Topic 09

Basic Analysis Concepts for Digital Imagery

(overhead)

Classification is the process by which we **partition image data** into a set number of groups (classes) based on the values of the pixels in one or more image channels. In simplest terms, we can group pixels based on values or ranges in a single spectral channel of an image (**density slice**). More complex classification techniques rely on calculation of statistical **probabilities of membership** of pixels in different classes. It is this latter subject area that will form the bulk of our discussions in this class.

Errors - the two errors we must be cognizant of in classification are those of omission and commission. **Omission errors** occur when an area not assigned to the correct class (omitted from the class). **Commission errors** are just the opposite; pixels are assigned to the wrong class.

You may also hear of errors referred to as **producer** versus **user** accuracy. Consider the example given by Story and Congalton. An area is classified into three land-use types; forest, water, urban. Samples are drawn to assess the accuracy and an error matrix (contingency table) is produced**.** **(overhead)**

	Reference Data			
	Forest	Water	Urban	Row Total
Forest	28	14	15	57
Water	1	15	5	21
Urban	1	1	20	22
	30	30	40	100

Producer vs. User Accuracy

<u>Columns</u>: The producer of the map would use the number of correct cells divided by the total reference cells (column total) to generate a figure of map accuracy.

<u>Rows</u>: A user of the data would take the map and go to an area assuming it was the cover type labeled on the map (the rows).

The samples not correctly classified as the reference were omitted (producer's accuracy) while the samples wrongly placed into a category from the field perspective were committed (user's accuracy). Calculation of the two types of accuracy give very different answers.

<u>Producer's Accuracy</u>	**<u>User''s Accuracy</u>**
F = 28/30 = 93%	F = 28/57 = 49%
W = 15/30 = 50%	W = 15/21 = 71%
U = 20/40 = 50%	U = 20/22 = 91%

A general method of accuracy reporting is to give the **overall**; the sum of the diagonals divided by the total sample, in this case **63/100 = 63%.** But some may give the **average** of one or the other **193/3 = 64% (producer's), 211/3 = 70% (user's).** Be aware of biases in accuracy reporting because large classes with high accuracy can overshadow smaller but more important classes in the overall accuracy figure. We'll revisit these concepts when we talk about field data collection.

Signatures - classification techniques all seek to develop a statistical definition of each class called a **signature**. **Signatures** are composed of the **mean spectral value** from each channel and the associated **standard deviation** (variance: defines the spread of values about the mean). Signatures are used to determine **class membership** by labeling each pixel as the class it most closely resembles. This is generally based on a calculation of **probability of membership**.

(overhead examples of two class distributions)

(overhead)

Classification procedures - fall into different categories based on the amount of input the user has to development of class definitions: **unsupervised**, **supervised**, **hybrids, and expert systems**. A common starting point is to use unsupervised approaches to explore the intrinsic variability in a data set. These processes are particularly useful in initial analysis of data for which the user has little familiarity.

<u>**Unsupervised classification algorithms**</u>

(overhead)

Clustering - The large size and complexity of remotely sensed images, coupled with lack of knowledge of overall partitionable variance in the imagery, make the procedures desirable for a first-inspection classification. The results may be then used to guide more in-depth analysis (**supervised classification or "guided clustering"**).

Clustering algorithms can perform in a variety of ways. The most common is to identify **natural groupings** in the data by partitioning in to **statistically homogeneous groups**. **Statistical homogeneity** means: having a tight distribution about a mean with the distribution being unimodal for all channels. These procedures try to minimize the variance of each cluster class and maximize the differences between clusters.

(overhead)

The most common clustering algorithm is based on the **modified k-means** algorithm that is called a **migrating mean procedure (isodata).** This is an **iterative procedure** that first identifies (or is **seeded**) with a set of **starting means** (think of them as points in feature space). Every pixel is compared to every mean location and is placed in the class that it resembles. Then the means are recalculated based on these groupings and the procedure is repeated. Thus, the means shift with each iteration until:

1) the number of specified iterations has been complete or
2) the amount of shift of all means is very small.

(overhead)

Some of the basic clustering parameters include: **# of classes, seed locations, iterations, convergence value, and max class variance.**

(overhead)

Spatial operators look are designed to identify spatially homogeneous areas (regions of contiguous pixels with similar values) as separate possible classes. This approach is used operationally in the national **forest inventory of Finland**. The idea is to find contiguous homogeneous regions (stands) in satellite imagery then use field plots to define the forest inventory information for each region.

Supervised classification **(overhead)**

In supervised classification processes, the **user supplies the information** to train the pattern recognition algorithm. This is accomplished by selecting representative areas of each class that the user wishes to identify. Steps in supervised classification include:

1. Select **training areas** for all cover classes in the image (in Imagine we will do this with AOI's).

2. Generate **class signatures** (statistical definitions) based on the pixels contained in each training area.

3. Compare the **class signatures** to determine if the classes are separable (use example of single distribution).

4. Based on the signature comparisons, delete, combine, expand, or split areas into meaningful classes that can be detected and are also useful in the application.

5. Run Classification

6. Evaluate and if needed re-run based on changes to training data.

The user may go through several iterations of refining signatures prior to actually running the classification on the entire data set. The next step is to set up and run the classification. The three most common techniques are: **parallelepiped, minimum distance (Euclidean), and maximum likelihood.** **(overhead)**

Parallelepiped (overhead)- this classifier is actually like using a density slice across several channels. Rather than the mean values, the ranges of values define a **class "volume**." All pixels that fall in this volume are assigned to that class. This algorithm seldom provides adequate results on its own because there is often **overlap in the class definitions**. This technique is often **used with the other two algorithms** in combination to provide the first cut at classifying all pixels in volumes that do not overlap. The areas of overlap are then resolved by use of the minimum distance or maximum likelihood processes.

(overhead)

Minimum Distance - this is based on Euclidean distance to mean class values and generally produces fairly good results. Problems arise when class boundaries vary in size due to different possible class variance. In these cases, pixels may be placed in the wrong class because of their proximity to the class mean in the decision space.

(overheads)

Maximum Likelihood - normally the Imagine user has the option to classify all pixels that do not have overlapping class decision boundaries based on parallelepiped technique then use maximum likelihood to classify remainder. This is done to save on processing times for large data sets.

Parameters that are set for the maximum likelihood procedure include: **a priori and threshold**. **A priori** defines to the algorithm which class to favor if there is a classification decision tie. That is, in order to **maximize the likelihood** of correct classification, favor classes that are known to be more frequently occurring in the study area.

(overhead)

Classification thresholds allow the user to pre-set the constraints on the decision boundary. This is normally accomplished in two phases. First, the user instructs the software to generate a **probabilities layer** that indicates the classification probability of each pixel. Next, the user can use this layer to threshold the resulting classification to mask out areas that have a specified uncertainty of mis-classification. **(overhead)**

<u>**Hybrid classifiers**</u>

Guided clustering - This type of clustering takes advantage of some user interaction to define areas to be clustered. One method of guided clustering is to utilize **user-defined training areas** of known composition and perform cluster analysis within these areas. This will sometimes result in spectral class (or subclass) definitions that could not otherwise be obtained due to time constraints or the complexity of the input image.

A variation on guided clustering is to perform a very basic classification of broad cover types: (forest, water, bare, etc.) then utilize these **regions as strata** in which to perform clustering to extract additional spectral subclasses of the broader features.

Region classification - This classification technique utilizes a theme map as a mask to define areas to characterize. Class statistics are produced for each distinct region in the theme map and these are used to then classify the imagery.

<u>**Classification Strategies**</u>

The image analyst has to make many decisions before a strategy can be selected to achieve a particular classification goal. The goal is key; in this case, it is the classes themselves.

(overhead)

Class definition - There are two basic types of classes that one would derive from image data: 1) **spectral** and 2) **information**. These types often overlap but just as often are mutually exclusive and even at odds with the final project objectives.

> **Information classes** are usually what we think of for the classification process. Most analysts with a mapping background have a tendency to form class definitions based on the utility of the land for various uses (i.e. **land use classes**). Examples of these are: **urban, crop land, forest land, range land**, etc. These are, in part, the classes associated with human influence on the landscape. Natural resource tend to think in terms of the vegetative cover (or lack of). Examples of these include: **forest, grassland, water, bare soil, scrub**.
>
> **Spectral classes**, on the other hand, are based on the internal statistical structure of the data. In the best case, spectral classes partition the variance of a total data set that is **normally distributed**. This is the assumption made: each channel of data follows an approximate normal distribution. This may be true for areas that are uniform or change at a predictable rate. But major changes in the image data (i.e. land to large water bodies in the NIR) will give a **multi-modal distribution**. Attempts to combine and process multiple data sets (**multi-temporal**) can also create a confounding effect.

(overhead)

Selecting classes - The analyst should consider a number of things in order to optimize the chances for success in developing a classification that achieves the project objectives.

1. Select imagery at a time of year that is appropriate to the classes
2. Define information classes that prove useful and compatible with spectral ones
3. Define classes that are separable and perhaps repeatable for subsequent years
4. Select data sets and classes appropriate to classifiers.

(overhead)

Classification optimization - a number of techniques can be used to improve on classification results by improving on class separation.

1. area stratification for sampling and classification (masks)
2. use of multiple data sets for temporal analysis
3. use of different data sets for different regions (splice in best results)
4. inclusion of spatial operators (texture)
5. use ancillary data (either pre- or post-classification stratification or enhancement)
6. use context (adjacency operators)

(overhead)

Other considerations involve the reduction or optimization of costs (both monetary and time).

1. Use only the essential data (data sets and/or area of interest).
2. Simple algorithms will reduce the total processing time
3. sampling techniques can improve efficiencies in estimation (systematic and/or stratified)

Field data concerns - the analyst should also take into account the field data that will have to be utilized both to train and verify the classification in the context of project goals. Sampling is generally in direct proportion to the coverage of each class.

(overhead)

Classification strategies - there are numerous possibilities in developing a classification.

1. **Unfamiliar area** - sometimes the analyst uses a clustering strategy to identify preliminary cluster classes as an exploratory measure. These cluster classes may for the basis for more detailed analysis or serve as the final product.

2. **Sequential classification** - this may be based on knowledge of phenological sequences that can be identified in multidate imagery. Examples are the different green-up and senescence periods for vegetation.

3**. Hierarchical** - this is probably the most frequently used technique. The data set is first partitioned into several **broad but distinct land-cover classes**. These are in turn used to stratify the data set for further analysis through either more training or clustering or combinations of procedures.

Sometimes modifications are made to the above used in concert. For example, the analyst may derive a general map of land cover then subject the forest class to cluster analysis to attempt to identify subclasses that may be related to species or canopy characteristics.

<u>Estimation</u> **(overhead)**

Estimation is the process by which we transform remote sensing information to estimates of specific resource parameters. This is accomplished by various modeling techniques. The general steps are:

1. selection and calibration of imagery classification
2. conversion of classification based on correlation to physical parameters
3. implementation of model based on class correlations

Examples of estimate procedures: **(overhead)**

1. LAI in ecological work
2. Biomass
3. future condition models (based on current condition and ancillary data such as climate, soils etc.)