CHAPTER 7

7 CLASSIFICATION TECHNIQUES FOR REMOTE SENSING DATA

When satellite images are taken to get useful information from it different classification techniques has to be used. This classification require the need of either a photo interpreter (a person who because of its experience with dealing with satellite images can be able to interpret the information from the images) or use of qualitative technique which requires the use of computer software to carry out the classification task. This classification technique either uses clustering algorithm to divide the image into different classes or use more efficient method which combine the knowledge of the analyst and computer technology into one method of supervised classification. Both the unsupervised and supervised classification techniques will be explained in detail in the following report along with their limitations and selection.

Since photo interpretation is based upon the ability of the human analyst-interpreter to assimilate the available data, only three or so of the complete set of spectral components of an image can be used readily. Yet there are seven bands in ‘Landsat thematic mapper’ data and many for imaging spectrometer data. It is not that all of these would necessarily need to be used in the identification of a pixel; rather, should they all require consideration or evaluation, then the photo-interpretive approach is clearly limited [309].

7.1 Existing Satellite Image Classification Methods:

There are two types of Image classification methods available.

I. Unsupervised classification

Unsupervised classification is a means by which pixels in an image are assigned to spectral classes without the user having foreknowledge of the existence or names of those classes. It is performed most often using clustering methods. These procedures can be used to determine the number and location of the spectral classes into which
the data falls and to determine the spectral class of each pixel. The analyst then identifies those classes afterwards by associating a sample of pixels in each class with available reference data, which could include maps and information from ground visits. Clustering procedures are generally computationally expensive yet they are central to the analysis of remote sensing imagery. While the information classes for a particular exercise are known, the analyst is usually totally unaware of the spectral classes, or sub-classes as they are sometimes called. Unsupervised classification is therefore useful for determining the spectral class composition of the data prior to detailed analysis by the methods of supervised classification. There are basically two types of methods available under unsupervised classification.

a. K-Means Classifier  
b. ISO Data Classifier

### 7.1.1 K-Means Classifier

In statistics and data mining, k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean [303].

It is an iterative procedure.

- In first step, it assigns an arbitrary initial cluster vector.
- The second step classifications each pixel to the closest cluster.
- In the third step the new cluster mean vectors are calculated based on all the pixels in one cluster.
- The second and third steps are repeated until the "change" between the iteration is small. The "change" can be defined in several different ways; either by measuring the distances the mean cluster vector has changed from one iteration to another or by the percentage of pixels that have changed between iterations.
- The objective of the k-means algorithm is to minimize the within cluster variability. The objective function (which is to be minimized) is:
Sums of squares distances (errors) between each pixel and its assigned cluster centre.

\[ SS_{\text{distances}} = \sum_{x} [x - C(x)]^2 \]

Where, \( C(x) \) is the mean of the cluster that pixel \( x \) is assigned to.

Minimizing the \( SS_{\text{distances}} \) is equivalent to minimizing the Mean Squared Error (MSE). The MSE is a measure of the within cluster variability.

\[ MSE = \frac{\sum_{x} [x - C(x)]^2}{(N - c)b} = \frac{SS_{\text{distances}}}{(N - c)b} \]

It is similar to the expectation-maximization algorithm for mixtures of Gaussians in that they both attempt to find the centers of natural clusters in the data as well as in the iterative refinement approach employed by both algorithms.

Input: Data & number of clusters (K)

Figure 7.1 (a): Randomly guess locations of K cluster centers
Figure 7.1 (b): For each center – assign nearest cluster

Figure 7.1 (c): Repeat till convergence

Figure 7.1 (d): Repeat till convergence

Figure 7.1 (e): Classified clusters
**K-Means Classifier: Merits**
Process is simple hence fast
Works good for image clusters having same variance

**K-Means Classifier: Demerits**
Depends on initialization
Need to specify K
Sensitive to outliers
Sensitive to variations in sizes and densities of clusters
Not suitable for non-convex shapes

Does not apply directly to categorical data

### 7.1.2 ISO Data (Iterative Self-Organizing) Classifier

ISO Data stands for *Iterative Self-Organizing Data Analysis Techniques*. This algorithm allows the number of clusters to be automatically adjusted during the iteration by merging similar clusters and splitting clusters [310].

**Clusters are merged** if either the number of members (pixel) in a cluster is less than a certain threshold or if the centres of two clusters are closer than a certain threshold.

**Clusters are split** into two different clusters if the cluster standard deviation exceeds a predefined value and the number of members (pixels) is twice the threshold for the minimum number of members. [334]

The ISODATA algorithm is similar to the k-means algorithm with the distinct difference that the ISODATA algorithm allows for different number of clusters while the k-means assumes that the number of clusters is known a priori.

- Each pixel is compared to each cluster mean and assigned to the cluster whose mean is closest in Euclidean distance.
A new cluster center is computed by averaging the locations of all the pixels assigned to that cluster.

The Sum of Squared Errors (SSE) computes the cumulative squared difference (in the various bands) of each pixel from its cluster center for each cluster individually, and then sums these measures over all the clusters.

The algorithm will stop either when the # iteration threshold is reached or the max % of unchanged pixel threshold is reached.

**ISODATA Classifier: Merits**
- Clustering not geographically biased to any particular portion of the image
- Highly successful at finding inherent spectral clusters
- Results similar to a minimum-distance-to-means classifier

**ISODATA Classifier: Demerits**
- Analyst doesn’t know a priori number of spectral classes
- Number of iterations needed, can be time consuming
- Does not account for pixel spatial homogeneity
- Insensitive to variance/covariance

**II. Supervised classification**

Supervised classification is the procedure most often used for quantitative analysis of remote sensing image data. It rests upon using suitable algorithms to label the pixels in an image as representing particular ground cover types, or classes. Training site is required for each class for classification. A variety of algorithms is available for this, ranging from those based upon probability distribution models for the classes of interest to those in which the multi-spectral space is partitioned into class specific
regions using optimally located surfaces partitioned into class-specific regions using optimally located surfaces.

a. Minimum Distance Classifier  
b. Parallel Piped Classifier  
c. Maximum Likelihood Classifier  
d. Seeded Region Growing Classifier

7.1.3 Minimum Distance Classifier:

This classifier is also referred to as central classifier. This classifier is the simplest classifier.

- Analyst first computes mean of each training class.
- Next the distance (Euclidian) of each pixel from the mean is calculated.
- The pixel is assigned to that class whose distance is nearest to the mean (i.e. the Euclidian distance is minimum between the pixel and the mean).

The distance is defined as an index of similarity so that the minimum distance is identical to the maximum similarity. Figure 1 shows the concept of a minimum distance classifier.

The following distances are often used in this procedure.

![Figure 7.2: Minimum Distance classifier](image)

The minimum distance classifier is used to classify unknown image data to classes which minimize the distance between the image data and the class in multi-feature space.
Euclidian distance

In mathematics, the Euclidean distance or Euclidean metric is the "ordinary" distance between two points that one would measure with a ruler, and is given by the Pythagorean formula.

In the Euclidean plane, if \( p = (p_1, p_2) \) and \( q = (q_1, q_2) \) then the distance is given by

\[
d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}.
\]

The Euclidian distance is theoretically identical to the similarity index.

Minimum Distance Classifier: Merits

It is very easy to process

Minimum Distance Classifier: Demerits

Accuracy is less as it considers only the mean.

7.1.4 Parallel Piped Classifier:

The parallelepiped classifier (often termed multi-level slicing) divides each axis of multi-spectral feature space, as shown in an example in following figure [304].

Figure 7.3(a): Schematic concept of parallel piped classifier in three dimensional feature spaces
Figure 7.3(b): example of set of two-dimensional parallel-piped

Figure 7.3(c): parallel piped classification of correlated data showing region of inseparability.

The decision region for each class is defined on the basis of a lowest and highest value on each axis. The accuracy of classification depends on the selection of the lowest and highest values in consideration of the population statistics of each class. In the two-dimensional feature space this forms a rectangular box. We can have as many boxes as number of classes. All the data points which fall within the box are labeled to that class. In this respect, it is most important that the distribution of population of each class is well understood.

**Parallel Piped Classifier: Merits**

The parallelepiped classifier is very simple and easy to understand schematically. In addition the computing time will be a minimum, when compared with other classifiers.

**Parallel Piped Classifier: Demerits**
However the accuracy will be low especially when the distribution in feature space has dependency with oblique axes. Orthogonalization should be undertaken using principal component analysis, for example, before adopting the parallelepiped classifier.

7.1.5 Maximum Likelihood Classifier (MXL):

The maximum likelihood classifier is one of the most popular methods of classification in remote sensing, in which a pixel with the maximum likelihood is classified into the corresponding class. The likelihood $L_k$ is defined as the posterior probability of a pixel belonging to class $k$ [306].

$$L_k = \frac{P(k)*P(X/k)}{\Sigma P(i)*P(X/i)}$$

where $P(k)$ : prior probability of class $k$
$P(X/k)$ : conditional probability to observe $X$ from class $k$, or probability density function

Usually $P(k)$ are assumed to be equal to each other and

$\Sigma P(i)*P(X/i)$ is also common to all classes.

Therefore $L_k$ depends on $P(X/k)$ or the probability density function.

For mathematical reasons, a multivariate normal distribution is applied as the probability density function. In the case of normal distributions, the likelihood can be expressed as follows [307].

$$L_k(X) = \frac{1}{\sqrt{(2\pi)^n|\Sigma_k|}} e^{-\frac{1}{2}((X-X_k)^T\Sigma^{-1}(X-X_k))}$$

Where,

$n$: number of bands
$X$: image data of n bands
$L_k(X)$ : likelihood of $X$ belonging to class $k$
$X_k$ : mean vector of class $k$
$\Sigma_k$ : variance-covariance matrix of class $k$
$|\Sigma_k|$ is determinant of $\Sigma_k$
In the case where the variance-covariance matrix is symmetric, the likelihood is the same as the Euclidian distance.

Figure 7.4: concept of the maximum likelihood method.

**Maximum Likelihood Classifier: Merits**
It is an efficient method to classify objects.
The accuracy is high when large number of training sites is taken.

**Maximum Likelihood Classifier: Demerits**
- The maximum likelihood method has an advantage from the view point of probability theory, but care must be taken with respect to the following items. Sufficient ground truth data should be sampled to allow estimation of the mean vector and the variance-covariance matrix of population.
- The inverse matrix of the variance-covariance matrix becomes unstable in the case where there exists very high correlation between two bands or the ground
truth data are very homogeneous. In such cases, the number of bands should be reduced by a principal component analysis.

- When the distribution of the population does not follow the normal distribution, the maximum likelihood method cannot be applied.

### 7.1.6 Seeded Region Growing (SRG) Classifier:

Seeded region growing (SRG) that is introduced by (Adams and Bischof, 1994) is robust, rapid and free of tuning parameters. These characteristics allow implementation of a very good algorithm which could be applied to large variety of images [313].

SRG is also very attractive for semantic image segmentation by involving the high-level knowledge of image components in the seed selection procedure. However, the SRG algorithm also suffers from the problems of automatic seed generation and pixel sorting orders for labeling (Mehnert and Jackway, 1997; Fan et al., 2001a).

In conventional SRG technique the first step is to select a set of seed points as described in [315]. Seed point selection is based on some user criterion like pixels in a certain gray-level range, pixels evenly spaced on a grid, etc. An image is segmented into regions with respect to a set of q seeds. Given the set of seeds, S1, S2, ..., Sq, each step of SRG involves one additional pixel to one of the seed sets. Moreover, these initial seeds are further replaced by the centroids of these generated homogeneous regions, R1, R2, ..., Rq, by involving the additional pixels step by step. The pixels in the same region are labeled by the same symbol and the pixels in variant regions are labeled by different symbols. All these labeled pixels are called the allocated pixels, and the others are called the unallocated pixels. Let H be the set of all unallocated pixels which are adjacent to at least one of the labeled regions [316].

\[
H = \left\{ (x, y) \in \bigcup_{i=1}^{q} \left[ R_i \right] \left| N(x, y) \cap \bigcup_{i=1}^{q} R_i \neq \emptyset \right. \right\}
\]

Where, N(x, y) is the second-order neighborhood of the pixel (x, y) as shown in above equation. For the unlabeled pixel (x, y) ∈ H, N(x, y) meets just one of the labeled image
region $R_i$ and define $\phi(x, y) \in \{1, 2, \ldots, q\}$ to be that index such that $N(x, y) \cap R_{\phi(x, y)} \neq \phi$. 
$\delta(x, y, R_i)$ is defined as the difference between the testing pixel at $(x, y)$ and its adjacent labeled region $R_i$. $\delta(x, y, R_i)$ is calculated as

$$
\delta(x, y, R_i) = |g(x, y) - g(X^c_i, Y^c_i)|
$$

Where $g(x, y)$ indicates the values of the three color components of the testing pixel $(x, y)$, $g(X^c_i, Y^c_i)$ represents the average values of three colors components of the homogeneous region $R_i$, with $g(X^c_i, Y^c_i)$ the centroid of $R_i$.

<table>
<thead>
<tr>
<th>$(x-1, y-1)$</th>
<th>$(x, y-1)$</th>
<th>$(x+1, y-1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(x-1, y)$</td>
<td>$(x, y)$</td>
<td>$(x+1, y)$</td>
</tr>
<tr>
<td>$(x-1, y+1)$</td>
<td>$(x, y+1)$</td>
<td>$(x+1, y+1)$</td>
</tr>
</tbody>
</table>

**Figure 7.5: Identification of pixels around the seed pixel**

If $N(x, y)$ meets two or more of the labeled regions, $\phi(x, y)$ takes a value of $i$ such that $N(x, y)$ meets $R_i$ and $\delta(x, y, R_i)$ is minimized.

$$
\phi(x, y) = \min_{(x,y) \in H} \{\delta(x, y, R_j) | j \in \{1, \ldots, q\}\}
$$

This seeded region growing procedure is repeated until all pixels in the image have been allocated to the corresponding regions. The definition of above Equation ensures that the final classification will divide the image into a set of regions as homogeneous as possible on the basis of the given constraints. SRG algorithm is robust, rapid and free of tuning parameters. It can correctly separate the regions that have the same properties we define. The concept is simple and we only need a small numbers of seed point to
represent the property we want. We can also choose multiple criteria at the same time. It performs well with respect to noise. However, SRG algorithm requires the mechanism to automate the seed selection procedure [340].

**Seeded Region Growing (SRG) Classifier: Pros**

The concept of seed point marking makes the analyst free from marking entire polygon as the seed point grows as a training signature.

**Seeded Region Growing (SRG) Classifier: Cons**

Existing Algorithms are dependent on user intervention. The problems of automatic seed generation and pixel sorting orders for labeling (Mehnert and Jackway, 1997; Fan et al., 2001a) isolate from the existing methods.

- The seed points are required to be selected by a user or a person who has the prior knowledge of the site. The person has to collect co-ordinates (Latitude & Longitude) of the feature on ground and bring these attributes with crop details at Image processing terminal. Hence this is a time consuming process.

- In Existing Algorithm, threshold is considered for Training sites generation (seed point based) and same threshold is used for image classification technique. The result can be good for photographs but degrades for satellite images having variability in terms of intensity and volume of pixels.

- SRG Technique considers the regions specified in seed points by a person. Entire image may be or may not be considered that leads to accuracy concern.