Identify Kelp in the BC Coast Using Sentinel-2 Satellite Imagery

by

KELP

Department of Computer Science

A Project Report Submitted to the Computer Science of University of Victoria

Supervisor:
Yvonne Coady

Department of Computer Science

May, 2021

Declaration

We Group KELP do hereby declare that this Project Report is original and has not been published and/or submitted for any other degree award to any other University before.

#	Names	Student Number	Signature
1	Xi Sun	V0025479	
2	Ting Liu	V00045129	
3	Zheng Yin	V00915261	

Date: May 20, 2021

Contents

Declaration **i**

1 Literature Review **1**
1.1 Background 1
1.2 Main Objective 1

2 Methodology **3**
2.1 Sentinal-2 Satellite Data 3
2.2 Atmospheric Correction and NDVI 3
2.3 Distributed Computation 5
2.3.1 Comparison of Multiple Cloud Platform 5
2.3.2 Platforms Selection 7
2.4 Visualization 7

Chapter 1

Literature Review

1.1 Background

Coastal infrastructures based on seaweed and kelp provide a very important ocean ecosystem to countries around the world, including providing a living environment for marine life, protecting the coast from erosion, providing carbon storage, supporting biodiversity and fisheries. Usually the health of seaweed is an indicator of the overall assessment of the health of offshore ecosystems.

On the Pacific coast of Canada and around the world, seaweed including kelp forest provides natural habitat for a large number of creatures. However, the ecology of seaweed and kelp is now being destroyed, threatened by climate change, industrial development and eutrophication. Therefore, the monitoring of the distribution of marine macrophytes is very important to understand and quantify the changes in the placement of seaweeds and kelp.

1.2 Main Objective

In British Columbia, ocean-oriented activities provide 8% of B.C. jobs and 7% of the province's GDP, while at the same time aquaculture from British Columbia presents half of the aquaculture production in Canada [9]. Sustainable seaweed production is the next big thing in aquaculture, so seaweed production should be important for British Columbia.

Secondly, monitoring the kelp also helps to monitor the ocean environment, for seaweed and kelp are sensitive to ocean current, temperature, and creatures. Observing the kelp changing can make a solution of some elements that kelp demands are changed, for example, informal sea temperature change, and biological invasion. Kelp forest disappearing might lead to a biological disaster which should be alerted. For the Ministry of Environment and Climate Change Strategy in British Columbia, information about ocean environment change should affect the strategy of climate change, so new acts should be applied [7]. Our project should combine the information about the seaweed and kelp in the ocean, like a timeline of kelp distributed, which data should be easier to read and understand.

As we know human activity will deeply change the underwater environment, which we believe seaweed and kelp are key elements to evaluate the effectiveness of human activity in the environment. For example, over fishing and pollution will quickly destroy the kelp forest than the natural change, if we see there is the large size of kelp disappear on 1-2 year satellite image, the government should be alert that there is a serious problem with the ocean environment.

In conclusion, our view should help British Columbia in protecting the ocean environment, formulating an aquaculture plan, and reacting to long-term climate change which will find a balance between humans and nature in climate change.

Chapter 2

Methodology

2.1 Sentinal-2 Satellite Data

Satellite remote sensing technology has been used to predict and detect the global distribution of marine macrophytes in optical shallow waters. The Sentinel-2 satellite imagery is going to be used in our project. The Sentinel-2 satellite is equipped with 12 multi spectral imaging (MSI) sensors. It is able to collect a variety of data types, including optical signals, high-resolution images, every 2-5 days from coastal areas with 10 meters as the basic pixel unit. SENTINEL-2 data are acquired on 13 spectral bands in the visible and near-infrared (VNIR) and Short-wavelength infrared (SWIR) spectrum[8] (Figure 2.1).

However, in our project, the distribution of seaweed and kelp does not occupy a large area like forests and lakes. So we will mainly study on high-resolution data. At a spatial resolution of 10m, we can get four bands: B02, B03, B04, B08 from sentinel-2 data, the center wavelengths of which are 490 nm (Blue), 560 nm (Green), and 665 nm (Red) and 842nm (Near infrared). These are the only wavebands that utilize high spatial resolution in our research.

2.2 Atmospheric Correction and NDVI

We know that the absorption and scattering of light by atmospheric components such as atmospheric molecules, aerosols and water vapor affect the signal received by the satellite sensor. Therefore, before processing the raw data, we must perform atmospheric correction. Previous work with Sentinel-2 in papers[14][13][15] explored the use of ACOLITE to atmospherically correct Sentinel-2 data for button habitat identification. The full tile image was

Band name	Resolution (m)	Central wavelength (nm)	Band width (nm)	Purpose
B01	60	443	20	Aerosol detection
B02	10	490	65	Blue
B03	10	560	35	Green
B04	10	665	30	Red
B05	20	705	15	Vegetation classification
B06	20	740	15	Vegetation classification
B07	20	783	20	Vegetation classification
B08	10	842	115	Near infrared
B08A	20	865	20	Vegetation classification
B09	60	945	20	Water vapour
B10	60	1375	30	Cirrus
B11	20	1610	90	Snow / ice / cloud discrimination
B12	20	2190	180	Snow / ice / cloud discrimination

Figure 2.1: 13 spectral bands

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + \mathrm{Red})}$$

Figure 2.2: Calculation of NDVI

corrected using the dark spectrum fitting approach. The assessment of this approach is beyond the scope of our discussion in this paper. Anyway, this correction approach will be used in our project. From the corrected data, we will use the surface reflectance of band 4 and band 8 to calculate NDVI. NDVI (the Normalized Difference Vegetation Index) is one of the most widely used indexes to measure vegetation (Figure 2.2). It was developed by NASA scientist Compton Tucker [12]. Its values range from +1.0 to -1.0.

All pixel values equal to or greater than 0.4 are considered as vegetation, seaweed and kelp in our content [11]. On the contrary, All pixels with NDVI <0.4 were used to classify some non-vegetated habitats. In further analysis, we can use other algorithms such as machine learning to classify the selected areas to determine whether they belong to beaches, mud or other mixed areas. Nevertheless, this is not within the scope of our project.

2.3 Distributed Computation

Due to the huge amount of satellite data, when calculating NDVI for all pixels, it is impossible for us to carry on the computation on only one computer, which will be very time-consuming. If you need to process all the satellite imagery that started from 2015, it may not be completed within a few months. Therefore, we need to adopt a distributed computation framework.

Sentinel-2 imagery can be divided into multiple small blocks (split by pixel), and then these small blocks can be mapped to the distributed node for computation. After determining the NDVI of all pixels in the selected area , we are able to filter out the pixels that meet the condition: NDVI is less than the threshold 0.4, and apply the reduce method to get all pixels we need. Using the MapReduce framework like Hadoop on Google Colab, Amazon EMR, we can process tons of data at the same time. It will be a great practice on a distributed system that is also the main motivation and purpose for this project. To determine suitable cloud platforms for our project become an urgent topic.

2.3.1 Comparison of Multiple Cloud Platform

Arbutus Cloud

Arbutus Cloud supports the virtualization of computer hardware platforms. Users will generally build one or more virtual machines (VM), then log into the VM, install and run the software applications needed. These applications may be various, from "CPU-intensive analysis of particle physics data" to "a web service directed towards scholars of literature and the humanities".

The advantages of Arbutus are the user can fully handle the collection of installed software (the "software stack") and the VM can be easily duplicated. If the user wants to launch the VM again somewhere, what they need is to take a "snapshot" of it. "This makes it easy to replicate or scale up a service, and to recover from (for example) a power interruption."

The disadvantage is that users must have relevant experience in software installation and computer management[3]. Moreover, Arbutus's support for midware and software stacks is not very excellent, so the user often needs to build them one by one, which results in a lot of work and difficulty.

Google Cloud: Colab

Google Colaboratory (Colab) aims to provide a Jupyter-based cloud online tool to help realize the education and research of Machine Learning. Colab mainly has the following three advantages. The first one is “No configuration required”. Colab is a Jupyter notebook environment that can be used without any settings, and it runs completely in the cloud. The second advantage is “Free use of GPU”. Users can conveniently run Keras, TensorFlow, PyTorch, and other frameworks for deep learning and application development. The last one is “Easy sharing”. In order to facilitate the deployment of the project, Colab notebooks can be stored in Google Drive and shared, just like using Google Docs or Sheets[6].

Google Cloud: TPUs

Tensor Processing Units (TPUs) are “custom-developed application-specific integrated circuits (ASICs)” used to stimulate machine learning tasks. Machine learning applications require extensive use of linear algebra computation, and TPUs can speed up this computing performance. Especially when training large and complex neural network models, TPUs greatly shorten the accuracy time, which can be completed in just a few hours, while these models often took weeks to train on other hardware platforms[4].

Azure: Cognitive Services Tasks

Azure Cognitive Services are cloud-based service with REST API and client library SDK. It can help users add cognitive intelligence into applications to build various cognitive solutions, which include viewing, hearing, speaking, understanding and even making decisions[10]. Via Docker containers provided by Cognitive Services, users can use the same API provided locally by Azure and flexibly move Cognitive Services closer to the data for conformity, preservation or other working reasons[2].

Azure: Data Lakes

Azure Data Lakes solutions, which include both storage and processing, contain all the functions needed to enable developers, data scientists and analysts to easily “store data of any size and shape and at any speed, and for all types of processing and analysis across platforms and languages.” Because the data is stored in its original format, the data will never be discarded. Moreover, users can search the data and build their own queries. Data Lakes may be faster than traditional ETL tools, because it can store unstructured and semi-structured data. It is more flexible than a data warehouse[5].

AWS: Amazon EMR

Amazon EMR is a cloud platform for processing vast amounts of data using open source tools. By automating time-consuming tasks, such as provisioning capacity and tuning clusters, big data environments can be easily built, operated, and escalated with EMR[1].

EMR has high fault tolerance and is easy to use, customize, and operate. It works well for managing analyses that use multiple tools, such as Hadoop and Spark. It is flexible and can scale cluster capacity according to job specifications.

2.3.2 Platforms Selection

According to the above overview of each platform, we believe Arbutus and Colab are the most suitable platforms for our project. First of all, our project does not need machine learning and cognitive intelligence. Second, Data Lakes and EMR are a professional service model, which comes with a new set of tools and services that require users to be trained before using. However, Colab is a Jupyter notebook environment and can be used without any settings. It runs completely in the cloud and can be easily shared. We can operate our project on Colab from the beginning to the end. These advantages of Colab make up for the deficiencies of Arbutus and fit our project well.

2.4 Visualization

Essentially, we will rasterize these pixels to a new image to visualize the distribution of seaweed. Further, we will use some kind of map-based interface to render the image, add the layer to the global map [1] or build a webserver [2]to show the distribution of seaweed and kelp over time.

[1]https://www.hectaresbc.org/app/habc/HaBC.html
[2]www.algaeexplorer.ca

References

[1] Amazon emr.

[2] Azure cognitive services containers.

[3] Cloud - CC Doc.

[4] Cloud tensor processing units (tpus).

[5] Data lakes.

[6] Google colaboratory.

[7] Ministry of environment and climate change strategy.

[8] Sentinel2 – sentinel-2 products.

[9] Take a dive into the 5 reasons we exist.

[10] What are azure cognitive services?

[11] Laurent Barillé, Marc Robin, Nicolas Harin, Annaëlle Bargain, and Patrick Launeau. Increase in seagrass distribution at bourgneuf bay (france) detected by spatial remote sensing. *Aquatic Botany*, 92(3):185–194, 2010.

[12] Parul Pandey. Vegetation index calculation from satellite imagery.

[13] Quinten Vanhellemont. Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the landsat and sentinel-2 archives. *Remote Sensing of Environment*, 225:175–192, 2019.

[14] Quinten Vanhellemont and Kevin Ruddick. Acolite for sentinel-2: Aquatic applications of msi imagery. In *Proceedings of the 2016 ESA Living Planet Symposium, Prague, Czech Republic*, pages 9–13, 2016.

[15] KL Wilson and E Devred. Exploratory analysis to retrieve biogenic habitat and bottom substrate type from high-resolution satellite imagery in atlantic canada. *Can. Tech. Rep. Fish. Aquat. Sci*, 3337(56), 2019.