***HyperGix***

Building a User-Friendly, Open-Source Hyperspectral Imagery Analysis Desktop Application with Python and PyQt5

Master's Project

Department of Computer Science

Montclair State University

John Gallino

Fall 2021

Advisor: Dr. Stefan Robila

## Abstract

### Contents

Abstract……………………………………………………………………………………………………………………………………. 2

List of Figures………………………………………………………………………………………………………………………..…. 3

1. Introduction…………………………………………………………………………………………………………………….. 4
   1. Complications of Hyperspectral Imaging..………………………………………………………………… 6
   2. Popular Hyperspectral Software Packages…………………………………………………………………9
   3. Current State of Hyperspectral Satellite Imaging………………………………………………………
2. HyperGix Overview
3. System Requirements………………………………………………………………………………………………………
4. System Design and Structure……………………………………………………………………………………………
   1. Module System………………………………………………………………………………………………………..
   2. Code Structure…………………………………………………………………………………………………………
   3. Downloading and Handling Hyperion Data………………………………………………………………
   4. Handling Other HSI Files……………………………………………………………………………………………
   5. Creating a Custom Spectral Library…………………………………………………………………………..
   6. Pixel Classification…………………………………………………………………………………………………….
5. Conclusions and Future Directions…………………………………………………………………………………..

References…………………………………………………………………………………………………………………………………

Appendix I - Installation of HyperGix …………………………………………………………………………………………

Appendix II - Source Code and Modularity………………………………………………………………………………..

## List of Figures

Figure 1. Advanced Land Imager and Hyperion band comparison

Figure 2. Example spectral signatures of three minerals

Figure 3. GERBIL software interface

## 1. Introduction

Spectral imaging is the art and science of capturing electromagnetic (EM) information across the EM spectrum. This spectrum includes visible light, but also infrared, ultraviolet, microwaves, x-rays, gamma rays and radio waves. Wave type is determined simply by its frequency and wavelength. Radio waves are the largest and have the lowest frequency while gamma rays are at the opposite end of the spectrum with wavelengths as small as atoms. [] All physical objects reflect these waves in different amounts. Plants appear green not because they absorb green light, but because they reject it and reflect it away back towards our eyes. However, human eyes can only see within the visible light spectrum and cannot measure the light reflected in other frequencies without special tools. Spectral imaging was developed to capture and display that type of information, using specialized film or digital sensors sensitive to specific frequencies of interest.

Hyperspectral imaging (HSI) is a more recent advancement within this field. Rather than capturing light in a few choice frequency ranges, hyperspectral sensors aim to capture reflectance values across a wide and continuous expanse of the EM spectrum – usually beginning with ultraviolet and visible light and expanding to the near infrared and shortwave infrared frequency ranges.

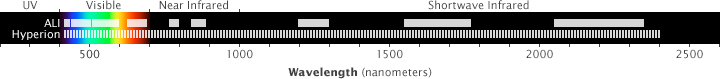


Figure 1 Comparison of the continuous Hyperion sensor hyperspectral frequency range and the Advanced Land Imager (ALI) sensor's more limited multispectral range. Human visible light range is also represented.

Materials all have a unique reflectance pattern, like a fingerprint. A material like silica reflects energy across the EM spectrum in a very different way than water vapor, or soil, or iron. Thus, by utilizing hyperspectral imaging sensors, scientists can log these patterns – referred to as a *spectral signature* – from known material samples, store them in a database known as a *spectral library*, and then refer to that database in the future to identify unknown materials in the field [1].

Chart, line chart

Description automatically generated

Figure 2 Example spectral signatures of three minerals

This method of data capture can be used not only in a lab setting, but in aerial capture from planes or satellites. The Hyperion sensor, mounted aboard the Earth Observering-1 satellite, captured hyperspectral images of the Earth’s surface from space from 2000 to 2017 [3]. Data collected from this sensor is possibly the largest open source of hyperspectral imagery available to the public as of this writing, with over 80,000 images available for download from the United States Geological Survey’s EarthExplorer online tool [4]. Aerial capture of HSI is particularly useful for mining, agriculture, and many fields of earth science [1]. HSI technology can be used to measure the health of crops, rainforests, or coral reefs, or to see through the smoke of a forest fire [5].

1. **Complications of Hyperspectral Imaging**

Because hyperspectral imaging is a relatively young technology, it is still hindered by several issues that present challenges to data scientists. One issue facing researchers is the size and quantity of the data captured. Compared to typical digital images which contain three values per pixel (red, green, and blue), hyperspectral images often contain hundreds of values per pixel. Scans from the Hyperion sensor, which records roughly 220 bands, contain 220 values for every pixel in the image [3]. This results in a single hyperspectral image containing 200 to 400 megabytes of data. Working with a satellite that captures images 24 hours a day, this quickly adds up to many terabytes of data that must be stored on site or transmitted back to Earth. For data scientists, this immense amount of information is sometimes more than necessary to perform material classification or other analytical tasks, so they rely on dimensionality reduction techniques that isolate the most relevant information.

The outstanding spectral resolution of hyperspectral images means spatial resolution is often sacrificed. In factory settings, hyperspectral image sensors used along assembly lines to assess product quality are often recording images of only a few hundred pixels []. The spectral information of these few pixels is usually enough to identify bruises on fruit or sugar content in potatoes. *Spatial resolution* is defined as the real-world area covered by a single pixel in an image [6]. Satellite hyperspectral imagery does not have nearly the spatial resolution one might find on Google Maps Satellite View. White lines on roads are clearly visible from Google’s satellites in most areas, indicating a spatial resolution of only a few inches. In contrast, a single pixel in a Hyperion scan represents an area of 30 square meters [3]. Thus, when speaking about satellite hyperspectral imagery, it is often impossible to distinguish any structure smaller than a large building or highway as of this writing. This limitation can be addressed to an extent by bringing the sensor closer to the subject (i.e., flying the sensor lower to the ground or using a plane or unmanned drone instead of satellites).

Once one has a hyperspectral scan of their research area, what is next? A farmer may be interested to know if one area of their soy field has higher iron content than another. A biologist may be scanning a forest for the presence of a particular damaging fungus. A city planner may be interested in measuring the growing number of buildings in a particular area over the past decade. Central to all these tasks is the classification of pixel data. Considering HSI’s poor spatial resolution, a single pixel in an image usually does not contain a uniform material, but a blend of several different materials. Determining the makeup of this combination is known as ‘spectral unmixing’ and is yet another challenge data scientists must reckon with [7].

For much of the past few decades, the lack of data standards also posed a challenge to researchers. As of this writing, there is no agreed upon standard for hyperspectral image data. This means HSI sensor manufacturers are recording data in whatever matter they see fit. While some sensors record every band of a single pixel before moving to the next pixel, others may record every pixel of one band before proceeding to the following band. [8] These differences in methodologies create headaches for anyone developing software or algorithms to manipulate this data. The IEEE Standards Association, the group that establishes such rulings, has been working on this problem since 2018 [9] , but have yet to complete their work as of December 2021.

Regarding ­­data analysis, the large number of variables when capturing data out in the field make data processing especially challenging. Although materials do have their own unique reflectance pattern, this pattern can be influenced by a variety of factors in the field, including time of day, time of year, angle of sunlight, intensity of sunlight, cloud cover, atmospheric haze, angle of the surface, and differences in light and shadow within the image. Furthermore, the HSI sensor in use may be more sensitive to some areas of the EM spectrum than others, further distorting captured reflectance values. In my own project, I took certain steps to minimize this issue, which will be discussed shortly.

1. **Popular Hyperspectral Software Packages**

Chart

Description automatically generatedToday there are many software packages that can read and analyze hyperspectral data. Many of these packages are provided by the same companies that manufacture HSI cameras. Examples include Resonon with its *Spectronon* software [10] and Prediktera with its own software suite [11]. This software not only controls the HSI cameras remotely, but usually include several processing and data analysis tools to make sense of collected data. Common features are Principal Component Analysis (PCA), clustering algorithms for classification and quantification, and the implementation of machine learning. *ENVI*, produced by L3 Harris Geospatial, claims to be the industry standard for image processing and analysis among scientists and GIS professionals [12]. GIS, or Geographic Information Systems, is a field of study that often goes hand in hand with aerial and satellite image analysis. *QGIS*, a popular open-source GIS software package, includes installation of *GDAL* – an open-source library of tools to read and manipulate geospatial imagery that has been in development since 1998 and upon which many newer libraries rely on [13] [14]. *MultiSpec* is another open-source program developed by Purdue University which reads hyperspectral images and provides several statistical tools to cluster and classify the pixel data. [15] One final open-source software of note is *GERBIL* [16]. Produced by German researchers, *GERBIL* tackles the challenging task of visually representing hyperspectral information to the user and includes several interesting visualization features.

Figure 3 GERBIL software includes several innovative data visualization techniques

1. **Current State of Hyperspectral Satellite Imaging**

In September 2021, NASA launched Landsat 9 – the latest generation of imaging satellites in the Landsat program which first started in 1972. [https://landsat.gsfc.nasa.gov/landsat-9/landsat-9-overview] This satellite, like its predecessors, will contribute thousands of images of the Earth’s surface to the US Geological Survey which are vital to tracking the distribution of species and ecological changes occurring over its many years in service. However, Landsat 9 does not include a hyperspectral sensor. In fact, it records in only 11 bands, which is sufficient for its purposes of detecting vegetation, cloud cover, and thermal infrared radiation. This goes to show how specialized the field of hyperspectral imaging can be. NASA did maintain a hyperspectral sensor (Hyperion) in Earth’s orbit aboard The Earth Observing-1 satellite from 2000 to 2017, but it has yet to reintroduce a new hyperspectral sensor to take its place.

Other nations are taking the lead, however, with the German Environmental Mapping and Analysis Program (EnMAP) in development and set to launch some time in 2022[]. The Italian Space Agency (ASI) launched its own hyperspectral imaging satellite, PRSIMA, in March of 2019, with the intention of eventually sharing its data with the public[]. HySIS is a hyperspectral imaging satellite currently in orbit that was launched by the Indian Space Research Organization (ISRO) in November 2018[]. China has several hyperspectral imaging satellites in orbit currently. OHS-01, 02, 03, and 04 were launched together in September 2019[https://www.tandfonline.com/doi/pdf/10.1080/10095020.2020.1838957]. All four satellites can capture data in 256 bands. These joined Gaofen 5 (GF 5) which was launched in 2018 as part of the China High-definition Earth Observation System (CHEOS). A second Gaofen 5 (GF 5-02) was launched in September 2021. []

These satellite missions are public knowledge, but it is safe to assume there are other hyperspectral sensors currently in orbit in use by various military forces around the world. Hyperspectral technology can prove useful in detecting “camouflaged or hidden military targets, such as missile launch sites or testing facilities for nuclear weapons.” [https://www.scmp.com/tech/science-research/article/1900103/hidden-man-fuelling-chinas-military-ambitions-xiang-libin] The American TacSat-3 satellite, launched in 2009 and out of service since 2012, at the time provided real-time hyperspectral data in over 400 spectral bands to US forces on the ground. The China Commercial Remote-sensing Satellite System (CCRSS), of which little information is known publicly, was scheduled to launch in 2016 and can allegedly outperform the TacSat-3. [https://www.popsci.com/china-to-launch-worlds-most-powerful-hyperspectral-satellite/]

## 2. HyperGix Overview

HyperGix is an open-source, user-friendly desktop application to experiment with and learn about hyperspectral imagery. It is written in Python 3 and uses the PyQt GUI framework for its interface. The backend of HyperGix is a SQLite database included with the application that stores data about the user’s hyperspectral files, as well as spectral information to be discussed in detail later in this report.

HyperGix aims to be intuitive and easy to use, with little instruction necessary. Presently it consists of three modules: Image Viewer, Spectra Manager, and USGS Search. Using HyperGix, the user can organize and view hyperspectral files downloaded to their system, in either RGB (red, green, blue), single-band or Normalized Difference Vegetation Index (NDVI) views. The user may also perform Principal Component Analysis (PCA) on an image, define custom material classes and assign individual pixels from hyperspectral images to those classes to create a training set for classification tasks.

If the user does not have access to hyperspectral files, or is inexperienced in where to find them, the USGS Search interface allows them to simply search for a geolocation and query available hyperspectral scans to download from the United States Geological Society’s database of Hyperion satellite scans captured from the years 2000 to 2017.

## 3. HyperGix System Requirements

HyperGix was written and designed for Windows 10 desktop machines. Though both Python 3 and the PyQt library are supported on Mac, Linux, and older versions of Windows, no testing has been performed with HyperGix on those systems and support cannot be guaranteed.

HyperGix requires the installation of the Geospatial Data Abstraction Library (GDAL) to operate. GDAL is a free, open-source translator library and command-line tool to read and manipulate geospatial data files such as those used by HyperGix. Installation of GDAL can be complex, and it is not included with HyperGix. A guide to installing it on Windows 10 can be found in the Appendix. Information about GDAL can be found at [**www.gdal.org**](http://www.gdal.org)**.**

HyperGix includes an installation of Python 3.6, as well as several Python libraries listed below. These libraries are already packaged with the software so no action should be required from the user. However, in the event of an issue, any of the libraries below can be easily acquired from the Python Package Library (pypl.org) and installed with PIP.

* brokenaxes 0.5.0
* gdal 3.0.2
* geocoder 1.38.1
* matplotlib 3.4.3
* numpy 1.21.2
* pyqt 5.9.2
* requests 2.26.0
* spectral python 0.22.2
* sqlite 3.36.0

The HyperGix USGS Search feature also requires login credentials to the USGS EROS Registration System. Users may register an account at **ers.cr.usgs.gov/register**

For demonstration purposes, login credentials are already included with the HyperGix download. However, if the user intends to use HyperGix extensively, it is encouraged they register their own account and input the credentials when prompted.

## 4. HyperGix Design and Structure

### 4.1 Module System

HyperGix version 1.0 consists of three modules separated into tabs at the top of the program interface. These modules are also separated in the code as distinct classes – qViewer, qProfileManager, and qHypbrowser – which are defined in qviewer.py, qmanager.py, and qhypbrowser.py respectively. The main application is defined in main\_pyqt.py. This file includes code for the HyperGix application window and tab structure (MyNotebook class) as well as the many signals and slots used by PyQt for the different modules of the app to correspond with each other. Developers are encouraged to continue this system of separating modules into distinct classes and adding a tab to the main interface that instantiates that class. It is required that all classes that intend to use the signal and slots system inherit from the PyQt QWidget or QFrame classes.

**Image Viewer**

Most of the time in HyperGix will be spent in the Image Viewer module. This module is dedicated to viewing and managing the user’s hyperspectral image library and the files therein. Upon opening HyperGix, any suitable Hyperion images contained in the **downloads** folder in the HyperGix directory will be automatically added to the Scan Library list on the left side. Hyperion images must have the file extension **.L1R** to be recognized and added to the image library automatically.

Early in development, an error inherent to Hyperion image header files was discovered in which the ‘header offset’ value is mistakenly recorded as zero when it should be 2502, causing the images to display incorrectly. HyperGix automatically handles and corrects this error so the images display accurately in the software.

HyperGix also allows for importing supported hyperspectral images from external directories. The location of the imported file on the user’s computer is stored in the SQLite database so moving the file to the ‘downloads’ folder is not necessary. At present, supported filetypes are limited to **.lan, .hdr, .L1R, .tif, .hd5, and .he5**, though importing other filetypes can be attempted. HyperGix will attempt to use the GDAL library to convert an unsupported file into something it can read, but success is not guaranteed due to the lack of current standards in hyperspectral data storage.

The **Remove From Library** button in the Image Viewer module will remove information about a spectral image from the internal database, but it will not delete the file from the user’s system. A future update may include a pop-up window to present this option.

Image files can be viewed in the application by double-clicking their name in the Scan Library list. Information about the file is displayed in the File Properties panel in the lower left side of the Image Viewer module. This information is included in the image file itself and can vary by file format. For Hyperion scans, it includes the unique Scene ID assigned by the USGS, along with capture date and time, and the dimensions of the image. Hyperion images all consist of 242 bands and 256 columns (or samples) but the number of rows may vary.

The center area of the Image Viewer module allows the user to view an individual hyperspectral scan, pan and zoom on it, and save still frames of an area to their computer. It also allows switching between three viewing modes.

**RGB Mode** – a full color representation of the captured area. By default, bands 50, 27, and 17 are used for red, green, and blue values respectively but the user may select any bands for these values.

**Single Band Mode** – view a single band of the hyperspectral image. Some bands may include noisy data or no data at all. When hovering the cursor over the image, single-band values for the pixel currently under the cursor are shown in the upper right corner of the central panel.

**NDVI Mode** – Normalized Difference Vegetation Index is a graphical indicator of live green vegetation based on the nature of vegetation to reflect infrared light while absorbing red light. It provides an approximation of vegetation health in a region. HyperGix uses green to indicate areas of lush vegetation, red to indicate inorganic materials or dead vegetation, while black usually indicates bodies of water.

By double clicking a spot within the image, a entire spectral profile of the single pixel clicked is presented on the right side of the Image Viewer module as a line graph. For Hyperion scans, a break is included on the X-axis to indicate the omission of bands 58 to 76 which are known to contain no data. [<https://developers.arcgis.com/python/sample-notebooks/landcover-classification-using-hyperspectral-imagery-and-deep-learning/>] Near the bottom of the chart are bars that represent the areas of blue, green, red and near-infrared light among the spectral bands.

Below the Pixel Spectra chart are two buttons. **Add Pixel to Profile** will store the most recently viewed pixel spectra in the database, along with an assigned material class selected by the user. **Calculate PCA** will perform principal component analysis. *This process may take several seconds depending on the size of the file and may appear to freeze up the application.* Please be patient while the extensive calculations are performed. This process will produce a covariance matrix as well as the number of principal components necessary to represent 99% of the variance of the image data. In most cases, this is less than 5. The individual principal components can be viewed in the central panel by adjusting the band slider at the bottom of the panel.

**Spectra Manager**

TO-DO LIST

* USGS login credential input

APPENDIX

Installing GDAL

<https://towardsdatascience.com/python-and-gdal-installation-automated-for-windows-10-f22686595447>

Sample AVRIS and Hydice image downloads