Statistical Machine Learning for Spectral Image Processing

William F. Basener

August 22, 2024

This document was typeset with the help of KOMA-Script and LaTeX using the kaobook class.

Preface

This is a textbook on the processing of spectral images using statistical machine learning.

By a ***spectral image***, we mean a digital image with multiple bands. A digital image acquired by a standard digital camera has three bands associated with red, green, and blue (RGB) colors. More generally, each band in a spectral image is the quantity of light detected in a given range of wavelengths. A spectral image is generally stored on a computer as 3-dimensional array with shape m, n, b, where m is the number of rows, n is the number of columns, and b is the number of bands. Each pixel is a vector of length b.

Each pixel in the image is a vector of values, and each value is a physical measurement of the light in a specific wavelength range that was reflected from or emitted by the area on the ground or object corresponding to the pixel. The number of values in the pixel vector is the number of bands in the image. The word *spectral* implies that the pixel vector is a physical-chemical-biological properties of the object or area corresponding to the pixel. The focus of this book is determining spectral (physical-chemical-biological) properties of materials in the pixels from the image.

By ***statistical machine learning*** we mean computer code or software that takes data as input and outputs useful desired information. The relationship (or function) that predicts (or estimates) output values from input data is called the machine learning ***model***. A model is a function defined by a formula with parameters (or a set of rules), and the values for the parameters are determined by providing a set of input-output pairs called ***training data*** and estimating values for the parameters that optimize some criteria such as accuracy.

The term *statistical* in statistical machine learning emphasizes that we describe statistical assumptions and relationships within the models. This is important in spectral imagery as these relationships are often correspond to physical and chemical properties of materials and image collection. The term *learning* in statistical machine learning emphasizes that the model learns the parameters from training data rather than having the function explicitly defined. The parameters often have physical meaning. Consideration of the physical meaning of the parameters can be a guide toward building more accurate models, and the values of the parameters can provide physical-chemical information in addition to accurate predictions.

The output information may be specific physical-chemical-biological properties, for example the condition of a road surface, chemical content of soil, health of vegetation, or specific polymer type of an object. Or it may be less specific, for example grouping of pixels in an image into clusters of undetermined types (***unsupervised learning*** or ***clustering***), detection of pixels that have unusual or anomalous spectra (***anomaly detection***), or detection of pixels that are likely to contain an object or material matching a known target spectrum (***target detection***). The process of determining a specific class for the material in a pixel is called ***identification***. There is some overlap in these terms, but detection usually means detecting pixels satisfying some criteria form a whole image, whereas identification usually means identifying ery specific information for a given pixel.

Often the area measured by a pixel in a spectral image consists of more than one material. The process of determining the materials that are present and their relative abundances in the pixel is called *spectral unmixing* or just *unmixing*. The process of unmixing is an important goal in itself, for example determining the mixture of minerals that are present on the earth's surface for the purpose of understanding the geology or locating ore deposits for mining. The process of unmixing is also an important step in a general process of clustering, classification, detection, or identification.

Most of these processes are determined by the application goals, properties of the materials of interest, and characteristics of the image collection. Important characteristics of the image collection include the number of bands, wavelength ranges of the bands, spatial resolution, and signal to noise ratio of the sensor.

Machine learning models are often categorized by whether the output is continuous-valued (***regression***) or discrete-valued (***classification***). They are further categorized by assumptions about the type of relationship between input data and output variables. For regression models this may mean whether the relationship follows a linear formula, follows a polynomial formula, or has some less-structured set of rules. For classification, this may mean rules and formulas for assigning class labels and correspondingly whether the separation surfaces between classes is linear or nonlinear. A model structure or formula often corresponds to assumptions about the underlying probability distributions for the classes and parameters.

A primary goal of this book is equipping the reader to:

1. Open and visualize spectral image data.
2. Understand the phenomenology an application using spectral imagery.
3. Determine a machine learning model with assumptions suited to the application and phenomenology.
4. Build and train the model using Python.
5. Evaluate the model performance using appropriate metrics.
6. Evaluate whether a dataset satisfies the assumptions of the model.
7. Visualize the model properties and output.
8. Write Python software to practically implement a model and

William F. Basener

Contents

Introduction 1

1.1 Digital Images

A digital image is an image collected by a sensor that is stored on a computer. The image consists of a grid of pixels, each of which is a the intensity value for a discrete location in the image.

A digital image of the eye of a dog is shown in Figure Figure 1.1. This image has 15 rows and 10 columns of pixels. This is a grayscale image, meaning that each pixel has a single value that is used for the brightness of the pixel. The numerical value for each pixel is shown in red in this image.

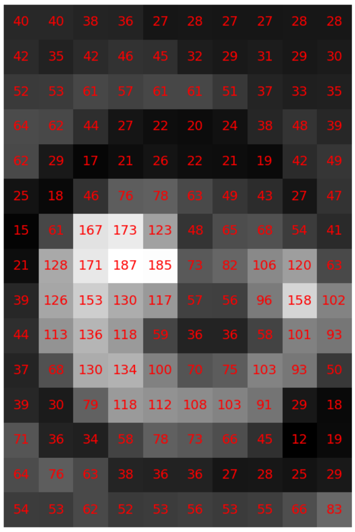

Figure 1.1: A grayscale image of the eye of a dog. The numerical value for each pixel is shown in red. The pixel values are 8-bit integers ranging from 0 to 255.

The number of bands in an image is the number of numerical values for each pixel. A grayscale image has a single band, and is stored in computer memory as an $m \times n$ array, where m is the number of rows and n is the number of columns. A standard color image has three bands, one band for each standard visual color (red, green, blue, abbreviated RGB). A **spectral image** is a digital image that has more than one band, usually three or more, and is stored in computer memory as an array with shape $m \times n \times b$, where b is the number of bands. While an RGB image is technically a spectral image, the term *spectral* implies that the pixel values represent measurements of light that provide physical and chemical information about the objects in the image.

A digital image is stored in memory by storing the numerical values for the pixels consecutively. There are different conventions for how the values are read and reorganized into the 2-dimensional or 3-dimensional array. The convention for how the sequential numerical values are reorganized into a multidimensional array is usually not important, and we interact with the image data as a variable representing the array.

However, the method for storing each number in the image array as a sequence of binary numbers, called the **data type** of the array, determines computational speed in processing, the amount of memory needed to store the image, and the degree of accuracy. Most images whose primary use is displaying on a screen or printing are stored as unsigned 8-bit integer arrays (equivalently, have an unsigned 8-bit integer data type). An unsigned 8-bit integer is a value that can be stored using 8 bits of memory, and takes on $2^8 = 256$ possible integer values ranging from 0 to 255. The word *unsigned* here indicates that we consider only positive integers, but often this data type is called 8-bit integers. The image in Figure **??** is an 8-bit image. For a grayscale image, a 0 represents a black pixel and 255 represents a white pixel.

In an 8-bit color image, each pixel has three values represented as a vector $[R, G, B]$, called the RGB value for the pixel, providing the amount of red, green, and blue in the pixel. For example, a pixel with the values $[0, 255, 0]$ will be green. The following code creates a GUI (Graphical User Interface) for selecting a color and printing the RGB value for that color. The GUI created by this code is shown in Figure Figure 1.2.

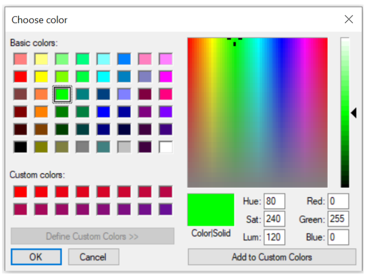

Figure 1.2: A GUI for selecting a color and returning the associated RGB value.

```python
from tkinter import colorchooser

# Create the GUI to select the color, and return the color code.
color_code = colorchooser.askcolor(title ="Choose color")
print(color_code)
```

```
((0, 255, 0), '#00ff00')
```

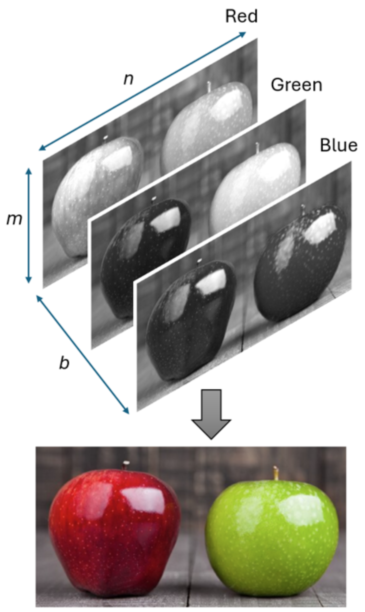

Figure 1.3: The 2-dimensional color visualization of an image of two apples (bottom) from the 3-dimensional array (top).

A color image is stored in memory as an $m \times n \times b$ array which we call usually call `Im`. The first two coordinates are called the spatial dimensions, and the third coordinate is the spectral dimension. The pixel at coordinate x, y is the vector `Im[x,y,:]` of length b. For each $i \in \{0, \ldots, b-1\}$, the array `Im[:,:,i]` is the i-th band of the image, and is often displayed as a grayscale image. Figure Figure 1.3 shows an RGB color image of two apples, first showing the individual bands as individual grayscale images and then the 2-dimensional color image.

Unsigned 8-bit integer images take up relatively little space in memory, and are fast to load and display. The 256 available values provide sufficient variation in color for visualization, but 8-bit integer arrays are of limited use for representing precise measurements of physical quantities that are the primary interest in spectral imaging. Unsigned 16-bit integers are sometimes used for spectral images because they provide $2^16 = 65,536$ possible values, ranging from 0 to 65,535. The most common data type for spectral images is 32-bit float, which allows floating point (decimal) numbers with about 7 digits of precision.

The images we focus on in this book are in the context of remote sensing, which in general is the process of taking measurements about an object without making contact. Remote sensing can be subdivided into the categories of passive and active. Active remote sensing is remote sensing where the energy is emitted (e.g. sending a radar wave, laser pulse, underwater sonar, etc.) and the energy returning from the object are measured. Passive remote sensing is remote sensing where energy comes from external sources (e.g.light from the sun, infrared light from a hot object, etc.) and the measurement is made of the energy after it interacts with the object.

1.2 Light

In spectral imaging, the pixel values are measurements of light coming from a material that can be used to infer information about the material. The behavior of light is determined by both relativistic physics (for example the speed of light is constant) and quantum physics (for example light behaves as discrete particles that we call photons). These theories are expansive and beautiful, but almost completely beyond the scope of this book. In this section we describe some physical principles of light that are useful for understanding spectral imagery. The details of the specific physical laws are not important for the rest of the book, but we include them here to provide a more complete explanation of the physics underlying observations in spectral imagery.

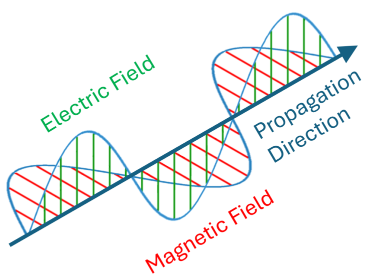

Figure 1.4: Light propagates as electric field waves and magnetic field waves. These waves are perpendicular to each other and perpendicular to the direction of propagation.

Light is energy that propagates through space as a wave. This light wave consists of waves of electric and magnetic fields, called electromagnetic radiation, as shown in Figure Figure 1.4. James Clerk Maxwell first conjectured in 1865 that light is an electromagnetic wave in his famous equations. Various forms of electromagnetic radiation, which differ only in wavelength, are depicted in Figure Figure 1.5. The full range of types of electromagnetic radiation is called the electromagnetic spectrum.

Light travels at a constant speed of $c = 3 \times 10^8$, which was implied by Maxwell's work in 1865. The frequency of light, ν (pronounced noo), is the number of waves that pass a fixed point in space during a fixed unit of time. This is is related to the wavelength λ by the formula

$$\nu = c/\lambda.$$

For example, light with a wavelength of $\lambda = 600nm = 6 \times 10^{-7}m$ has a frequency of

$$\nu = (3 \times 10^8)/(6 \times 10^{-7}m) = 5 \times 10^1 41/s.$$

The connection between light and wavelength was implied by Maxwell's equations in 1865, and was experimentally validated by Heinrich Hertz in 1886 when he showed that abrupt changes in the electromagnetic field around a wire (caused by changes in the current in the wire) create a form of radiation that can be measured at a distance. The waves generated by Hertz were radio waves, with a wavelength of about 4 meters. This is still how radio waves are transmitted (alternating current in a wire generates the the wave) and received (radio waves incident on an antenna cause an alternating current).

Electromagnetic radiation that we perceive as visible light has a wavelength in the range from approximately 400nm to 700nm. Light is detected in human eyes by cone cells (that detect color) and rod cells (that are more sensitive to intensity than cone cells but do not detect color). Human eyes normally have three types of cone cells that (approximately) detect blue, green, and red light. Blue light corresponds (approximately) to light in the 450nm-500nm range, green corresponds to light in 500nm-550nm, and red corresponds to light in 600nm-700nm. The spectral response function is the function providing the amount of response (or sensitivity) of a receptor as a function of wavelength. Spectral response functions for the three types of cones in the human eye, on average, are shown in Figure Figure 1.6. The ranges designated for the colors blue, green, and red vary substantially depending whether we are measuring with reference to the peak sensitivity of the different types of cones in the human eye, wavelengths that are perceived as most unique, or other criteria. Reasonable single wavelength values for these colors are 460nm (blue), 550nm (green), and 640nm (red).

A vector of values representing the amount of light in a sequence of wavelengths is called a **spectrum**. Each pixel in a spectral image is a spectrum for the location represented by the pixel.

When light is incident on a material, it can either be **reflected** off the material, **transmitted** through the material, or **absorbed** in the material in the form of kinetic energy. The fraction of light reflected off a material is a function of wavelength is called a **reflectance spectrum** and the fraction of light transmitted a material is called a **transmission spectrum**.

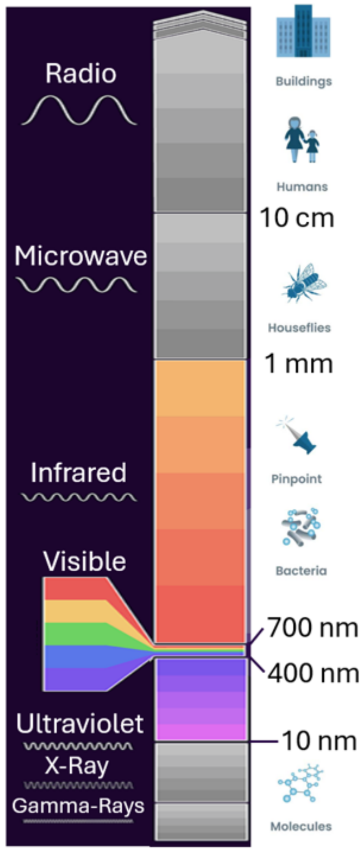

Figure 1.5: The wavelength for various types of radiation across the electromagnetic spectrum. To the right of each category of radiation is an object of approximate size of the wavelength.

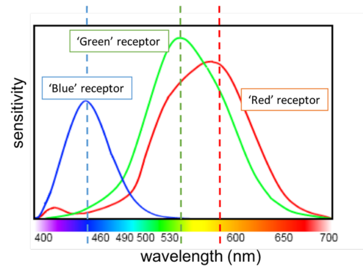

Figure 1.6: The spectral response functions of the three cones in the human eye across wavelengths of light.

Definition 1.2.1 *A **reflectance spectrum** is a vector $\mathbf{x} = (x_1, ..., x_n)$ where x_i is the percent of light reflected by a material at a fixed wavelength or wavelength range.*

In addition to reflecting or transmitting incident light, all objects **emit** light. This emission of light is observable for hot objects; for example iron and steel start glowing red around 1000°F (500° C), and as temperature increases the brightness increased and color changes through orange, yellow, then white at around 2200°F (1200° C). The amount of light emitted by a material as a function of wavelength is called an **emission spectrum**.

Definition 1.2.2 *An* ***emission spectrum*** *is a vector* $\boldsymbol{x} = (x_1, \ldots, x_n)$ *where* x_i *is an amount light emitted by a material at a fixed wavelength or wavelength range.*

A **blackbody** is an idealized theoretical object that absorbs light in all wavelengths. The emission spectra for a blackbody depends on the temperature of the blackbody according to Planks law, which says that the radiance of a blackbody at temperature T at wavelength λ is

$$B_\lambda(\lambda, T) = \frac{2\hbar c^3}{\lambda^5} \frac{1}{e^{\hbar c/(\lambda K_B T)} - 1}$$

where $\hbar$ is Plank's constant, c is the speed of light, and K_B is the Boltzman constant. Plots of this radiance spectrum for a blackbody at various temperatures are shown in Figure Figure 1.7. The sun emits light close to a blackbody with temperature 5777K, which is called the apparent temperature of the sun.

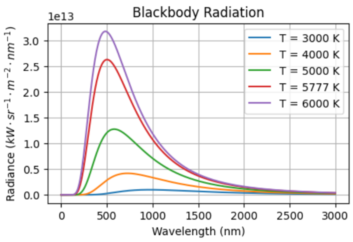

Figure 1.7: Blackbody emission spectra for different temperatures.

When light encounters and object, there are a number of phenomena that can occur that depend on the elements and bonds present. This is the fundamental principle that makes spectral imaging possible; measurement of light after interaction with matter at multiple wavelengths provides insight into the structure of the matter.

An early observation that provided insight into a relationship between light an matter was observations of light hitting a metal surface that cause electrons to be emitted (called the photoelectric effect) and furthermore:

1. The kinetic energy of electrons emitted does not change when light intensity increases.
2. The kinetic energy of electrons emitted increases when the frequency increases (wavelength decreases).
3. The number of electrons emitted increases when light intensity increases.
4. The number of electrons emitted does not change when the frequency increases (wavelength decreases).

This contradicted classical electromagnetism, which increasing intensity of light should correspond to increasing energy, and thus electrons emitted with increasing energy. This observation led Einstein to propose (in 1905) that light has both wave-like and particle-like properties, that intensity in light is proportional to the number of photons, and that energy of light each photon is proportional to the frequency. The energy of an individual photon E_{photon} is determined by Plank's law,

$$E_{\text{photon}} = \hbar\nu$$

where $\hbar$ is Plank's constant. Max Plank first proposed this law in 1900 when studying the formula for blackbody radiation.

The principle that photons of light at a fixed wavelength (or frequency) have a fixed energy determines much of the phenomenology in light-matter interaction.

Light with a long wavelength, like radio waves, have low energy that is not sufficient to cause an interaction with matter. This explains why radio waves travel through solid objects, and why AM signals (with lower frequency that FM) travel better through solid objects than FM signals do.

Microwaves are waves in the short-wavelength range of radio waves (hence the prefix micro). Inside a microwave oven, microwaves with various wavelengths are generated. The fluctuations in the electric field in microwaves cause molecules in food that have non-symmetric charge distribution (polar molecules) to rotate, aligning with the changing field. This movement results in kinetic energy in the food. Microwaves are also useful in satellite communications, active remote sensing like radar systems (determine speed of a vehicle, locations of aircraft), and a form of imagery called synthetic aperture radar (SAR).

In SAR imagery, a radar wave is emitted and the return is measured and reconstructed to create an image. SAR imagery is less intuitive to interpret visually in comparison to visual spectral imagery, but the SAR microwaves can pass through clouds, smoke, rain, dust, and snow, and they contain additional information about the objects. As a general rule, microwaves with longer wavelengths (lower frequency) have less data throughput and require a larger antenna, but are less susceptible to interference from weather and fading with distance. The image resolution of a SAR image is related to the ratio of the wavelength to antenna size, so lower wavelengths provide higher resolution SAR images but are more susceptible to weather interference. Table Figure **??** lists some common SAR bands and the applications of SAR imagery acquired with these bands.

Application	**K-band**	**X-band**	**C-band**	**S-band**	**L-band**	**P-band**
Wavelength (cm)	0.75-2.4	2.4-3.75	3.7-7.5	7.5-15	15-30	30-100
Penetration:						
crop canopy (low-moderate)			Yes	Yes		
crop canopy (high)					Yes	Yes
Hydrometeors or rain core				Yes	Yes	Yes
Precipitation			Yes	Yes	Yes	Yes
Cloud, fog, dust, smog, smoke		Yes	Yes	Yes	Yes	Yes
Dry alluvium		Yes	Yes	Yes	Yes	Yes
Dry snow or ice		Yes	Yes	Yes	Yes	Yes
Wet soil		Yes	Yes	Yes	Yes	Yes
Mapping:						
Canopy mapping	Yes	Yes				
Flooded grass	Yes	Yes	Yes			
Flooded reed & brush			Yes			
Flooded canopy					Yes	Yes
Sea ice monitoring	Yes	Yes	Yes	Yes	Yes	Yes
Oil spill mapping	Yes	Yes	Yes	Yes	Yes	Yes
Earth surface mapping	Yes	Yes	Yes	Yes	Yes	Yes
Flood mapping	Yes	Yes	Yes	Yes	Yes	Yes
Soil moisture monitoring	Yes	Yes	Yes	Yes	Yes	Yes

Table 1.1: Common microwave bands used for synthetic aperture radar imagery and associate applications, adapted from ArcGIS Pro documentation.

Imaging using the infrared, visible, and ultraviolet wavelengths of light is collectively called **electro-optical (EO)** imaging. Electro-optical imaging uses lenses to focus light on a detector array that converts incident light energy into electrical energy.

Incident infrared light is absorbed by a molecule when the frequency of the light resonates with a resonant frequency of a vibrational model of a bond within the molecule. A vibrational model is simply a type of vibration that can occur among the bonds in a molecule. Some types of vibrational modes are shown in Figure Figure 1.8. The resonant frequencies for molecules can be modeled physically in computational chemistry software using Hooke's law (which is the law for vibrations in springs). Understanding that a given molecule will have fixed frequencies of light that it will absorb (actually, photons with wavelength near the resonant frequency will be absorbed) is important for understanding spectral imagery, but specific associations between molecules and vibrational frequencies is not a focus of this book. I a frequency x corresponds to a vibrational mode of a molecule, then integer multiples of x ($2x$, $3x$, etc.) will also be vibrational modes. In terms of wavelength, this means that if wavelength λ is a vibrational mode (light at this wavelength is absorbed), then light at wavelengths $\lambda/2$, $\lambda/3$, etc. will also be absorbed. These additional wavelengths are called overtones, and the range of wavelengths absorbed for the overtone λ/n becomes broader as n increases.

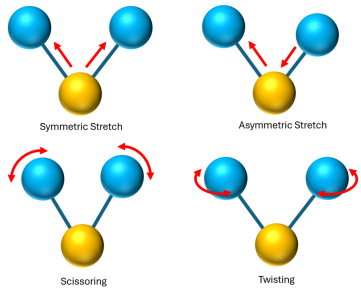

Figure 1.8: Some of the vibrations that can occur within a molecule as a result of incident infrared light.

Incident visual and ultraviolet light on a molecule can cause electrons to move up in energy levels (shells) within an atom. This movement of an electron will occur if the energy of the photon (which is proportional to its frequency) is equal to the amount of energy needed to move the election. Since the energy levels, and amount of energy needed to move electrons between them, are fixed, this means that specific wavelengths of light will be absorbed and the rest of the light will be transmitted through or reflected off the material.

The values in a measured reflectance
transmission
emission spectrum of light from an object depends on the chemistry of the material. The process of inferring the properties of the object from the spectrum is called **spectroscopy**. Lab spectroscopy involves measuring spectra in controlled setting, obtaining detailed a spectral information for a single material sample. Spectral image processing is sometimes called *spectroscopy write large* because it enables a spectroscopy analysis across a large area.

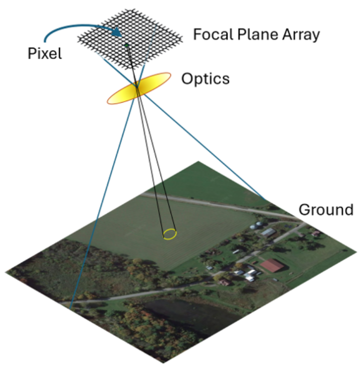

Figure 1.9: The collection of an image using a focal plane array.

1.3 Sensors and Image Collection

The physical instrument that measures light to create a spectral image is called an **imaging sensor** or just sensor. Light enters the sensor through a lens. The lens focuses the light onto a 2-dimensional grid of detectors called a focal plane array (FPA), as depicted in FigureFigure 1.9. The light incident on each detector is coming from a location on the ground which can be represented by an ellipse as shown in Figure Figure 1.9. The detector outputs an electric charge that is proportional to the intensity of the incident light, which is then converted to a pixel value.

The sensitivity of a detector to incoming light is called the *quantum efficiency* of the detector, which is defined as the percent of electrons generated per photon of light. Detectors can have a quantum efficiency in the 50%-80% range.

In collection of a panchromatic image (or grayscale image), each detector measure the light from a location on the ground across a broad range of wavelengths. The most common method for collecting a visual RGB color image is by placing a filter in front of each detector, so that each detector measures only light in a single wavelength range. A common pattern for the color filters is a Bayer pattern or Bayer filter, shown if Figure Figure 1.10. The measured colors are then combined to create the color pixels in an image.

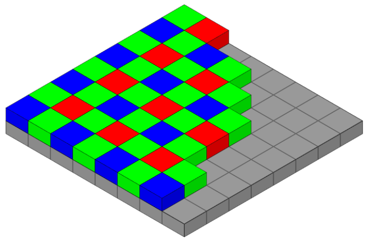

Figure 1.10: A Bayer pattern, which is an arrangement of color filters on a FPA for creating a color image. (Image by CBornett.)

A sensor designed for collecting a large number of bands usually collects one row of pixels at a time as shown in Figure Figure 1.11. The sensor moves in the direction perpendicular to the row as shown. Light for the row of pixels enters the sensor through a slit, passes through optics that focus the light and disperse it through either a prism or diffraction grating. The 1-dimensional row of light entering the sensor is spread across the 2-dimensional focal plane array. One dimension of the FPA collects the spatial dimension of the row and the other dimension of the FPA collects the spectral dimension. As the sensor moves, for example on an aircraft, successive rows are collected and stacked to create an image. This configuration is called a **pushbroom sensor**. The other common sensor configuration uses a mirror to move the side-to-side rather than the motion of the sensor. In both cases, the light enters through a slit and is dispersed across the FPA with one spatial and one spectral dimenskion.

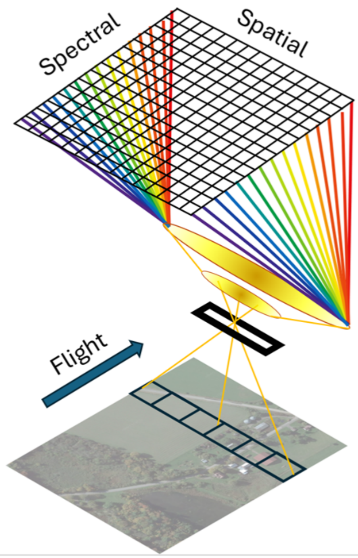

Figure 1.11: A pushbroom spectral image collection configuration. The sensor collects one row of pixels at a time, spreading the light out spectrally across the FPA and moving in the direction indicated to collect an image.

1.4 Viewing Spectral Images

Appendix

Notation

The next list describes several symbols that will be later used within the body of the document.

c Speed of light in a vacuum inertial frame

h Planck constant

Greek Letters with Pronunciations

Character	Name	Character	Name
α	alpha *AL-fuh*	ν	nu *NEW*
β	beta *BAY-tuh*	ξ, Ξ	xi *KSIGH*
γ, Γ	gamma *GAM-muh*	o	omicron *OM-uh-CRON*
δ, Δ	delta *DEL-tuh*	π, Π	pi *PIE*
ϵ	epsilon *EP-suh-lon*	ρ	rho *ROW*
ζ	zeta *ZAY-tuh*	σ, Σ	sigma *SIG-muh*
η	eta *AY-tuh*	τ	tau *TOW (as in cow)*
θ, Θ	theta *THAY-tuh*	υ, Υ	upsilon *OOP-suh-LON*
ι	iota *eye-OH-tuh*	ϕ, Φ	phi *FEE, or FI (as in hi)*
κ	kappa *KAP-uh*	χ	chi *KI (as in hi)*
λ, Λ	lambda *LAM-duh*	ψ, Ψ	psi *SIGH, or PSIGH*
μ	mu *MEW*	ω, Ω	omega *oh-MAY-guh*

Capitals shown are the ones that differ from Roman capitals.

Alphabetical Index