Supervised Classification

Thematic Mapping with ERDAS Imagine

Jessica Garrett

CVEN 5389; Advanced Remote Sensing

Spring 2016

1. Introduction

Thematic mapping is a data reduction process defined by image classification where each pixel is assigned to some *class* or *theme.* Given a multispectral image, data is transformed from a numerical set to a categorical, qualitative description. The final output image then describes land cover use for an area of interest. Common classes include urban/built-up land, agricultural land, water, forestland, etc. These classes could be further divided into subcategories as water could be defined as rivers, lakes, reservoirs, etc. See Appendix A for the standard Anderson classification scheme which establishes a hierarchal flow for thematic labeling and mapping.

2. Overview of Classification

The process of image classification is generally divided into two steps. The first involves training a classifier to identify various themes (also referred to as *signatures*) based on pixel characteristics. After the classifier has been trained, each pixel within the entire image is then labeled. There is also an optional pre-step where the multispectral image is first transformed to a *feature image.* One common technique seen here is the *Principle Component Transformation* as dimensionally is reduced along with computational costs. Such transformations can also isolate unwanted noise associated with atmospheric or topographic effects, or other uncorrelated features in an image. These problematic features can then be removed from the dataset.

Classification can be executed via *supervised* or *unsupervised.* Under supervised classification, the process of training the classifier is guided by the user. In other words, the analyst selects a small subset of features and then manually classifies them accordingly. The pixel-to-class assignment is then computer generated via algorithm of choice (maximum likelihood, nearest neighbor, nearest mean, etc.). Unsupervised classification, however, is a “blind” process in the sense that grouping of clusters or features are computer generated initial based on pixel properties, such as DN values. The user then assigns labels to the classifications produced by a select algorithm, such as the K-Means clustering algorithm.

Lastly, it is worthy to note that classification algorithms fall under one of two categories. *Parametric* methods assume a common statistical distribution within each class. A convenient model is the Gaussian distribution, for example. The maximum-likelihood algorithm mentioned earlier is a commonly used parametric algorithm which establishes class boundaries based on a specific probability spread. In contrast, *nonparametric* algorithms carry no assumptions regarding probability. The use of a different algorithm will likely alter the resulting classification map.

3. Data and Software

We will now divert our focus to a specific multi-image and its land type classification. Gathered from Colorado View, our image of concern is a three band TIF image of the north east portion of Denver. Using the remote sensing software ERDAS Imagine, we will create and analyze a thematic map using a non-parametric supervised classification method.

3.1 Priming Data

The Colorado Landsat tile reference map divides the state into, overlapping, parcels. The area of interest is just a small portion of Denver County. Provided by Denver Open Data, a counties shape file is placed over the TIFF image. We can see that Denver is located within the bottom left corner. Using the Under the Raster Tab, we can create image subset; i.e. a clip of a particular region. Note that all images represented in a 3-2-1 band viewing.

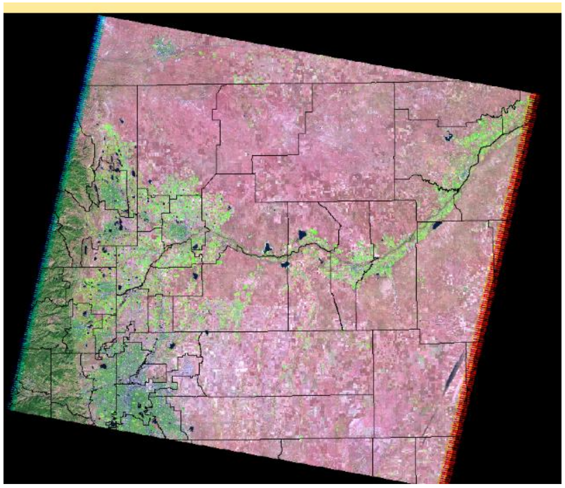

Figure 1

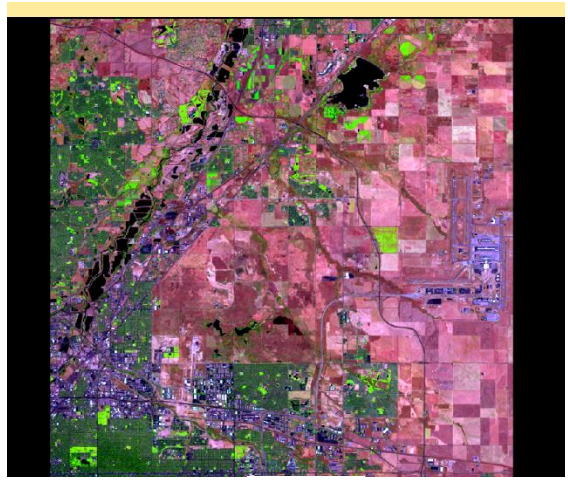

Figure 2

4. Supervised Classification

4.1. Defining Signatures

Now that we have the desired image, the initial step is to train the classifier by selecting signatures. We can see that the bottom left corner is north Denver and just north of that is a more industrial area (Commerce City). DIA is in the middle right surrounded by agricultural/open lands. 13 signatures were defined, such as urban, industrial, agricultural, open spaces and water. Collection methods included digitalized polygon, neighborhood and feature space collection.

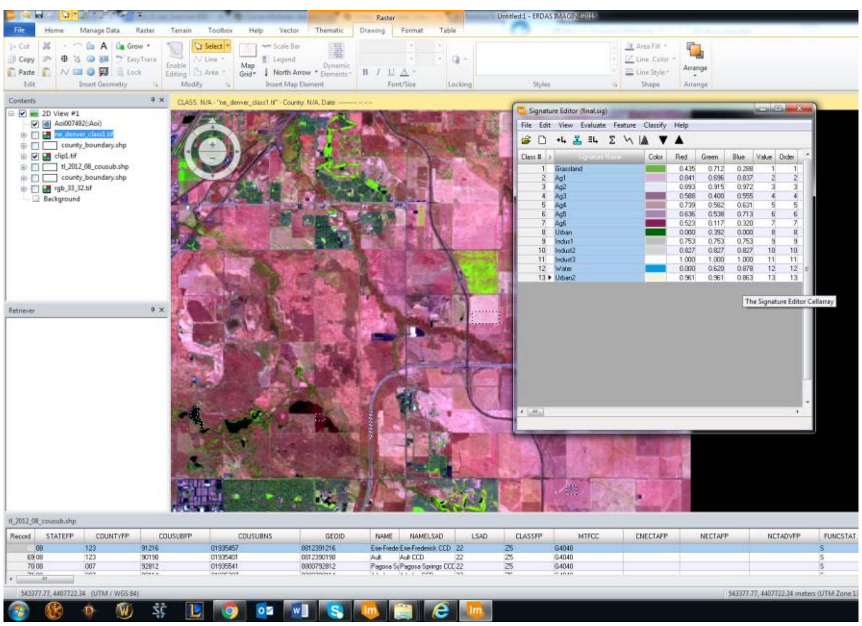

Figure 3

4.2. Evaluate Signatures

After the signatures have been established for the classifier, it is important to evaluate potential accurately of theses classifications. There are numerous ways to assess the signature set. We will look at the error matrix, ellipses, histograms as well as the degree of separabliliy between classes.

4.2.1. Error Matrix

It is unlikely that every pixel within the AOI will be assigned to a class it was trained. We can generate a contingency or error matrix that gives light to how well our sample signatures train the classifier for the entire image. The matrix shows how many pixels were assigned to each signature based on training samples (reference data). Ideally, the data would be take the form of a diagonal matrix as the training set would fully map to the classified. We can see this with Industrial 1, Industrial 2 and the Agriculture 6 classes. In contrast, the training sample for Agriculture 1 generated classes outside of the data it trained. Data was mapped to other agricultural classes, however.

ERROR MATRIX														

				Reference Data										

Classified														
Data	GrsInd	Ag1	Ag2	Ag3	Ag4	Ag5	Ag6	Urban	Indus1	Indust2	Indust3	Water	Urban2	Row Total
----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------
GrsInd	100	0	0	0	0	0	0	5	0	0	0	0	0	105
Ag1	0	1377	7	1	13	3	0	0	0	0	0	0	0	1401
Ag2	0	10	468	0	0	0	0	0	0	0	0	0	0	478
Ag3	0	17	0	598	4	6	0	0	0	0	0	0	0	625
Ag4	0	36	20	8	557	0	0	0	0	0	0	0	1	622
Ag5	0	14	0	39	0	120	0	0	0	0	0	0	0	173
Ag6	0	0	0	0	0	0	108	0	0	0	0	0	0	108
Urban	0	0	0	0	0	0	0	675	0	1	0	3	0	679
Indus1	0	0	0	0	0	0	0	0	24	0	0	0	0	24
Indust2	0	0	0	0	0	0	0	2	0	35	0	1	0	38
Indust3	0	0	0	0	0	1	0	0	0	0	25	0	1	27
Water	0	0	0	0	0	0	0	0	0	0	0	173	0	173
Urban2	0	4	0	0	26	2	0	0	0	0	0	0	118	150
Col Total	100	1458	495	646	600	132	108	682	24	36	25	177	120	4603
				----- End of Error Matrix -----										

Table 1

4.2.2. Plot Ellipses over Feature Space

Here, we can see a pairwise representation of all bands. Plotting ellipses for each class over the feature space provides more information about where particular classes fall within the spectrum. Note that each ellipse is defined by both the mean and standard deviation of the class that it represents. Looking at bands one and two, there is minimal overlapping of the ellipses, indicating high separability and a clear distinction among signatures.

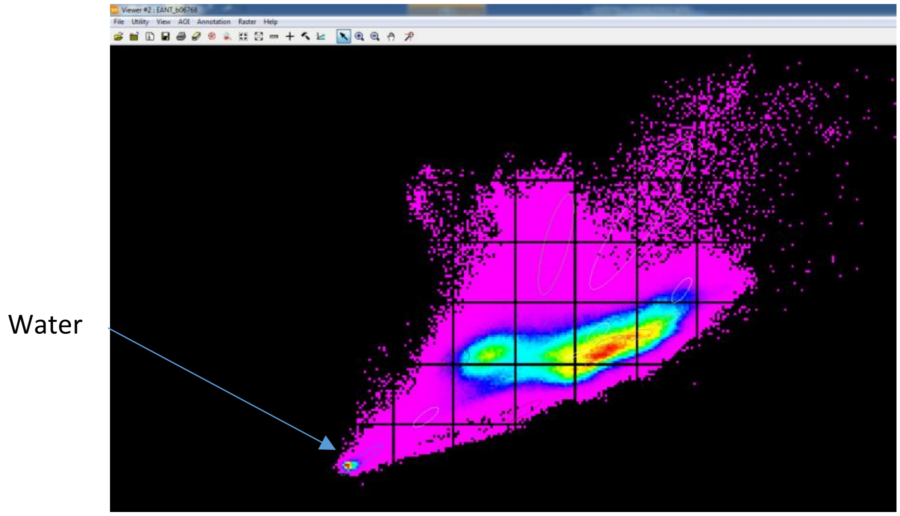

Figure 4

4.2.3. Histograms

Histograms show, although in a discrete domain, a general idea of data distribution of a class. By studying the spread of each class, we can evaluate whether the class was established properly. If there appears to be more than one distribution within a single signature, we would likely consider dividing the class into further signatures. Looking at all 13 signatures for each band, most tend to follow a single distribution (there is only one curve present). There were a few exceptions, however. The class "grassland" and "industry1" do not follow a single distribution. All of the agricultural signatures, however, displayed a very distinct single distribution.

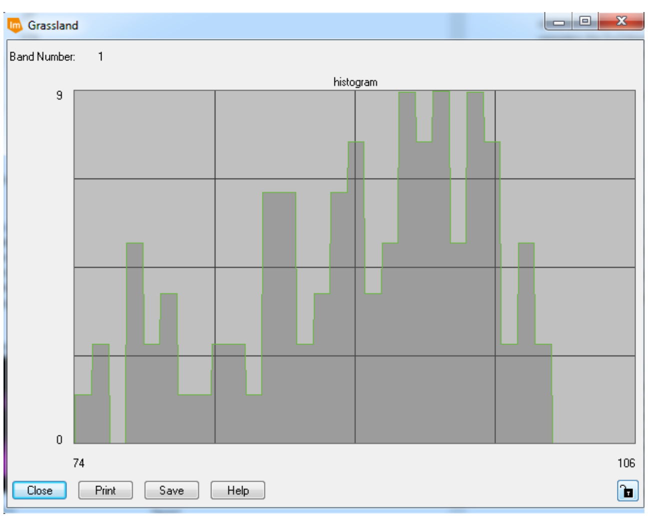

Figure 5: Grasslands

Figure 6: Industry 1

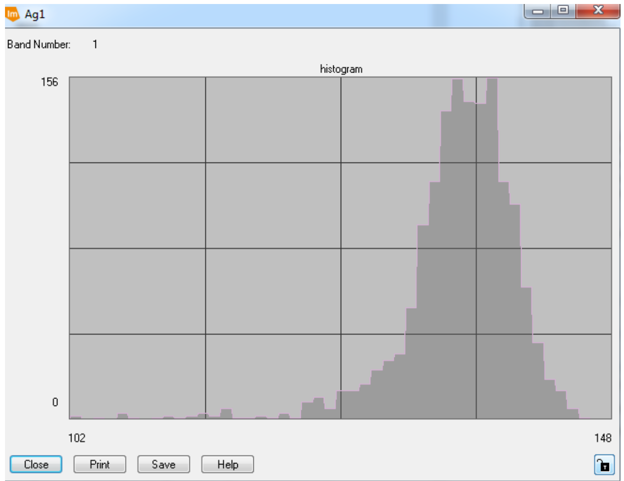

Figure 7: Agriculture 1

4.2.4. Separability

Another analysis tool to consider is the separability report. Here, we compute the statistical distances between each class. Mathematically, there are many ways to define distance. ERDAS Imagine considers the Euclidean Distance between means, divergence (based on likelihood ratios), transformed divergence or the Jeffries-Matusita distance. Table 2 shows the pairwise separability between all combinations of signatures. The overall matrix shows high separability among all classes. Agriculture 1 and Agriculture 2 show a slightly lower measure of separability, yet is still very acceptable. In general, the agricultural classes displayed a slightly lower separability distance among themselves compared to other class combinations within the matrix.

Separability CellArray

Distance Measure: Jefferies-Matusita
Using Layers: 1 2 3
Taken 3 at a time
Best Average Separability: 1406.32
Combination: 1 2 3

Signature Name	1	2	3	4	5	6	7	8	9	10	11	12	13
Grassland 1	0	1414.21	1414.21	1414.07	1414.2	1413.99	1414.21	1412.8	1410.17	1414.21	1411.5	1414.21	1413.12
Ag1 2	1414.21	0	1294.73	1407.24	1337.47	1361.02	1414.21	1414.21	1414.21	1414.21	1414.18	1414.21	1412.42
Ag2 3	1414.21	1294.73	0	1414.11	1399.43	1412.85	1414.21	1414.21	1414.21	1414.21	1414.21	1414.21	1414.17
Ag3 4	1414.07	1407.24	1414.11	0	1392.77	1299.24	1413.52	1414.21	1414.17	1414.21	1412.39	1414.21	1393.42
Ag4 5	1414.2	1337.47	1399.43	1392.77	0	1391.97	1414.21	1414.21	1414.21	1414.21	1414.21	1414.21	1376.87
Ag5 6	1413.99	1361.02	1412.85	1299.24	1391.97	0	1414.21	1414.21	1414.12	1414.21	1391.94	1414.21	1333.67
Ag6 7	1414.21	1414.21	1414.21	1413.52	1414.21	1414.21	0	1414.21	1413.81	1414.21	1414.21	1414.21	1414.15
Urban 8	1412.8	1414.21	1414.21	1414.21	1414.21	1414.21	1414.21	0	1414.21	1412.87	1414.21	1413.72	1414.21
Indus1 9	1410.17	1414.21	1414.21	1414.17	1414.21	1414.12	1413.81	1414.21	0	1414.02	1409.29	1414.21	1413.95
Indust2 10	1414.21	1414.21	1414.21	1414.21	1414.21	1414.21	1414.21	1412.87	1414.02	0	1414.21	1413.61	1414.21
Indust3 11	1411.5	1414.18	1414.21	1412.39	1414.21	1391.94	1414.21	1414.21	1409.29	1414.21	0	1414.21	1413.41
Water 12	1414.21	1414.21	1414.21	1414.21	1414.21	1414.21	1414.21	1413.72	1414.21	1413.61	1414.21	0	1414.21
Urban2 13	1413.12	1412.42	1414.17	1393.42	1376.87	1333.67	1414.15	1414.21	1413.95	1414.21	1413.41	1414.21	0

Table 2

4.3. Results

See figure 8 for the resulting non-parametric classification map of north east Denver and surrounding suburbs.

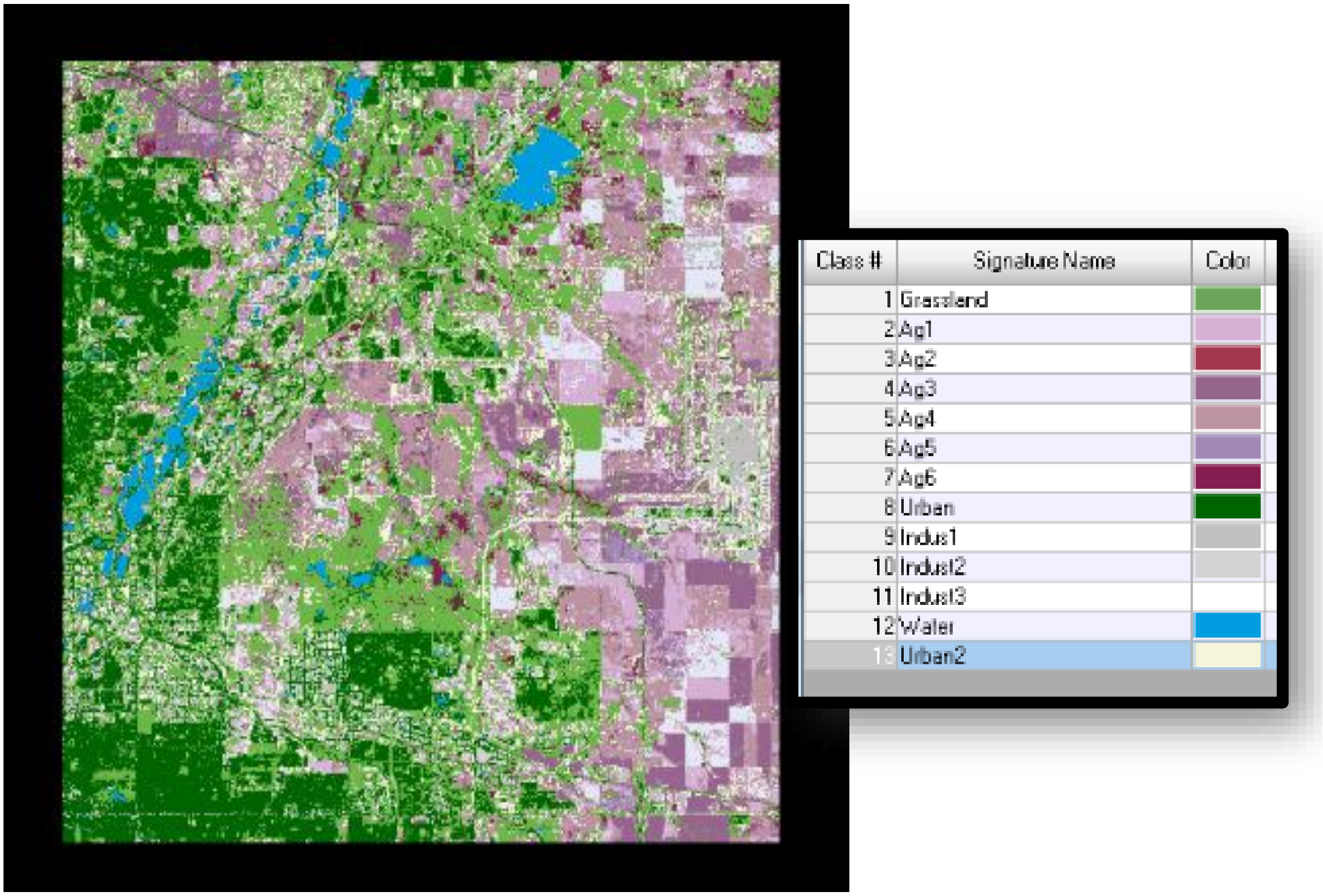

Figure 8: Supervised Classification

5. Conclusions

We saw through image classification the type of land cover for north east Denver. The top right portion of the map largely consisted of agricultural lands, while the land closer to Denver was classified as urban industrial. We can also see water coverages, like the South Platte River as well as the lake in city park located in the bottom right corner. DIA can clearly by identified as an “industrial class”. Considering the results of various evaluation analysis, the resulting thematic map can be deemed as acceptable.

6. Works Cited

"Colorado Data." *Colorado View*. N.p., n.d. Web. 12 May 2016.

"Denver Open Data Catalog." *Denver Open Data Catalog*. N.p., n.d. Web. 12 May 2016.

Schowengerdt, Robert A. Remote Sensing: Models and Methods for Image Processing. Academic Press, 2007. Print.

7. Appendix A

The Anderson classification scheme was first developed in 1976 to categorize land-coverage and usage. It established is a much needed standard for land classification and thematic labeling. By grouping land data with similar characteristics into *class signatures*, we can build a hierarchal scheme ranging from Level I to Level IV, where detail increases with level. Level I usually consists of LANDSAT data, Level II can be described with high-altitude data (having a scale less than 1:80,000), Level III is medium-altitude data (1:20,000 – 1:80,000) and Level IV typically represents low-altitude data (scale more than 1:20,000). For example, there are nine classifications of Level I data. See table below for complete listings of Level I and Level II categories.

GENERAL CATEGORIES	SECONDARY CATEGORIES
1. Urban or built-up land	11. Residential 12. Commercial Services 13. Industrial 14. Transportation, Communications 15. Industrial and Commercial 16. Mixed Urban or Built-Up Land 17. Other Urban or Built-Up Land
2. Agricultural land	21. Cropland and Pasture 22. Orchards, Groves, Vineyards, Nurseries 23. Confined Feeding Operations 24. Other Agricultural Land
3. Rangeland	31. Herbaceous Rangeland 32. Shrub and Brush Rangeland 33. Mixed Rangeland
4. Forest land	41. Deciduous Forest Land 42. Evergreen Forest Land 43. Mixed Forest Land
5. Water	51. Streams and Canals 52. Lakes 53. Reservoirs 54. Bays and Estuaries
6. Wetland	61. Forested Wetlands 62. Non-forested Wetlands
7. Barren land	71. Dry Salt Flats 72. Beaches 73. Sandy Areas Other than Beaches 74. Bare Exposed Rock 75. Strip Mines, Quarries, and Gravel Pits 76. Transitional Areas 77. Mixed Barren Land
8.Tundra	81. Shrub and Brush Tundra 82. Herbaceous Tundra 83. Bare Ground 84. Wet Tundra 85. Mixed Tundra
9.Perennial snow and ice	91. Perennial Snowfields 92. Glaciers

https://www.teachengineering.org/collection/cub_/lessons/cub_images/cub_environ_lesson08_fig2.jpg