Feature Space Interactive

American Museum of Natural History's
Center for Biodiversity and Conservation

Published 18 May 2009

Table of Contents

1. Introduction

The feature space interactive is designed to teach basic concepts about what feature space is and how it is used in image classification. When we use the term "feature" in the context of "feature space" we refer to data layers that comprise a data set. When the data set is a multispectral image the features are the individual bands. Other data layers, such as elevation, slope, aspect can also be features in a data set. A subset of a Landsat image will be used for the interactive but the same concepts apply regardless of the type of features.

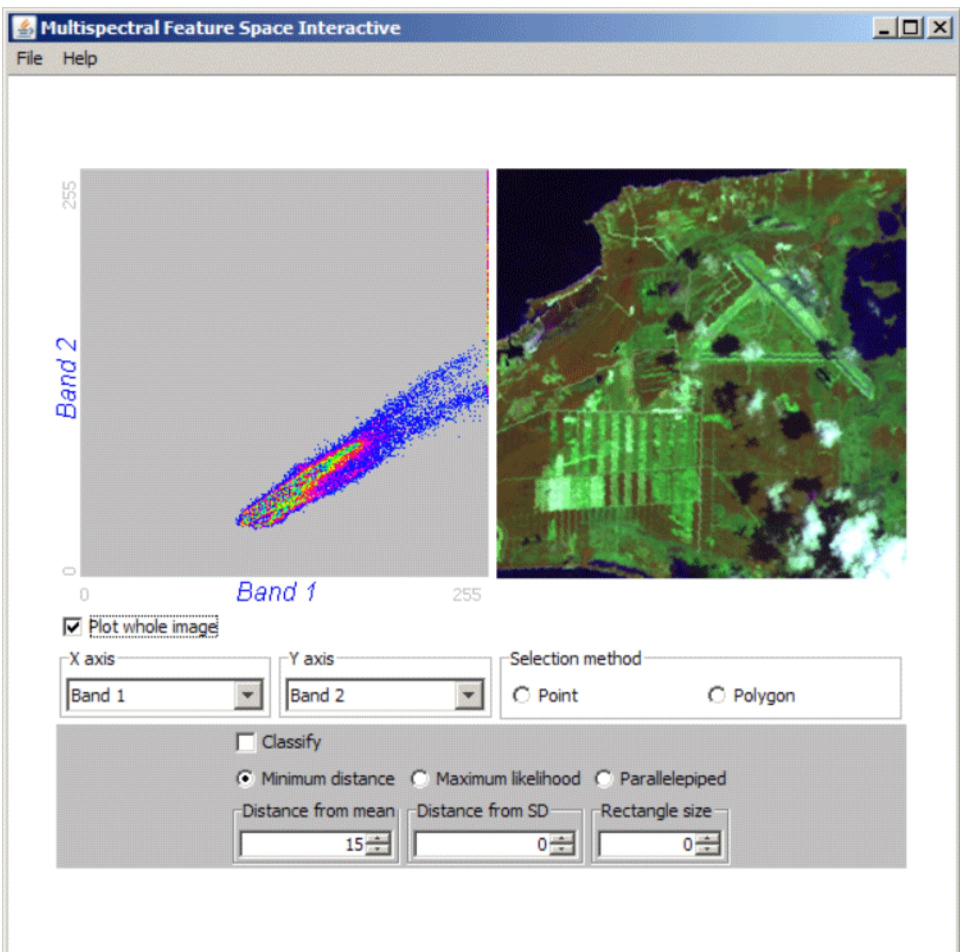
Data in an image are arranged as a grid of pixels. When we look at a satellite image the pixels are organized in a grid so that objects such as lakes, rivers, and forests appear in a way similar to the way they would in a photograph or on a map. We can also view image data using a feature space reference system. In a feature space plot the axes represent the range of possible values for a specific feature. In our case the axes represent the range of pixel values which for Landsat images range from 0 to 255. The simplest feature space plot uses two axes (X and Y) and this is what is used in this interactive. Three-dimensional plots are also possible for visualization purposes. When used in an image classification algorithm the feature space can have several dimensions, one dimension for each feature. In other words if you are processing a 7 band image the algorithm will be working in 7-dimensional feature space.

2. Displaying an image and a feature space plot

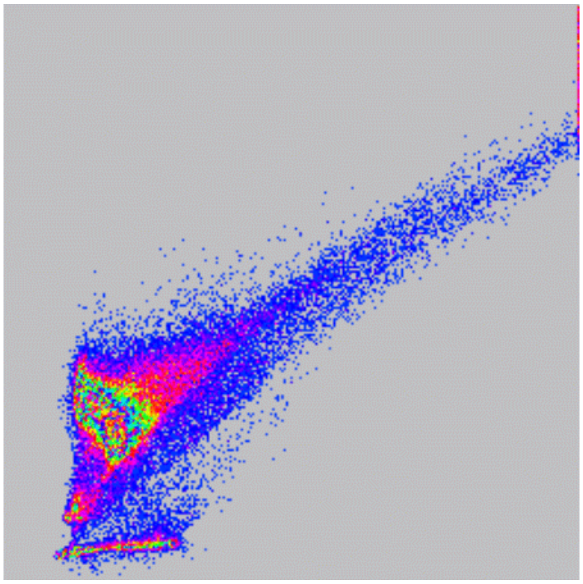
When the interactive opens you will see two display windows, one on the left for the feature space plot and one on the right displays a subset of a Landsat ETM+ image.

The image can be colored by selecting bands to display in the red, green, and blue channels by clicking ***File* → *Band Selection***. This launches the **Band Selection** window. Click on the **Band Selection** drop-down boxes to select the Landsat image band to display in the **Red**, **Green**, and **Blue** channels. You can see the colors in the image change as you make different selections. Click OK when the selection is finished.

To display the feature space plot for the image click on the **Plot whole image** check-box. This will display a multicolored cloud of points. The colors represent the frequency of occurrence for a particular pixel value pair. Blue is for value pairs with a low frequency of occurrence and magenta is used for value pairs with a high frequency, with several color in between. The units for each axis are the digital number (DN) (i.e., pixel value) for the image band. For example, an object, such as water, that has a low DN in each band will occur in the lower left of the feature space plot and a feature with a high DN for both bands will occur in the upper right . Clicking on the image or the feature space plot will highlight the corresponding pixel in the other display.



When started the feature space plot displays Band 1 on the X axis and Band 2 on the Y axis. This produces a fairly linear plot since these two bands are closely correlated. To see a plot from two bands that are not closely correlated select Band 3 for the X axis and Band 4 from the Y axis.



3. Exploring feature space

To see where features in an image occur on a feature space plot click on the **Point** radio button and then click somewhere on the image. As you click on the image you will see a colored “+” cursor appear at the corresponding point in the feature space plot. Click around the image to get a sense for how different image features appear in 2-dimensional feature space.

Clicking on the feature space plot will highlight the locations in the image where those pixels occur. Clicking on green points of the plot will highlight several pixels in the image and clicking on the blue points will highlight only one of just a few pixels in the image.

To see how groups of pixels appear click on the **Polygon** radio button. Draw a polygon on the feature space plot by holding down the left mouse-button. Releasing the mouse button to close the polygon. All of the image pixels corresponding to the value point pairs enclosed by the polygon will be highlighted on the image. When you draw a polygon on the image, the corresponding feature space points will be displayed in the plot window and the other points will be removed.

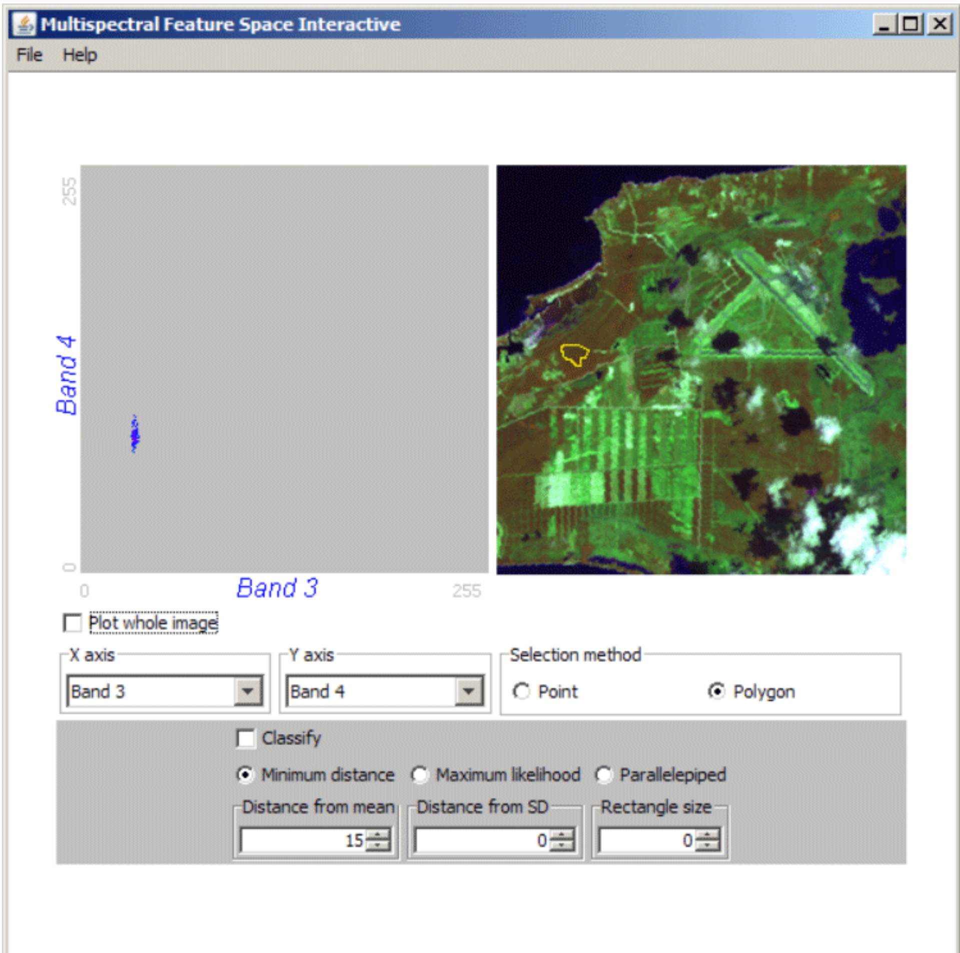
Distance in feature space is measured in the units of the axes, in our case pixel value (DNs). For example, the distance between two diagonal corners of the feature space plot (i.e., the point X=0, Y=255 to X=255, Y=0) is 360 (i.e., SQRT[255^2 +255^2]). Grasping the concept of feature space distance is important to understand how the classification algorithms described below work.

Use the Point and Polygon tools and display different bands on the X and Y axes to explore the image and get a good understanding of how an image and its feature space are related.

4. Image classification algorithms

Most image classification algorithms work by defining objects in feature space. We will demonstrate how three different algorithms can be used. The three algorithms we use in this interactive are: Minimum distance, Maximum likelihood, and Parallelepiped. For this interactive you will define a training area on the image using the Polygon drawing tool. A training area is a group of contiguous pixels that represent the object you want to classify.

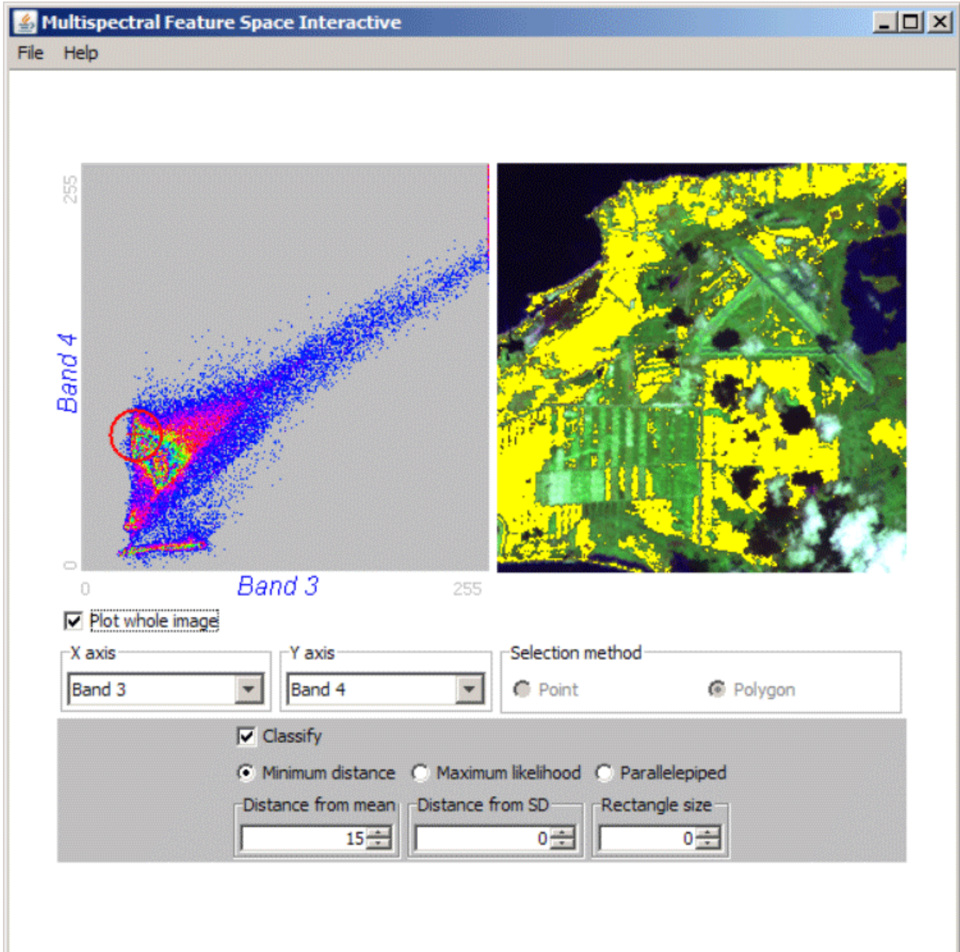
To see how this works, draw a polygon around the forest area on the image (see figure below). When you close the polygon the corresponding points will be shown in the feature space plot window. Statistics such as range, mean, and standard deviation are calculated for each of the feature space bands and the co-variance between bands is also calculated. These statistics are used to define objects in the feature space plot.



4.1. Minimum distance

The minimum distance algorithm determines if a pixel in an image is associated to a particular class by measuring the distance in feature space between the location of the pixel being classified and the location of the mean for each of the training areas.

Click on the **Classify** check-box then click on the “Minimum distance” radio button. This will display a circle in the feature space plot window and it will highlight all of the corresponding pixels in the image. The center of the circle will be located at the mean of the pixels in the forest area object you defined and the radius of the circle is defined by the **Distance from mean** text box. You can change the size of the circle by clicking the up or down arrows or typing in a new distance value. The circle can be thought of defining the forest object in the feature space plot.



Clicking on the **Plot whole image** check-box will show how the object defined by the circle is positioned relative to the other points in the feature space plot.

4.2. Maximum likelihood

Clicking on the **Maximum likelihood** radio button the circle will change to an ellipse. The size of the ellipse is defined by standard deviations from the mean. With a maximum likelihood classification distance is measured in units of probability, standard deviations from the mean. You can change the size of the ellipse by clicking the up or down arrows or typing in a new standard deviation value.

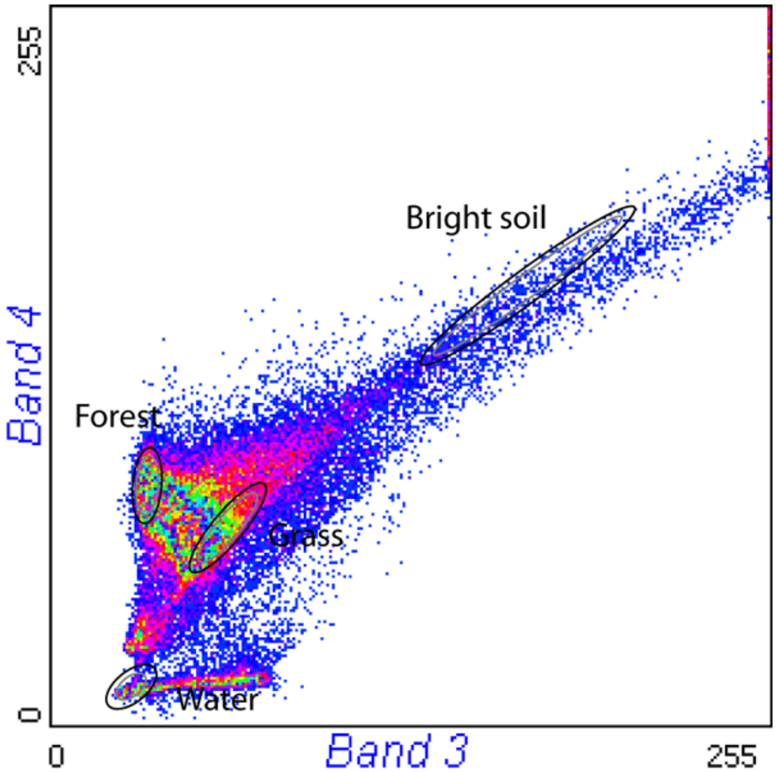
4.3. Parallelepiped

Clicking on the **Parallelepiped** radio button the ellipse will change to a rectangle. The dimensions of the rectangle are defined by the range of values in the two bands displayed in the feature space plot. Clicking the up or down arrows or typing in a new rectangle size value will change the size of the rectangle. A size value of 0 corresponds to a rectangle that fully encloses the points defined by the training area. Positive size values increase the length and width (in increments of pixel values) of the rectangle and negative values decrease the size.

5. Classifying the image

In this interactive you can work with a single training area at a time. To Classify an entire image you would select several training areas, usually a few for each class you want to classify.

In image classification the statistics from the training areas and an algorithm are used to determine which training area a pixel, in feature space, is closest to. Each pixel in the image is compared to all of the training areas that have been defined. The pixel's class assignment is determined by the “closest” training area. For example, the figure below shows some objects that have been defined on a feature space plot.



In this figure four objects are defined based on four training areas that were defined on the image to represent; water, grass, forest, and bright soil. The ellipses represent the pixels assigned to each object area using a maximum likelihood classifier with 1 standard deviation (gray) and 3 standard deviations (black) from the mean. If we want to classify all of the pixels in the image using only these four training areas then each of the points in the feature space plot must be assigned to one of the four classes based on distance. Since we used a maximum likelihood classifier, distance is measured as standard deviations from the mean. Clearly there are a lot of gaps in the feature space plot that should be filled in with additional training sites. In most classification projects dozens of training areas are defined to eliminate the large gaps between objects in feature space.