A REPORT

ON

Earth Observation through Satellite Imaging Data

BY

Name of the Students	*ID. No.*
Anushka Bhattacharjee	2020A7PS0003P
Urvashi Sharma	2020A7PS0017P
Sarthak Shah	2020A7PS0092P

AT

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

(November 2022)

TABLE OF CONTENTS

Acknowledgment

We are grateful to Dr. Navneet Goyal of the Foundations of Data Science Faculty, for giving us the opportunity to work on this Project and for the guidance given for this study. Your useful advice and suggestions were really helpful to us during the project's completion.

Abstract

This is the report for the project done for the course ‘Foundations of Data Science under the guidance of Dr. Navneet Goyal. The assignment's goal is to understand the Satellite imaging physics and how key satellite systems store, download, and pre-process satellite picture data using an Application.

Introduction To Satellite Imaging

When we are in an airplane we have the firm tendency of sitting by the window side not only to blankly stare at the blue sky but because its wonderful to look at the miniature streets,the skyscrapers,the cars,the reflection of sun in a body of water.It was hard to even imagine what the Earth looked like from above for the majority of human history. It wasn't until the crew of Apollo 17 captured what has become one of the most famous photos of the Earth: the Blue Marble. This image has become one of the most iconic images of the Earth. This has been made possible because of something called "Satellite Images."

Need Of Satellite Images-There are currently approximately 2,000 satellites in orbit around the Earth, each of which possesses a unique set of capabilities. Imaging obtained from satellites is utilised in a variety of disciplines today, including meteorology, conservation, geology, agriculture, cartography, intelligence, and warfare, amongst others.The free availability of data from satellite constellations like MODIS, Landsat, and Sentinel has democratized access to timely satellite images of the entire world. This is one of the reasons why there is a lot of interest in Earth Observation through Satellites at the moment.Cloud service providers like Amazon Web Services and Google Cloud are storing satellite data for free, which is further boosting the use of satellite.

But how does Satellite Images help in various domain like vegetation ,warfare?

To answer this question one must know the “Physics”behind satellite Imaging

Satellite utilizes remote sensing. "Remote sensing" is the method of assessing the physical qualities of distant objects by utilizing reflected or emitted energy. Through the use of electromagnetic radiation as a medium of interaction, it is possible to determine the surface of the Earth, its features, and an evaluation of their geophysical qualities.

Electromagnetic Radiation as medium of interaction

The entire range of light that can be seen or sensed, from radio waves to gamma rays, is referred to collectively as the electromagnetic spectrum. The varied wavelengths of light that are absorbed or reflected by the various things. When it comes to Earth Observation, certain parts of the spectrum have varying degrees of importance, both in terms of the kinds of information that we are able to collect and the amount of geospatial data that can be obtained. Sensing in the visible and infrared ranges is responsible for the vast majority of GDA's output. The ultraviolet range encompasses the wavelengths that are most relevant to Earth Observation because of their utility in the field. Radiation from ultraviolet light has the potential to expose certain characteristics of minerals and the atmosphere. Microwaves are located at the other end of the useful range for Earth Observation; they are able to, among other things, provide information on the surface roughness and the moisture content of soils.

The "visible component" of the electromagnetic spectrum, which is composed of wavelengths that produce color (blue, green, and red), represents only a relatively small portion of the total EM wavelength range. Only a very small amount of the overall electromagnetic wavelength range is considered to be the "visible section" of the spectrum. These wavelengths are the ones that produce color. Infrared refers to the portion of the electromagnetic spectrum that lies beyond red light and possesses longer wavelengths (IR). For instance, analyzing near-infrared (and mid-infrared) radiation allows us to differentiate between different types of vegetation as well as the level of stress that plants are experiencing. This method is significantly more effective than relying on color alone. Because it causes a feeling similar to that of "heat," infrared radiation with a wavelength that is greater than 3 meters. The sensation of something being hot is not produced by near- or mid-infrared radiation. The peak wavelength of the thermal emissions coming from the surface of the Earth, which are 288 degrees Kelvin, is 10 meters.

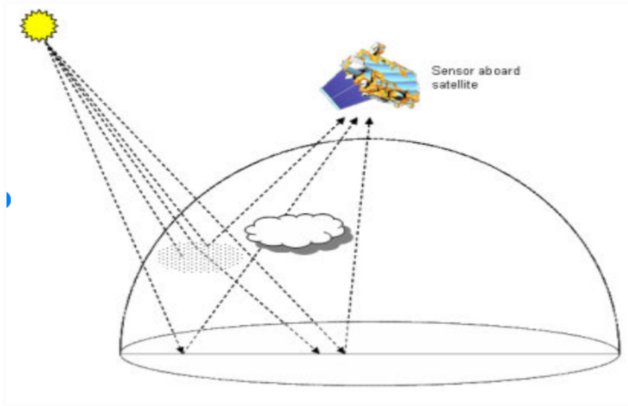


Fig-1 Diagram shows a sensor abroad satellite capturing photographs of the Sun's radiation on Earth and how different wavelengths are considered.

Sensing methods used in Satellite Imaging:

There are different kinds of sensors and different methods for acquiring satellite images. Here are some of the ways that satellites and their sensors can vary.

Active Sensing--An active sensor is a sensing device which requires an external source of power to operate.For example,when we take a picture with its flash turned on the camera sends its own source of light.After it illuminates the target the camera captures the reflected light back to the lens.

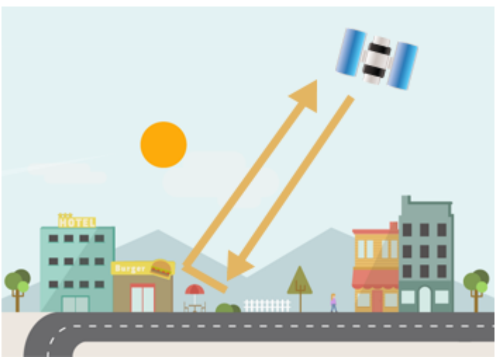


Fig:Active Sensing

Passive Sensing-Passive sensing uses the natural emitted light by the sun.

For example,Cameras become passive sensors when photographer does not use flash.Here, the camera does not send any source of light.It uses the natural light emitted by the sun.

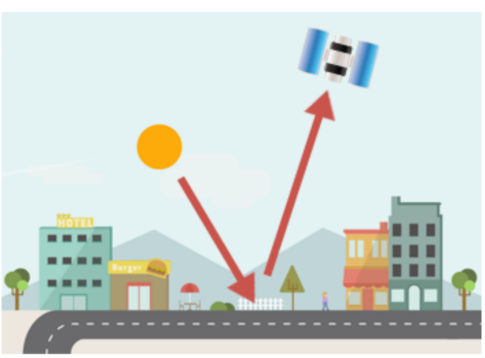


Fig:Passive Sensing

Need Of Satellite Resolution

Resolution is the key feature satellite providers emphasise, although it's not the only one. Accuracy determines image quality.Resolution is the smallest size an image can represent. Higher resolution equals smaller pixels, which adds detail.Resolution plays a role in how data from a sensor can be used. Resolution can vary depending on the satellite’s orbit and sensor design.

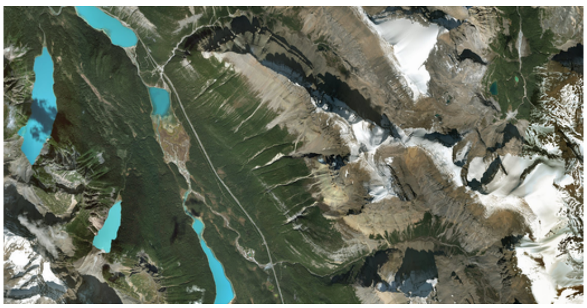


Fig-High resolution Image

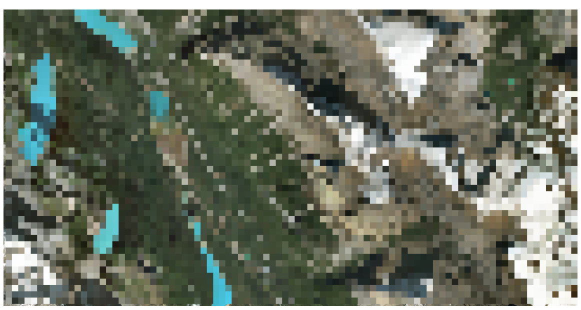


Fig-Low resolution image

Sensor resolution types in remote sensing

There are four types of resolution to consider for any dataset—radiometric, spatial, spectral, and temporal.

1)Spectral Resolution-The capacity of a sensor to distinguish finer wavelengths, or having more and smaller bands, is known as spectral resolution. Numerous sensors are categorised as multispectral, which means they have 3–10 bands. Some sensors, which are referred to as hyperspectral, have hundreds or thousands of bands. The spectral resolution is finer when the wavelength range for a given band is smaller. The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), for instance, collects data over 224 spectral channels. The right cube displays the data's level of detail. Differentiating between rock and mineral types, flora types, and other traits is possible at this level of specificity.

2)Spatial resolution-The size of each pixel in a digital image and the portion of the Earth's surface every pixel represents together constitute spatial resolution. For instance, the Moderate Resolution Imaging Spectroradiometer (MODIS) observes most bands with a spatial resolution of 1 km; each pixel corresponds to a 1 km x 1 km region on the ground. Bands from MODIS also have a 250 m or 500 m spatial resolution. One can see more information when the resolution is finer (lower number).

3)Radiometric Resolution-The number of bits that represent the energy captured in each pixel, or the radiometric resolution, represents the quantity of information contained inside. Each bit stores a power-2 exponent. For instance, an 8 bit resolution is 28, which means that the sensor may store data in 256 different digital values (0–255). Because there are more values available to store information, there is better ability to distinguish between even minute variations in energy with increased radiometric resolution.

4)Temporal resolution- The amount of time it takes a satellite to complete an orbit and return to the same observation region is known as the temporal resolution. This resolution is influenced by the orbit, the features of the sensor, and the swath size. The temporal resolution is substantially higher for geostationary satellites since they rotate at the same speed as the planet. The temporal resolution of polar orbiting satellites can range from one day to sixteen days. As an illustration, the MODIS sensor on NASA's Terra and Aqua satellites has a temporal resolution of 1-2 days, enabling the sensor to see how the Earth changes throughout the day.

Need Of satellite Imagery

The need for satellite imagery arises from the need for accurate and comprehensive information about the Earth's surface and atmosphere. It is a valuable resource for a wide range of applications and industries.

Types of Satellite Imagery

There are three major types of satellite imagery

1)**Visible Imagery**-Visible satellite pictures can only be viewed during the day, since clouds reflect the light from the sun.

Application-Visible imagery is highly useful for observing the development of thunderstorms. As the developing thunderstorms will be visible by satellite even before radar picks them up.

2)**Infrared Imagery-**Clouds can be seen on infrared satellite images throughout both day and night. Clouds are recognised by satellite sensors that measure heat radiating off of them, rather than by sunlight reflecting off of them. The sensors also track heat emitted from the earth's surface.

3)**Water Vapour Imagery-**Satellite images of water vapour show how much moisture is present in the upper atmosphere. The highest humidities reveals the whitest area whereas the dry portions are the darkest. This kind of photography can be quite helpful in identifying potential areas for severe rainfall.

Data Processing, Interpretation, and Analysis

Remote sensing data acquired from instruments aboard satellites require processing before the data are usable by most researchers and applied science users. Most raw NASA Earth observation satellite data (Level 0, see data processing levels) are processed at NASA's Science Investigator-led Processing Systems (SIPS) facilities. All data are processed to at least a Level 1, but most have associated Level 2 (derived geophysical variables) and Level 3 (variables mapped on uniform space-time grid scales) products.

Our concern: Since we know that satellite images mainly aims to track and identify various activity.So the concern is to produce quality images.The images are to be processed in such a way that we can get the best information out of it.

Hindrance of clouds in the overall quality of satellite images-

Clouds block out the view of the Earth's surface, they can cause information to be lost in satellite photographs. Clouds can obscure the view of the underlying landscape, vegetation, and other objects in a satellite image. Extracting important information from the image, such as land cover, land use, or geological features, may become challenging or impossible as a result.

Clouds can also result in low contrast and poor image quality in satellite photographs, among other issues. Due of this, it may be more difficult to identify and categorize specific aspects in the image, which may result in less accurate information being produced.We use cloud masking techniques to detect and remove clouds from the photos in order to lessen the impact of clouds on the photographs. These techniques are not always error-free, thus the final photos might still contain some clouds.

How do we eliminate the influence of clouds?

Cloud Masking Algorithm:-

Clouds are recognised and eliminated from satellite photos using cloud masking algorithms. To separate clouds from other elements in the image, these algorithms often combine spectral, spatial, and temporal data.

For instance, the algorithm might utilise each pixel's brightness and reflectance to identify clouds since they are frequently brighter and more reflecting than the terrain beneath them.

Additionally, compared to other features, clouds often have a more consistent texture and

structure, therefore the algorithm may use texture and spatial data to identify clouds from other features.
When the clouds have been located, the algorithm can remove them from the image by masking them off, either by setting the pixels to a no-data value or by filling them with a value that depicts the ground beneath. By reducing the obscuration brought on by the clouds, this can enhance the image's quality and utility.

The choice of technique may be influenced by the type of satellite, the image resolution, and the particular application of the produced data. There are numerous alternative algorithms and approaches for cloud masking in satellite images. The simple threshold, the normalised difference cloud index (NDCI), and the cloud probability approach are a few popular cloud masking algorithms.

Our Implementation Towards Solving The Concern

Now for Identifying The clouds from the image we have implemented a Artificial Neural Network model for classification. Here it is just a simple model to classify a image as clouds or water. However, it can be expanded.

So for the model firstly we needed a dataset:-

We found the dataset from this website:-

https://git.gfz-potsdam.de/EnMAP/sentinel2_manual_classification_clouds

About Dataset

Images were collected from the satellite Sentinel 2.

The European Space Agency (ESA) created the Sentinel-2 satellite project to deliver high-resolution optical pictures of the Earth's surface. Two satellites, Sentinel-2A and Sentinel-2B, were launched for the mission in 2015 and 2017, respectively.

The multispectral imaging system on board Sentinel-2 satellites is capable of taking pictures in 13 spectral bands, spanning the visible to the short-wave infrared spectrum. Depending on the spectral band, the instrument has a spatial resolution of 10–60 metres and can capture images that cover an area of 290 km x 290 km.

How to access the data?

Users can access Sentinel-2 data for free, and it is utilised for many different purposes, including mapping, monitoring land cover, and environmental monitoring. The information is also utilised to assist the Copernicus programme, which offers users all around the world services for global environmental monitoring.

Sentinel-2 data is available for download on the Scientific Data Hub. So from here they downloaded images covering various parts of the world.

Techniques of storing the Satellite Data

Depending on the particular requirements of the satellite system and the kind of data being gathered, a variety of approaches can be employed to store satellite images. Among the usual techniques for archiving satellite photos are:

Data compression: Because satellite photos are frequently very large, they are typically compressed to save storage space. Algorithms for data compression can be applied in a variety of ways, including lossless and lossy techniques. While lossy algorithms surrender part of the visual detail to minimise the file size, lossless algorithms maintain all of the original data.

After the data has been compressed, it is often put into a database or another sort of data structure to make it simple to access and analyse. The data may need to be indexed in a variety of ways, such as by location, date, or other pertinent factors.

Data storage on satellites: Due to their often low storage capacities, satellites are frequently made to only save the most recent or significant data. Normally, this information is sent to Earth in real time or on a regular basis, where it can be kept on ground-based systems for longer-term archiving.

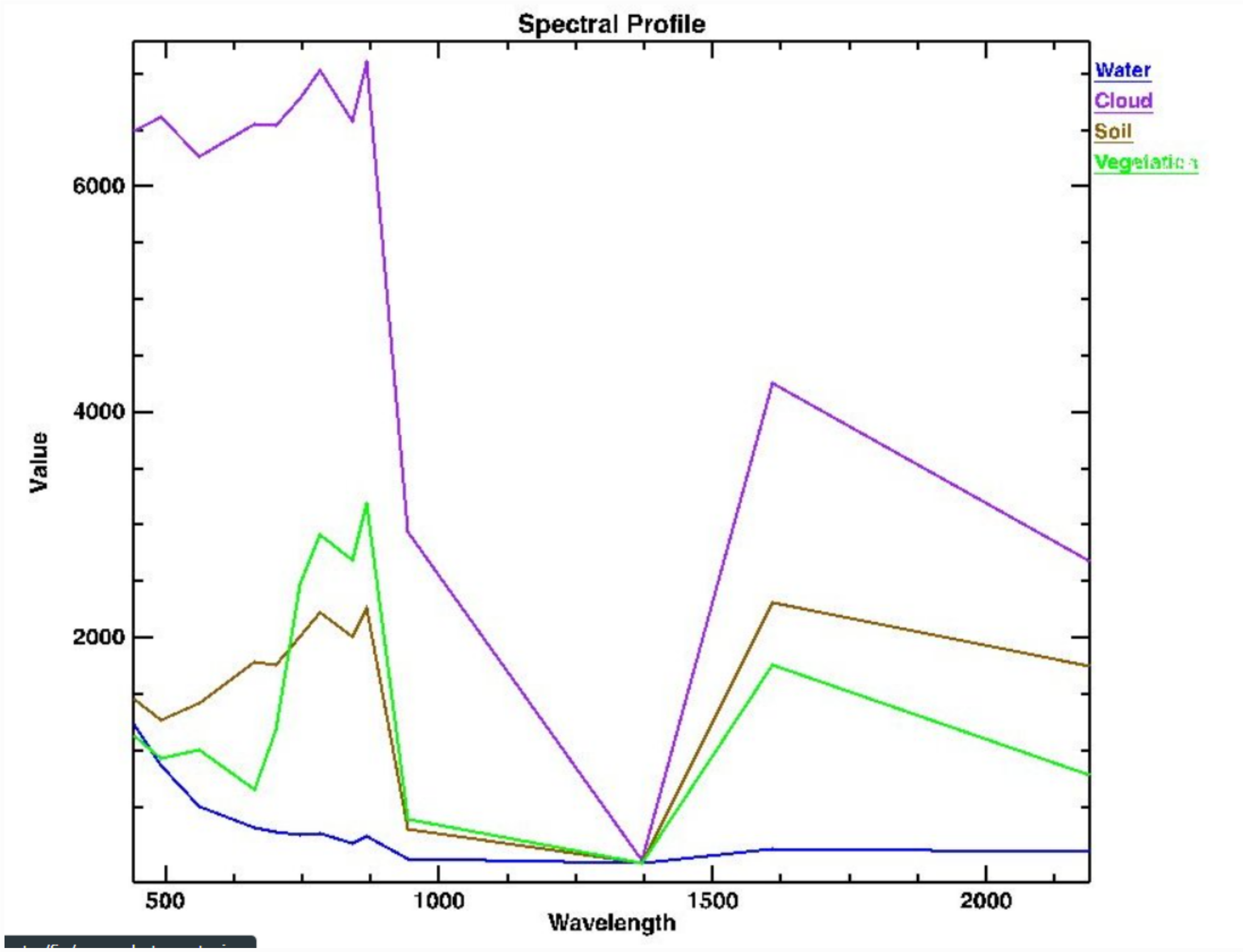
The information is often broadcast back to Earth in real-time or on a regular basis and kept on ground-based servers for satellites that do not have enough storage space. These servers could be situated at the satellite's control facility or at a distant site that is networked to the satellite. Scientists and other users can then access the data and do any necessary analyses.

Data Preprocessing & Our application

By means of different spectral tools, granule pixels are selected and classified into one of the following six classes:

Class	Coverage
cloud	opaque clouds
cirrus	cirrus and vapor trails
snow	snow and ice
shadow	shadows from clouds, cirrus, mountains, buildings, etc
water	lakes, rivers, seas
clear-sky	remaining: crops, mountains, urban, etc

And following graph shows four different spectral profiles from a Sentinel-2 image.



So here is our implementation using this dataset.

1)Firstly as the dataset was huge so we uploaded it on drive for better performance on google colab.

```
from google.colab import drive
drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

2)Loading the dataset:-

```
# import required libraries
import h5py as h5
import numpy as np
import pandas as pd
#this file I will attach in the zip and while running you can change the path accordingly as here I have used it by uploading it to the drive.
f = h5.File("/content/gdrive/MyDrive/3-1/3-1/FoDS/20170710_s2_manual_classification_data.h5", "r")
print(f)
# Get and print list of datasets within the H5 file
datasetNames = [n for n in f.keys()]
for n in datasetNames:
        print(n)
# extracting spectra data from the H5 file
#class_ids
class_ids=f['class_ids']
class_idsData=class_ids[:]
print(class_idsData)
#class_names
class_names=f['class_names']
class_namesData=class_names[:]
print(class_namesData)
#spectra
spectra=f['spectra']
spectraData=spectra[:]
print(spectraData)
#classes
classes=f['classes']
classesData=classes[:]
print(classesData.shape)
```

3)Extracting the spectral data of the images which were classified amongst clouds or water:-

```
# here I am coverting the data to a csv with 14 columns where first 13 being the spectra data and last column being the label (cloud[1] or water[0] )
# firstly making the corresponding list from first 1 million samples of spectra data.
temp_data=np.array(spectraData).tolist()[0:1000000]
final_data=[]
for i in range(1000000):
  x=classesData[i]
  if x==50:
    x=1
    temp_data[i].append(x)
    final_data.append(temp_data[i])
  if x==20:
    x=0
    temp_data[i].append(x)
    final_data.append(temp_data[i])
```

```
df=pd.DataFrame(final_data)
df.to_csv('data.csv', index=False)
```

4) Splitting the data into Training and Testing Dataset:-

```
dataset = pd.read_csv('data.csv')
#Splitting the data into x and y , x being spectra data, y being labels
X_train = dataset.iloc[:, 0:13].values
Y_train = dataset.iloc[:, 13].values
# Feature Scaling
# Applying Z-score normalization
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
# Splitting the dataset into the Training set and Test set
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X_train, Y_train, test_size = 0.2, random_state = 42)
```

5) Implementing the model:-

Model Features:-

3 Layers(Input , output , 1 Hidden Layer)

Input size :- 13 neurons

Hidden layer- 20 neurons, Activation Function :- Relu

Output - 1 neuron, Activation Function :- Sigmoid

Adam Optimiser was used and cross-entropy loss function

```
# Importing the Keras libraries and packages
import keras
from keras.models import Sequential
from keras.layers import Dense

#from keras.layers import Dropout
from keras import regularizers

# Initialising the ANN
classifier = Sequential()

# Adding the input layer and the first hidden layer
#  adding l1 regulisar to prevent overfitting
classifier.add(Dense(units = 20, kernel_initializer = 'uniform', activation = 'relu', input_dim = 13, kernel_regularizer=regularizers.l1(0.001)))

# Adding the second hidden layer
# Adding the output layer
classifier.add(Dense(units = 1, kernel_initializer = 'uniform', activation = 'sigmoid'))

# Compiling the ANN
classifier.compile(optimizer = 'adam', loss = 'binary_crossentropy', metrics = ['accuracy'])
start_time = time.time()

# Fitting the ANN to the Training set
classifier.fit(X_train, Y_train, batch_size =1024, epochs = 100, shuffle= True )

print("--- %s seconds ---" % (time.time() - start_time))
```

6) Finally tested the model on test data and here are the accuracy results:-

```
#Making predictions

# Predicting the Test set results
y_pred = classifier.predict(X_test)
y_pred=np.rint(y_pred)

#Evaluating the model
from sklearn.metrics import accuracy_score
accuracy_score(Y_test,y_pred)
```

```
1320/1320 [==============================] - 2s 1ms/step
0.9999289705464532
```

Conclusion

So with the help of this classification of clouds and other elements(here water) cloud masking can be appropriately done through setting their spectral data to null or reproducing the terrain data instead of cloud data. Overall it's quite important to prevent loss of data through masking the clouds.