time a similar input is flown into the network it is drawn into the vicinity of the previous best matching unit. Basically, it 'updates' the parametric reference vectors of the BMU and its neighbours in accordance with equation 4.3:

$$m_i(t + 1) = m_i(t) + hci(t) \ast [x(t) - m_i(t)]$$

(4.3)

Where \( t \) the discrete time coordinate, \( m_i \) the parametric reference vector, \( hci \) the neighbourhood kernel (which decides which parametric reference vectors need to be updated, i.e. the BMU and its neighbours) and \( x \) an input data vector at time t.

The main parameters to be set prior to running the algorithm are the size of the network, the topology type of the network (hexagonal or square), the neighbourhood function (which decides which neighbouring neurons will get modified along with the Best matching unit, this is contained in the \( hci \) kernel) and the learning rate (which describes the degree of modification of the weights of the best matching unit and its neighbours and is also contained in the \( hci \) kernel).

The working principle is best demonstrated with an example (based on the schematic example presented in [78]). Let's assume a three by three neural network and a random input dataset of 4 vectors of dimensionality three (see figure 4.5). These vectors could for instance represent the RGB values of four pixels. In feature space these vectors form two clusters (vector 1 and 4; vector 2 and 3) representing two different colour types (for instance yellowish and purplish colours). As such the result should be a SOM with two distinct clusters. The neurons of the three by three network are first initialised with parametric reference vectors (also commonly referred to as weights) of dimensionality three. Then the vectors are flown into the network one by one. The first vector finds its BMU on the 2nd row first column. That neuron then gets the exclusive right to modify its weights towards the input vector. Additionally the weights of its single-connected neighbours are slightly modified also (the neighbourhood type is defined by the user before the training). The second vector finds its BMU in the upper right corner and adjusts its weight and that of its neighbours accordingly (using equation 4.3). When the third vector is flown into the network it is drawn into the vicinity of the previous BMU due to its similar nature with the 2nd vector. Its BMU becomes the 2nd row last column. The weights are adjusted accordingly but this includes neurons whose weights were previously modified by vector 1 and 2. Note how the middle neuron becomes blank. This is because
members of two different clusters have tried to pull it in either direction. This type of neuron is known as a separator neuron, a neuron that separates clusters [78]. Finally the last vector is flown into the network and it is drawn into the vicinity of the BMU of the first vector. The result is a map that self organized itself into two clusters.

![Figure 4.5: Example of a SOM training. Here a three by three neural network is trained with four input vectors which in feature space form two distinct clusters. Schematic based on example in [78].](image)

Examples of applications were Kohonen Self Organising Maps were used are the classification of landforms on Mars [79], the visualisation of human movement in attribute space and automatic road extraction from remotely sensed imagery [78].

### 4.3 Pros and Cons of existing detection methods

In order to select detection methods for the two case studies the benefits, pit falls and applicability of the methods have to be considered. Table 4.1 revises the general advantages and disadvantages that come with the three aforementioned detection classes. Note that there can be some exceptions, it is but a generalised table of pros and cons, and thus not descriptive of each individual method that fall under these classes. Also note, methods that fall under one of the three classes may also fall under another of the classes (such as motion detection which can fall under both the template matching class and the morphological and thresholding class).

Recalling the research question, what is strived for is a method/or methods which requires the least user input, and simultaneously provides a usable result for both case studies. Important for the identification of microphytobenthos is to make sure it is not confused with macroalgae which can also be present in the images (the problem will be further elaborated in the following section). As such a method that is good in dealing with images that possess different targets (which however possess similar features) would be an asset. For the detection of birds the most significant problem that arises is the camouflaging coats most wading birds possess (lack of contrast). A method or a combination of methods that do not only use colour as a means for detection is advisable.
<table>
<thead>
<tr>
<th>Detection Class</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphological &amp; Thresholding techniques</td>
<td><em>Good in handling multi-modal data</em>&lt;br&gt;<em>Fast once training is performed</em>&lt;br&gt;<em>Can handle large training sets</em>&lt;br&gt;<em>Can be sensitive to quantization</em>&lt;br&gt;<em>Can work well in noisy environments and situations of occlusion (position)</em>&lt;br&gt;<em>Can perform tasks a Linear program cannot perform</em>&lt;br&gt;<em>Little to no human intervention or assistance necessary after the training</em>&lt;br&gt;<em>Fast once training is performed</em>&lt;br&gt;<em>Good in handling multi-modal data</em>&lt;br&gt;<em>Can perform tasks a Linear program cannot perform</em>&lt;br&gt;<em>Little to no human intervention or assistance necessary after the training</em>&lt;br&gt;<em>Fast once training is performed</em>&lt;br&gt;<em>Good in handling multi-modal data</em>&lt;br&gt;<em>Can perform tasks a Linear program cannot perform</em>&lt;br&gt;<em>Little to no human intervention or assistance necessary after the training</em>&lt;br&gt;<em>Fast once training is performed</em></td>
<td><em>Problems can arise with low contrast images</em>&lt;br&gt;<em>Can be sensitive to noise and quantization</em>&lt;br&gt;<em>Dependent on object shape and size</em>&lt;br&gt;<em>A threshold has to be determined</em>&lt;br&gt;<em>Can be sensitive to noise and quantization</em>&lt;br&gt;<em>Dependent on object shape and size</em>&lt;br&gt;<em>A threshold has to be determined</em>&lt;br&gt;<em>Can be sensitive to noise and quantization</em>&lt;br&gt;<em>Dependent on object shape and size</em>&lt;br&gt;<em>A threshold has to be determined</em>&lt;br&gt;<em>Can be sensitive to noise and quantization</em>&lt;br&gt;<em>Dependent on object shape and size</em>&lt;br&gt;<em>A threshold has to be determined</em>&lt;br&gt;<em>Can be sensitive to noise and quantization</em>&lt;br&gt;<em>Dependent on object shape and size</em>&lt;br&gt;<em>A threshold has to be determined</em></td>
</tr>
</tbody>
</table>
4.4 Case studies

To prevent duplication with work done by NIOZ and IMARES and to organise the research constructively, two targets were selected in chapter 3. The two targets that were selected were microphytobenthos and birds. Two case studies have been formulated for these targets, these are:

1. Case Study I: Microphytobenthos cover
2. Case Study II: Bird number determination

The distinction between the case studies is important as the methodologies of the respective case studies differ. A case study describes the objective, the problems involved, the target details and the strategy of a specific research problem. In each of the case studies the properties of the target under question will be elaborated and potential strategies (i.e. the detection method(s) most suitable for the target in question) will be laid out.

4.4.1 Case Study I: Microphytobenthos cover

The aim of this case study involves the automatic detection of microphytobenthos. The objective, limiting factors, details on the target and strategies for detection are mentioned below.

Objective

*Detecting the microphytobenthos cover using the Multispectral camera of the Argus BIO platform.*

Problems to be encountered

- Low contrast and blurry images
- Different illumination across the image and between different images
- Distinguishing between macro and micro algae

Microphytobenthos properties

On muddy substrates and sandbars, diatoms (a group of microalgae) are the dominant microphytobenthos [80, 81] and as such are what will be essentially detected. Diatoms are the main primary producers along side dinoflagellates and cyanobacteria and thus play a major role in the ecology of the Oosterschelde. Diatoms are often described as microalgae, mostly unicellular but can cluster together to form colonies. They are encased in cell walls made of silica called frustules which are naturally present in a variety of shapes. As they are microbenthos, i.e. smaller than 0.1 mm, they are not visible individually. On the field one can visually recognise groups of diatoms by their gold brownish colour film on top of the surface known as biofilm or EPS (extracellular polymeric substance) which is excreted by diatoms. For an indication of how individual diatoms look like and the biofilm see figure 4.6.
An important property of diatoms is their distinct photosynthetic pigments. These pigments are chemical compounds which reflect only certain wavelengths of visible light, which makes them appear a certain colour. What is more important is their ability to absorb certain wavelengths, the energy of which is used to photosynthesise \[84\]. Diatom pigments consist mostly out of chlorophyll-a and chlorophyll-c, beta-carotene, fucoxanthin and diadinoxanthin. The property of chlorophyll-a (which is a principal photosynthetic pigment and is common to all algae) is that it has absorbance maxima in the violet band (at around 431-33 nm) and in part of the red band (at around 670 nm) and shows high reflectance in the NIR (at around 720 nm) and green bands (550 nm) \[69, 85\]. As chlorophyll-a is only present in plants (or cyanobacteria) it is often used as a proxy for biomass for benthic algae and phytoplankton \[86\]. However, as chlorophyll-a is common to all algae, macroalgae present similar reflectance properties as microalgae (diatoms) and this makes it hard to distinguish the one from the other \[18\].

Fucoxanthin is a slightly more characteristic pigment as its presence, other than its high concentration in diatoms, is found elsewhere only in brown (macro) algae. Fucoxanthin spectral properties shows high absorbance at around 450 nm (in the blue band)\[87, 88\]. Diadinoxanthin displays high absorption maxima at 420 nm, 444 nm and 474 nm \[3\]. The absorbance spectra of fucoxanthin, chlorophyll-a and diadinoxanthin can be seen in appendix C. Because fucoxanthin is found elsewhere only in brown macroalgae it has been used as a taxonomic pigment marker of diatoms (as for instance in \[89\]). Nevertheless, its presence in brown macroalgae makes fucoxanthin less suitable as a pigment marker for sites where both diatoms and brown macroalgae are present.

That is why fucoxanthin is often combined with another marker such as chlorophyll-a (a pigment marker for algae) to obtain details of a specific algal class. The fucoxanthin to chlorophyll-a ratio is a common proxy for the diatom biomass (representing a specific algal class from the algae group). Sediments dominated by diatoms typically possess ratio values of 0.6 to 0.8 \[90, 91\]. The pigment ratio fucoxanthin to chlorophyll-a representative of diatoms show their highest relative abundance at the end of March \[92\].

The disadvantage of using pigment concentrations as proxies for diatom biomass is that in situ sampling is required. In situ sampling for pigments is often limited or not feasible due to limited accessibility of the study site, which is why the potential of remote sensing to map

*Figure 4.6: A visual of individual diatoms (left) and diatom biofilm cover on a mudflat [82, 83]*
benthic algae has often been researched. According to literature remote sensing can provide for a quantitative estimation of biomass (i.e. pigment concentration) from reflectance data [69]. For phytoplankton in the sea or canopies this has typically been achieved with non linear regressions between biomass (of the pigment chlorophyll-a) and vegetation indices such as the NDVI (bands obtained from satellite imagery). The non linearity is because the NDVI for such targets saturates for high chlorophyll-a values. The range of NDVI observed on mud substrates does not reach the level of saturation and can be assumed to possess a linear relationship with chlorophyll-a.

Aside from their photosynthetic pigments another defining feature of diatoms is their motion. Diatoms follow a vertical migration rhythm. They surface when the tide is gone (upward migration) and move down about an hour before the tide comes (downward migration). There is a lot of discussion surrounding the matter as to what controls this vertical migration and how, it being either a diurnal rhythm (controlled by light intensity), the tidal rhythm, both of these combined, salinity and/or temperature [93, 94, 95]. However, it should be considered whether vertical migration of diatoms is visible in the images from the Argus Bio platform in the first place. Seasonal occurrence of diatoms on the Galgeplaat is also noticeable.

To sum up, the most distinctive features of diatoms are:

1. Spectral reflectance properties of microphytobenthos pigments (chlorophyll-a and fucoxanthin)
2. Biofilm colour (gold-brownish)
3. Chlorophyll-a and fucoxanthin have been used as proxies for microphytobenthos biomass
4. Vertical migration
5. Diatom relative abundance highest at the end of March

**Preprocessing strategies**

Preprocessing the images is generally advisable as the raw images can be blurry and occasionally possess different brightness at the edges. A method that does this is for instance the homomorphic filtering method mentioned earlier in subsection 4.2.1. Homomorphic filtering normalises the across brightness in a digital image and simultaneously enhances the contrast [46]. The images are also subjected to different ambiance conditions (as they are captured at different times) and as such a particular colour normalisation method should be applied when processing the images to make them comparable. Examples of colour normalisation techniques are grey world normalisation, histogram equalisation, histogram specification or radiometric normalisation methods [96, 97, 98]. Note that there is a difference between across brightness normalisation and colour normalisation. Across brightness normalisation normalises the brightness across one image. Colour normalisation normalises the brightness across multiple images with respect to the brightness of a selected reference image.
Detection strategies

Both the Morphological & Thresholding techniques class and the Machine Learning algorithm show potential for this case study (the Template matching techniques being unsuitable as diatom biofilms do not have a particular fixed shape). For the Morphological & Thresholding class, strategies that can be used to detect microphytobenthos is the use of vegetation indices or the use of false colour composites mentioned in chapter 4. Attention should be paid to the small time lag between the RGB and NIR images from the Multispectral camera (macroalgae might have moved slightly due to small surface water currents or wind). Also, attention should be paid to the presence of macroalgae whose spectral properties are not much different from microphytobenthos. Taking advantage of the colour of the diatom biofilm in this case would be beneficial to detect only the diatoms. However, despite the high contrast with respect to green macroalgae the colour of the diatom biofilm is quite similar to that of red-brown macroalgae. Creating a mask for the macroalgae prior to the detection of diatoms could thus opt for a solution for better distinction between the microalgae and macroalgae.

Another approach is the use of a Machine learning algorithm, the SOM (self Organising Map) method shows potential. This because it is good in finding underlying relationships of multivariate datasets. In this way it helps the user to gain more insight into which variables exactly contribute or aid human perception in distinguishing between macroalgae and microalgae.

Validation strategies

Referring back to chapter 3, it was advised by IMARES and NIOZ to use images that have little macroalgae present to be able to validate the result properly. The reason for this is that in situ percentage cover of microphytobenthos was not estimated when covered by macroalgae. If, on the other hand, the presence of macroalgae could be quantified, then that could be used as a quality measure of the algorithm’s microphytobenthos outcome (i.e. the more macroalgae the less reliable the detection output).

Macroalgae cover is highest in August, while diatoms peak around the end of March as aforementioned [92]. To capture the diatom seasonal behaviour and to be able to validate it with the in situ monthly data, images from each month of the year need to be used. This means there will be some macroalgae present in some of the images and thus it should be made sure that microphytobenthos is not confused with macroalgae. A clear vertical migration behaviour of diatoms is not visible in any of the images from morning to evening and a such it should be considered whether vertical migration can be monitored at all using images from the multispectral camera.

The output will need to be validated with the actual field samples (ground truthing). It would thus be advisable to choose images taken on the sampling day or images close to the sampling day. Also, to be able to compare the results of different months, the corresponding images should be normalised (calibrated) prior to the detection process. To summarise the approach for microphytobenthos we have the following strategies:
4.4 Case studies

- Preprocessing to remove noise and across image illumination effects and enhance contrast
- Radiometrically Normalising images prior to the detection process
- Approach 1: Use RGB and NIR bands (and extracted vegetation indices) to detect diatoms
- Approach 2: Use a SOM to detect diatoms
- Use of images on or close to the in situ sampling days
- Validate with the in situ sampled data

4.4.2 Case Study II: Bird Number determination

The aim of this case study concerns the automatic detection of the number of foraging birds. A bird census typically includes the species of the birds as well. Due to time constraints identifying bird species will not be the primary focus. The objective, limiting factors, details on the target and strategies for detection are mentioned below.

Objective

*Detecting the number of foraging birds in an image frame using a few consecutive images of the BIO pan/tilt/zoom camera of the Argus BIO platform.*

Problems to be encountered

- Low resolution, low contrast and blurry images
- Dealing with camera movement (due to wind) in video data
- Dealing with the size variation of the target in oblique and high oblique images
- Different illumination across the image and between different images
- Reflections of birds in the surface water
- Birds whose coats are camouflaging
- Dealing with resting birds or birds in mid air (they don’t count as foraging birds)
- Dealing with birds which occlude each other
- Birds are not rigid objects (shape changes with movement)

Bird Properties

The bird species counted currently on the Galgeplaat are the foraging wading birds. Resting waders (inactive), gulls and terns were not included in the bird count. In general, the presence of birds is either betrayed by their striking colour(s) or by their movement; the latter being the dominant factor for birds with camouflaging colours.

Colour becomes perceivable in this case for either of the following reasons: 1. the bird colour produces contrast with respect to the background (as is the case for white gulls with respect to the grey background) or 2. the bird’s colours produce contrast with each other (i.e. the
Oystercatcher’s black and white coat and distinct red beak and legs).

Movement becomes perceivable due to the contrast perceived when intensities, textures and/or shapes shift between consecutive frames. Birds on the Galgeplaat are there to forage or rest. Foraging comes with specific motions such as a 2D horizontal movement over the plain (the wading) and vertical movement of the bird’s head motion (the feeding). This introduces the earlier stated problem of a non rigid target body. Wading speed can also be used to differentiate between birds in flight (high velocity) and foraging birds (low velocity).

Other striking features of wading birds are their distinctive legs and beaks (i.e. waders such as the Eurasian curlew and the Grey plover have long legs and long beaks). Texture too, of the birds coats, can be descriptive of their species (take for instance the spotted nature of the Dunlin and the Eurasian curlew in comparison with the texture of an Oystercatcher). Colour of bird coats can also be descriptive of a species (i.e. gulls are generally white) but attention should be paid to using colour for bird species with camouflaging coats. The size of a bird also says something about the species, i.e. compare the large Eurasian Curlew with the smaller in size Dunlin. Here too attention should be paid to the fact that Bio pan/tilt/zoom images are highly oblique and as such depending on the position of the bird in the image the bird size is variable. Also, bird numbers peaks in October (which is also why all bird counting by Habitat Advies was performed in the month of October from 2009 to 2011).

Noting these features down we have that the target’s distinctive properties are:

- Colour
- Shape (beaks, neck, legs)
- Size
- Texture (i.e. the oyster catchers coat)
- Foraging motion
- Bird numbers peak in October

**Preprocessing strategies**

Preprocessing the images as mentioned earlier is generally advisable as the raw images can be blurry and occasionally possess different brightness at the edges. For across brightness normalisation and contrast enhancement the same method as proposed for case study I can be used, i.e. the homomorphic filtering method. Similarly, one of the earlier colour normalisation techniques can be used to normalise the brightness across multiple images by using the brightness of a reference image.

**Detection strategies**

The use of colour, position and bird size show potential in the detection of bird numbers in an image. However, considering the lack of contrast in some of the images and the camouflaging nature of most wading birds, it is advisable to use them in combination with a motion detection technique. This could be used on a mere two consecutive images or many consecutive
images of the same area. Because the Bio pan/tilt/zoom camera only scans over 5 quadrants around the Argus Bio platform it was arranged that the settings of the camera be changed to additionally take consecutive snapshots of the same scene as well. This was done in the time between the traditional consecutive scannings and resulted into image sequences over a time span of 10 seconds (2 frames per second) for a few hand selected scenes.

Recalling the background subtraction principle a few showed potential for this particular application. Most of the techniques mentioned require many frames to produce a reasonable background model. However, the available image sequences are only 20 frames in length which is much less than the number of frames of the videos these methods were tested on. The MoG approach and Multimodal Mean approach are considered to possess the best performance for detecting the foreground (in particular for images with dynamic backgrounds). However, MoG based background subtraction is also known for its high demand in CPU memory and both methods require a sufficient number of frames to model the background (a training phase). Because the Bio pan/tilt/zoom camera images do not really possess dynamic backgrounds (such as waving trees) and due to the small number of frames available per scene it was decided not to use a Gaussian Mixture model based background subtraction or Multimodal mean adaptive backgrounding.

Frame differencing shows potential in this case because aside from its low computational load does not require a training phase to model the background (it makes use of the previous frame). It does require the setting of a particular threshold.

Median filtering, the approximate median background subtraction method or the weighted average adaptive filtering method also come to mind to estimate the background. However, approximate median background subtraction requires a sufficient amount of frames to fully converge to the actual median (a pixel can be incremented at most by the number of frames). Since only a few consecutive frames are available this method is not appropriate. Median filtering works better with many frames but it is more robust for few frames than approximate median background subtraction. The use of weighted mean adaptive filtering could also be considered as the learning rate can be adjusted such that older observations are not discounted as fast.

The selected method(s) will have to be adapted to fit this particular application. The use of colour and object size can be incorporated in the motion detection techniques. Using texture of the birds coats shows less potential as the size of the birds in the images are too small to detect i.e. individual specks or dots on the birds coats.

Validation strategies

Since there is no actual in situ data for the particular area at which snapshots are taken validation will mostly be done through visual inspection. A confusion matrix (also referred to as error matrix or contingency matrix) is commonly employed to visualise and organize information used to assess the quality of a classification method [99]. A confusion matrix contains information about actual and classified targets. It displays the number of false positives, false
negatives, true positives, true negatives, and one can extract the error rate (number or incorrect detections/total number of detections) and the overall accuracy rate per class (number of correct detections/total number of detections) [100]. A confusion matrix or the extraction of true positives, false positives and negatives is done per frame instead of per video sequence. Perhaps also useful is to extract the number of accurately identified frames, the number of frames where wrongly identified birds are present and the number of frames were birds were present but not detected. This could be a useful indication when assessing the performance of the methods used.
This chapter explains the methods used during the procedure of diatom detection. It starts with the specifics of the raw images and the in situ sampled data. It is then followed by setting up the test image set to be used and the actual diatom detection procedure. This detection procedure has been split up into three task blocks, see figure 5.1 for clarification. The input being the raw images from the Multispectral camera of the Argus BIO platform and the output being: i) The classification result (visual output) and ii) A table containing information on the detected target(s) that can be used to compare to the in situ data (tabular output).

The aim of this chapter is to primarily determine the feasibility of automatic diatom detection using the available terrestrial imagery from the Argus BIO platform. This will be done by evaluating the quality of the output of the used approach(es) to detect diatoms. The chapter is additionally focused on the sub questions concerning the detection method and software used with particular focus on the practicability and suitability of the two (see chapter 1).

5.1 Raw image Specifications

For this case study the images of the Multispectral (MS) camera were used. This is because i) it is the only camera that possesses two Near Infra Red bands and ii) because it is the only camera that points towards the field plot where the in situ sampling takes place. The former is important because the NIR band is required for most vegetation indices and the latter is important in order to be able to validate the detection output. The raw multispectral images have several properties (also mentioned in chapter 2). These are quickly summarized below:
The MS camera captures the RGB channel (400 nm-700 nm).
The MS camera captures the NIR channel centered at 775 nm.
The MS camera captures the NIR channel centered at 860 nm.
There is a small time delay between the RGB, NIR 1 and NIR 2 images (1 to 4 seconds).
The spatial resolution of the images is 1384 (horizontal) by 1036 (vertical) pixels.
The pixel size is 6.45 $\mu$m.
The images possess 4 text bands with specs of the image, camera and location.

The text bands present on the images (two at the top corners of the image and two at the bottom corners of the image) describe the date and time the image was taken, the time reference system (UTC in this case), the location, some camera specifics (tilt and zoom) and the image id.

### 5.2 The in situ sampled data

To validate the output of the automatic detection algorithm in situ sampled data will be used. The field sampling days held in 2011 were:

- March 22nd
- May 4th
- June 17th
- July 20th
- August 15th
- September 27th
- October 26th
- November 15th
- December 19th

The in situ data consist of pigment concentration samples and estimations of the macroalgae coverage analysed in six grids that are spread over the field plot. The pigments we speak of here are the photosynthetic pigments mentioned earlier in chapter 4.

The multispectral images display the field plot right below the tower, however, the in situ pigment sampling and coverage estimation is performed in six grids whose location is spread over this field plot. See figure 5.2 for the location of these grids.
5.2 The in situ sampled data

Figure 5.2: Location of the six sampling grids in the plot below the Argus tower
For each grid the presence of a variety of photosynthetic pigments are tested. In our case only two pigments are of importance, fucoxanthin and chlorophyll-a. The ratio of fucoxanthin to chlorophyll-a is a common proxy for diatoms. Biomass (a.k.a pigment concentrations or pigment ratios), as explained in section 4.4.1, can be quantitatively estimated from the NDVI. Figure 5.3 shows the average pigment concentration over time of chlorophyll-a and fucoxanthin. For the month December the data about the pigment concentration was not yet available. Figure 5.4 displays the fucoxanthin to chlorophyll-a ratio over time. Note the high abundance of diatoms in March, September and November in this figure. Also interesting for ecological purposes is the range of the ratio values (0.22 to 0.49) which is considerably smaller than observed by [90, 91] for sediments dominated by diatoms (refer back to section 4.4.1). This is possibly due to different environmental variables of the Galgeplaat site as compared to the sites of the literature (other intertidal sediments of the Oosterschelde and Westerschelde).

![Figure 5.3: In situ measured Pigment concentrations over time. Pigment concentrations of December were not yet available.](image)

For each grid also an estimation (by visual inspection) is made of the amount of macroalgae cover in each cell of each grid. That data is used to extract a percentage cover value per grid for each sampling day. Figure 5.5 displays the macroalgae coverage estimated from the in situ data over time (i.e. it shows the average of the six grids outcomes for each sampling day). The error bars indicate the ratio of the standard deviation of the 6 grid outputs to the square root of 6 (the amount of grids).

In situ cover estimation was done for the diatom biofilm as well but with no estimations for the months in which macroalgae were abundant (the summer months June to September).
5.2 The in situ sampled data

As a result it was decided not to use the in situ diatom coverage estimations as too few months remain for proper validation (i.e. only in situ pigment data will be used). Coverage estimations will only be used to validate the macroalgae cover detected in the images.
5.3 Test Image set

For the detection of microphytobenthos (which comes down to the detection of diatoms for the Galgeplaat) a set of images was selected from the year 2011 whose date is on or near to the in situ field sampling days held in 2011. This allows for proper validation of the detection output with the in situ data. The field sampling days of 2011 were mentioned in the previous section. It was attempted that the capturing time of the test set images all be close to each other.

The selected pictures (RGB, NIR1, and NIR2) for the test set and their details can be seen in table 5.1. Note the three different times for each date, this results from the small time delay between the capturing of the RGB and NIR images by the Multispectral camera. Note also that the images are not selected on the actual in situ sampling date. This is because the sampling crew is present in the images on the sampling dates or because the time of capture is too far off from the other test set images capturing times. The selected images can be seen in table 5.2. A bird is present in the two NIR images of August 16th. For June there were no images for June 16th and in the images of June 18th the mudflat is submerged under water. The December images have poor quality hence the reason why the selected image of December is 4 days before the actual sampling day.

Figure 5.5: Average (of the six grids) in situ percentage coverage estimations of macroalgae.
### Table 5.1: Test set Case Study I: dates, times and corresponding remarks

<table>
<thead>
<tr>
<th>ID</th>
<th>Date</th>
<th>Time</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23-03-2011</td>
<td>13:00:45, 46, 48 PM</td>
<td>High season of diatoms</td>
</tr>
<tr>
<td>2</td>
<td>05-05-2011</td>
<td>12:30:42, 43, 44 PM</td>
<td>Presence of sand ripples</td>
</tr>
<tr>
<td>3</td>
<td>15-06-2011</td>
<td>10:24:42, 43, 45 AM</td>
<td>Presence of macroalgae</td>
</tr>
<tr>
<td>4</td>
<td>19-07-2011</td>
<td>12:30:42, 43, 45 PM</td>
<td>Presence of macroalgae</td>
</tr>
<tr>
<td>5</td>
<td>16-08-2011</td>
<td>11:30:43, 44, 45 AM</td>
<td>Presence of macroalgae &amp; bird</td>
</tr>
<tr>
<td>6</td>
<td>28-09-2011</td>
<td>10:30:42, 43, 44 AM</td>
<td>Presence of brown algae</td>
</tr>
<tr>
<td>7</td>
<td>27-10-2011</td>
<td>10:42:41, 43, 44 AM</td>
<td>Presence of brown algae</td>
</tr>
<tr>
<td>8</td>
<td>16-11-2011</td>
<td>11:00:44, 46, 47 AM</td>
<td>Presence of brown algae</td>
</tr>
<tr>
<td>9</td>
<td>15-12-2011</td>
<td>12:24:42, 46, 49 PM</td>
<td>Presence of 3 poles (upper right)</td>
</tr>
</tbody>
</table>
Table 5.2: Image test set for Case Study I from March to December 2011 (April Excluded). Columns from left to right: RGB (400-700 nm), NIR1 (775 nm) & NIR2 (860 nm)

<table>
<thead>
<tr>
<th></th>
<th>March</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
<th>October</th>
<th>November</th>
<th>December</th>
</tr>
</thead>
</table>
5.4 Image Preprocessing

Preprocessing methods of images typically concentrate on dealing with non uniform illumination and colour correction [101]. Dealing with non uniform illumination refers to the removal (filtering) of the light source’s effect on the images. Colour correction usually refers to the equalisation of the colour means (i.e. linearly translating the image histogram to the desired colour mean). Nevertheless, these two methods alone are often not enough to facilitate the detection/classification process, the methods only process the colour of images but do nothing to enhance the contrast of the imagery.

To deal with the varying across image illumination and the low contrast of the raw images it was decided to use four out of the nine preprocessing steps which were described in [101]. This was done as its objective (image preprocessing to facilitate segmentation) is quite similar to our objective (image preprocessing to facilitate detection/classification). Not all of the preprocessing steps mentioned in that paper were used due to the different nature of the images ([101] deals with underwater images) but the order of the steps taken is the same. An additional step was added which removes the text bands present in all of the Bio platform images. The steps executed can be seen in figure 5.6.

In the following these steps will be quickly described. For further details on the order of preprocessing steps refer to paper [101].

![Figure 5.6: Detailed Preprocessing task block](image)

5.4.1 Removing image text bands

As mentioned earlier the multispectral images possess text bands, two present at the top corners of the image and two present at the bottom corners of the image. To avoid taking these text bands through the calculation process 8 pixel rows from the top and 8 pixels rows
from the bottom of the image were cropped away (by using the imcrop function of Matlab’s image processing toolbox).

5.4.2 Colour space Conversion

This step consists of the colour space conversion from RGB to YCbCr. The reason why converting to YCbCr colour space is beneficial is that it requires only the processing of the luminance band (Y) instead of the processing each individual RGB band. As such it saves computation time. After processing it can then be converted back to the RGB colour space.

5.4.3 Resizing data size with Symmetrical extension

Resizing is necessary as the next step makes use of a fast Fourier transform (FFT). A fast Fourier transform (also known as the Cooley-Tukey FFT algorithm) in essence is a faster method for calculating the discrete Fourier transform (DFT). It speeds up the process from $O(n^2)$ to $O(n \log(n))$ by splitting the problem in half [102]. The requirement for this recursive algorithm is that the data vectors have a size of a power of 2 [103]. As such the images that will be used should be resized to squares whose size is a power of 2 (i.e. the columns and the rows should be resized to a length that is a power of 2). This ‘resizing’ is usually done by padding the image with zeros. However, this creates an artificial discontinuity at the image borders. This is a problem as it can lead to the creation of boundary artefacts.

In general, convolution processes (which is multiplication in the Fourier domain) can produce boundary artefacts due to discontinuities at the image borders. This is also known as the boundary value problem (or the border effect), which is a result of the missing values around the boundary and results in the error propagating through the entire restored image [104]. Being aware of this is important because the homomorphic filtering process is subject to this boundary value problem.

This problem can be dealt with a variety of methods, examples being periodic, reflective (symmetric) and anti reflective (anti symmetric) boundary condition methods [104]. The periodic boundary condition method in essence repeats the signal (in our case the image) at the borders [105]. In other words it assumes the left border and the right border of the image are connected, and the same goes for the top and the bottom border of the image. As this in general is not the case for images it has the disadvantage that it creates artificial discontinuities (in the same way zero padding produces artificial discontinuities). Symmetric extension has the advantage over periodic extension in that it preserves continuity at the image borders [106]. It basically reflects the image pixels at the edges. This is the method used in [101]. The disadvantage with symmetric extension is that it retains only the continuity of the image and not of its gradient [104]. The anti symmetric boundary condition method on the other hand, does manage to do this but at the cost of a higher computation time. Our purpose is to simply enhance and remove the illumination effects present in the images. That is why symmetric extension was the chosen method (which is computationally less expensive than anti symmetric padding). The length of the largest dimension of the image was used to determine the next power of 2, setting the basis for the resizing procedure.
5.4.4 Homomorphic filtering

Homomorphic filtering is used to correct for non uniform illumination, image enhancement and dynamic range compression as mentioned in chapter 4 [46, 47]. It is a method especially developed for convolved and non linearly related signals and makes use of the frequency domain. Homomorphic filtering is based on the illumination-reflectance model which assumes that the reflectance and the illumination components possess a multiplicative relationship (see equation 5.1) [46].

\[ f(x, y) = i(x, y) \ast r(x, y) \]  \hspace{1cm} (5.1)

Where \( f(x, y) \) is the image, \( i(x, y) \) is the illumination component and \( r(x, y) \) is the reflection component with \( (x, y) \) being the row and column coordinates of the image pixels. A simple transformation in the Fourier domain however does not make these components separable yet, that is why a natural logarithm of the image is taken prior to the Fourier transformation, to make the relationship additive (see equation 5.2) [47]. As the natural logarithm of zero is not defined a value of 1 is added to all pixels in the image to avoid log zero problems.

\[ \ln(f(x, y)) = \ln(i(x, y)) + \ln(r(x, y)) \]  \hspace{1cm} (5.2)

In the frequency domain the components then should become separable (in theory and based on the earlier reflectance illumination model assumption) and filtering can be performed by multiplying the Fourier transform by a linear filter. In general illumination is characterised by low spatial variations (in the centre of the frequency domain) whereas reflectance is characterised by the higher frequencies (outer area of the frequency domain). Therefore a high pass filter is required. Several high pass filters exist, such as a Gaussian high pass filter, but in this case use will be made of a modified Butterworth high pass filter (which saves more computation time compared to the original Butterworth filter). This was tested in [46] and found more beneficial than the Gaussian filter for two reasons. Primarily, it allows for a steeper slope (parameter \( n \)) than the Gaussian bell curve, and secondly because it is easier to adjust the transition point/the cut off (parameter \( a \)) of the filter response [46]. Equation 5.3 shows the modified Butterworth high pass filter.

\[ BF(u, v) = 1 - \frac{1}{1 + \left( \frac{u^2 + v^2}{a} \right)^n} \]  \hspace{1cm} (5.3)

Where \((u, v)\) are the coordinates in the Fourier domain, \( a \) is the parameter that controls the transition point of the filter and \( n \) is the parameter that controls the steepness of the filter (and is a positive power of 2). After filtering the inverse Fourier transform is taken as well as its exponential to get an output image whose illumination component has been filtered out. Figure 5.7 shows the homomorphic filtering process [46].

It should be noted though that the reflectance and illumination components in actuality are not entirely separable in the Fourier space, as such the filtered output might still have some illumination effects present. Also attention should be paid to the fact that this model bases itself on the assumption that the illumination and reflectance spectra are spatially isotropic.
Figure 5.7: The homomorphic filtering process, where Image(x,y) is the original image, ln is the natural logarithm, FFT is the fast Fourier transform, BF(u,v) is the Butterworth high pass filter, IFFT is the inverse FFT, exp is the exponential and filtered_image(x,y) is the output.

(uniform in all directions). Figure 5.8 displays the frequency domain of the natural logarithm of the image captured on August the 16th. It is visible that the spectra are not entirely isotropic (perceivable from the slight elliptical behaviour). Nevertheless, a circular high pass filter with the proper starting parameters should be a good approximation to remove the low frequency components.

Figure 5.8: Fast Fourier transform of the natural logarithm of the Y (luminance) band of the RGB to YCbCr converted image taken August the 16th. The centre has the low frequencies (the illumination) and the outer areas are the higher frequencies (the reflectance)

So the next step is to set the right starting parameters for the Butterworth filter. Figure 5.9 is a visual of the modified Butterworth high pass filter BF(u,v) with different starting parameters. It is a cross section of the filter. Also note that in figure 5.9 the maximum response of the filter is 2 and the minimum response is 0.5. This was proposed by [107] and used by [101], and is achieved by amplifying (i.e. multiplying) the Butterworth high pass filter with a factor 1.5 and adding an offset of 0.5.

Figure 5.9: A visual of the modified Butterworth high pass filter BF(u,v) with different starting parameters.

By looking back at figure 5.8 one can derive proper starting parameters, in this case a was chosen to be 100 000 (transition point from the spectral centre in pixel numbers) and n was
chosen to have an order of 4 (steeper slope). Figure 5.9 displays this filter in blue.

\[
BP(\omega) = \frac{1}{1 + (\frac{\omega}{\omega_c})^n}
\]

**Figure 5.9:** The modified Butterworth filter frequency domain response with different starting parameters. Parameter \( a \) controls the transition point (the cut off) and parameter \( n \) controls the slope steepness (the higher \( n \) is the steeper the slope).
5.4.5 Contrast Stretching and Equalization

Contrast stretching is a common procedure used to improve the contrast in the image [101]. In essence, it adjusts the colour ranges in the images such that the ranges become the same (i.e. the image range is transformed into a desired range of values for all images). In Matlab contrast stretching is automatically performed during conversion back to the RGB colour space (to the range 0 to 255). Contrast stretching can be calculated with a fairly simple equation (see equation 5.4):

\[ \text{ContrastStretched}(x,y) = \frac{\text{Im}(x,y) - \min(\text{Im}(x,y))}{\max(\text{Im}(x,y)) - \min(\text{Im}(x,y))} \times 255 \]  

(5.4)

Where \( \text{Im}(x,y) \) represent the colour value at position \((x,y)\), \( \min(\text{Im}) \) is the minimum colour value and \( \max(\text{Im}) \) is the maximum colour value in the respective band. Because the outcome is between 0 and 1 it is additionally multiplied by 255 to transform it back into the RGB range (this is all done automatically during the YCbCr to RGB conversion in Matlab).

In [101] colour equalization is also performed after contrast stretching. Colour equalization refers to the linear translation of the image histogram to a desired colour mean. This can be achieved by adding for instance the difference between the colour band mean and the desired mean to each pixel of the image. This step however, as [101] states, does not do much for the actual detection process, it simply visually produces a more pleasant image (i.e. applying this to an image where the blue colour is more prominent will appear as an image were the colours are balanced out). See Figure 5.3 for a visual of the image preprocessing procedure after each step.

Summarizing the previous, what has been achieved with the preprocessing procedure is image enhancement and across brightness normalization per image. This does not mean that images with initially differing illumination look similar after preprocessing. That is known as colour normalisation or radiometric normalisation and will be explained later on.
The input image with text bands removed

The extracted Y band after RGB to YCbCr conversion

The Y band after symmetric extension

The Y band after Homomorphic filtering

Conversion back to RGB colour space

Table 5.3: A visual of the image preprocessing procedure steps.
5.5 Image Processing

Two processing approaches were used to detect the diatoms in the images: i.) diatom detection achieved by masking the macroalgae using Maximum likelihood classification prior to the detection and ii.) detection of diatoms using an artificial neural network known as SOM (Self Organising Map). The former is a supervised method conceptualised to deal with the problem of distinguishment between macroalgae and microalgae by masking the unwanted target. The latter is an unsupervised method which aids in finding underlying relationships between the variables that contribute to the distinguishment of a particular target.

5.5.1 Macroalgae Masking with Maximum Likelihood Classification

For this approach, processing of the images consists of defining the proper Regions of Interest (ROI), masking the macroalgae present in the images and determining the coverage of diatoms in the selected ROI with the proper vegetation indices. This was the first approach followed to detect diatoms. The maximum likelihood classification method was chosen due to its familiarity and ease of use in the image processing sector and its property of taking the variability of classes (targets) into account. The input for this approach are the preprocessed images produced in the previous section. The steps executed are:

1. Defining the ROI’s for the sampling grids
2. Masking the macroalgae cover
3. Detection procedure of diatoms in ROI’s
4. Results

A detailed version of the processing task block of this approach can be seen in figure 5.10.
Defining the ROI

As mentioned earlier in section 5.2 the in situ sampling takes place in six local grids. To properly validate the outcome the algorithm should thus be able to extract data from the locations bounded by these six grids. These are the Regions Of Interest (ROI) and are required for both the detection approaches used. In the field these grids are positioned with the use of a measuring tape (not GPS) which means the locations of the grids can show a small variation between the different field sampling visits (in the range of a few cm). That is why a slightly bigger rectangle (bigger than the grid size) was used to define the grid ROI’s. Inside these regions the detection of diatoms will be performed to enable direct comparison with the in situ sample data.

The images taken on March 22nd (a field sampling day) were used as a basis to define the grid ROI’s. See figure 5.11 for the final defined grid ROI’s. If the outcome is favourable (i.e. the outcome is comparable to the in situ data) then the algorithm can be applied to the entire field plot.
The objective is to detect the microphytobenthos (i.e., diatoms) present in the field plot. However, depending on the season a lot of macroalgae might be present also (their presence peaks in August). Distinguishing between macroalgae and microphytobenthos is challenging as they both contain chlorophyll-a. After having inspected vegetation indices (NDVI, MSAVI, IRGVI and LIRGVI) it became clear that using these indices alone would not be enough to make a clear distinction. This is because the vegetation indices for less dense macroalgae often coincided with the vegetation indices of dense diatom presence. An option considered then was to mask the macroalgae prior to detecting diatoms, to avoid the selection of vegetation index values that do not belong to the diatom class. The vegetation index used for the vegetation index output of the algorithm (to compare to the in situ pigment data) is the NDVI. The NIR band used to compute the NDVI was the one centered around 775 nm. The NIR centered at 775 nm was selected because it is closer to the chlorophyll-a high reflectance peak centered at 720 nm.

For the ‘masking’ procedure advantage was taken of false colour composite images (combining the NIR, Red and Green bands). These images display macroalgae in bright red and diatoms in a more dull red. To make the images comparable (and thus be able to classify the macroalgae) the false colour composite (FCC) images had to be normalised prior to the classification. Here we refer to colour normalisation with respect to a selected reference image and not to across brightness normalisation in which the brightness across one image is normalised (as was done in the preprocessing section). As mentioned in section 4.4.1 there are many colour normalisation techniques. In this particular case normalization was done following the radiometric normalization method of [98] because this method does not require...
the separate computation of each band’s cumulative histogram [98]. It essentially normalizes each band of the image by using the mean of each band and the standard deviation of each band with respect to the means and standard deviations of a reference image. So for each band (in this case the NIR band at 775 nm, the Red and Green bands) the pixel value is determined by Equation 5.5:

\[
NormBand(x, y) = m_{ref} + \sigma_{ref} \star \frac{Band(x, y) - m}{\sigma}
\] (5.5)

Where \( m_{ref} \) and \( \sigma_{ref} \) are the means and standard deviations of the reference image’s three bands, \( Band(x, y) \) is the pixel value for the band under question, \( m \) and \( \sigma \) are the mean and standard deviation of pixel values of the band under question and \( NormBand(x, y) \) the normalized band pixel values. In this particular case the reference FCC image chosen was the one of September as its bands means were closest to the overall time series bands mean, i.e. the yearly bands mean observed (in this case for the 9 months described by the image test set).

The next step is the classification of macroalgae using maximum likelihood classification which is a supervised classification method (see section 4.2). An image from August was chosen to serve as a basis for the class training (this because it contained both targets clearly). Two classes were trained with hand selected ROI’s, representing the macroalgae and non macroalgae classes.

The masks created are then converted into binary masks with the uncovered mud pixels having a value of 1 and covered mud (i.e. macroalgae) given a value of zero. See figure 5.12 for the original false colour composite field plot and the masked field plot (yellow areas representing zero values). Combining this binary mask with the grid ROI’s (by multiplying the binary masks with each other) gives the area to be used for diatom detection (i.e. the area that is not covered by macroalgae and is in one of the six sampling grids).

\[\text{Figure 5.12:} \text{ Left: the original False colour composite image, Right: the masked macroalgae (yellow) and the uncovered mud}\]
Detection of Diatoms in the ROI’s

The output will have to be comparable to in situ data for proper validation. The in situ data consists of vegetation percentage cover estimations and pigment concentration samples. The output will thus consist of two parameters, the computed percentage cover of vegetation and the average diatom NDVI observed in the six grids. The latter being the parameter that will be compared to the pigment concentration data. The average diatom NDVI is also indicative of the thickness of the biofilm. Percentage cover does not say much about the thickness of the biofilm, but in combination with the average diatom NDVI observed one can say something about the thickness of the biofilm. Section 4.4.1 mentions how vegetation indices can be used for quantitative estimation of the biomass (the pigment concentration).

The output of the detection consists of a visual of the classification and a tabular output for each image. The tabular size is six by four, with the six rows representing each grid and the four columns representing four different parameters. The first column describes the average diatom NDVI value per grid (note that NDVI is between -1 and 1). The second column describes the percentage coverage of diatoms with respect to the uncovered mud area, the third with respect to the grid area. The final column indicates the percentage of macroalgae present with respect to the grid area.

Results Macroalgae masking approach

Examples of outputs can be seen in figure 5.13. This figure shows the output of the algorithm for the image of June. There is a visual and a tabular result for each of the input images. The tabular output is a three dimensional matrix (6 by 4 by 3) with the row representing the grid number, the first column representing the average diatom NDVI, the second column the percentage coverage of diatoms with respect to the uncovered mud area, the third column the percentage coverage of diatoms with respect to the grid area and the last column the percentage of macroalgae with respect to the grid area. The third dimension indicates the image number.

At closer inspection of the classified images it is observed that not all diatoms are classified correctly in the grids. Particularly for the images of March, July, November and December the classification is ambiguous. This is probably due to two reasons. Primarily, due to the high presence of water in those images. The presence of water on the subsurface (and thus on top of diatoms) often results into negative NDVI values. The diatoms in this approach were identified based on the NDVI variable (and also on the assumption that diatoms should possess positive NDVI values). Because this was not the case the images were misclassified. In this case letting the algorithm take into account the negative values could opt as a solution. Nevertheless, the result then is an image in which most bare sand is classified as diatoms. Another reason for the misclassification is most likely the additional presence of red macroalgae such as in the image of November. Note also, that the macroalgae masking approach has the disadvantage that its overall performance relies on the initial choice of sample pixels representing a particular class.
Figure 5.13: Display of the detection output, a Visual and a Tabular output. The six rows of the table represent the six grids in the image and the four columns the average diatom NDVI present in each grid, the percentage of diatoms with respect to the uncovered mud area, the percentage of diatoms with respect to the grid area and the macroalgae percentage coverage respectively.
5.5.2 Self Organising Maps

The macroalgae masking approach is dependant on the efficiency of the macroalgae masking step and the assumption that diatom presence is indicated by positive NDVI values. Due to the similar nature of the spectral properties of macroalgae and microalgae the masking step at times was inaccurate (for instance when the dense microalgae were classified as macroalgae and masked). Additionally, the presence of water on the subsurface inhibited this approach as diatoms covered by a layer of water possessed a negative NDVI.

It was then considered to use the SOM approach as a tool to provide some insight as to what other features aside from the NDVI help contribute to the distinguishment of targets. Indeed, the SOM has been claimed to be an excellent tool in the exploratory phase of data mining [108]. In addition, the use of Self Organizing maps for image segmentation is particularly useful for clustering multiple targets as mentioned in chapter 4. Image segmentation application examples of SOM is colour image segmentation [109, 110], hyperspectral image classification [111] and segmentation of medical images [112, 113]. The approach was also selected due to its unsupervised nature (i.e. user input is not required) as automatic detection is what is striving for and its software package availability (open source software). Summarizing, the reasoning behind using a self organising map for the detection of diatoms is:

- It is an excellent tool in the exploratory phase of data mining
- Useful for the detection of multiple targets
- Open source software package
- Its unsupervised nature

The SOM approach uses a multivariate input of which its variables may or may not contribute to the distinguishment of targets. The objective is to determine which variables (or properties) contribute the most to us being able to perceive diatoms, macroalgae or other targets present in the image. At first sight it appears that the only properties that distinguish diatoms from macroalgae is their colour and their high and lower NDVI values. This can be prone to ambiguity as the colour of red-brown macroalgae as opposed to the more common green macroalgae is in fact highly similar to that of diatoms and low NDVI can also indicate less dense macroalgae. However, other characteristics can be perceived as well, such as the strong contrast of macroalgae with respect to its background (which allows us to perceive object boundaries) or the homogeneity of the diatom biofilm.

The SOM method finds underlying relationships between variables by 'learning' with a training input and attempts to cluster the targets in the other images accordingly. This 'learning' is done in an initial phase known as the training phase in which the neurons in the artificial network compete with each other for the exclusive right to respond to a particular input sample (as was earlier explained in Chapter 4) [76].

After the training of the SOM network the trained neurons will be further clustered with a hierarchical clustering method, as was done in [79]. The clusters will then be analysed to give
insight about the variables, their relationships and their contribution to a particular cluster. This analysis will then be used to reconfigure (if necessary) the input data or parameters to give an optimal result.

This section starts with details concerning the package used and the image training set to be used. It then continues with the selection and creation of the multivariate input. This is followed by an example run (with a particular variable and parameter configuration) to illustrate the working principle behind the SOM approach. This is then followed by an analysis of different runs (the different configurations) performed. The selection of the final input variables and parameters is then decided for based on the analysis of the runs. This section then ends with the results of the SOM approach. The section layout is as follows:

1. The SOMPAK package
2. The image training set
3. Selection and creation of a Multivariate input
4. Correlation of the input variables
5. Example run: SOM training & Hierarchical clustering
6. Observations of the runs & selected configuration
7. Results SOM

For a visual of the processing steps occurring in the SOM Approach see figure 5.14.
For the SOM method the public domain software package SOMPAK was used. This was developed by the Laboratory of computer and information science department of the University of Technology of Helsinki in 1992 [77]. The package splits the SOM computation procedure into 4 programs, each of which require the input of several parameters. The four programs and their required input parameters are listed below:

- Initialization program
  - randinit or lininit
  - Network dimensions
  - Network topology
  - neighbourhood function
- Training Program
  - rlen
  - alpha or alphatype

**The SOMPAK package**

**Figure 5.14:** Detailed Processing task block for the SOM approach. Note that its iterative.
An introduction to the SOMPAK package and the above parameters is given in Appendix D.

The Image training set

The SOM method is divided into the training phase and the application phase (in which the trained network in our case is used for image classification). The training phase requires an example input. This could be for instance the pixels of one particular image (where each pixel is descriptive of multiple variables at that specific image location) or the pixels of several images (such as images for each month for a year). The latter is advisable as it is more representative of the to be detected targets throughout the year. The use of one image on the other hand has the advantage (aside from a shorter training time) that an image can be selected were all targets to be detected are certainly present. Both input types were tried out. From here on the use of 12 images for training the SOM will be referred to as input type I and the training of the SOM with one image will be referred to as Input type II.

The 12 images (monthly representatives) chosen for the training of the SOM are shown in table 5.4. These images are preprocessed with homomorphic filtering (for enhancement) and radiometrically normalised with respect to the image brightness of April. The image of April was chosen because its RGB and NIR means deviated the least from the yearly RGB and NIR mean for the images of the training set (see figure 5.15). The test image set to be used for validation is also radiometrically normalised but with respect to the image of September for the same reason mentioned previously. Coming back at the training image set, only the pixels of a cropped section of the images will be the input vectors to train the SOM. A cropped region is selected to reduce the overall training time of the SOM. Note that each pixel will describe multiple variables. The selection of variables is explained in the following section.

For the other type (the use of 1 image for the training of the network) the image of July was selected as it clearly possessed all the targets to be detected (macroalgae, diatoms, bare soil, oysters).
Figure 5.15: The average RGB and NIR values of each image displayed over time. The yearly means for RGB were 148.6, 149.1 and 144.8 respectively, and for NIR the yearly mean was 147.3. The Month of April is closest to these yearly means and is selected for that reason as the reference image for radiometric normalisation.

Table 5.4: Image Training set details for training the SOM.

<table>
<thead>
<tr>
<th>Image number</th>
<th>Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16-01-2012</td>
<td>12:00:45, 47, 49 PM</td>
</tr>
<tr>
<td>2</td>
<td>25-02-2012</td>
<td>12:00:45, 46, 48 PM</td>
</tr>
<tr>
<td>3</td>
<td>24-03-2011</td>
<td>12:00:45, 46, 48 AM</td>
</tr>
<tr>
<td>4</td>
<td>19-04-2011</td>
<td>12:00:45, 47 PM</td>
</tr>
<tr>
<td>5</td>
<td>10-05-2011</td>
<td>12:30:44, 45, 47 AM</td>
</tr>
<tr>
<td>6</td>
<td>08-06-2011</td>
<td>12:30:44, 45, 46 AM</td>
</tr>
<tr>
<td>7</td>
<td>22-07-2011</td>
<td>13:00:43, 44, 45 AM</td>
</tr>
<tr>
<td>8</td>
<td>20-08-2011</td>
<td>12:00:42, 43, 45 AM</td>
</tr>
<tr>
<td>9</td>
<td>03-09-2011</td>
<td>12:30:43, 45, 47 AM</td>
</tr>
<tr>
<td>10</td>
<td>14-10-2011</td>
<td>12:30:42, 44, 45 AM</td>
</tr>
<tr>
<td>11</td>
<td>18-11-2011</td>
<td>12:30:43, 44, 46 AM</td>
</tr>
<tr>
<td>12</td>
<td>13-12-2011</td>
<td>10:30:42, 44, 45 PM</td>
</tr>
</tbody>
</table>

Selection and creation of a Multivariate Input

It is important to mention that the selection of input variables is the most significant step in the SOM approach, as they have the largest impact on the final output of the image classification [114]. Thus, for this particular case study, having insight into the nature of diatoms and macroalgae is of great importance. Furthermore, note that the detection procedure is an iterative process (as is visible in the block chart in figure 5.14). This means a variety of runs were performed and analysed before the setting of the final parameters and input data.
variables.

The initial variables that were considered as possible candidates for the input dataset (i.e. contributors that aid in the detection of diatoms and macroalgae) and are retrievable from the data available (the Argus BIO platform images) are:

1. rgb (the chromaticity components)
2. A degree of "Orangeness" and/or "Greenness" parameter
3. NIR (the Near infrared attribute)
4. Y (the luminance component)
5. Intensity gradients
6. A textural attribute (such as the local standard deviation of the image intensities)
7. NDVI (the normalised differential vegetation index values)

They were derived by inspecting the working principle behind human perception. Take for instance the rgb chromaticity components. The most striking feature at first sight (i.e. the most apparent contributor) to human perception being able to detect diatoms or macroalgae from a raw RGB image is their colour (the orange-brownish colour of the biofilm and the dark green colour of the macroalgae). As such it seems only reasonable that the RGB attributes of an image can contribute to the distinction of diatoms. Note though, that RGB images are formed by the combination of the red, green a blue intensities. This combination does not say much about the contribution of each band to a pixel colour. As such the rgb chromaticity components are more potential candidates, as they represent the proportions of each band to the visualised pixel colour [115]. The rgb chromaticity components are defined as:

\[
\begin{align*}
    r &= \frac{R}{R + G + B} \\
    g &= \frac{G}{R + G + B} \\
    b &= \frac{B}{R + G + B}
\end{align*}
\]

Where R,G,B are the RGB intensities from the raw image. The denominator describes the total intensity of a particular pixel (i.e. the sum of the RGB channels). In the same sense a parameter describing the degree of "orangeness" (or actually orange-brownness for diatoms) or "greenness" (or actually black-greenness for green macroalgae) of a pixel could help in the SOM recognising diatoms (high orangeness) and macroalgae (high greenness). Such a parameter can be created by computing for each pixel the euclidean distance in RGB space of its rgb values from a standard orangeness or greenness value. The smaller the distance the higher the degree of orangeness or greenness of the pixel. Standard colours can be used for comparison or an average of a user selected patch of the orangeness of diatoms and greenness of green macroalgae. The latter means that the user has to provide a priori knowledge to the detection method. The selected decimal code (rgb triple in range 0 to 255) in this case for the standard orangeness value is (160, 145, 110) and the selected decimal code for the greenness
value is \((62, 61, 66)\).

The Near Infrared attribute is a potential candidate as it is sensitive to varying vegetation biomass and land-water boundaries [116]. This may be true for images with very green vegetation (i.e. canopies) but in this case the NIR images, while clearly indicating land water boundaries, do not indicate a very high contrast between macroalgae and microalgae. As such its degree of potential is uncertain. The luminance component indicates how bright a particular surface will appear. And this is dependent on the properties of the surface that is reflecting the light, i.e. a diffuse surface will produce a different brightness than that of a flat surface. As such it might be reasonable to assume that oyster banks (diffuse surface) will produce a different luminance value than for instance macroalgae (not diffuse). It might not contribute directly to the detection of diatoms but it can aid in the proper clustering of the other targets present in the image.

Next, objects can also visually be distinguished from one another by perceiving the contrast they produce with respect to each other. Take for instance the high contrast between soil and green macroalgae. That is why a variable describing intensity change (local gradients in intensity) appears as a potential attribute. However, an intensity gradient describes not the object itself, but its boundary. It is important to distinguish between human an computer perception in this case. If a contrast variable (or an edge variable for that matter) is introduced to the SOM the SOM will treat all pixels with similar contrast as one object. In other words it will classify the boundaries of the objects as a separate object. This is not what is strived for and thus this candidate is eliminated as a candidate.

Aside from the contrast characteristic, human perception also makes use of texture to identify objects. An example being distinguishing an oyster bank (that has high textural content) from bare soil (homogeneous textural content). This textural attribute could be described by using a local standard deviation filter, i.e. each pixel is represented by the standard deviation of its neighbourhood pixels. Inspecting such a local standard deviation image however showed that the textural attribute of oyster banks is very similar to that of macroalgae and bare soil possessing water ripples. As such is not a very potential candidate to input into the SOM.

The NDVI as a candidate was considered as it says a lot about the vegetation present in the image (see chapter 4 for further detail on the NDVI). The presence of organic and non organic matter is clearer than in the NIR images. Note that the location of the targets (macroalgae, diatoms, oysters) is not taken into account, i.e. the SOM will work location independent.

From the above the selected candidate variables became:

1. rgb (the chromaticity components)
2. The Orangeness parameter and/or Greenness parameter
3. NIR (the Near infrared attribute)
4. Y (the luminance component)
5. NDVI (the normalised differential vegetation index values)
Prior to creating a multivariate input dataset the selected variables have to be normalised (commonly to a range 0-1). This is important because the SOM algorithm makes use of the Euclidean distance as a measure of which neuron is the best matching unit (the winning neuron). This poses a problem when the ranges of the variables vary considerably (RGB is between 0 and 255 while NDVI is between -1 and 1). That is why normalisation of the variables is advised prior to input into the SOM package\[78\]. The normalisation to the range 0-1 is computed as indicated in equation 5.6:

\[
\text{NormalisedRange}(x,y) = \frac{Im(x,y) - \min(Im(x,y))}{\max(Im(x,y)) - \min(Im(x,y))}
\]  

(5.6)

Where \(Im(x,y)\) the respective pixel value in the image, \(\min(Im(x,y))\) and \(\max(Im(x,y))\) the maximum and the minimum of the pixel values in the entire image.

Prior to normalisation however, it is also necessary to analyse the distributional behaviour of each variable. If a variable possesses a highly skewed distribution (or displays power law behaviour) normalising the variable will not fix the fact that the distribution’s ‘tail’ (the less dense points) will be given more influence than the denser part of the distribution. In such a case a logarithmic transformation of the variable prior to normalisation is the common choice [78]. From the candidate variables only the orangeness parameter showed highly skewed behaviour in its distribution. This variable was thus logarithmically transformed prior to normalising to a range 0-1. The NIR and luminance variables did not have a skewed distribution but did show a slight bimodal behaviour.

**Correlation of the input variables**

A thorough analysis of the input variables is required to be able to improve the clustering and thus to achieve a better detection of the different targets present in the images. This section is concerned with interpreting the influence of the selected candidate input variables.

The eight candidate input variables are the NDVI, NIR, Y, Orangeness parameter, the r, g, b chromaticity components and the Greenness parameter. Table 5.5 shows a correlation matrix of the eight candidate variables.

Several correlations stand out. Note for instance the high negative correlation between the NDVI and the luminance variable and the high correlation between the NIR and NDVI variables. The latter is logical considering the NDVI uses the NIR band for its computation. The reasonably high negative correlation between the luminance and the NDVI indicates that using both variables for the SOM input may be redundant. The r with the g chromaticity component and the Greenness component with the NDVI have a reasonably high negative correlation as well. Note the very high correlation between the Greenness parameter and the Luminance component (0.98). This is logical as according to the physiology of the human eye the green band is the largest contributor to human perception perceiving luminance [117]. That is also why the green photo sensors in digital cameras according to the Colour imaging
array patent of Bayer (1976) are referred to as the luminance sensitive elements whereas the red and blue are referred to the chrominance sensitive elements [118]. Because the correlation is nearly 1 it was decided not to use the Greenness parameter as the use of both the greenness and luminance variables is redundant.

**Table 5.5:** Correlation matrix of the eight input variables; the NDVI (a vegetation index), NIR (the near infrared band), Y (the luminance component), Orangeness (a parameter describing the degree of orangeness) and the rgb chromaticity components.

<table>
<thead>
<tr>
<th></th>
<th>NDVI</th>
<th>NIR</th>
<th>Y</th>
<th>Orangeness</th>
<th>r</th>
<th>g</th>
<th>b</th>
<th>Greenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVI</td>
<td>1</td>
<td>0.52</td>
<td>-0.66</td>
<td>-0.07</td>
<td>-0.18</td>
<td>-0.32</td>
<td>0.07</td>
<td>-0.60</td>
</tr>
<tr>
<td>NIR</td>
<td>0.52</td>
<td>1</td>
<td>0.23</td>
<td>0.21</td>
<td>0.06</td>
<td>0.07</td>
<td>-0.06</td>
<td>0.24</td>
</tr>
<tr>
<td>Y</td>
<td>-0.66</td>
<td>0.23</td>
<td>1</td>
<td>0.45</td>
<td>0.08</td>
<td>0.30</td>
<td>-0.06</td>
<td>0.98</td>
</tr>
<tr>
<td>Orangeness</td>
<td>-0.07</td>
<td>0.21</td>
<td>0.45</td>
<td>1</td>
<td>-0.18</td>
<td>-0.13</td>
<td>0.13</td>
<td>0.54</td>
</tr>
<tr>
<td>r</td>
<td>-0.18</td>
<td>0.06</td>
<td>0.08</td>
<td>-0.18</td>
<td>1</td>
<td>0.38</td>
<td>-0.52</td>
<td>0.00</td>
</tr>
<tr>
<td>g</td>
<td>-0.32</td>
<td>0.07</td>
<td>0.30</td>
<td>-0.13</td>
<td>0.38</td>
<td>1</td>
<td>-0.09</td>
<td>0.17</td>
</tr>
<tr>
<td>b</td>
<td>0.07</td>
<td>-0.06</td>
<td>-0.06</td>
<td>0.13</td>
<td>-0.52</td>
<td>-0.09</td>
<td>1</td>
<td>-0.04</td>
</tr>
<tr>
<td>Greenness</td>
<td>-0.60</td>
<td>0.24</td>
<td>0.98</td>
<td>0.54</td>
<td>0.00</td>
<td>0.17</td>
<td>-0.04</td>
<td>1</td>
</tr>
</tbody>
</table>

**SOM training & Hierarchical clustering**

The purpose of this subsection is to explain the SOM training and hierarchical clustering procedure using an example run. As mentioned before, a variety of combinations of input variables and parameters were tested. Their parameter settings and input types are noted down in a table in section 5.5.2. In the following, the procedure of one of the runs in the table (Run 1) will be explained. This run used the 12 training images (Input type I) and the seven variables (without the greenness variable) to create the multivariate input. This part is divided into the training of the SOM and the hierarchical clustering of the trained SOM.

### I. Training the Self Organising Map:

A linear appointment of reference vectors to the neurons was selected to initialise the network (the lininit command). This linearly orders the reference vectors in accordance to the two principal eigenvectors of the input data, i.e. the two components with the largest percentage explaining the total variance. It is favourable as it does not require and additional ‘ordering’ phase as is required when a random initialization is selected (the randinit command).

A hexagonal network of 30 by 30 neurons (i.e. 900 neurons) was selected in the configuration of Run 1. The number of input samples is 1384572 (i.e. the number of pixels to be used for training). Note that the amount of neurons is considerably smaller than the amount of input samples. This means the data will be compressed to a certain degree. That is why it is important to take into account the quantization error after the training. The quantization error is the Euclidean distance between the input data and the trained network ($\|x - mc\|$). If it is too large it is an indication that the network should be increased or the number of iterations should increase for better fine-tuning of the map. The learning rate alpha was set to 0.05, the radius to 10 and the neighbourhood function was set to Gaussian. The running
length is set to 1.4 million (which means 15428 samples will be reused in the training phase).

A U-matrix and a Sammon visualisation of the trained network can be seen in figure 5.16. The black dots in the U-matrix indicate the position of the neurons. The darker a cell is the larger the (Euclidean) distance between the neighbouring cells is. Dark lines/ridges in the Sammon visualisation indicate the presence of clusters (they are cluster separators as explained in chapter 4). Inspecting the U-matrix it appears quite homogeneous. The Sammon map too possesses unclear cluster separators. This indicates that more iterations or a larger alpha parameter is required for the SOM to fully organise itself. The quantization error experienced by the data with respect to the training images was 0.0283 which may be considered reasonable considering the range of the variables (0-1). The quantization error with respect to the test data was 0.0272 which is less than the quantization error with respect to the training images. In other words the trained network describes the test data better than the training data. This is usually not the case. This may be because the test data does not represent all months of the year, i.e. clusters present in the training images might not be present in the test images.

![Figure 5.16: U-matrix (left) and Sammon visualisation (right) of the 30 by 30 trained network. Homogeneity indicates the lack of cluster presence. Dark ridges in the Sammon map are cluster separators.](image)

II. Hierarchical clustering of the trained network:

The training of the network is followed by the hierarchical clustering of the trained network. This two step approach of calculating a SOM followed by the Ward clustering method (a hierarchical clustering method) of the trained SOM was conducted by [79]. In essence this basically further clusters similar neurons. A representation often used to illustrate the arrangement of the clusters of a hierarchical clustering is a dendrogram (a particular tree diagram). The number of clusters to extract can then be experimentally decided for. This basically ‘cuts’ the dendrogram there were it would result into the amount of clusters specified. The number of clusters to extract was set to 10, twice as large as the amount of different targets expected to be present in the images (Oysters, macroalgae, diatoms, bare soil, red macroalgae).

During hierarchical clustering some information can be lost. The so-called cophenetic correlation coefficient can be used in that case to assess the quality of the hierarchical clustering. The
The cophenetic correlation coefficient is a measure of how faithfully the dendrogram preserves the pairwise distances between the original unmodelled data points (in our case the best matching units of the trained network). In general a high cophenetic correlation coefficient (0.6 to 1) shows that the dendrogram is a good representation of the unmodelled data. It is in a way a value that indicates the amount of distortion that has occurred during the clustering (high cophenetic coefficient low distortion) [119]. Several hierarchical clustering methods were tried out aside from Ward clustering. Two methods in particular stood out, this was Ward clustering and Unweighted average distance clustering. These produced the highest cophenetic coefficients. The latter had the higher cophenetic coefficient (0.77) for this configuration. Despite this the Ward clustering (cophenetic coefficient of 0.59) resulted in better clustering of the images (visual observation). While the cophenetic correlation coefficient has widely been used for assessing how faithfully the clustering method preserves the initial dissimilarity matrix, it is not always a suitable approach for determining the best clustering method [120, 121]. So despite its slightly lower cophenetic coefficient Ward clustering was chosen as the preferred hierarchical clustering method, and will be used from here on.

![Dendrogram of Ward clustering](image)

**Figure 5.17: Dendrogram of Trained network**

The dendrogram result of the Ward clustering method for this particular run can be seen in figure 5.17. Note how the clusters are indicated with mere numbers. By analysing the produced dendrogram and the average, minimum and maximum values of each cluster one can interpret which cluster number represents which target. Once Ward clustering is performed a labelled trained network is produced (each neuron is given a specific cluster appointment). This is then used to cluster the pixels of all the test images accordingly.
Observations of the runs and the selected final configuration

In total seven runs were performed to assess the influence of the different SOM parameters and the two input types. For the influence of different variable configurations another four significant runs will be presented which affected the setting of the final configuration the most. The specifications and performance of the seven runs are shown in table 5.6. The specs and performance of the four runs with different variable configurations are shown in table 5.7.

In table 5.6 each run has a change in a particular parameter. Between the first and the second run (column 1 and 2) the change is the number of iterations to train the network. The number of iterations selected in Run 2 is the initial selected number of runs (1,400,000) times a factor of four. Note how the training time of Run 2 (1562 seconds) is approximately the training time of Run 1 times a factor of four (384*4=1536 seconds). One can then assume that the training time of the network is linearly proportional to the number of iterations selected. Also note the decrease in the quantization error between Run 1 and 2. This indicates that a longer training time is indeed necessary for the network to fully organise itself. The percentage change between the quantization error with respect to the training data and the quantization error with respect to the test data is positive as opposed to the first run. This is most likely due to the fact that the network was more able to organize itself than that of Run 1 due to its higher number of iterations. As such it becomes a better representative of the training data than that of the test data which is expected.

In Run 3 the changed parameter is the learning rate alpha. The increase of the learning rate does not seem to have an effect on the overall training time or the quantization error for that matter. Inspecting the U-matrix however it appears less homogeneous than that of Run 1 indicating, as it should, that the alpha parameter influences the formation of clusters in the network (the higher the alpha parameter the bigger the modification of the BMU and its neighbours).

In Run 4 the dimension of the network is increased to 40 by 40. This produced an increase in training time and a decrease in quantization error. The latter is logical considering the data is described by more neurons than as is the case for the smaller network. The radius parameter was decreased in Run 5 which resulted into a slight decrease in the quantization error as compared to Run 1. In general the radius should be around 1/4 of the network dimension radius (i.e. 10 for a 40 by 40 network and around 8 for a 30 by 30 network). In Run 6 the neighbourhood type was changed to bubble (a step function) instead of a Gaussian function. A substantial decrease in the quantization error and the training time was discerned.

Finally, in Run 7 the input type was changed (using only one image for training instead of a set of images). Three major changes were observed. Primarily the decrease in the quantization error with respect to the training data. Secondly, the U-matrix visualisation is significantly less homogeneous than that of Run 1. This is probably due to the clear presence of all targets to be detected in the one image. And third, the percentage increase observed is significantly larger than those of the other runs (an increase of 54.5%). This is probably due to the fact that when one image is used it can only recognise targets present in that image. That means a slight
variation of targets (i.e. the image possessed only green macroalgae and no red macroalgae) in the test images will be ambiguously clustered as they are not known to the trained network.

Table 5.6: Specifics and performance of the different SOM runs

<table>
<thead>
<tr>
<th>Specs</th>
<th>Run 1</th>
<th>Run 2</th>
<th>Run 3</th>
<th>Run 4</th>
<th>Run 5</th>
<th>Run 6</th>
<th>Run 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input type</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>nr. of variables</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>network dimensions</td>
<td>30 by 30</td>
<td>30 by 30</td>
<td>30 by 30</td>
<td>40 by 40</td>
<td>30 by 30</td>
<td>30 by 30</td>
<td>30 by 30</td>
</tr>
<tr>
<td>iterations</td>
<td>1,400,000</td>
<td>5,600,000</td>
<td>1,400,000</td>
<td>1,400,000</td>
<td>1,400,000</td>
<td>1,400,000</td>
<td>1,400,000</td>
</tr>
<tr>
<td>radius</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>alpha</td>
<td>0.05</td>
<td>0.05</td>
<td>0.5</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Neighbourhood</td>
<td>Gaussian</td>
<td>Gaussian</td>
<td>Gaussian</td>
<td>Gaussian</td>
<td>Bubble</td>
<td>Gaussian</td>
<td>Gaussian</td>
</tr>
<tr>
<td>Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time to train</td>
<td>384 sec</td>
<td>1562 sec</td>
<td>379 sec</td>
<td>692 sec</td>
<td>383 sec</td>
<td>178 sec</td>
<td>379 sec</td>
</tr>
<tr>
<td>QE w.r.t training data</td>
<td>0.0283</td>
<td>0.0254</td>
<td>0.0283</td>
<td>0.0242</td>
<td>0.0272</td>
<td>0.0239</td>
<td>0.0222</td>
</tr>
<tr>
<td>QE w.r.t test data</td>
<td>0.0272</td>
<td>0.0260</td>
<td>0.0262</td>
<td>0.0245</td>
<td>0.0270</td>
<td>0.0253</td>
<td>0.0343</td>
</tr>
<tr>
<td>% change QE</td>
<td>-3.9</td>
<td>2.4</td>
<td>-7.4</td>
<td>1.2</td>
<td>-0.7</td>
<td>5.8</td>
<td>54.5</td>
</tr>
</tbody>
</table>

To summarize, the observations drawn from these seven runs about the SOM parameters are:

- Increasing the dimension of the network decreases the quantization error with respect to the training data.
- Increasing the dimension of the network increases the training time of the network.
- Increasing the number of iterations decreases the quantization error with respect to the training data.
- The number of iterations is proportional to the training time of the network.
- The alpha parameter and the number of iterations play a role in the cluster formation in the SOM.
- Decreasing the radius parameter slightly decreases the quantization error with respect to the training data.
- Using a step function for the neighbourhood function (bubble instead of Gaussian) significantly decreases the training time and the quantization error experienced.
- The use of one image for training (Input type II), increases the quantization error experienced between the training data and the test data significantly.

Table 5.7 shows the specs and performance of the four runs with different variable configurations. Run A excludes the NIR variable, Run B the Y variable, Run C the NIR and luminance variable and Run D the NIR and red chromaticity variable. Aside from these runs also a run was performed excluding the rgb chromaticity components but including the greenness component. The clustering of this run was very unsatisfactory. A run was also performed without the orangeness parameter but the results were still lesser than when the orangeness parameter was included for the training.

The quantization error in table 5.7 does not say much as the fewer variables are used the smaller the quantization error will be (less variables to be compressed to the same amount...
of neurons). Interpretation was thus done mostly through visual inspection and analysis of the dendrograms and cluster values. The clustering result of Run A was the best amongst the four runs, with that of Run D being the worst. This makes clear that the red variable is required and the NIR variable individually as input to the SOM is not really necessary (but the NIR is required to compute the NDVI). The runs without the luminance variable (Run B and C) also performed quite unsatisfactory indicating that the luminance variable is necessary for proper SOM clustering. For Run A to D also networks of 40 by 40 were tested but no particular difference was visible in their clustering result with respect to the 30 by 30 networks.

Table 5.7: Specifics and performance of the different SOM runs

<table>
<thead>
<tr>
<th>Specs</th>
<th>Run A</th>
<th>Run B</th>
<th>Run C</th>
<th>Run D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input type</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>nr. of variables</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>network dimensions</td>
<td>30 by 30</td>
<td>30 by 30</td>
<td>30 by 30</td>
<td>30 by 30</td>
</tr>
<tr>
<td>iterations</td>
<td>5,600,000</td>
<td>5,600,000</td>
<td>5,600,000</td>
<td>5,600,000</td>
</tr>
<tr>
<td>radius</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>alpha</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time to train</td>
<td>1443 sec</td>
<td>1440 sec</td>
<td>1407 sec</td>
<td>1422 sec</td>
</tr>
<tr>
<td>QE w.r.t training data</td>
<td>0.0181</td>
<td>0.0218</td>
<td>0.0127</td>
<td>0.0168</td>
</tr>
<tr>
<td>QE w.r.t test data</td>
<td>0.0181</td>
<td>0.0221</td>
<td>0.0132</td>
<td>0.0169</td>
</tr>
<tr>
<td>% increase QE</td>
<td>0</td>
<td>1.4</td>
<td>3.9</td>
<td>0.6</td>
</tr>
</tbody>
</table>

**Final Configuration SOM Input**

Having observed the different settings of the runs the final selection of parameters, variables and input type for the detection of diatoms were set. These final configurations can be seen in table 5.8. The variable NIR was excluded.

Table 5.8: Specifics of the selected final settings for the SOM training for the detection of diatoms.

<table>
<thead>
<tr>
<th>Specs</th>
<th>final configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input type</td>
<td>I</td>
</tr>
<tr>
<td>nr. of variables</td>
<td>6</td>
</tr>
<tr>
<td>network dimensions</td>
<td>30 by 30</td>
</tr>
<tr>
<td>iterations</td>
<td>5,600,000</td>
</tr>
<tr>
<td>radius</td>
<td>8</td>
</tr>
<tr>
<td>alpha</td>
<td>0.5</td>
</tr>
<tr>
<td>Neighbourhood</td>
<td>Bubble</td>
</tr>
</tbody>
</table>

These configurations were decided for a couple of reasons. Primarily, the increase of the alpha parameter and iteration number decrease the quantization error and the homogeneity of the
trained network respectively. Secondly, the selection of 6 variables instead of 7 variables was based on the interpretation of the results of Run A to D. A 30 by 30 network was selected as it possesses a shorter training time than that of a 40 by 40 network and because no particular improvement was detected when using the 40 by 40 network dimension. Input type I was chosen due to its superior performance in terms of the quantization error with respect to the test data. A step function was chosen for the neighbourhood function instead of a Gaussian because of the significant decrease in training time and quantization errors.

Table 5.9 displays the cluster averages of the ten classes. High NDVI values indicate dense vegetation cover as opposed to low NDVI values. Negative NDVI values indicate water presence. In the Macroalgae masking approach this produced misclassification of diatoms (as water often covers the diatoms). High luminance indicates bright surfaces whereas low luminance values indicate dark surfaces. Low orangeness values (i.e. small distance to the standard orangeness value selected) indicate high degree of orangeness. The degree of r, g, b chromaticity per cluster is indicated by the magnitude of the values.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>NDVI</th>
<th>Y</th>
<th>Orangeness</th>
<th>r</th>
<th>g</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.19</td>
<td>193</td>
<td>5.02</td>
<td>0.334</td>
<td>0.336</td>
<td>0.329</td>
</tr>
<tr>
<td>2</td>
<td>-0.13</td>
<td>164</td>
<td>4.56</td>
<td>0.332</td>
<td>0.336</td>
<td>0.330</td>
</tr>
<tr>
<td>3</td>
<td>-0.08</td>
<td>140</td>
<td>4.03</td>
<td>0.331</td>
<td>0.335</td>
<td>0.331</td>
</tr>
<tr>
<td>4</td>
<td>-0.10</td>
<td>117</td>
<td>3.63</td>
<td>0.337</td>
<td>0.334</td>
<td>0.327</td>
</tr>
<tr>
<td>5</td>
<td>0.15</td>
<td>118</td>
<td>3.51</td>
<td>0.350</td>
<td>0.335</td>
<td>0.313</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
<td>96</td>
<td>4.08</td>
<td>0.336</td>
<td>0.329</td>
<td>0.332</td>
</tr>
<tr>
<td>7</td>
<td>0.27</td>
<td>84</td>
<td>4.35</td>
<td>0.345</td>
<td>0.328</td>
<td>0.323</td>
</tr>
<tr>
<td>8</td>
<td>0.35</td>
<td>41</td>
<td>5.06</td>
<td>0.349</td>
<td>0.316</td>
<td>0.326</td>
</tr>
<tr>
<td>9</td>
<td>0.87</td>
<td>11</td>
<td>5.39</td>
<td>0.173</td>
<td>0.220</td>
<td>0.387</td>
</tr>
<tr>
<td>10</td>
<td>0.71</td>
<td>14</td>
<td>5.36</td>
<td>0.526</td>
<td>0.240</td>
<td>0.137</td>
</tr>
</tbody>
</table>

Clusters 1 and 2 possess the highest luminance. In the dendrogram too they are paired under the same branch. After inspection of the clustered images it is clear these two clusters represent bare soil (with cluster 1 representing the highly illuminated soil). Clusters 7 to 10 possess the highest NDVI values and the lowest luminance values and degree of orangeness, indicating these clusters represent green macroalgae. Particularly clusters 8 to 10 represent the denser green macroalgae. In the dendrogram cluster 6 and 7 are paired under the same branch. Surprising is the low average NDVI value of cluster 6. After inspection of the clustered images this is because cluster 6 can represent shaded areas as well (low luminance, low NDVI). The high degree of orangeness of Cluster 5 and 4 indicates they represent either diatoms or red macroalgae. From inspection of the clustered images cluster 5 mostly represents dense diatom cover and occasionally confused with red macroalgae. Cluster 4 appears to represent the less dense diatom cover (despite its negative NDVI). Cluster 3 possesses a high luminance value with respect to cluster 4 and 5 and also the next highest degree in orangeness. From visual inspection this is in particular bare soil like clusters 1 and 2 but with an orange tint (very thin diatom film perhaps). Figure 5.18 shows the dendrogram with the numbered clusters assigned to the targets.
5.5 Image Processing

Figure 5.18: Dendrogram with the clusters appointed to targets

Results SOM

Figure 5.19 shows sections of some of the clustered image outputs for the selected test images. The complete set of clustered images (months March, May, June, July, August, September, October, November and December) can be seen in Appendix E. Several peculiarities were observed; the test images of months May and December have a higher detection of diatoms and macroalgae than is actually present in the images. Both have falsely clustered diatoms and macroalgae occurring in the lower corners of the image. Perhaps this is indicative of a lens distortion of some kind, perhaps that of chromatic aberration (distortion caused when the lens fails to focus all colours to the same convergence point). Also clear is that occasionally cluster 5 represents red macroalgae instead of diatoms (see bottom row figure 5.19).
Figure 5.19: Sections of the clustered image outputs for the selected training images June, August and September. Displayed are the clusters macroalgae (dark green), Bare soil (from white to beige indicating the decreasing brightness of the soil) and diatoms (from light orange to dark indicating the less dense to dense diatom cover).
5.6 Validation

A tabular output is given for each image (shown in Figure 5.20). Like the macroalgae masking approach it has six rows indicating the sampling grid number. The columns indicate the average NDVI, the diatom percentage coverage with respect to the grid area and the macroalgae percentage coverage with respect to the grid area respectively. The general outcome is a table of three dimensions (6 by 3 by 9), with the 3rd dimension describing the image number.

<table>
<thead>
<tr>
<th>Grid #</th>
<th>Avg NDVI</th>
<th>Diatoms [%]</th>
<th>Macroalgae [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1049</td>
<td>41.2</td>
<td>14.4</td>
</tr>
<tr>
<td>2</td>
<td>0.0746</td>
<td>20.9</td>
<td>18.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0550</td>
<td>24.2</td>
<td>13.5</td>
</tr>
<tr>
<td>4</td>
<td>0.0796</td>
<td>24.3</td>
<td>21.1</td>
</tr>
<tr>
<td>5</td>
<td>0.0533</td>
<td>26.0</td>
<td>15.1</td>
</tr>
<tr>
<td>6</td>
<td>0.0778</td>
<td>19.7</td>
<td>18.6</td>
</tr>
</tbody>
</table>

Figure 5.20: Tabular output of the SOM approach. The rows indicate the grid number. The columns indicate the average NDVI, the diatom percentage coverage with respect to the grid area and the macroalgae percentage coverage with respect to the grid area respectively. The third dimension describes the image number.

5.6 Validation

The image test set was selected such that the dates the images were taken were close to the in situ sampling dates for this purpose. Three types of validation can be performed. The first form of validation is a comparison between the in situ cover estimations of macroalgae and the computed macroalgae cover of both approaches (these are time series plots). Because in situ diatom cover estimations were incomplete this type of validation is not possible for diatoms. That is why the second type of validation compares the diatom percentage cover to the fucoxanthin to chlorophyll-a ratio values time series (indirect validation). The third form of validation uses the average diatom NDVI or computed diatom percentage coverage of each grid and plots it against characteristic of diatoms pigment concentrations or ratios. Finally, to assess the representativeness of the six grids for the entire field plot the average of the grids will be compared to the average of the entire plot.

This section is divided accordingly:

1. In situ macroalgae cover estimation and computed macroalgae cover comparison
2. Computed diatom cover and pigment ratio seasonal behaviour
3. Computed diatom NDVI versus in situ pigment data
4. Grid representativeness
5.6.1 In situ macroalgae cover and computed cover comparison

As aforementioned this type of validation is more suitable for validating the macroalgae percentage coverage computed rather than that of diatoms. This because for the months with high abundance of macroalgae no diatom coverage estimations were made.

In figure 5.21 a time series is displayed of the in situ estimations made for macroalgae cover and the computed percentage cover from the two approaches (macroalgae masking approach and SOM approach). The error bars indicate the ratio of the standard deviation of the six grid outputs to the square root of 6 (the amount of grids). The computed and in situ values displayed for each month represent the average value computed from the six grids. Observing the plot two aspects are worrying. Primarily, the computed percentage covers do not ideally follow the behaviour of the in situ estimations (and are not within the computed error bars) and secondly, there appears to be a vertical bias between the two detection approaches. The latter may be due to the dependency of the macroalgae masking approach on the representativeness of the selected reference pixels for the macroalgae class. These were selected from one image (August) which served as a basis for the class training (this because it contained both targets clearly) as opposed to the SOM approach which used 12 images for training itself.

Regardless of this vertical bias it is clear the behaviour of the computed data does not ideally

Penelope Rammos Master of Science Thesis
follow the in situ cover estimation or fall within the range of the error bars. On the other hand, when examining how the error is computed one could question whether this error is more a measure of the error involved due to random sampling rather than an indication of the quality of the data. This error alone is perhaps too optimistic in describing the quality of the data for both the in situ data and computed data. This because it does not take into account three aspects; I. the human subjectivity involved during the in situ estimations of macroalgae, II. the error of the two detection approaches themselves and III. the possible error involved due to the daily variety of macroalgae cover.

The month of November is an example where the accuracy of the in situ data is debatable. When compared to the in situ data the computed data for the month of November is considerably higher for both approaches. The month of November has a high presence of red-brown macroalgae (this can be seen in figure 5.22). Observing the November image, the computed macroalgae for this month appears more or less correct as the red-brown macroalgae indeed seems to cover about 20% of the surface. On the other hand, zooming into one of the sampling grids (see figure 5.23), one can see that diatoms are quite dominant in this month as well. It becomes obvious that distinguishing between the brown macroalgae and the diatom biofilm is a crafty task using visual inspection alone. Indeed, the sampler (from NIOZ) stated that estimating macroalgae and microalgae was toughest for the month of November due to the targets’ similar colours. The question that arises here is, which is the more accurate approach to macroalgae cover detection, the in situ macroalgae estimation or the computed macroalgae cover? In situ data is subject to human error whereas computed cover is subject to errors produced by the detection algorithms.

![The RGB image of November. Note the high presence of red-brown and decaying macroalgae.](image)

**Figure 5.22:** The RGB image of November. Note the high presence of red-brown and decaying macroalgae.
Figure 5.23: Sampling grid image for the month November 2011 used for pigment sampling and estimation of macroalgae.

In further inspection of the plot, the month of March produces a high amount of computed macroalgae cover while there is virtually no macroalgae present (see Appendix E). For the SOM approach this could be due to cluster 6 which in general represents macroalgae but in this particular image appears to represent shaded areas as well. Another explanation could be the presence of water on the surface. The images that have the biggest amount of negative NDVI (which is indicative of water presence) are the images of March, July and November, which also happen to be the months significantly differing from the in situ data. Figure 5.24 shows the previously displayed computed and in situ cover time series of macroalgae together with the water cover percentage for each month. The water cover percentage was computed by finding all pixels which possess a negative NDVI and a NIR value less than 150. The presence of negative NDVI pixels seems to have a bigger influence on the macroalgae masking approach than on the SOM approach with the exception of the month of July. For the month of July the computed values of both approaches are lower than the estimated value obtained in situ. Inspection of the clustered images does not give much reason as to why this difference is so large (up to almost 20%). Some less dense macroalgae pixels are clustered as diatoms in both approaches but not enough to account for such a large difference. Here too it is debatable whether this is because of potential human error in the in situ estimations or as a result of the methods used for detection. Continuing, one can indirectly say something about the robustness of the two detection methods in dealing with images where the target is absent, such as the images of May and December in which no macroalgae is present (with the exception of some macroalgae caught in the oysterbanks). An overestimation is observed for both approaches for these two days, perhaps the result of negative NDVI (for the image of May) or a lens distortion of some kind.
Another source of error in the detection output could be the error involved from potential daily variety of macroalgae cover. The test images were selected such that they are close to the actual sampling day (refer to section 5.3) but none of them are on the actual sampling day. They are a day or up to 4 days away from the actual sampling day. In figure 5.21 no account is taken as to how much the macroalgae cover can vary daily. To get an idea of how the macroalgae cover can vary daily, additional images captured a day before and a day after where selected. For some of the test images getting a suitable image a day before and after was not possible, for these images 2 days before or 2 days after were selected. The specifics of these images can be seen in appendix F. Figure 5.25 displays the average coverage of macroalgae detected and the standard deviation observed between the three days (the day before, the test image and the day after) for each month. Observing the plot one can see the standard deviation can occasionally exceed 20%, i.e. there is a high variation of macroalgae coverage daily. As such it should be considered whether using the test images (which are not on the actual sampling day) are suitable enough for performing validation. Note that the variance observed could be attributed to the detection algorithms as well. To see whether this is the case or not, detection was performed on 20 minute interval images for March 23rd additionally to assess the consistency of the detection methods. Between these 20 minute interval images no variation is observed in coverage through visual interpretation. A constant detection output would thus say something about the consistency of the detection methods (by relying on the accuracy of our visual interpretation). Because on March the 23rd barely
any macroalgae is present this was done for diatom detection and its results will be shown in the next section.

The consistency of the methods is postponed to the next section but one can additionally assess the suitability of the two detection approaches for macroalgae cover by analysing the goodness of fit with respect to the in situ data. This is of course under the presumption that the in situ data is accurate. Coming back at figure 5.21, no particular advantage of one approach over the other is perceived visually. Therefore, their performance was assessed by using 4 error quantification techniques. These are:

- The Model Efficiency [122, 123]
- The percentage Model bias [124]
- The Cost function [125]
- The Coefficient of determination [124]

The Nash Sutcliffe model efficiency (see equation 5.7) measures the ratio of the model error to the variability of the data [124].

Penelope Rammos

Master of Science Thesis
\[ ME = 1 - \frac{\sum_{n=1}^{N} (D_n - M_n)^2}{\sum_{n=1}^{N} (D_n - \bar{D})^2} \] (5.7)

D is the original data (in this case it would be the estimated in situ data), M is the corresponding model output (i.e. the computed percentage cover output from the detection procedure), \( \bar{D} \) is the mean of the original data and N is the number of samples (in this case it is 9, corresponding to the amount of months data is available). The model efficiency is categorised as follows; excellent when \( ME > 0.65 \), very good when \( 0.5 < ME < 0.65 \), good when \( 0.2 < ME < 0.5 \) and poor when \( ME < 0.2 \) [124, 126].

The Percentage model bias indicates whether the model systematically underestimates or overestimates the observations. The closer to zero the better the model, i.e. when \( Bias < 10 \) excellent, when \( 10 < Bias < 20 \) very good, when \( 20 < Bias < 40 \) good and when \( Bias > 40 \) poor [124, 126]. The percentage model bias is given in equation 5.8.

\[ Bias = \frac{\sum_{n=1}^{N} (D_n - M_n)}{\sum_{n=1}^{N} D_n} \times 100 \] (5.8)

The Cost function is indicative of the goodness of fit between two sets of data. It is a ratio of the difference between data and model output to the variance of the data. The function is described in equation 5.9. The output is dependent on the variance of the data set but in general terms the smaller the value the better.

\[ CF = \frac{1}{N \times \sigma_D^2} \times \frac{\sum_{n=1}^{N} (D_n - M_n)^2}{\sum_{n=1}^{N} (D_n - \bar{D})^2} \] (5.9)

The Coefficient of determination (commonly symbolized by \( R^2 \)) is the square of the linear correlation coefficient (between the data and the model output) and has a range from 0 to 1. When close to zero the fit is poor, when 1 the fit is excellent. It is, like the cost function, a measure of the goodness of the fit. The computed error quantifications for both approaches can be seen in table 5.10.

The Model efficiency of the SOM approach is classified as 'very good', and that of the macroalgae masking approach as 'good'. The negative sign of the bias produced by the macroalgae masking approach indicates that in overall this method overestimates the in situ cover estimations. It also has a magnitude between 20 and 40% and is thus classified as a 'good' model bias. The positive sign in the Bias for the SOM approach indicates that overall this method underestimates the in situ cover estimations. The magnitude for this approach is
smaller than 10 and can thus be classified as an ‘excellent’ model bias. These performance measures point to the SOM method as the better approach for macroalgae detection. Yet, as previously mentioned, this is still under the presumption that the in situ data is accurate.

### 5.6.2 Computed diatom cover and pigment ratio seasonal behaviour

As mentioned before the in situ diatom cover estimations were unfit and thus the previous validation was not possible for diatoms. What can be done however is comparing the behaviour of the computed diatom cover over time with the fucoxanthin to chlorophyll-a behaviour pigment ratio over time (referring to section 5.2, figure 5.4). Figure 5.26 shows the computed diatom cover over time behaviour for the SOM approach (top) and the macroalgae masking approach (bottom). The values (units) themselves are not comparable but a similar seasonal pattern can serve as a partial validation.
Figure 5.26: Computed diatom percentage cover over time of the SOM approach (top) and the macroalgae masking approach (bottom).
The first observation is the high dissimilarity of the computed diatom cover seasonal behaviour of the macroalgae masking approach (bottom of figure 5.26) with respect to the pigment ratio seasonal behaviour (figure 5.4). The SOM approach on the other hand shows similar patterns with the pigment ratio behaviour. From months March to May and July to November the two present a similar behaviour. The month June on the other hand shows deviation from the pigment ratio behaviour. The pigment ratio behaviour (top plot) indicates an increase in diatoms with respect to the month of May, whereas the computed diatom cover indicates a decrease with respect to the month of May. In both the highest abundance is observed in the month of March. The distinguished peak and valley observed from the months August to October are present in both plots. The high dissimilarity of the macroalgae masking approach points to the SOM approach as the better one for diatom detection. Note though, that potential daily variation of diatom cover is not taken into account here, nor is any error resulting from the detection approaches.

To test the consistency of the detection methods 10 images captured at 20 minute intervals on March the 23rd where additionally selected. These visually show no apparent change in diatom cover. A constant or near constant detection of diatom cover would indicate the detection is consistent. If the diatom cover varies significantly this could point to either: i. there really is a variation of diatom cover (such as the vertical migration rhythm of diatoms observed in literature) even if it is not visually interpretable with human sight or ii. the detection is inconsistent. Figure 5.4 shows the diatom cover behaviour at 20 minute intervals from 11:24 AM to 14:24 PM.

The SOM shows a slightly decreasing behaviour whereas the macroalgae masking approach shows increasing behaviour of the diatom cover. The maximum deviation (max-min) between the 20 min interval detection observed for the SOM approach is 4.3%, for the macroalgae masking approach this is 7.9%. If no diatom cover variation between 20 minutes actually occurs this would mean the SOM approach is more consistent than the macroalgae masking approach in its detection. If there is a vertical migration rhythm of diatoms (and presuming it can be observed with the Multispectral camera) then according to literature one should observe the diatoms migrating downwards about an hour before the tide comes, which is around 14:30 for this day. In that case the macroalgae masking approach performs poorly for diatom detection as it shows an increase in diatom cover before the tide. Taking this into account and that the computed diatom cover seasonal behaviour of the macroalgae masking approach earlier showed high dissimilarity with the pigment ratio behaviour, it appears the SOM detection approach outperforms the macroalgae masking approach for diatom detection.
Figure 5.27: Diatom cover behaviour at 20 minute intervals from 11:24 AM to 14:24 PM for the SOM and macroalgae masking approach.
5.6.3 Computed diatom NDVI versus in situ pigment data

As mentioned earlier in section 4.4.1, vegetation indices can be used for quantitative estimation of the biomass. For phytoplankton in the sea or canopies this has typically been achieved with non linear regressions between biomass (of the pigment chlorophyll-a) and vegetation indices such as the NDVI (computed by using the red and NIR spectral bands of satellite imagery). This non linearity is because the NDVI for such targets saturates for high chlorophyll-a values. The range of NDVI observed on mud substrates, however, does not reach the level of saturation and can be assumed to possess a linear relationship with chlorophyll-a.

To test whether such a linear relationship is practicable for diatom biomass estimation the scatter plot in figure 5.28 was plotted. The data consists of each of the 6 grid average diatom NDVI’s and pigment concentration values over all eight months for which pigment concentration samples were available. Figure 5.28 additionally displays the corresponding least square linear fits of both approaches. Note how the fits are virtually parallel to the x-axis indicating that the NDVI variable is constant regardless of the pigment concentration. This behaviour in retrospect is logical considering the small range of diatom NDVI’s (from 0.038-0.153 for the macroalgae masking approach and from 0.051-0.198 for the SOM approach). This is approximately 7% of the total NDVI range (-1 to 1). This (indirectly) indicates that using the NDVI for quantitatively estimating diatom biomass on mud substrates may be ineffective.

Also important to note is that the NDVI is more susceptible to the chlorophyll-a pigment which is present in both diatoms and macroalgae than to fucoxanthin (only present in diatoms and brown macroalgae). This is because chlorophyll-a has absorbance maxima in the violet band (at around 431-33 nm) and in part of the red band (at around 670 nm) and shows high reflectance in the NIR (at around 720 nm), the red and NIR band being the defining variables of the NDVI. Fucoxanthin has high absorbance in the blue band centered at around 450 nm (see appendix C) and not the red band. Using the NDVI is thus not suitable for estimating the fucoxanthin pigment concentration.
Figure 5.28: NDVI versus Chlorophyll-a for the two approaches. The macroalgae masking approach is indicated by dark green and the SOM Approach by light green.

The use of the percentage coverage per grid as opposed to the use of the NDVI then appears as a more suitable candidate for estimating the diatom biomass. Because of the large deviation of the macroalgae masking approach from the fucoxanthin to chlorophyll-a time series only the SOM data was used. Figure 5.28 shows the scatter plot of the computed diatom percentage cover and the corresponding fucoxanthin to chlorophyll-a ratio values. Added is a least squares fit. The y-intercept is 12.8 and the slope is 19.2 (i.e. first degree polynomial of $19.2x + 12.8$). It indicates a proportional relationship between the amount of diatom cover and the fucoxanthin to chlorophyll-a pigment ratio (i.e. the higher the diatom cover the higher the ratio).
Figure 5.29: Diatom percentage cover versus fucoxanthin to chlorophyll-a ratio and a corresponding linear fit.

5.6.4 Grid representativeness

The presence of a sampling error is inevitable when random sampling is performed. That is why it is important to see whether the six sampling grids accurately represent the entire plot. Figure 5.30 shows the computed macroalgae and diatom cover for the entire plot and the mean of the six grids for each month. Table 5.11 indicates the average deviation for each variable for both approaches. With average deviation we refer to the average of the differences between the means of the six grid values and the entire plot cover values of each month.

The average and maximum deviations of the SOM approach for both diatoms and macroalgae is smaller than that of the Macroalgae masking approach (i.e. a smaller sampling error). This might be contributed to the fact that the macroalgae masking approach is more affected by negative NDVI values than the SOM approach. The SOM approach takes advantage of more variables and thus appears less influenced by negative NDVI values. Another reason may be the dependency of the macroalgae masking approach on the representativeness of the selected reference pixels for the diatoms.
5.7 Discussion

The results confirm that automatic detection of diatoms using multispectral imagery is indeed possible. At the current phase of development however it cannot be claimed fully automatic yet. The Macroalgae masking approach requires the selection of reference pixels representative of the macroalgae class. The SOM approach requires proper parameter settings for the training of the network. On the other hand, once the representative pixels for the macroalgae masking approach have been selected or the proper parameters have been set for the SOM approach no additional effort is required on the users part. For reasons of clarity this section is divided into the following four subsections:

- Distinguishing between macroalgae and diatoms
- Method suitability
- In situ data & Validation
- General recommendations

Table 5.11: Average and maximum deviation between the entire plot value and the mean of the 6 grids value for both approaches.

<table>
<thead>
<tr>
<th>Cover type</th>
<th>Macroalgae masking approach</th>
<th>SOM approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg Deviation</td>
<td>Max deviation</td>
</tr>
<tr>
<td>Macroalgae cover [%]</td>
<td>8.07</td>
<td>14.67</td>
</tr>
<tr>
<td>Diatom cover [%]</td>
<td>3.20</td>
<td>8.00</td>
</tr>
</tbody>
</table>

Figure 5.30: Grid representativeness for the computed values of macroalgae cover for the maximum likelihood masking approach (approach 1) and the SOM approach (Approach 2).
5.7.1 Distinguishing between macroalgae and diatoms

The problem in the distinction between macroalgae and microalgae experienced during the detection procedure reflects the notifications in paper [18] where the problem was also stated. The primary source of difference between the two targets in the available Argus BIO platform images was their low and higher NDVI values and their colour (not the case for red-brown macroalgae). To add more distinct properties it would be nice to add the bands which are descriptive of the fucoxanthin and diadinoxanthin pigments absorption maxima. Fucoxanthin displays high absorption around 450 nm and diadinoxanthin displays high absorption also around that wavelength (447 nm and 477 nm). Both have reflectance in the NIR region. In that case perhaps using the blue band (450 to 490 nm) may be beneficial for diatom detection, i.e. instead of using the red band in the NDVI, one could use the blue band (see equation 5.10).

\[
NDVI_{blue} = \frac{NIR - B}{NIR + B}
\]  

(5.10)

Figure 5.31 shows a comparison of the NDVI with the red band (traditional) and the NDVI with the blue band. Note how the \(NDVI_{blue}\) indicates diatom presence there were the traditional NDVI fails to detect diatoms. On the other hand, in images were there is high water presence the traditional NDVI barely differs from the NDVI that uses the blue band. As such using this \(NDVI_{blue}\) may not necessarily improve results for the months that have lots of water presence. Its performance under different conditions (i.e. different atmospheric conditions or with the presence of other algae with similar absorbance maxima in the blue range) will need to be assessed in future work.
Figure 5.31: Comparison NDVI and NDVI that uses the blue band instead of the red band. First column represents the original RGB image, the middle column the traditional NDVI and the last column the NDVI that uses the blue band instead of the red band.
Another suggestion could be the addition of the band centered around 586 nm which is used to describe Meleders specific index mentioned in chapter 4. The paper on this specific index ([69]) describes it significantly outperforms other indices for the detection of diatoms. The effectiveness of this index on mud substrates however is unknown.

5.7.2 Method Suitability

The results indicate that proper distinction is inhibited by the presence of water (this more for the macroalgae masking approach than the SOM approach) and the presence of the similarly coloured brown-red macroalgae. Despite this, the error quantification parameters indicate that both approaches perform reasonably good for macroalgae cover, with the SOM approach outperforming the macroalgae masking approach. On the other hand, the computed cover clearly does not ideally follow the behaviour of the in situ cover estimations nor is within the computed errorbars. The suitability of using the four error quantification parameters to assess the goodness of fit in this case seems dubious. The same goes for the used error bars (ratio of standard deviation observed over the six sampling grids to the square root of six), which is more a measure related to the variance observed over the six sampling grids than a measure of the total uncertainty involved in the detection procedure. Note that these measures for the goodness of fit are useful only as far as the in situ data can be trusted. The robustness of the methods in dealing with images where the target is absent can indirectly be assessed by looking at the images of May and December in which no macroalgae is present (with the exception of some macroalgae caught in the oysterbanks). An overestimation is observed for both approaches for these two days, perhaps the result of negative NDVI (for the image of May) or lens distortion of some kind.

Computed diatom cover was indirectly validated by comparing its seasonal behaviour with that of the fucoxanthin to chlorophyll-a ratio. The behaviour of the SOM diatom cover detections showed the higher similarity with the fucoxanthin to chlorophyll-a ratio behaviour and thus points to the better approach for diatom detection. Diatom cover was also computed for 20 min spaced images to test the consistency of the macroalgae masking and SOM approach. These images visually showed no apparent change in the diatom cover, thus a constant or near constant detection should indicate the detection is consistent. SOM showed a maximum deviation of 4.3% and the macroalgae masking approach a maximum deviation of 7.9%. Account should be taken whether the vertical migration of diatoms is the cause of the small variation observed during these 20 minute intervals. If this is the case then the SOM detection outperforms the macroalgae approach as according to literature the diatoms migrate downwards an hour before the tide, this only being observed for the SOM detection results. Taking this into account and that the computed diatom cover seasonal behaviour of the macroalgae masking approach showed high dissimilarity with the pigment ratio behaviour, it appears the SOM detection approach is the better choice of the two selected approaches for diatom detection.

Next, using the NDVI to estimate diatom biomass seems ineffective due to the small NDVI range of diatoms, regardless of the approach used. The use of diatom coverage shows more potential, particularly when plotted against the fucoxanthin to chlorophyll-a ratio. Interesting to note is that the pigment ratio values using the in situ sampling data range from 0.22 to...
0.49 which is considerably lower than observed in other literature such as in [90, 91]. Different environmental conditions may be the cause for this.

Overall, the results point towards the use of a self organising map as the more reliable method for both macroalgae and diatom detection. This could be due to various reasons, an example being that the macroalgae approach is dependent on the selected reference pixels of the macroalgae class and that it seems to be more influenced by the presence of negative NDVI than the SOM approach. Weighing the outputs accordingly (i.e. giving more influence to detection outputs of images with low water presence or red-brown macroalgae) could improve both approaches. Either way, the influence of exterior parameters such as the presence of water or red brown macroalgae on the detection results should be further looked into. It is recommended to perform a sensitivity analysis on the detection methods for the following reasons; i. to further assess the robustness of the detection methods and ii. to be able to fully understand the relationship between the variables involved. These are required to support decision making in future work and to further develop the detection method for the better (i.e. remove model errors, reduce influence of several variables, etc).

Several aspects should be looked into during the sensitivity analysis; the influence of the homomorphic filtering process on the detection output, the influence of the SOM parameters on the detection output, the influence of the selected reference pixels on the detection output of the macroalgae masking approach and the influence the time lag between the NIR and RGB images might have on the detection output. Assessing the influence of homomorphic filtering can be done comparing detection results that were performed on a set of images without and with the images preprocessed priorly with homomorphic filtering. An additional way to assess whether homomorphic filtering is really necessary for the detection process is to purposefully corrupt images, perform homomorphic filtering on these, extract their frequency domain and compare it with the frequency domain of the original images. If the difference is small than enhancing the imagery with homomorphic filtering may be redundant. The influence of the SOM parameters on the visual clustering, training time and the quantization error with respect to the test data and training data has been assessed, but it is recommended the influence of the parameters on the actual detection results is looked into as well (i.e. how big is the variation observed in the results when a variable is changed?). The influence of the time lag between RGB and NIR images is hard to assess unless the camera can be set such that the images are captured simultaneously. If that is not possible, images could be selected were a significant change is observed between the NIR and RGB images (presence of a bird in the RGB image but not in the NIR image for instance) and compared to results of RGB-NIR images captured shortly before or after this event occurred. Here too, the assessment of the influence of the involved variables and parameters, is based on the truthness of the in situ data.

5.7.3 In situ data & Validation

Proper validation, which is needed for determining the suitability of detection methods, only goes as far as the quality of the in situ data. The issue at hand is "to what extent can I trust my in situ data?". In particular for the in situ macroalgae cover estimations there is a certain
human error involved (or subjectivity), an error which is hard to quantify. If many different people were to estimate the macroalgal cover than statistically one could extract an error attached to this "subjectivity". This may however prove impractical. Another note to mention is the used errorbars for the in situ data. They are derived by taking the standard deviation of the six sampling grids average values divided by the square root of six. This seems to be too optimistic an error for macroalgal cover estimation and appears to be more a measure for the sampling error than the uncertainty involved in the estimation procedure.

The sampling error induced by the random sampling in the six grids is another important factor to note. The grid representativeness of the six grids was tested by performing the detection algorithm of both approaches on the entire field plot visible in the image and the six square areas representative of the grids were sampling was performed. A sampling error is present but does not exceed a deviation higher than 8% for the macroalgal masking approach and 7.2% for the SOM approach for macroalgae. The maximum deviation is even lower for both approaches for diatom cover, i.e. the sampling error is more present for macroalgal cover computation than diatom cover computation. The irregular positioning of the macroalgae on the field plot could account for this. However, the spread of diatoms is not entirely regular over the plot either. Perhaps the irregular positioning of macroalgae over the plot is larger than that of diatoms.

To assess the error that comes with random sampling, subjective cover estimation and pigment sampling it might be recommendable that cover estimation for both macroalgae and diatoms be performed in a controlled environment. That could also give more insight on the behaviour of diatom vertical migration (and whether it can be detected with a multispectral camera). Experimenting in a controlled environment would be additionally beneficial in assessing which variables have the most impact on the detection of diatoms and macroalgae (i.e. presence of water, dark lighting conditions, presence of red brown macroalgae, etc).

5.7.4 General recommendations

A practical addition to the detection process would be the filtering out of unfit images for detection (i.e. the images that contain rain droplets on the lens, contain the shadow of the platform, have too bright light due to the sun’s reflection on the water surface or are too blurry due to movement of the camera due to wind). Also a nice addition would be the filtering out of RGN-NIR images that are too different (i.e. the NIR image contains a bird whereas the RGB doesn’t, or as a result of camera movement). This is practical only if the time lag between RGB and NIR images indeed has a large impact on the detection output.

Another recommendation for future reference is proper camera calibration to determine the intrinsic and extrinsic parameters behind the captured images. Intrinsic parameters include the focal length, the principal point, the image size, sensor format and lens distortions. Extrinsic parameters refer to the coordinate reference system transformations between real world coordinates and camera coordinates. These parameters are important if the results need to be compared with detections obtained with other cameras and in determining how lens distortions (if consistently present) propagate through the detection procedure.
5.8 Conclusions

It is confirmed that automatic detection of diatoms using multispectral imagery is indeed possible, albeit occasional faltering. In short, two aspects inhibit the proper detection of diatoms, 1. the presence of water on the surface and 2. the similar colour of diatoms with red-brown macroalgae. The presence of water overpowers the presence of diatoms, i.e. the signal to noise ratio in this case is too low to properly detect diatoms. The SOM approach is less influenced by the presence of water due to its dependency lying also in other variables. Nevertheless, the SOM approach too is affected to some degree by the presence of red-brown macroalgae. The results on the other hand, are quite satisfying considering that with human inspection of the RGB images the diatom biofilm is also hard to distinguish from the red macroalgae.

The main observations made are:

- Water presence & similarly coloured red-brown macroalgae negatively affect the quality of the detection output
- The SOM approach outperforms the Macroalgae masking approach for macroalgae detection
- The SOM approach outperforms the Macroalgae masking approach for diatom detection
- A sampling error is present, and it is smaller for the SOM approach
- The range of diatom NDVI's is too small to properly estimate diatom biomass
- Diatom cover shows more potential in estimating the diatom biomass than diatom NDVI

Aspects that should be taken into account in future research are:

- The influence of homomorphic filtering on the detection output
- The influence of man made objects on the detection output (such as the pole in the images)
- The influence of the slight time lag between the NIR and RGB image snapshots on the detection output

Recommendations for future work are:

- The use of more in situ data for proper validation
- Performing the detection in a controlled environment
- The addition of a band centered around 568 nm to try out Meleder’s specific index
- The use of the blue band for the NDVI
- Weighing the outputs accordingly (i.e. giving more influence to detection outputs of images with low water presence or red-brown macroalgae)
- The addition of an algorithm prior to the preprocessing step that filters out irreparable images
- The addition of an algorithm that filters out RGB-NIR images that are too different (due to presence of bird, camera movement)
- Perform a sensitivity analysis to gain more insight into the variables that affect the detection
This chapter explains the methodology used for testing the feasibility of bird number detection on the Galgeplaat using Argus Bio Platform imagery. Details will be given about the raw images used for this case study and will be followed by the image preprocessing and processing methods. The input will be raw consecutive images from the Bio pan/tilt/zoom camera and the output will be a count of the birds detected in each frame.

The aim of this chapter is to determine the potential and feasibility of automatic bird number detection using the available terrestrial imagery from the Argus BIO platform. This will be done by evaluating the quality of the output of the used approach(es) to detect the birds. The chapter is additionally focused on the sub questions concerning the practicability and suitability of the detection method and software used. Note, the focus of the thesis was primarily on diatom detection. The remainder of the time was used to determine the feasibility of bird detection on the Galgeplaat. As such this chapter is more of a feasibility study rather than a case study.

6.1 Raw Image Specifications

For the bird case study the images of the Bio pan/tilt/zoom camera were used. This is because the images of the Argus cameras and the Multispectral camera rarely possess birds. The reason for the lack of birds in the images of the Argus cameras is the sand nourishment to which the cameras are directed. However, for 2011 in the month of October (high season) birds can be seen from images of the fourth Argus camera (see figure 6.1). The use of consecutive images is important because of the camouflaging nature of most wading birds. As mentioned in chapter 2 the Bio pan/tilt/zoom camera scans over five quadrants around the tower. However the snapshots taken of the quadrants may not be suitable for bird detection. This because a bird can walk from one snapshot image to the next (and thus result in double detection) and because motion of the bird cannot be taken advantage of. The settings of the camera were changed such that in between the traditional scanning of the five quadrants the
camera would also take 20 consecutive snapshots of some hand selected scenes with different
digital zooms where birds have generally been observed. It should be noted that this was
done only for the month of June 2012, which is not the season in which bird numbers peak
(October).

Figure 6.1: Birds following the shore line. Image captured from the fourth Argus camera on the
platform October 2011. Note the muddy raindrop on the lens.

Four sets of image series (each 20 frames long) were selected for this case study. The frame
rate is two frames/second. The first two possess the same zoom (close scale), the third a
smaller zoom (far scale) and the fourth the smallest scale (farthest scale). The zoom values
(9000, 6500 and 4000) do not say much about the focal length. Through empirical measure-
ments done by Deltares however, they were correlated to the horizontal field of view (HFOV)
in degrees. these values can be seen in Table 6.1. Do note that these values are approxima-
tions.

The first video possesses one bird, the second three birds and the third possesses nine birds.
For the fourth video sequence the ground truth is unknown, but it appears there are approxi-
mately 20 birds present. Because the ground truth is unknown for this video it will not be
used to assess the suitability of the methods. It will serve purely to see how the detection
deals with this zoom scale. See Table 6.1 for all details of the four video sequences and fig-
ure 6.2 to see an individual frame of the first 3 video sequences. Figure 6.3 shows the first
frame of VID4. The yellow squares indicate the location of the birds present in the image. In
VID2 there is a reasonable amount of water on the surface producing reflections of the birds.
In the far scale image (VID3) also birds with white and black coats are present (perhaps
oystercatchers or ducks). Their coats make it seem there are two small birds next to each
other instead of one bird. VID4 contains highly oblique images (i.e. the horizon is visible).
Table 6.1: Test set Case Study II: Number of birds, dates, times and Zoom details.

<table>
<thead>
<tr>
<th>Series ID</th>
<th>Nr. of birds</th>
<th>Date</th>
<th>Time</th>
<th>Zoom</th>
<th>HFOV</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>VID1</td>
<td>1</td>
<td>19-06-2012</td>
<td>10:40</td>
<td>9000</td>
<td>4.6</td>
<td>clear sky</td>
</tr>
<tr>
<td>VID2</td>
<td>3</td>
<td>20-06-2012</td>
<td>07:10</td>
<td>9000</td>
<td>4.6</td>
<td>clear sky, bird reflections</td>
</tr>
<tr>
<td>VID3</td>
<td>9</td>
<td>20-06-2012</td>
<td>09:10</td>
<td>6500</td>
<td>11.9</td>
<td>clear sky, birds at shoreline</td>
</tr>
<tr>
<td>VID4</td>
<td>unknown</td>
<td>20-06-2012</td>
<td>08:25</td>
<td>4000</td>
<td>22.5</td>
<td>clear sky, horizon visible</td>
</tr>
</tbody>
</table>
Figure 6.2: First frames of the first three selected video sequences. From top to bottom: VID1, VID2, VID3. The yellow square indicates the position of a bird (the ground truth). The frames have already undergone preprocessing.
6.2 Image Preprocessing

As was done in the diatom case study the images were preprocessed using the homomorphic filtering method. Text bands were removed in these images as well. The transition point parameter alpha and the steepness parameter of the high pass filter were respectively set to 1000 and 2. This because the theoretically low frequency components (in the centre of the frequency domain) are less dominant than as observed in the images of the Multispectral Camera.

While the consecutive images were only half a second apart radiometric normalisation (see chapter 5 for explanation) was performed with reference to the first frame to avoid any sudden illumination changes due to cloud overpass. The frames in figure 6.2 and figure 6.3 are already preprocessed.

6.3 Bird Detection

As a consequence of the small number of frames available for bird detection, methods such as MoG and Multimodal mean, were ruled out. The toolbox used and approaches followed will be laid out in this section. Three methods for this case study were tested for their applicability to the available study set:
1. Frame Differencing
2. Background subtraction using a median filtered image as background
3. Background subtraction using a weighted mean image as background

Refer to section 4.2 for an explanation of background subtraction techniques.

6.3.1 Output requirements

Current bird monitoring of the Galgeplaat (see section 1.1) produces the maximum number of foraging birds for each compartment, the average number of foraging birds for each compartment and the foraging time in minutes for each compartment. The foraging time will not be an output in this case study as only video sequences of 10 seconds were used. The laid out compartments for bird counting on the Galgeplaat will not be used for validation as the Bio pan/tilt/zoom camera was not specifically directed to them for the video sequences. The output of the detection can be either 1. bird count per frame and 2. bird count per video sequence. For practical purposes the latter is more beneficial because less storage space is required and it provides a better overview for the ecologist. It is possible that birds walk in and out of the video during the video sequence. As such giving a quality measure (i.e. an uncertainty) along with the outcome will be practical. The bird count per frame can aid in assessing the suitability of the methods used.

6.3.2 The Matlab Computer Vision system toolbox

For creating the video sequences (avi format) and for the video processing use was made of the Computer Vision system toolbox of Matlab. The toolbox provides tools for the design and simulation of computer vision and video processing systems [127, 128]. The setting up of relevant system objects (such as output lay out and the use of size as a parameter) was based on the example codes (tutorials) of car tracking in Mathworks [129].

6.3.3 Frame Differencing

In this approach the previous frame is used as the background model. This means only for 19 out of the 20 frames a bird count will be made. Prior to the subtraction both the current frame and the previous frame are converted to a grey scale image, this is useful because it requires the processing of only one band instead of three. A threshold has to be set manually to determine what magnitude of the frame difference image (current frame minus previous frame) is considered foreground (i.e. a moving object). When the foreground is determined using the user defined threshold (producing a binary image) it is additionally filtered with a 2D median filter to remove speckles and then morphologically closed (refer to section 4.2.1 for an explanation) with a disk shaped structuring element (10 pixel diameter for the close scale video sequences and 8 for the far scale video sequence). Morphological closing is required because frame differencing is subjected to the aperture problem. After these filtering operations an image results which possesses several ‘blobs’. Only the blobs bigger than a particular size (60 pixels for the close scale video sequence and 10 for the far scale video sequence) are then selected as bird candidates, i.e. size is a decision parameter in deciding whether a blob is a
bird or not. The bird pixel sizes were experimentally derived. For the Matlab code refer to Appendix G. Figure 6.4 shows a section of a frame and its counterpart frame difference with respect to the previous frame. Figure 6.5 shows a flowchart of the steps taken for the frame differencing approach.

**Figure 6.4:** An original section of a frame of VID2 and its counterpart frame difference with respect to the previous frame.

**Figure 6.5:** Processing task block for the frame differencing approach.

### 6.3.4 Background subtraction with a median filtered background image

This approach in essence is similar to frame differencing except that it uses a median filtered image as the background image. The background image is produced by taking the median of the pixel values over all the available frames. Then an image should be produced in which moving objects are filtered out (provided that the birds keep moving). 2D median filtering (for noise removal) and morphological closing was performed for this approach as well with
the same parameter values as in the previous section. The noise filtering and morphological closing was then also followed by the removal of blobs less than 60 pixel large or 10 pixels large depending on the scale of the video. Figure 6.6 shows the steps taken for this approach. For the Matlab code refer to Appendix G.

![Diagram](image)

**Figure 6.6:** Processing task block for background subtraction with a median filtered background image.

### 6.3.5 Background subtraction with a weighted mean image

In this approach the background is modelled by use of an adaptive filter (refer to equation 6.1 in section 4.2). It basically takes a percentage of the current frame pixels and a percentage of the background image at the previous time step to update the background image. Since in the earlier two approaches the reflections of the birds were sometimes also counted as potential blobs it was decided to take advantage of the standard deviation of the pixel values over time as well. The higher the standard deviation the higher the potential that the pixel is a foreground pixel. When the standard deviation is small it is more likely the respective pixel represents a bird reflection (less contrast) rather than the actual moving target. The weighted mean and standard deviation adaptive filters equations were obtained from [55] were they were used for the detection of people in video images. Refer to equation 4.1 in section 4.2 for the computation of the weighted mean image (using the weighted mean adaptive filter). See equation 6.1 for the computation of the standard deviation image (using the standard deviation adaptive filter).

\[ \sigma_t^2 = \alpha * (P_t - \mu_t)^2 + (1 - \alpha) * \sigma_{t-1}^2 \]  

(6.1)

Where, \( \sigma_t \) the standard deviation image at time \( t \), \( \alpha \) is the learning rate (values 0 to 1), \( P_t \) the image pixel values at time \( t \) and \( \mu_t \) is the mean image computed up to frame \( t \). For an example of a weighted mean image see figure 6.7. Note how some birds can vaguely be seen in it. This is because for those birds too little movement was observed. For an example of a
standard deviation image see figure 6.8.

![Image of weighted mean image for VID3](image_url)

**Figure 6.7**: Example of a weighted mean image for VID3.
Because there are only a few frames available a low learning rate was selected (alpha=0.08). This makes sure that the previous background image or standard deviation image is not discarded very quickly. The background subtraction was performed on all three channels as opposed to the previous two methods, following the approach of [55]. This was done with the rationale that a change observed in any of the colour channels is an indicator that the pixel is a foreground pixel. The background subtraction thus produces three difference images (one for each colour). From these three difference images the maximum of each pixel was selected (change in any colour channel shows potential that the pixel is a foreground pixel). Two thresholds were selected, one for thresholding the difference image and one for thresholding the standard deviation image. Basically, if a pixel in the 'difference image' is bigger than the difference threshold AND the same pixel in the standard deviation image is bigger than the standard deviation threshold than it is considered foreground. The difference threshold was set to 40 for all video sequences and the standard deviation threshold was set to 17 or 35 (17 for VID1 & VID3 and 35 for VID2). The values were manually extracted by observing the difference and standard deviation images. For VID1 the standard deviation threshold was set low because the camouflaging coat of the curlew in the frame produces a smaller spread over time for bird pixels. For VID3 the standard deviation threshold was set low because the birds movement is less apparent (and is overlapping) due to the scale of the image frames. Morphological closing of the produced binary foreground image was performed in this method too. Figure 6.9 shows the processing task block for this approach. Table 6.2 shows all the parameter settings for the three selected methods.
6.3 Bird Detection

**Figure 6.9:** Processing task block for background subtraction with a weighted mean background image.

**Table 6.2:** The parameter settings for each used method. FD stands for the Frame differencing approach, MF stands for the Median filtered background method and WM refers to the Weighted Mean method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameters</th>
</tr>
</thead>
</table>
| FD     | \( T_{diff} = 40 \)  
|        | \( MinBlobsize = 60 \) (close scale) or \( MinBlobsize = 10 \) (far scale)  
|        | \( DiskRadius = 10 \) (close scale) or \( DiskRadius = 8 \) (far scale) |
| MF     | \( T_{diff} = 40 \)  
|        | \( MinBlobsize = 60 \) (close scale) or \( MinBlobsize = 10 \) (far scale)  
|        | \( DiskRadius = 10 \) (close scale) or \( DiskRadius = 8 \) (far scale) |
| WM     | \( T_{diff} = 40 \)  
|        | \( T_{std} = 17 \) (VID1 and VID3) or \( T_{std} = 35 \) (for VID2)  
|        | \( MinBlobsize = 60 \) (close scale) or \( MinBlobsize = 10 \) (far scale)  
|        | \( DiskRadius = 10 \) (close scale) or \( DiskRadius = 8 \) (far scale) |

\( T_{diff} \) refers to the frame difference threshold, \( MinBlobsize \) refers to the selected minimum 'blob' size depending on the zoom of the video sequence (in pixels), \( DiskRadius \) refers to the radius of the morphological closing structuring element in pixels and \( T_{std} \) refers to the threshold set for the standard deviation image.
6.4 Results

The number of birds detected for each frame of VID1, VID2 and VID3 can be seen in Table 6.3. Figure 6.10 shows example outputs of the bird detection with a green cross indicating the position where the method believes it detected a bird. In the 'frame differencing' method and the 'background subtraction with a median filtered image', occasionally the reflection of a bird was detected as a bird (this was the case for VID2, where there was a sufficient layer of water on the surface). The third video sequence (VID3), with the most birds of the three videos, produced the most erroneous detections. This may be due to the limited movement visible in the video. Notable is that in the far scale video no waders are visible (only gulls with their white coats and ducks with partially white coats). This is a problem because it is desirable to perform bird detection (foraging waders) over a great area (far scale). Also worth mentioning is that the centroids of the detected birds did not always coincide with the birds’ main bodies (sometimes it was the legs or the head). This is important because this indicates the presence of the aperture problem. It means that the 'blobs’ observed are not representative of the entire bird’s body. An example being for instance if only the chest of the bird is detected. Since size is used for determining the candidacy of a 'blob’ this can have influence on the output (i.e. a bird being ruled out as a candidate because it is too small).
Table 6.3: Bird count per frame for each of the three methods for all three video sequences. GT stands for Ground Truth, i.e. the actual number of birds present in the video sequences.

<table>
<thead>
<tr>
<th>Method</th>
<th>Series ID</th>
<th>Frame number</th>
<th>GT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</td>
<td></td>
</tr>
<tr>
<td>FD</td>
<td>VID1</td>
<td>- 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VID2</td>
<td>- 3 3 3 3 3 3 3 4 3 3 3 3 3 3 3 3 3 2 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VID3</td>
<td>- 7 3 5 4 8 9 5 9 7 7 5 8 9 8 8 9 8 7 6 9</td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>VID1</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VID2</td>
<td>3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VID3</td>
<td>9 9 8 7 7 9 8 7 5 7 7 6 7 8 8 9 8 8 7 9</td>
<td></td>
</tr>
<tr>
<td>WM</td>
<td>VID1</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VID2</td>
<td>3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VID3</td>
<td>9 9 9 9 9 9 9 8 9 9 9 9 8 10 9 10 9 9 9</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 6.10:** Visual examples of the bird detection output from the weighted mean background image approach. A green cross indicates what the algorithm has identified as a bird. The bird count at that particular frame is shown at the top left corner.
6.5 Method Suitability

For VID4, the weighted mean background subtraction method outperformed the other two methods as well. However, it is clear that beyond a certain latitude in the image the detection is failing (note that ships in the far horizon are counted as birds as well). In figure 6.11 a visual example is shown of the detection output for VID4. The horizontal line indicates where the method is performing good (below the line) and where the method is performing poorly (above the line). The line’s position was estimated based on visual observation of the image as the ground truth is unknown. The horizontal line is not far from the centre of the image. The distance in the real world from the platform to where the centre of the image is pointing to can be found by using the tilt of the camera and the height of the camera (refer back to section 2.5.3). The height of the camera is 16.895 metres, and the observed tilt for VID4 is 6.75 degrees downwards from the horizontal. Using the Pythagorean theorem the distance between the platform and the real world objects at the centre of the image is 142.75 metres. Thus, beyond approximately 150 metres the detection is faltering.

![Image](image.jpg)

**Figure 6.11:** Visual example of the bird detection output for VID4. The horizontal white line indicates where the detection performs good (below the line) and where it performs poorly (above the line).

6.5 Method Suitability

In this section the purpose is to determine the most suitable method for bird detection when using 20 frame video sequences. This was done using only the videos for which the ground truth is known (VID1, VID2, VID3). Table 6.4 shows the average true positives, false positives and false negatives for each video sequence for all three methods. These were extracted from the tables in appendix H which possess the number of true positives, false positives and false negatives per frame. A true positive (TP) represents the number of correctly identified
birds in a frame. A false negative (FN) represents the number of not identified birds per frame. A false positive (FP) represents all the wrongly identified birds per frame, such as a reflection of a bird being counted as a bird.

For VID1 the background subtraction using a median filtered background or a weighted mean background the detection produced entirely correct detections for all frames despite the camouflaging nature of the curlew’s coat. Detection for VID2, using background subtraction with a median filtered background or a weighted mean background, outperformed the frame differencing approach again but the median filtered background approach did possess one frame with erroneous detection. All these methods performed the least good for the far scale video (VID3) but with the weighted mean (WM) background approach outperforming the other two methods significantly. Even though WM is the only method for VID3 that possesses false positives (for 3 out of the 20 frames) its true positives average (8.9) is the least deviating from the actual number of birds (9) present in the video. Considering the previous, the background subtraction using a weighted mean background method from the three methods tested is the most promising for bird detection using these particular video sequences.

<table>
<thead>
<tr>
<th>Series ID</th>
<th>Occurrence type</th>
<th>FD</th>
<th>MF</th>
<th>WM</th>
</tr>
</thead>
<tbody>
<tr>
<td>VID1</td>
<td>TP</td>
<td>0.89</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>FN</td>
<td>0.11</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VID2</td>
<td>TP</td>
<td>2.95</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>0.053</td>
<td>0.05</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>FN</td>
<td>0.053</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VID3</td>
<td>TP</td>
<td>6.95</td>
<td>7.55</td>
<td>8.90</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>0</td>
<td>0</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>FN</td>
<td>2.11</td>
<td>1.45</td>
<td>0.15</td>
</tr>
</tbody>
</table>

**Table 6.4:** Average true positive (TP), false positive (FP) and false negative (FN) for each method for the three video sequences.

### 6.6 Proposed post processing of results

As mentioned earlier (see section 6.3.1) it is more practical to obtain one bird count per video sequence rather than a bird count for each frame. This means that using all the counts of each frame a value should be extracted that is most likely to represent the actual bird value present in the video sequence. Typically this would be the arithmetic mean, assuming that the errors of the measurements are random, i.e. the error of a given measurement is as likely to be above as well as below the mean [130]. In this way the random errors of the measurements cancel each other out as you sum them in the process of computing the arithmetic mean.

Based on statistical convergence, the variability of the estimator (the mean), epsilon, is defined by equation 6.2.
\[ \epsilon^2 = \frac{1}{N} \left( \frac{\sigma}{X} \right)^2 \] (6.2)

where \( \epsilon \) the variability of the estimator, \( N \) the number of measurements, \( \sigma \) the experimental standard deviation of the finite sample and \( X \) the true mean. Since computing the true mean requires an infinite number of samples, \( X \) is estimated with the arithmetic mean. The variability of the estimator is inversely proportional to the number of measurements, and directly proportional to the relative fluctuation of the random variable \( \left( \frac{\sigma}{X} \right) \) [131]. Note, the equation assumes that the uncertainty of the mean is smaller than the uncertainty of a single measurement. Epsilon is an indicator of how confident you are that the actual mean is within the range \( [\bar{x} - \epsilon \bar{x}, \bar{x} + \epsilon \bar{x}] \).

In our case \( N \) equals the number of frames (20). Assuming the errors are randomly distributed then logically the larger the number of frames the smaller the uncertainty of the mean. Take the output counts of VID3 for the weighted mean background approach. The arithmetic mean is exactly 9, the experimental standard deviation is 0.459. Using equation 6.2 it can then be found that \( \epsilon = 0.0114 \), i.e. we are confident that the actual mean lies within 1.14% of the arithmetic mean (between 8.9 and 9.1). Producing a non integer output can be confusing (there is no such thing as 8.9 birds) that is why it is best to round the arithmetic mean to its nearest integer, the upper bound (9.1) to the nearest integer greater than or equal to the upper bound and the lower bound to the nearest integer less than or equal to the lower bound. That would produce an output of 9 birds \( \pm 1 \). In cases when there is no spread (standard deviation equal zero), as can be observed for VID1 and VID2 for the same method, then typically the count of 1 frame should suffice. However, this is all based on the following assumptions:

- The error is random (above and below the mean)
- No birds walk in or out of the field of view during the recording of the video sequence
- The detection method produces equal uncertainties for videos of different zoom
- The number of birds in the field of view has no influence on the uncertainty of a single measurement

The first assumes that the selected detection method produces random errors (i.e. it occasionally overestimates the bird number and it occasionally underestimates the bird number in a video sequence). While this happens to be the case for VID3 for the weighted mean background method this may not necessarily be the case in general (see the median filtered background method output which for VID3 generally underestimates the number of birds). This would mean computing the variability of the estimator in this case would produce a value lower than the actual number of birds present.

Also it is assumed that the bird number is constant in the video sequence, i.e. no birds walk in or out the field of view during the 10 seconds of recording. This (birds walking in and out of the field of view) is not observed in the three video sequences observed but it is probable. In such a case an additional uncertainty should be added representative of the probability of
this event occurring.

Next, using just the variability of the estimator, it is assumed that the zoom factor of the
video has no effect on the uncertainty of the output. From the three videos used, the far
scale video (least zoom) was the most prone to erroneous detections (observed for all three
methods). This is probably as a result of using size of the birds as one of the parameters for
detection. In far scale images the birds are very small, as such it is more likely noise will be
interfering in the process of detection.

Finally, note that equation 6.2 does not take into account the number of birds present in a
video sequence. However, the larger the number of birds in a video sequence the higher the
probability of occlusion (birds occluding each other and thus producing an underestimated
count). As such it is reasonable to assume that the uncertainty of the frame count will be
higher if many birds are present in the frame than when there are fewer birds present. This
is thus an additional uncertainty to be added on top of the uncertainty of the mean.

6.7 Discussion

Detection of bird numbers on the Galgeplaat can be achieved using consecutive snapshots
images of the Bio pan/tilt/zoom camera of the Argus BIO Platform. The methods used
do require a priori knowledge (for the setting of thresholds) from the user and as such they
cannot be referred to as fully automatic. However, after the thresholds have been set (the
difference image threshold and the standard deviation image threshold) no additional effort
on the user part is required.

From the three videos tested the far scale video sequence (VID3) produced the most erroneous
detections for all three methods. This is also the case for VID4 although this is based on
visual observation (ground truth unknown). This 'faltering' from the detection methods on
far scale videos appears to be the result of the limited movement in the far scale videos or as
a result of the small sizes of the birds. Noise is filtered out using a 2D median filter on the bi-
nary image that is produced after thresholding the difference image (between the background
image and the current frame image). The produced blob sizes in the far scale images for birds
however are not significantly larger than the noise blobs observed in the binary images. In
that case, if a median filter is used it should have a small diameter otherwise bird blobs may
be filtered out. If no median filter is used to filter out noise the detection will become less
robust against the presence of noise (resulting probably into an overestimation of the bird
count). As size is one of the deciding parameters in the selection of potential bird blobs after
the noise removing median filtering it is essential this be taken into account in future work
(for instance by using an additional deciding parameter for bird detection or another noise
filtering method).

Also important to note is that in the two far scale videos (VID3 and VID4) only birds with
white or highly contrasting coats can visually be seen. This poses a problem as many of the
waders that are of ecological interest have camouflageing coats. It is even more so a problem

Penelope Rammos  Master of Science Thesis
as remotely sensed bird detection will only be useful to ecological studies if it covers a large ground area, i.e. far scale video sequence are prerequisites for future development. Since waders with camouflaging coats are not visible in far scale videos unless their movement is large it might be advisable to add an additional thermal imaging camera on the Argus BIO platform. Thermal imaging cameras detect radiation in the infrared range (9 to 14 µm) and show potential for bird detection as they make warm objects (i.e. birds) stand out against the cooler environment. If it where to be purchased attention will have to be paid to the range to the target and the size of target to make a proper selection of the lens focal length [132]. What should be considered on the other hand, is that if a thermal image were to be used it is still no means to distinguish the different species from each other (unless the different species have different thermal emissivities). In far scale videos, size of the bird still shows the most potential to distinguish between species (in particular for wading birds with camouflaging coats). For close scale images shape and colour of the birds also show potential in distinguishing between species (in the far scale images, beaks, legs, and colour patterns are not visible). Perhaps a suggestion for future work is for the camera to automatically zoom into the areas in the video where it detects movement such that colour and shape of birds can be used additionally to distinguish the bird species.

From the three methods used the background subtraction using a mean weighted and standard deviation image produced the better results. The use of the standard deviation image was particularly effective in removing false positive detections (VID2) caused by reflection of the birds in the water. On the other hand the threshold for the standard deviation image appears to be different in VID2. The reason may be the higher contrast. For future work it would be beneficial if the threshold for the standard deviation image could be extracted automatically from the image. This could perhaps be achieved by taking advantage of the cumulative distribution function of the standard deviation image. A pitfall with all three methods used is that waves if visible in the video will be counted as moving objects as well.

The weighted mean background image produced by the method, although a good approximation, could on the one hand be better achieved with more frames (i.e. 100 frames or more) or by increasing the time interval between frames. The wading movement observed in consecutive frames often overlaps, resulting in a difference image where the bird as a whole is not always visible (aperture problem). This being the main reason why the output centroids are not always at the centre of the bird’s body. Decreasing the frame rate is the other option because it would reduce the overlapping of movement observed in consecutive frames (i.e. 1 frame/2 seconds instead of 2 frames/second). Another yet option would be to make use of time exposure images (currently only produced by the Argus cameras on the platform) to use as the background image. In that way the need for many frames to create a good background image could be avoided entirely. Using more frames on the other hand, means that more advanced techniques such as MoG or Multimodal mean can be used to perform bird detection.

Perhaps also beneficial for future research is for the algorithm to automatically set the bird size threshold (a decisive parameter) by making use of the zoom of the video and the vertical position of the bird in the image. Logically, the bird size depends on the scale of the image and due to the oblique nature of the video frames (perspective in the image) also on where the bird is standing in the image. When the bird is positioned at the top of the image it is
further from the camera than when it is positioned at the bottom of the image. Determining and incorporating the mathematical relationship between the zoom, the bird position in the image and the bird size would be a good investment because aside from making the process more automatic it would also provide an increased robustness in distinguishing different bird species. In other words, it would be more robust against problems occurring when multiple bird species of different size (i.e. goose versus dunlin) are present in the same frame but at different vertical positions in the frame. A large goose when positioned at the top of the image (far away from the camera) would appear the same size as a Dunlin positioned at the bottom of the image (closer to the camera). In order to relate bird size in pixels to real world bird size it is necessary to incorporate the focal length and the distance to the camera in the mathematical relationship as well. Possible lens distortions should be taken into account as well. All these camera parameters can be obtained through camera calibration. If highly oblique videos are to be used (VID4) then it is advised the detection is performed below a certain vertical position in the image. This could be below the horizon observed in such images, or in case of VID4 everything that is within 150 metres from the platform.

Next, no use was made of the birds velocities. This could be useful for rejecting flying birds or resting birds (which are not counted in current bird monitoring). It would also be useful for tracking of the birds (i.e. with a Kalman filter) the birds (and thus reduce occlusion problems). Next to velocity it might also be useful to take into account the periods of high and low tide. It was observed that the majority of birds were wading along the shoreline and walking with it as it resided during low tide. Making sure the camera follows the receding waters could thus be useful.

It is logical the bird count output should be accompanied by a quality measure. For the moment, the used ‘variability of the estimator’ is a reasonable measure for an uncertainty range output. However it does not incorporate the uncertainties produced by birds moving in and out of the field of view, the effect of the image scale on the output count or the number of birds in the scene. More videos need to be tested to assess the influence of the previous on the detection output. Also, the degree of influence of different atmospheric conditions should be researched (fog, or the presence of frost on the mudflat).

6.8 Conclusions

The aim of this chapter was to determine the potential and feasibility of automatic detection of bird numbers using imagery from the Argus Bio Platform. It shows potential for ecological monitoring as it would reduce the costs significantly (the costs involved reached an annual sum of 17.826 Euro in 2007 [1]) and because it could be performed any time of the year (not just October as is currently the case). The costs involved using the Argus Bio platform imagery would be mostly comprised of platform maintenance. Observing the results, detection of bird numbers is feasible but requires more testing on more video sequences for several reasons: primarily, to extract the right uncertainties that come with problems such as birds walking in and out of the field of view and the influence of the zoom and bird numbers on the detection procedure. Secondly, further testing is also required to determine the robustness of the detection procedure (i.e. how does it work on videos were no birds are present) and whether
6.8 Conclusions

it produces random errors. If it does not produce random errors, does it mostly underestimate or overestimate the ground truth? This need for more testing is required regardless of the approach selected to perform bird detection. A recommendation for future work is the use of videos with more frames, in that case more advanced motion detection techniques such as MoG and Multimodal mean could be used for bird detection.

The main observations made are:

- Background subtraction using a weighted mean image performed the best on the three test videos
- The standard deviation image was useful in avoiding false positive detections (reflection of birds in water)
- Output centroids were not always at the centre of the bird’s body (aperture problem)
- In the far scale videos only (white) gulls were visible

Aspects that should be taken into account in future research are:

- The influence of the zoom on the detection procedure (how big is the uncertainty involved)
- The influence of the bird number (higher probability of occlusion) on the detection procedure
- The influence of atmospheric conditions on the detection procedure (fog, frost)

Recommendations or suggestions for future work are:

- Testing on more videos
- Determining & incorporating the mathematical relationship between the zoom, the bird position in the image and the bird size
- The addition of a thermal imaging camera on the platform
- The use of time exposure images to use as a background image for background subtraction
- Making use of the birds velocities to reject flying birds or resting birds
- Increase the time interval between frames to 0.5 frames per second
- Increase the length of the video to obtain more frames (i.e. 5 minute videos instead of 10 second videos)
Chapter 7

Conclusion

The aim of this thesis was to assess the potential of the Argus Bio platform for ecological monitoring of the Galgeplaat mudflat in the Oosterschelde. To determine the platform’s potential the main research question was formulated as:

*Is it possible to automatically detect the main biological features (Birds and Benthic organisms) on the Galgeplaat mudflat from the available set of terrestrial images from the Argus-Bio platform and validate it with the available in situ data?*

Two case studies were formulated to answer this question; 1. diatom detection and 2. bird number detection.

For the detection of diatoms two methods where used; I. maximum likelihood classification combined with the masking of the macroalgae (the undesired target) and II. Kohonen’s self organising maps (SOM). Identification of diatoms was best achieved with the Self organizing map (SOM) approach. Using SOM for macroalgae identification alone, on the other hand, may be excessive as the maximum likelihood approach (lower learning curve) produced similar results for this undesired, in this thesis, target. However, taking into account that SOM clusters all the targets in the image simultaneously (diatoms, macroalgae and other), it is the more promising approach if both targets (diatoms and macroalgae) are needed for ecological monitoring. Recommendations amongst others include the use of more in situ data for proper validation, performing the detection in a controlled environment, the use of the blue band for the NDVI, the addition of a band centered around 568 nm to try out Meleder’s specific index and performing a sensitivity analysis to gain more insight into the variables that affect the detection process.

For the detection of bird numbers three low complexity background subtraction techniques were used; I. frame differencing, II. background subtraction using a median filtered background image and III. background subtraction using a weighted mean background image.
The latter was the most promising method for the three 20 frame videos used for testing. The method made additional use of a standard deviation image to filter out false positives occurring due to bird reflections on surface water. The bird counts per frame from the detection were then used to extract an average bird count for the video. The uncertainty was provided with a basic uncertainty measure; the ‘variability of the estimator’. Further testing is required to improve the uncertainty measure that will accompany the bird count output of the detection. Recommendations amongst others include the incorporation of a mathematical relationship between the zoom, the bird position in the image and the bird size, adding a thermal camera to the platform, the use of the birds velocities to rule out resting and flying birds and the use of videos with more frames such that more advanced techniques may be tested also.

The overall conclusion is yes, it is possible to detect biological features using the Argus Bio platform. In the current phase of development however, the process cannot be claimed fully automatic yet, as a priori knowledge from the user’s part is still required. A priori knowledge for diatom detection using the SOM approach are for instance the standard orangeness or greenness value to which the pixels are compared to and the proper parameter settings of the SOM (although these are required only once). For bird detection the a priori knowledge required are for the setting of the thresholds of the difference image, the standard deviation image and the bird size. Important to note is that the answer to the research question greatly depends on the ecological target to be detected as well. Continuing, it is in my opinion essential for future research that an algorithm be developed to filter out the unfit images (images were raindrops on the lens are visible or where the image is too blurry for any kind of detection). Some existing example methods are the partial blur image detection method and blur analysis framework of [133] suitable for detecting different blur types (motion blur, out of focus blur) and images with blurred regions (i.e. raindrops), or the no-reference blur metric of [134] which is based on the spread of edges observed in an image. In this way, no effort or time is required prior to the detection process to remove the irreparable images.
### Table of monitoring costs Galgeplaat

**Figure A.1:** Table of Costs of the experimental Sand Nourishment on the Galgeplaat 2007 to 2011 [1]

<table>
<thead>
<tr>
<th></th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Morfologie</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singlebeam</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Multibeam</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RTK</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Laser</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>€ 30,000</td>
</tr>
<tr>
<td><strong>Ecologie</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bodemdieren</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monstorname</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Analyse</td>
<td>-</td>
<td>€ 18,000</td>
<td>18,000</td>
<td>18,000</td>
<td>18,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>€</td>
<td>€</td>
<td>€</td>
<td>€</td>
</tr>
<tr>
<td>Vogellevelling</td>
<td>17,826</td>
<td>-</td>
<td>€ 9,000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Musselen</td>
<td></td>
<td>€ 51,100</td>
<td>35,900</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Zwevend stof</strong></td>
<td></td>
<td>€ 46,000</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Hydrodynamica</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroomsnelheid</td>
<td>-</td>
<td>€ 12,000</td>
<td>12,000</td>
<td>-</td>
<td>12,000</td>
</tr>
<tr>
<td>Golfhoogte</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix B

## Inventory of Birds

**Table B.1:** Inventory of Birds of the Middle region of the Oosterschelde

<table>
<thead>
<tr>
<th>Image</th>
<th>Name</th>
<th>Latin name</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Shelduck" /></td>
<td>Shelduck</td>
<td>Tadorna tadorna</td>
<td>Duck</td>
</tr>
<tr>
<td><img src="image2.png" alt="Goldeneye" /></td>
<td>Goldeneye</td>
<td>Bucephala clangula</td>
<td>Duck</td>
</tr>
<tr>
<td><img src="image3.png" alt="Oystercatcher" /></td>
<td>Oystercatcher</td>
<td>Haematopus ostralegus</td>
<td>Wader</td>
</tr>
<tr>
<td><img src="image4.png" alt="Ringed Plover" /></td>
<td>Ringed Plover</td>
<td>Charadrius hiaticula</td>
<td>Wader</td>
</tr>
<tr>
<td><img src="image5.png" alt="Mallard" /></td>
<td>Mallard</td>
<td>Anas platyrhynchos</td>
<td>Duck</td>
</tr>
<tr>
<td><img src="image6.png" alt="Crested Grebe" /></td>
<td>Crested Grebe</td>
<td>Podiceps cristatus</td>
<td>Grebe</td>
</tr>
<tr>
<td><img src="image7.png" alt="Widgeon" /></td>
<td>Widgeon</td>
<td>Anas penelope</td>
<td>Duck</td>
</tr>
<tr>
<td><img src="image8.png" alt="Dunlin" /></td>
<td>Dunlin</td>
<td>Calidris alpina</td>
<td>Wader</td>
</tr>
<tr>
<td>Image</td>
<td>Name</td>
<td>Scientific Name</td>
<td>Category</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td>-----------------</td>
<td>----------</td>
</tr>
<tr>
<td><img src="image1.png" alt="Eurasian Curlew" /></td>
<td>Eurasian Curlew</td>
<td>Numenius arquata</td>
<td>Wader</td>
</tr>
<tr>
<td><img src="image2.png" alt="Redshank" /></td>
<td>Redshank</td>
<td>Tringa totanus</td>
<td>Wader</td>
</tr>
<tr>
<td><img src="image3.png" alt="Grey Plover" /></td>
<td>Grey Plover</td>
<td>Pluvialis squatarola</td>
<td>Wader</td>
</tr>
<tr>
<td><img src="image4.png" alt="Bar-tailed Godwit" /></td>
<td>Bar-tailed Godwit</td>
<td>Limosa lapponica</td>
<td>Wader</td>
</tr>
<tr>
<td><img src="image5.png" alt="Greylag Goose" /></td>
<td>Greylag Goose</td>
<td>Anser Anser</td>
<td>Goose</td>
</tr>
<tr>
<td><img src="image6.png" alt="Brent Goose" /></td>
<td>Brent Goose</td>
<td>Branta bernicla</td>
<td>Goose</td>
</tr>
<tr>
<td><img src="image7.png" alt="Sanderling" /></td>
<td>Sanderling</td>
<td>Calidris alba</td>
<td>Wader</td>
</tr>
<tr>
<td><img src="image8.png" alt="Great Cormorant" /></td>
<td>Great Cormorant</td>
<td>Phalacrocorax carbo</td>
<td>Sea bird</td>
</tr>
<tr>
<td><img src="image9.png" alt="Red-breasted Merganser" /></td>
<td>Red-breasted Merganser</td>
<td>Mergus serrator</td>
<td>Duck</td>
</tr>
<tr>
<td><img src="image10.png" alt="Ruddy Turnstone" /></td>
<td>Ruddy turnstone</td>
<td>Arenaria interpres</td>
<td>Wader</td>
</tr>
<tr>
<td><img src="image11.png" alt="Common Snipe" /></td>
<td>Common Snipe</td>
<td>Gallinago gallinago</td>
<td>Wader</td>
</tr>
<tr>
<td><img src="image12.png" alt="Pied Avocet" /></td>
<td>Pied Avocet</td>
<td>Recurvirostra avosetta</td>
<td>Wader</td>
</tr>
<tr>
<td><img src="image13.png" alt="Arctic Tern" /></td>
<td>Arctic Tern</td>
<td>Sterna paradisaea</td>
<td>Sea bird</td>
</tr>
</tbody>
</table>
Appendix C

Absorbance spectra Fucoxanthin, Chlorophyll-a and Diadinoxanthin

![Absorbance spectra for the pigments Fucoxanthin and Chlorophyll-a](image)

**Figure C.1:** Absorbance spectra for the pigments Fucoxanthin and Chlorophyll-a [2]
Figure C.2: Absorbance spectra diadinoxanthin [3]
Appendix D

Introduction to SOMPAK

D.1 SOMPAK parameters

For the SOM method the public domain software package SOMPAK was used. This was developed by the Laboratory of computer and information science department of the University of Technology of Helsinki in 1992 [77]. The package splits the SOM calculation procedure into 4 programs, each of which require the input of several parameters. The 4 programs and their required input parameters are listed below:

- Initialization program
  - randinit
  - lininit
  - Network dimensions (ydim & xdim)
  - Network topology (topol)
  - neighbourhood function (neigh)

- Training Program
  - rlen
  - alpha
  - alphatype
  - radius

- Quantization accuracy program
  - qerror
  - qetype

- Monitoring Program
  - visual
  - umat
  - sammon
Initialization Program
Prior to the training a network has to be set up. Each neuron in the neural network gets assigned a parametric reference vector $\mu$ in $\mathbb{R}^n$ space. This basically is a vector which 'points' to a particular distinct region in the Input data space [135]. This appointment of reference vectors can be done either randomly (i.e. the program initializes the reference vectors to random values) or linearly (in which the reference vectors are orderly organized along a two-dimensional subspace spanned by the 2 principal eigenvectors of the input data vectors). These are described by the parameters randinit and lininit.

The network dimension (i.e. the amount of neurons that will be trained) is described by the parameters $xdim$ and $ydim$. The size of the network should not be too small (which can result in a high quantization error as the data will be ‘described’ by only a few neurons) but also not too large (increases the computation time significantly). The network topology can be either hexagonal or rectangular, although the common standard is a hexagonal grid.

Finally, an important parameter is the neighbourhood function which is either a Gaussian or a Bubble function (a step function). This function decides which neighbouring neurons will get modified along with the Best matching unit (i.e. the winning neuron).

Training program
Training can either be done in 2 phases (the ordering and the fine tuning phases) if random initialization is selected for the initialisation of the network or in 1 phase if linear initialization is selected. The parameter $rlen$ describes the amount of steps during the training, i.e. the running length. If the running length is smaller than the amount of samples then it means that not all samples will be flown into the network during the training. If the amount of steps is selected too big then some input samples will be reused (randomly picked and flown into the network again). The parameter $alpha$ describes the initial learning rate and is between zero and 1. During the training this decreases linearly to zero. The parameter $Alpha$ in essence describes the degree of modification of the best matching unit and its neighbours. The radius parameter describes the initial radius of the training area in the SOM algorithm and decreases linearly to 1 during the training. It is generally advised to use a radius of size approximately $1/4$ of the network dimension.

Special Note: The learning rate function is by default linear but can also be set to an inverse-time type function (the latter being advisable for large maps and long training runs for a more balanced fine-tuning) [77].

Quantization accuracy program
$qerror$ evaluates the quantization error, i.e. how much information is lost due to the compression experienced by the data. It is the difference between the best matching units in the neural network and the actual input data. It is possible to set this to compute a weighted quantization error with the $qetype$ parameter.

Monitoring programs
The monitoring program consists of different visualization techniques. The two most pop-
ularly used visualizations of the trained network are the Sammon map and the U-matrix (commands `sammon` and `umat`). Both visualize the distances between the reference vectors of neighbouring neurons but in a different way. These representations of the trained neural network are often hard to interpret, particularly for datasets that contain more than 3 dimensions. Nevertheless, they do indicate something about the quality of the training and the presence of clusters. If they are too uniform it means that there are no clear clusters present in the dataset (or that more iterations are needed). As such they are useful tools to use to analyse the input dataset for the presence of clusters or the eventual need of reconfiguration of input parameters into the SOM package.

The most significant function in this program, is the function `visual`, because this uses the trained and labelled neural network to classify/cluster the input data (and the untrained input data also). The output is a list of coordinates to the best matching unit in the network, a value for the quantization error experienced at each sample and a label that assigns it to a particular cluster.

### D.2 SOMPAK example

The following lines indicate example commands to train and use a SOM for classification. The windows command prompt is used to run the SOMPAK package. The input for the training is a .dat file (`inputtrain.dat`) with on the rows the samples (or the pixels in this case) and on the columns the input variables (i.e RGB, NDVI, etc). The output is a trained network (`trained.cod`) to be used for the classification of other images (`inputother.dat`).

1. Initialise a 30 by 30 network with hexagonal topology and a bubble neighbourhood function:
   ```
   lininit -din inputtrain.dat -cout inputtrain.cod -xdim 30 -ydim 30 -neigh bubble -topol hexa
   ```
2. Train the network for 100 iterations, a learning rate of 0.05 and a radius of 3:
   ```
   vsom -din inputtrain.dat -cin inputtrain.cod -cout trained.cod -rlen 100 -alpha 0.05 -radius 3
   ```
3. Obtain the quantization error of the network with respect to the training data:
   ```
   qerror -din inputtrain.dat -cin trained.cod
   ```
4. Visualise labelled image:
   ```
   visual -din inputother.dat -cin trained.cod -dout labelled.dat
   ```
5. Visualise sammon map for iteration 100 and make and eps file out of it:
   ```
   sammon -cin trained.cod -cout trained.sam -rlen 100 -eps 1
   ```
6. Visualise U-matrix:
   ```
   umat -cin trained.cod -o umatrix.eps
   ```
Appendix E

Clustered Test Images SOM
Table E.1: Clustered images of the SOM approach. Displayed are the classes macroalgae (dark green), Bare soil (white-beige) and diatoms (orange). The images are from March, May, and June (top to bottom).
Table E.2: Clustered images of the SOM approach. Displayed are the classes macroalgae (dark green), Bare soil (white-beige) and diatoms (orange). The images are from July, August and September (top to bottom).
Table E.3: Clustered images of the SOM approach. Displayed are the classes macroalgae (dark green), Bare soil (white-beige) and diatoms (orange). The images are from October, November, and December (top to bottom).
Appendix F

Specifics of images used for daily variance assessment of Macroalgae

Table F.1: Dates, times and corresponding remarks of the images used to assess the daily variance of macroalgae cover. Note that not all selected images are before and after the test image date. This was because of the presence of unfit images (blurry, the mudflat is completely submerged, platform shadow visible in the image, glare in the camera, raindrops on lens.)

<table>
<thead>
<tr>
<th>Date test image</th>
<th>Day before</th>
<th>Day after</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>23-03-2011</td>
<td>22-03-2011</td>
<td>24-03-2011</td>
<td>March 22nd bright light</td>
</tr>
<tr>
<td>05-05-2011</td>
<td>03-05-2011</td>
<td>06-05-2011</td>
<td></td>
</tr>
<tr>
<td>15-06-2011</td>
<td>13-06-2011</td>
<td>17-06-2011</td>
<td>June 13th and 17th high water presence</td>
</tr>
<tr>
<td>19-07-2011</td>
<td>16-07-2011</td>
<td>21-07-2011</td>
<td>July 16th and 21st high water presence</td>
</tr>
<tr>
<td>16-08-2011</td>
<td>15-08-2011</td>
<td>17-08-2011</td>
<td></td>
</tr>
<tr>
<td>29-09-2011</td>
<td>26-09-2011</td>
<td>27-09-2011</td>
<td></td>
</tr>
<tr>
<td>27-10-2011</td>
<td>28-10-2011</td>
<td>29-10-2011</td>
<td>October 29th blurry at edges</td>
</tr>
<tr>
<td>16-11-2011</td>
<td>17-11-2011</td>
<td>18-11-2011</td>
<td></td>
</tr>
<tr>
<td>15-12-2011</td>
<td>14-12-2011</td>
<td>18-12-2011</td>
<td>December 18th high water presence and visible water ripples</td>
</tr>
</tbody>
</table>
Appendix G

Matlab codes bird detection

%% FRAME DIFFERENCING %%

clear all
filename = 'farscale.avi'; % enter the name of the video file to be used for detection
hvfr = vision.VideoFileReader(filename, 'ImageColorSpace', 'RGB', 'VideoOutputDataType', 'uint8');

% A priori parameter settings
thresh = 40;
bg = step(hvfr); % read in 1st frame as background frame
bg_bw = rgb2gray(bg); % convert background to greyscale

% ----------------------- set frame variables -----------------------

Obj = VideoReader(filename);
nFrames = Obj.NumberOfFrames;
height = Obj.Height;
width = Obj.Width;
fg = zeros(height, width); % initiate foreground matrix

% ------------------------ set bird counter ------------------------

htextins = vision.TextInserter(...
    'Text', '%4d', ...
    'Location', [1 1], ...
    'Color', [255 255 255], ... % make the number appear white
    'FontSize', 12);
% Create system objects
%-------------------------Create system objects
-------------------------

% Construct video object to write on
wrobj = VideoWriter('Output_FD_farscale.avi'); % set name of output file
wrobj.FrameRate = 2; % Desired output frame rate
open(wrobj);

% Create System objects to display the results.
sz = get(0,'ScreenSize');
pos = [20 sz(4)-470 480 400];
hVideoOrig = vision.VideoPlayer('Name', 'Original', 'Position', pos);
pos(1) = pos(1)+400; % move the next viewer to the right
hVideoFg = vision.VideoPlayer('Name', 'Foreground', 'Position', pos);
pos(1) = pos(1)+400;
hVideoRes = vision.VideoPlayer('Name', 'Detection', 'Position', pos);

% Create a blob analysis System object (to select only birds bigger than a predefined size)
hblob = vision.BlobAnalysis(
    'CentroidOutputPort', true, ...
    'AreaOutputPort', true, ...
    'BoundingBoxOutputPort', true, ...
    'ExtentOutputPort', true, ...
    'OutputDataType', 'single', ...
    'NumBlobsOutputPort', true, ...
    'Connectivity', 8, ...
    'MinimumBlobAreaSource', 'Property', ...
    'MinimumBlobArea', 3, ...
    'MaximumBlobAreaSource', 'Property', ...
    'MaximumBlobArea', 10000);

% Create System object for drawing the bounding boxes around detected birds
hshapeins = vision.ShapeInserter(
    'BorderColor', 'Custom', ...
    'CustomBorderColor', [0 255 0]);

% Create System object for indicating centroids with a plus sign
hDrawMarkerPlus = vision.MarkerInserter(
    'Shape', 'Plus', ...
    'Size', 12, ...
    'BorderColorSource', 'Property', ...
    'BorderColor', 'Custom', ...
    'CustomBorderColor', [0 255 0]);

% --------------------- process video frames
-----------------------------------
BirdNumber=zeros(1,nFrames-1); % subtract one because first frame used as background
framenum=1;

while ~isDone(hvfr)
    fr = step(hvfr); % read in frame
    fr_bw = rgb2gray(fr); % convert frame to grayscale
    fr_diff = abs(double(fr_bw) - double(bg_bw)); % convert to double

    for j=1:width
        for k=1:height
            if (fr_diff(k,j) > thresh))
                fg(k,j) = 1;
            else
                fg(k,j) = 0;
            end
        end
    end

    bg_bw = fr_bw; % set background to current frame
    medianf = medfilt2(fg,[2 2]); % median filtering (remove speckles)
    se = strel('disk',8); % morphological closing on objects of foreground image
    closed = imclose(medianf, se);

    % Estimate the area and bounding box of the blobs in the foreground image
    closed = logical(closed);
    [area, centroid, bbox, extent] = step(hblob, closed);

    BL = extent>-1; % If extent not equal -1 then is blob. Sum these and you get the count
    count = int32(sum(BL)); % number of objects detected
    bbox(:, ~BL) = int32(-1);

    % Draw bounding rectangles around the detected bird (if desired uncomment below)
    image_out = uint8(fr);
    % image_out = step(hshapeins, image_out, bbox);

    % Draw centroids of birds
Matlab codes bird detection

```matlab
image_out = step(hDrawMarkerPlus, image_out, int32(centroid));

% settings for counter
image_out(1:20,1:40,:) = 0;  % black background for counter
image_out = step(htextins, image_out, count);  % insert count

% Display videos
step(hVideoOrig, fr);  % Original video
step(hVideoFg, uint8(fr_diff));  % Foreground
step(hVideoRes, image_out);  % Result

% save birdnumber per frame in vector
BirdNumber(1,framenum)=count;
framenum=framenum+1;

writeVideo(wrobj, image_out);
end

close(wrobj);

% Close the video file
release(hvfr);
release(hblob);
release(hVideoOrig);
release(hVideoFg);
release(hVideoRes);

%% BACKGROUND SUBTRACTION with WEIGHTED AVERAGE BACKGROUND %

clear all

filename = 'farscale.avi';  % enter the name of the video file to be used for detection

% create video file reader System object
hvfr = vision.VideoFileReader(filename, 'ImageColorSpace', 'RGB',
    VideoOutputDataType', 'uint8');

% A priori parameter settings
thresh = 40;  % threshold for frame differencing rgb space
thresh_std=17;  % threshold for std_image (35 for VID2,
               % 17 for VID1 & VID3)
a=0.09;  % learning rate of adaptive filters for
         % weighted mean & std image

% get background model using weighted averaging
[mu_image, std_image] = WeightedAvg(filename, a);
Sm=mean(std_image,3);
```

Penelope Rammos  Master of Science Thesis
Obj = VideoReader(filename);
nFrames = Obj.NumberOfFrames;
height = Obj.Height;
width = Obj.Width;
fg = zeros(height, width);  % initiate foreground matrix

htextins = vision.TextInserter( ...  
  'Text', '%4d', ...  
  'Location', [1 1], ...  
  'Color', [255 255 255], ...  % make the number appear white  
  'FontSize', 14);

% ----------------------- Create system objects  
-------------------------------

wrobj = VideoWriter('Output_FD_WA_Farscale.avi');  % set name of output file
wrobj.FrameRate = 2;  % frame rate 2fr/sec
open(wrobj);

sz = get(0,'ScreenSize');
pos = [20 sz(4)-580 580 300];
hVideoOrig = vision.VideoPlayer('Name', 'Original', 'Position', pos);
pos(1) = pos(1)+800;  % move the next viewer to the right
hVideoFg = vision.VideoPlayer('Name', 'frame difference', 'Position', pos);
pos(1) = pos(1)-800;
pos(2) = pos(2)-800;
hVideoRes = vision.VideoPlayer('Name', 'Detection Result', 'Position', pos);

% Create a blob analysis System object (to select only birds bigger than a predefined size)
hblob = vision.BlobAnalysis( ...  
  'CentroidOutputPort', true, ...  
  'AreaOutputPort', true, ...  
  'BoundingBoxOutputPort', true, ...  
  'ExtentOutputPort', true, ...  
  'OutputDataType', 'single', ...  
  'NumBlobsOutputPort', true,...  
);
Matlab codes bird detection

'Connectivity', 8, ... % neighbourhood connectivity
'MinimumBlobAreaSource', 'Property', ...
'MinimumBlobArea', 10, ... % setting the minimum blob size
'MaximumBlobAreaSource', 'Property', ...
'MaximumBlobArea', 10000);

% Create System object for drawing the bounding boxes around detected birds
hshapeins = vision.ShapeInserter( ...
    'BorderColor', 'Custom', ...
    'CustomBorderColor', [0 255 0]); % green bounding box

% Create System object for indicating centroids with a green plus sign
hDrawMarkerPlus = vision.MarkerInserter( ...
    'Shape', 'Plus', ...
    'Size', 16, ...
    'BorderColorSource', 'Property', ...
    'BorderColor', 'Custom', ...
    'CustomBorderColor', [0 255 0]);

% --------------------- Process the video frames ---------------------------
BirdNumber=zeros(1,nFrames); % initiate bird count vector
framenum=1;
while ~isDone(hvfr)
    fr = step(hvfr); % read in frame
    fr=double(fr); % convert values to double
    fr_r=fr(:,:,1); % extract rgb bands
    fr_g=fr(:,:,2);
    fr_b=fr(:,:,3);
    Dr=abs(fr_r-mu_image(:,:,1)); % find the difference images for each channel
    Dg=abs(fr_g-mu_image(:,:,2));
    Db=abs(fr_b-mu_image(:,:,3));
    D=cat(3,Dr, Dg , Db); % concatenate the difference images into same variable
    Dm=max(D,[],3); % extract maximum difference observed in rgb differences
    for j=1:width ...
for k=1:height
    if (Dm(k,j) > thresh) && Sm(k,j)>thresh_std
        fg(k,j) = 1;
    else
        fg(k,j) = 0;
    end
end

% morphological closing on objects of foreground image
se = strel('disk',8);
closed = imclose(fg, se);

% Estimate the area and bounding box of the blobs in the foreground image
closed = logical(closed);
[area, centroid, bbox, extent] = step(hblob, closed);

% If the extent is not equal -1 then is blob, sum these and you get the count
% (nr of blobs per frame)
BL = extent > -1;
count = int32(sum(BL)); % number of objects detected
bbox(:, ~BL) = int32([-1]);

% Draw bounding rectangles around the detected bird (if desired uncomment below)
image_out = uint8(fr);
%image_out = step(hshapeins, image_out, bbox);

% Draw centroids of birds
image_out = step(hDrawMarkerPlus, image_out, int32(centroid));

% settings for counter
image_out(1:20, 1:40, :) = 0; % black background for counter
image_out = step(htextins, image_out, count); % insert count

% Display videos
step(hVideoOrig, uint8(fr)); % Original video
step(hVideoFg, uint8(Dm)); % frame difference
step(hVideoRes, image_out); % Result

% save birdnumber per frame in vector
BirdNumber(1, framenum) = count;
framenum = framenum + 1;
writeVideo(wrobj, image_out);
end
close(wrobj);

% Close the video file
release(hvfr);
release(hblob);
release(hVideoOrig);
release(hVideoFg);
release(hVideoRes);

%% BACKGROUND SUBTRACTION with MEDIAN FILTERED BACKGROUND %%
clear all
filename = 'farscale.avi'; % enter the name of the video file to be used for detection
hvfr = vision.VideoFileReader(filename, 'ImageColorSpace', 'RGB', 'VideoOutputDataType', 'uint8');

% A priori parameter settings
thresh = 40; % threshold for frame diff
rgb space
MED = Medianfiltering(filename); % extract median background
bg_bw = rgb2gray(uint8(MED)); % convert background to greyscale

% ----------------------- set frame variables -----------------------
Obj = VideoReader(filename);
nFrames = Obj.NumberOfFrames;
height = Obj.Height;
width = Obj.Width;
fg = zeros(height, width); % initiate foreground matrix

% ----------------------- set bird counter -----------------------
htextins = vision.TextInserter( ... 
    'Text', '%4d', ... 
    'Location', [1 1], ... 
    'Color', [255 255 255], ... 
    'FontSize', 12);

% ----------------------- create system objects -----------------------
% Construct video object to write on
wrobj = VideoWriter('Output_FD_MF_Farscale.avi'); % set name of output file
wrobj.FrameRate = 2; % desired frame rate
open(wrobj);

% Create System objects to display the results
sz = get(0, 'ScreenSize');
pos = [20 sz(4)−470 480 400];
hVideoOrig = vision.VideoPlayer('Name', 'Original', 'Position', pos);
pos(1) = pos(1)+400; % move the next viewer to the right
hVideoFg = vision.VideoPlayer('Name', 'Foreground', 'Position', pos);
pos(1) = pos(1)+400;
hVideoRes = vision.VideoPlayer('Name', 'Detection', 'Position', pos);

% Create a blob analysis System object (to select only birds bigger than a predefined size)
hblob = vision.BlobAnalysis(...
    'CentroidOutputPort', true, ...
    'AreaOutputPort', true, ...
    'BoundingBoxOutputPort', true, ...
    'ExtentOutputPort', true, ...
    'OutputDataType', 'single', ...
    'NumBlobsOutputPort', true,...
    'Connectivity', 8,...
    'MinimumBlobAreaSource', 'Property',...
    'MinimumBlobArea', 3, ...
    'MaximumBlobAreaSource', 'Property',...
    'MaximumBlobArea', 10000);

% Create System object for drawing the bounding boxes around detected birds
hshapeins = vision.ShapeInserter(...
    'BorderColor', 'Custom', ...
    'CustomBorderColor', [0 255 0]);

% Create System object for indicating centroids with a plus sign
hDrawMarkerPlus = vision.MarkerInserter(...
    'Shape', 'Plus', ...
    'Size', 12,...
    'BorderColorSource', 'Property',...
    'BorderColor', 'Custom',..., 
    'CustomBorderColor', [0 255 0]);

% --------------------- Process the video frames ---------------------
BirdNumber = zeros(1, nFrames); % initiate bird count vector
framenum = 1;
while ~isDone(hvfr)
    fr = step(hvfr); % read in frame
    fr_bw = rgb2gray(fr); % convert frame to grayscale
    fr_diff = abs(double(fr_bw)−double(bg_bw)); % convert values to double

    Master of Science Thesis  Penelope Rammou
% selecting only pixels bigger than thresh as foreground pixels
for j=1:width
    for k=1:height
        if ((fr_diff(k,j) > thresh))
            fg(k,j) = 1;
        else
            fg(k,j) = 0;
        end
    end
end

% median filtering (remove speckles)
medianf = medfilt2(fg,[2 2]);

% morphological closing on objects of foreground image
se = strel('disk',8);
closed = imclose(medianf, se);

% Estimate the area and bounding box of the blobs in the foreground image
closed = logical(closed);
[area, centroid, bbox, extent] = step(hblob, closed);

% If extent not equal -1 then is blob. Sum these and you get the count
% (nr of blobs per frame)
BL = extent > -1;
count = int32(sum(BL)); % number of objects detected
bbox(:, ~BL) = int32([-1]);

% Draw bounding rectangles around the detected bird (if desired uncomment below)
image_out = uint8(fr);

% Draw centroids of birds
image_out = step(hshapeins, image_out, bbox);

% settings for counter
image_out(1:20,1:40,:) = 0; % black background for counter
image_out = step(htextins, image_out, count); % insert count

% Display videos
step(hVideoOrig, fr); % Original video
step(hVideoFg, uint8(fr_diff)); % Foreground
step(hVideoRes, image_out); % Result

% save birdnumber per frame in vector
BirdNumber(1,framenum) = count;
framenum = framenum + 1;
writeVideo(wrobj, image_out);
end

close(wrobj);

% Close the video file
release(hvfr);
release(hblob);
release(hVideoOrig);
release(hVideoFg);
release(hVideoRes);
Appendix H

Error types per frame

A true positive (TP) represents the number of correctly identified birds in a frame. A false negative (FN) represents the number of not identified birds per frame. A false positive (FP) represents all the wrongly identified birds per frame (i.e. a reflection of a bird identified as a bird). In binary object detection (object/non object) mentioning the true negative is redundant or without meaning. The true negative represents all the correctly rejected objects (i.e. a correctly identified bird means the background was correctly rejected as a bird). Table H.1 represents the error types for each video sequence per frame for the Frame differencing method. Table H.2 represents the same data but for the Median filtered background method and Table H.3 for the weighted mean background method.
Table H.1: True positives (TP), false positives (FP), and false negatives (FN) for each frame of the 3 video sequences.

<table>
<thead>
<tr>
<th>Frame number</th>
<th>VID1</th>
<th>VID2</th>
<th>VID3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frame Differencing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Series ID</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FN</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Penelope Ramos
Master of Science Thesis
Table H.2: True positives (TP), false positives (FP) and false negatives (FN) for each frame of the 3 video sequences

<table>
<thead>
<tr>
<th>Series ID</th>
<th>Occurrence type</th>
<th>Frame number</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>VID1</td>
<td>TP</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>FN</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>VID2</td>
<td>TP</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>FN</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VID3</td>
<td>TP</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>FN</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table H.3: True positives (TP), false positives (FP) and false negatives (FN) for each frame of the 3 video sequences

<table>
<thead>
<tr>
<th>Series ID</th>
<th>Occurrence</th>
<th>Frame number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>VID1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VID2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VID3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Weighed mean background method

Penelope Rammos Master of Science Thesis


Master of Science Thesis

Penelope Rammos


[103] P. Duhamel and M. Vetterli, “Fast fourier transforms: a tutorial review and a state of

Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pp. 505–508,
IEEE, 2008.


time-varying signal analysis,” *Advances in Electrical and Computer Engineering*, vol. 11.

Wesley*, vol. 16, no. 716, p. 8, 1996.

[108] J. Vesanto and E. Alhoniemi, “Clustering of the self-organizing map,” *Neural Networks, 


[110] M. Jaffar, M. Ishtiaq, A. Hussain, and A. Mirza, “Wavelet-based color image segmen-
tation using self-organizing map neural network,” in *Proceedings of International Con-

“Hyperspectral image classification using a self-organizing map,” in *Summaries of the

[112] P. Chang and W. Teng, “Exploiting the self-organizing map for medical image seg-
mentation,” in *Computer-Based Medical Systems, 2007. CBMS’07. Twentieth IEEE

[113] W. Teng and P. Chang, “Identifying regions of interest in medical images using self-
organizing maps,” *Journal of Medical Systems*, pp. 1–8, 2011.

selecting the best landing site for the exomars mission,” 2009.


[http://www.yale.edu/gsp/gis-files/remote_sensing_intro.pdf](http://www.yale.edu/gsp/gis-files/remote_sensing_intro.pdf), Accessed 22-04-
2012.


Penelope Rammos

Master of Science Thesis


