# GNR607
# Principles of Satellite Image Processing

Instructor: Prof. B. Krishna Mohan
CSRE, IIT Bombay
bkmohan@csre.iitb.ac.in

Slot 2

Lecture 27-32 Image Classification Methods

November 2020

IIT Bombay Slide 1

November 2020 Lecture 27-32 Image Classification

## Contents of the Lecture s

Image Classification

- Problem of Image Classification
- Supervised Classification Approaches - parametric
- Nonparametric Classifiers
- Feature Evaluation/selection
- Accuracy Assessment
- Unsupervised Classification

GNR607 Lecture 27-32 B. Krishna Mohan

## Concept of Image Classification

**Image classification - assigning pixels in the image to categories or classes of interest**

**Examples: built-up areas, waterbody, green vegetation, bare soil, rocky areas, cloud, shadow, …**





## Concept of Image Classification

Image classification is a process of mapping numbers to symbols

$f(x): x \rightarrow \Delta; x \in R^n, \Delta = \{c_1, c_2, \ldots, c_L\}$

Number of bands = n;

Number of classes = L

f(.) is a function assigning a pixel vector x to a single class in the set of classes $\Delta$

## Concept of Image Classification

- In order to classify a set of data into different classes or categories, the relationship between the data and the classes into which they are classified must be well understood
- To achieve this by computer, the computer must be ***trained***
- Training is key to the success of classification
- Classification techniques were originally developed out of research in ***Pattern Recognition*** field



IIT Bombay 

## Concept of Image Classification

Computer classification of remotely sensed images involves the process of the computer program ***learning*** the relationship between the data and the information classes

Important aspects of accurate classification

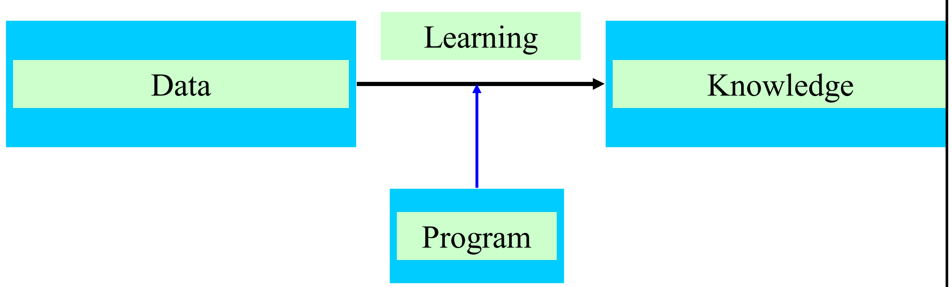
- Learning techniques
- Feature sets

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 6

# Concept of Learning

- Extract knowledge from data by learning mechanisms




GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 7

# Learning

Learning denotes changes in a system that enables it to perform the same task more efficiently the next time

e.g. Solving problems in mathematics and physics

Driving a car

Cooking!

Identifying classes of pixels given the data in different bands

GNR607 Lecture 27-32 B. Krishna Mohan

# Types of Learning

**Supervised Learning**

Learning process designed to form a mapping from one set of variables (data) to another set of variables (information classes)

A teacher is involved in the learning process

**Unsupervised learning**

Learning happens without a teacher

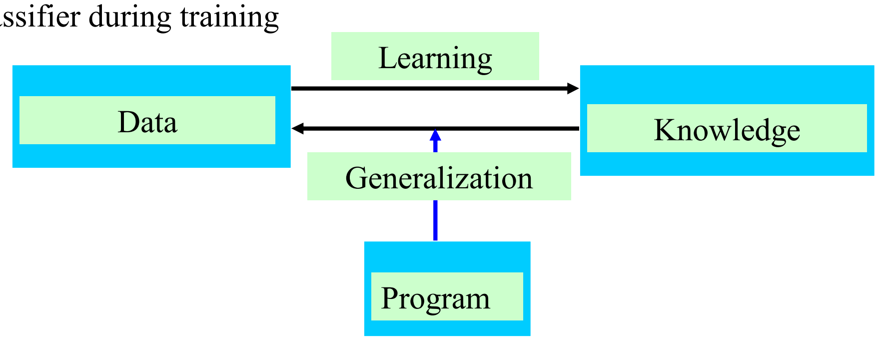
Exploration of the data space to discover the scientifc laws underlying the data distribution





# Generalization

Learning leads to recognition of unknown patterns not presented to the classifier during training



## Features

- Features are attributes of the data elements based on which the elements are assigned to various classes.
- E.g., in satellite remote sensing, the features are measurements made by sensors in different wavelengths of the eLecture romagnetic spectrum – visible/ infrared / microwave/texture features …





## Features

- In medical diagnosis, the features may be the temperature, blood pressure, lipid profile, blood sugar, and a variety of other data colLecture ed through pathological investigations
- The features may be qualitative (high, moderate, low) or quantitative.
- The classification may be presence of heart disease (positive) or absence of heart disease (negative)

## Supervised Classification

- The classifier has the advantage of an analyst or domain knowledge using which the classifier can be guided to learn the relationship between the data and the classes.
- The number of classes, prototype pixels for each class can be identified using this prior knowledge





## Unsupervised Classification

- When access to domain knowledge or the experience of an analyst is missing, the data can still be analyzed by numerical exploration, whereby the data are grouped into subsets or ***clusters*** based on statistical similarity

## Partially Supervised Classification

When prior knowledge is available

- For some classes, and not for others,
- For some dates and not for others in a multitemporal dataset,

Combination of supervised and unsupervised methods can be employed for ***partially supervised classification*** of images





## Supervised vs. Unsupervised Classifiers

Supervised classification generally performs better than unsupervised classification IF good quality training data is available

Unsupervised classifiers are used to carry out preliminary analysis of data prior to supervised classification

## Role of Image Classifier

The image classifier performs the role of a **discriminant** – discriminates one class against others

Discriminant value highest for one class, lower for other classes (**multiclass)**

Discriminant value positive for one class, negative for another class (**two class)**





## Discriminant Function

$g(c_k, \mathbf{x})$ is **discriminant function**, relating feature vector $\mathbf{x}$ and class $c_k$, k=1,…,L

*Denote $g(c_k,x)$ as $g_k(x)$ for simplicity*

**Multiclass Case**

$g_k(\mathbf{x}) > g_l(\mathbf{x}), l = 1,\dots,L, l \neq k \quad \mathbf{x} \in c_k$

**Two Class Case**

$g(\mathbf{x}) > 0\ \mathbf{x} \in c_1; \quad g(\mathbf{x}) < 0\ \mathbf{x} \in c_2$

## Example of Image Classification

**Multiple Class Case**

Recognition of characters or digits from bitmaps of scanned text

**Two Class Case**

Distinguishing between text and graphics in scanned document





## Prototype / Training Data

- Using domain knowledge (maps of the study area, experienced interpreter), small sets of sample pixels are selected for each class.
- The size and spatial distribution of the samples are important for proper representation of the total pixel population in terms of the samples

## Statistical Characterization of Classes

Each class has a conditional probability density function (pdf) denoted by p(**x** | $c_k$)

The distribution of feature vectors in each class $c_k$ is indicated by p(**x** | $c_k$)

We estimate P($c_k$ | **x**), the conditional probability of class $c_k$ given that the pixel's feature vector is **x**





## Supervised Classification Principles

- The classifier learns the characteristics of different thematic classes – forest, marshy vegetation, agricultural land, turbid water, clear water, open soils, manmade objects, desert etc.
- This happens by means of analyzing the statistics of small sets of pixels in each class that are reliably selected by a human analyst through experience or with the help of a map of the area

IIT Bombay Slide 22

## Supervised Classification Principles

- **Typical characteristics of classes**
  - Mean vector
  - Covariance matrix
  - Minimum and maximum gray levels within each band
  - Conditional probability density function $p(C_i|\mathbf{x})$ where $C_i$ is the $i^{th}$ class and $\mathbf{x}$ is the feature vector
- Number of classes L into which the image is to be classified should be specified by the user

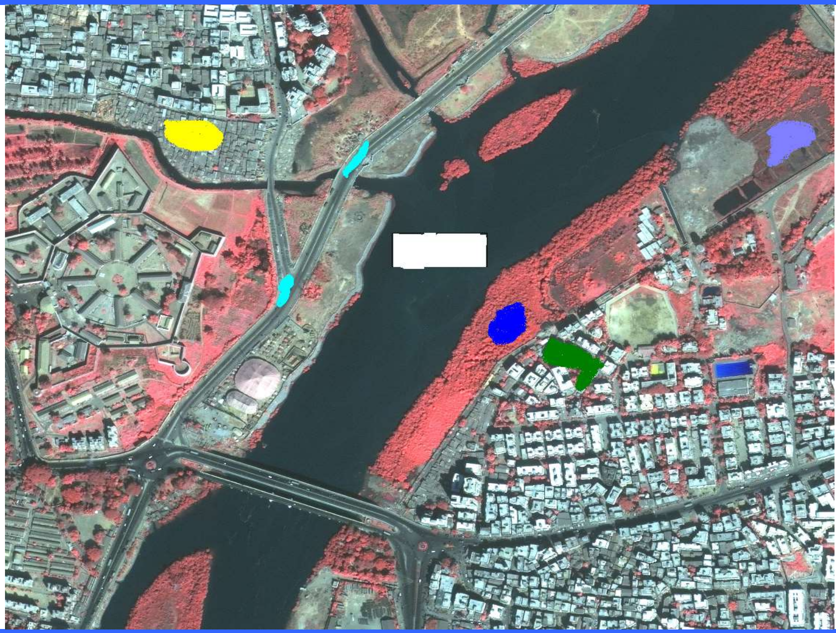
GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 23

## Prototype Pixels for Different Classes

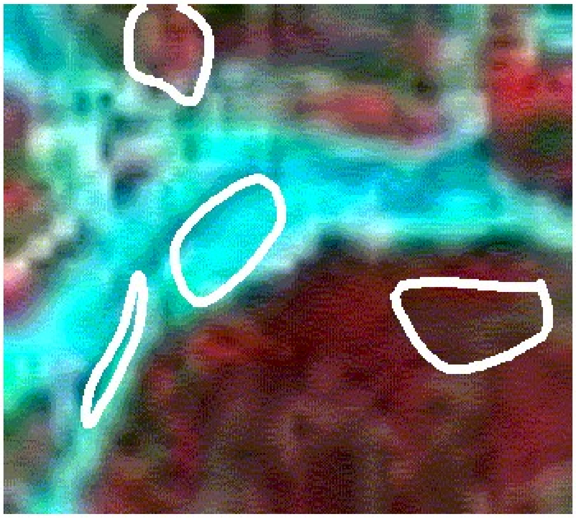
- The prototype pixels are ***samples*** of the population of pixels belonging to each class
- The size and distribution of samples are formally governed by the mathematical theory of sampling
- There are several criteria for choosing the samples belonging to different classes

GNR607 Lecture 27-32 B. Krishna Mohan









## Supervised Classification Algorithms

- There are many techniques for assigning pixels to informational classes, e.g.:
  - Minimum Distance from Mean (MDM)
  - Parallelpiped
  - Maximum Likelihood (ML)
  - Support Vector Machines (SVM)
  - Artificial Neural Networks (ANN)
  - …





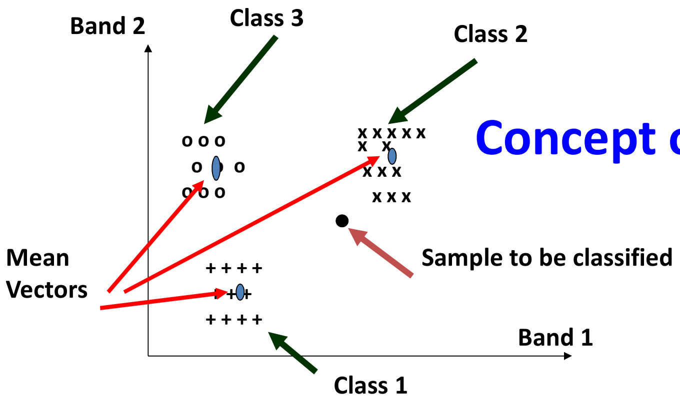
## Minimum Distance To Mean Classifier

- Simplest kind of supervised classification
- The method:
  - Calculate the mean vector for each class
  - Calculate the statistical (Euclidean) distance from each pixel to class mean vector
  - Assign each pixel to the class it is closest to

IIT Bombay Slide 28

- 2-D Feature Space

## Concept of MDM




GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 29

## Minimum Distance Classifier

- Algorithm
- Estimate class mean vector and covariance matrix from training samples
  - $m_i = \Sigma_{j \in Ci} X_j$ ; $C_i = E\{(X - m_i)(X - m_i)^T\} \mid X \in C_i\}$
  - Compute distance between X and $m_i$
  - $X \in C_i$ if $d(X, m_i) \leq d(X, m_j)\ \forall j$
  - Compute $P(C_k|X) = \dfrac{\frac{1}{d_k}}{\sum_{j=1}^{K} \frac{1}{d_j}}$ Leave X unclassified if $\max_k P(C_k|X) < T_{min}$
  -

GNR607 Lecture 27-32 B. Krishna Mohan

## Comments on MDM

- Normally classifies every pixel no matter how far it is from a class mean (still picks closest class) unless the $T_{min}$ condition is applied
- Distance between X and $m_i$ can be computed in different ways – Euclidean, Mahalanobis, city block, …

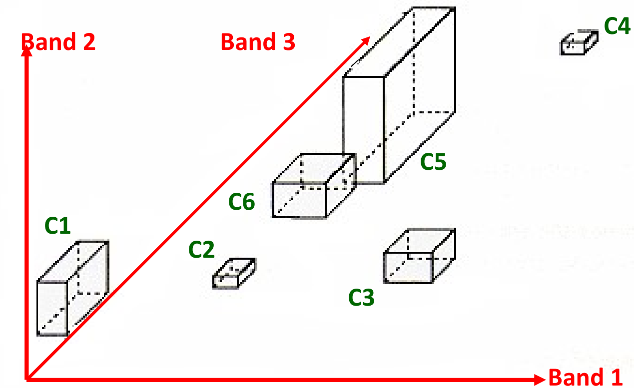




## Parallelepiped Classifier

- Assign ranges of values for each class in each band
  – Really a “feature space” classifier
  – Training data provide bounds for each feature for each class
  – Results in bounding boxes for each class
  – A pixel is assigned to a class only if its feature vector falls within the corresponding box

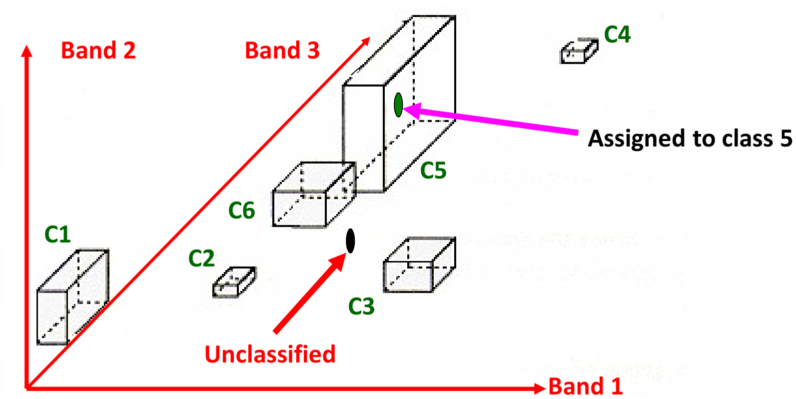
IIT Bombay Slide 32




Parallepiped Classifier

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 33




Parallepiped Classifier

GNR607 Lecture 27-32 B. Krishna Mohan

## Advantages/Disadvantages of Parallelpiped Classifier

- Does NOT assign every pixel to a class. Only the pixels that fall within ranges.
- Fastest method computationally
- Good for helping decide if you need additional classes (if there are many unclassified pixels)
- Problems when class ranges overlap—must develop rules to deal with overlap areas.





## Maximum Likelihood Classifier

- Calculates the likelihood of a pixel being in different classes conditional on the available features, and assigns the pixel to the class with the highest likelihood

## Likelihood Calculation

- The likelihood of a feature vector **x** to be in class Ci is taken as the conditional probability $P(C_i|\mathbf{x})$.
- We need to compute $P(C_i|\mathbf{x})$, that is the conditional probability of class $C_i$ given the pixel vector **x**.
- It is not possible to directly estimate the conditional probability of a class given the feature vector. Instead, it is computed indirectly in terms of the conditional probability of feature vector **x** given that it belongs to class $C_i$.





## Likelihood Calculation

$P(C_i|\mathbf{x})$ is computed using Bayes' Theorem in terms of $P(\mathbf{x}|C_i)$

$P(C_i|\mathbf{x}) = P(\mathbf{x}|C_i)\, P(C_i) / P(\mathbf{x})$

**x** is assigned to class $C_j$ such that

$P(C_j|\mathbf{x}) = \text{Max}_i\, P(C_i|\mathbf{x})$, i=1…K, the number of classes.

$P(C_i)$ is the prior probability of occurrence of class i in the image

$P(\mathbf{x})$ is the multivariate probability density function of feature **x**.

## Likelihood Calculation

- P(**x**) can be ignored in the computation of Max{P($C_i$|**x**)}
- If P(**x**|$C_j$) is not assumed to have a known distribution, then its estimation is said to be non-parametric estimation.
- If P(**x**|$C_j$) is assumed to have a known distribution, then its estimation is said to be parametric.
- The training data **x** with the class already given, can be used to estimate the conditional density function P(**x**|$C_i$)





## Likelihood Calculation

- P(**x**|$C_i$) is assumed to be multivariate Gaussian distributed in practical parametric classifiers.
- Gaussian distribution is mathematically simple to handle.
- Each class conditional density function P(**x**|$C_i$) is represented by its mean vector $\mu_i$ and covariance matrix $\Sigma_i$

$$p(\mathbf{x} \mid C_i) = \frac{1}{(2\pi)^{L/2} \mid \Sigma_i \mid^{1/2}} e^{-(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma_i^{-1} (\mathbf{x}-\boldsymbol{\mu}_i)}$$

## Assumption of Gaussian Distribution

- Each class is assumed to be multivariate normally distributed
- That implies each class has a mean $\mu_i$ that has the highest likelihood of occurrence
- The likelihood function decreases exponentially as the feature vector **x** deviates from the mean vector $\mu_i$
- The rate of decrease is governed by the class variance; Smaller the variance, steep will be the decrease, and larger the variance, slower will be the decrease.

**Since we are interested in the ordering of P($C_i$|X), taking logarithm and ordering log[P($C_i$|X)], does not change the position of the most likely class**





## Likelihood Calculation

- $$g_i(x) = -\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i) - \frac{L}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(C_i)$$
- $g_i$(**X**) = log[P($C_i$|**X**)]
- We assume that the covariance matrices for each class are different.
- The term $(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)$

  is known as the Mahalanobis distance between x and $\mu_i$ (after Prof. P.C. Mahalanobis, famous Indian statistician and founder of Indian Statistical Institute)

## Interpretation of Mahalanobis distance

- The Mahalanobis distance between two multivariate quantities x and y is

$$d_M(x,y) = (x-y)^t \Sigma^{-1}(x-y)$$

- If the covariance matrix is k.I, (I is the unit matrix) then the Mahalanobis distance reduces to a scaled version of the Euclidean distance.
- Mahalanobis distance reduces the Euclidean distance according to the extent of variation within the data, given by the covariance matrix $\Sigma$





## Advantages/Disadvantages of Maximum Likelihood Classifier

- Normally classifies every pixel no matter how far it is from a class mean
- Slowest method – more computationally intensive
- Normally distributed data assumption is not always true, in which case the results are not likely to be very accurate
- $T_{min}$ condition can be introduced into the classification rule to separately handle ambiguous feature vectors

## Computational Aspects of Supervised Classifiers

- Maximum likelihood classifier – highly computation intensive
- Storing certain terms pre-computed saves time
  - Inverse of covariance matrix
  - Matrix-vector products





## Maximum Likelihood Classifier

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i) - \frac{L}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(C_i)$$

What can be precomputed?

- Inverse of covariance matrix
- **($x - \mu_i$)** (a new image set can be computed with zero mean) that can be used in MDM, MLM
- $\ln|\Sigma_i|$, $\ln P(\omega_i)$ and $(L/2)\ln(2\pi)$

## Maximum Likelihood Classifier

- Computational cost increases nonlinearly with number of bands
- In case the data contains a large number of bands (original and derived) it is prudent to evaluate the bands for ability to separate pixels into different classes
- Fewer bands, less the size of training data needed





## Maximum Likelihood Classifier

- For N band data, covariance matrix size is proportional to $N^2$
- To determine the elements of covariance matrix, that many independent training samples are needed for each class
- Problem for high dimensional data like hyperspectral images

## Hughes Phenomenon

- Separability of classes improves with increased number of bands
- Higher number of bands requires more training samples to determine class statistics
- As number of bands increases, better class separability is nullified by need for more samples; classification accuracy actually starts decreasing
- This is known as Hughes phenomenon after the name of scientist who studied it





## Number of training samples

- For N band data, the number of training samples needed is of the order of N.(N+1)
- In practice, to deal with slightly erroneous samples and noise, the number of samples desired is of the order of 10N to 100N

## Nearest-Neighbor Classifier

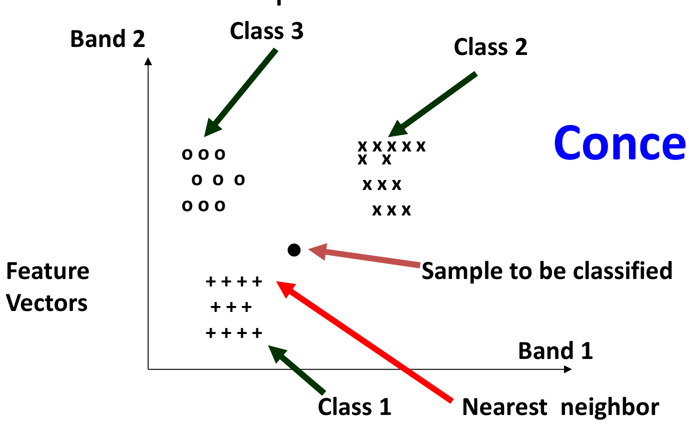
**Non-parametric in nature**

- The algorithm is:
  - Find the distance of given feature vector **x** from ALL the training samples
  - **x is assigned to the class of the nearest training sample (in the feature space)**
  - **This method does not depend on the class statistics like mean and covariance.**





- 2-D Feature Space




## Concept

## About the Terminology

- The ***nearest neighbor*** classifier is a general technique developed by Pattern Recognition practitioners. Image processing community adopted it.
- The nearest neighbor referred to in this context is the training sample with the most similar feature vector, i.e., at the smallest distance from the given pixel's feature vector. This has no relevance to where that pixel is in the image. Any pixel with the same feature vector will be identically classified, irrespective of its location.





## K-NN Classifier

- K-Nearest Neighbor ← Common terminology used in image smoothing and in classification
- In image smoothing, spatially adjacent pixels are referred
- In image classification, similar feature vectors in feature space are referred among the training data
- **Realize the difference and avoid confusion!**

## K-NN Classifier

- K-nearest neighbour classifier
- Simple in concept, time consuming to implement
- For a pixel to be classified, find the K closest training samples (in terms of feature vector similarity or smallest feature vector distance)
- Among the K samples, find the most frequently occurring class $C_m$
- Assign the pixel to class $C_m$





## K-NN Classifier

- Let $k_i$ be number of samples for class $C_i$ (out of K closest samples), i=1,2,…,L (number of classes)
- Note that
- The discriminant for K-NN classifier is
- $g_i(x) = k_i$
- The classifier rule is
- *Assign x to class* $C_m$ *if* $g_m(x) > g_i(x)$, *for all* $i, i \neq m$

## K-NN Classifier

- Let $k_i$ be number of samples for class $C_i$ (out of K closest samples), i=1,2,…,L (number of classes)
- Note that

$$\sum_i k_i = K$$

- The discriminant for K-NN classifier is
- $g_i(x) = k_i$
- The classifier rule is
- *Assign x to class $C_m$ if $g_m(x) > g_i(x)$, for all i, $i \neq m$*





## K-NN Classifier

- It is possible to find more than one class whose training samples are closest to the feature vector of pixel **x**. Therefore the discriminant function is refined further as

$$g_i(x) = \frac{\sum_{j=1}^{k_i} 1/d(x, x_i^j)}{\sum_{l=1}^{L}\sum_{j=1}^{k_l} 1/d(x, x_l^j)}$$

The distances of the nearest neighbours to the feature vector of the pixel to be classified are taken into account

## K-NN Classifier

- If the classes are in different proportions in the image, then the prior probabilities can be taken into account:

$$g_i(x) = \frac{k_i p(\omega_i)}{\sum_{l=1}^{L} k_l p(\omega_l)}$$

- For each pixel to be classified, the feature space distances to all training pixels are to be computed before the decision is made, due to which this procedure is extremely computation intensive, and is not used when the dimensionality (number of bands) of the feature space is large, e.g., with hyperspectral data.





## Spectral Angle Mapper

- Given a large dimensional data set, computing the covariance matrix, its inverse, and the distance for each pixel
- $(X - \mu)^T \Sigma^{-1} (X - \mu)$ is highly time consuming and if the covariance matrix is close to singular then its inverse can be unstable, leading to erroneous results
- In such cases, alternate methods can be applied, such as Spectral Angle Mapper

IIT Bombay Slide 60

## S.A.M. Principle

- If each class is represented by a vector $\mathbf{v}_i$, then the angle between the class vector and the pixel feature vector $\mathbf{x}$ is given by
- $\cos\theta = [\mathbf{v}_i.\mathbf{x}] / [|\mathbf{v}_i||\mathbf{x}|]$
- For small values of $\theta$, the value of $\cos\theta$ is large
- The likelihood of $\mathbf{x}$ to belong to different classes can be ranked according to the value of $\cos\theta$.

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 61

## S.A.M. Advantage

- The value of the vector would not be greatly affected by minor changes in $\mathbf{v}_i$ or $\mathbf{x}$.
- The computation is simpler compared to the Mahalanobis distance computation involved in ML method

GNR607 Lecture 27-32 B. Krishna Mohan

# Feature Evaluation Techniques

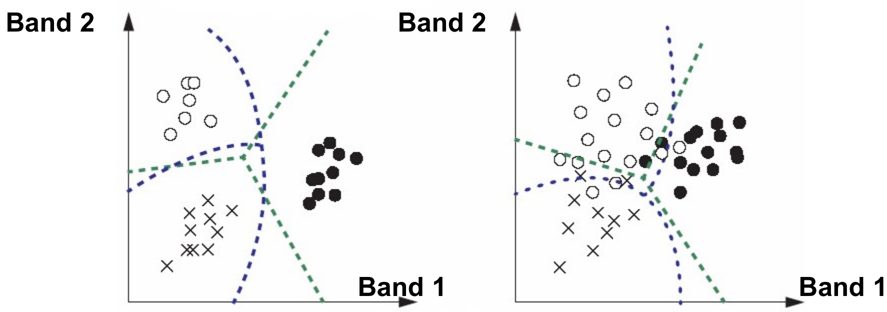


## Feature Evaluation

- **For a given set of classes and the training data, how well is each class separated from every other class?**
- How useful are the available features? Can we reduce the dimensionality of the dataset without compromising on accuracy of classification? In this case new data are ***generated from original data***
- The given image is ***then*** classified after training the classifier using training data
- The classification result is verified using test data

## Class separability in feature space

- Two generic cases

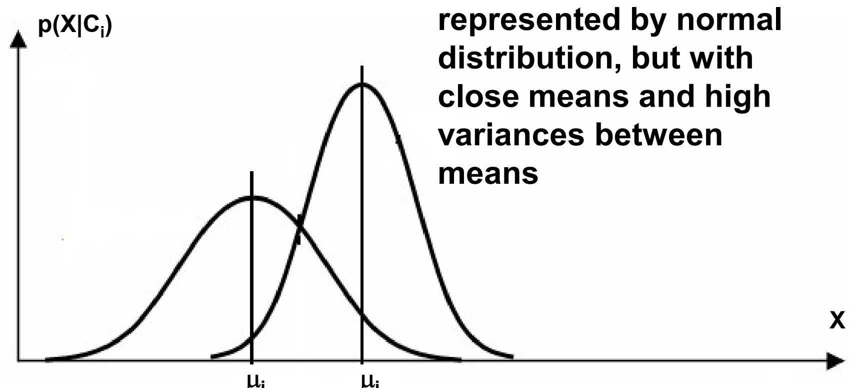



Well separated classes Poorly separated classes





## Low separability (with one feature)




Each class is represented by normal distribution, but with close means and high variances between means

## An example

- Separation of a bunch of fruits into different categories
- Useful features
  - Color
  - Size
  - Season of availability
  - Texture





## Divergence

Based on a measure of difference between pairs of classes

Consider the likelihood ratio for two classes

$$L_{ij}(x) = p(x \mid w_i) / p(x \mid w_j)$$

where $p(\mathbf{x}|w_i)$ and $p(\mathbf{x}|w_j)$ are the $i^{th}$ and $j^{th}$ spectral class probability distributions

$L_{ij}$ = 0 or infinity for highly separable classes

Let $L_{ij}' = \ln(p(\mathbf{x}|w_i)) - \ln(p(\mathbf{x}|w_j))$

# Divergence

- Divergence for a pair of classes is defined as

$$d_{ij} = E\{L'_{ij}(x) \mid w_i\} + E\{L'_{ji}(x) \mid w_j\}$$

- where

$$E\{L'_{ij}(x) \mid w_i\} = \int_x L'_{ij}(x) p(x \mid w_i) dx$$

- Therefore

$$d_{ij} = \int_x \{p(x \mid \omega_i) - p(x \mid \omega_j)\} \ln \frac{p(x \mid \omega_i)}{p(x \mid \omega_j)} dx$$





# Divergence

- Divergence is a ***distance metric***
- $d_{ij} = d_{ji}$
- For i=j, $d_{ii} = d_{jj} = 0$
- Divergence of a class with itself is zero.
- It can be shown that

  $d_{ij}(x_1, x_2, \ldots, x_n) > d_{ij}(x_1, x_2, \ldots, x_{n-1})$

This means that divergence increases with more features. This is true for statistically independent features

## Divergence for multidimensional Gaussian feature set

$$d_{ij} = 0.5Tr\{(\sum_i - \sum_j)(\sum_j^{-1} - \sum_i^{-1})+$$
$$0.5Tr\{(\sum_j^{-1} + \sum_i^{-1})(m_i - m_j)(m_i - m_j)^T$$

- The average divergence $d_{ave}$ is given by
- $d_{ave} = \sum_{i=1}^{L} \sum_{j=i+1}^{L} p(w_i)\, p(w_j)\, d_{ij}$





## Divergence for Feature selection

- Given L classes, B bands, and $B_1$ bands to be selected out of B, then the number of pairs to be evaluated using the divergence equation is

$$({}^{B}C_{B1} \cdot {}^{L}C_2)$$

which is very large even for modest size classification problem

For example, for a problem of ten classes, 7 bands in the input image, with 4 bands to be selected, the number of combinations is 1575.

## Bhattacharyya Distance

For normally distributed classes, the Bhattacharyya distance between two class distributions is given by

$$B = \frac{1}{8}(m_i - m_j)^T \left[ \frac{\Sigma_i + \Sigma_j}{2} \right]^{-1} (m_i - m_j) + \frac{1}{2} \ln \left[ \frac{\frac{|\Sigma_i + \Sigma_j|}{2}}{|\Sigma_i|^{1/2} |\Sigma_j|^{1/2}} \right]$$





## Jeffries Matusita (JM) Distance

- The JM distance is defined as
- $J_{ij} = 2(1-e^{-B})$, where B is the Bhattacharyya distance between the two classes
- The average JM distance $J_{ave}$ between the classes is given by

$$J_{ave} = \sum_{i=1}^{L} \sum_{j=i+1}^{L} p(w_i) p(w_j) J_{ij}$$

## Interpretation of JM distance

- As the separation between classes increases, the JM distance increases exponentially towards 2.
- This is a desirable behavior compared to the divergence measure
- Computationally JM distance is more expensive than Divergence
- Transformed divergence can also be used, with divergence replacing Bhattacharya distance in the formula for J-M distance





## Weight Factors

- Weighted divergence $W_{ij}$ is computed as

$$W_{ij} = \left[ \frac{\sum_{i=1}^{K-1} \left( \sum_{j=i+1}^{K} f_i f_j d_{ij} \right)}{\frac{1}{2} \left[ \left( \sum_{i=1}^{K} f_i \right)^2 - \sum_{i=1}^{K} f_i^2 \right]} \right]$$

where $f_i$ and $f_j$ are weights or prior probabilities of classes i and j.

## Other Measures

- There are several measures recently proposed, such as based on
- Genetic Algorithms
- Sequential Search Algorithms

- In practice, user can choose the pairs of classes of interest, and can see the values of divergence or JM distance only for those classes.



## Accuracy Assessment

## Relevance of Accuracy Assessment

- **Output of classification of remotely sensed images**
  - **Becomes input layer for a GIS**
  - **Accuracy part of metadata**
  - **Class-wise and overall accuracy need to be quantified**





## Accuracy Assessment

- Estimated using *ground truth* i.e., set of points whose class is known
- These points are known as test points
- Test points **should not** be used during the classifier training
- Test points should be in sufficient number for the accuracy assessment to be significant

## Choice of Test Points

- Test points should not be taken in the immediate neighbourhood of training data points
- High spatial dependence would bias the assessment in such a case
- Based on the accuracy of classification of the test data points, various quantitative estimates are defined.
- The degree of agreement between the classifier output and the ground truth is considered





## Classification Accuracy Assessment

- Error or *confusion* matrix is often deployed to derive measures of accuracy
- Used to derive the overall accuracy as well as *user's accuracy* and *producer's accuracy*

## Confusion matrix

| R ↓ C➔ | Forest | Indust. | Urban | Water | Total |
|---|---|---|---|---|---|
| Forest | **68** | 7 | 3 | 0 | **78** |
| Indust. | 12 | **112** | 15 | 10 | **149** |
| Urban | 3 | 9 | **89** | 0 | **101** |
| Water | 0 | 2 | 5 | **56** | **63** |
| Total | **83** | **130** | **112** | **66** | **391** |





## Measures of Accuracy from Confusion Matrix

- Overall Accuracy

  Number of test samples correctly classified / Total number of test samples
  Sum of diagonal elements / Sum of rows

- Producer's Accuracy (Omission errors)

  No. of samples of a class correctly classified / No. of samples of the class

- User's Accuracy (Commission errors)

  No. of samples that actually belong to a class / No. of samples assigned to that class

## Accuracy Computation

- Overall accuracy for this example
- Correctly classified samples =
- 68 + 112 + 89 + 56 = 325
- Total number of samples =
- 78+ 149 + 101 + 63 = 391
- Overall accuracy = 325 / 391 ~ 83%





## Producer's Accuracy

- Forest:
- Total reference samples = 83
- Correctly classified samples = 68
- Accuracy = 68/83 ~ 82%
- Industrial areas: 112/130 ~ 86%
- Urban areas: 89/112 ~ 79%
- Water: 56/66 ~ 85%

## User's Accuracy

- Industrial Areas
- Correctly classified samples = 112
- Samples assigned to Industrial Areas = 149
- User's Accuracy = 112 / 149 ~ 75%
- Forest: 68/78 ~ 87%
- Urban: 89 / 191 ~ 88%
- Water: 56/63 ~ 88%





## Kappa coefficient

- Kappa coefficient is defined as

$$\hat{\kappa} = \frac{N \sum_{i=1}^{r} C_{ii} - \sum_{i=1}^{r} C_{i+} C_{+i}}{N^2 - \sum_{i=1}^{r} C_{i+} C_{+i}}$$

R. Congalton

Cohen

$C_{i+}$ denotes the sum of the $i^{th}$ row elements; $C_{+i}$ denotes the sum of the $i^{th}$ column elements

$\kappa$ is interpreted as the index of correct classification after adjusting for chance agreement between the true and computed classes.

IIT Bombay Slide 113

## Kappa Coefficient for this example

$$\hat{\kappa} = \frac{N\sum_{i=1}^{r} C_{ii} - \sum_{i=1}^{r} C_{i+}C_{+i}}{N^2 - \sum_{i=1}^{r} C_{i+}C_{+i}}$$

= (85761/107575) ~ 0.80
(**Verify the calculations!**)

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 114

## Interpretation of Kappa Coefficient

- Kappa coefficient is used to compare the degree of consensus between raters (inspectors) in, for example, Measurement Systems Analysis.
- The degree of chance agreement between the raters and the measurement system is considered in the rating to assess the actual degree of agreement

GNR607 Lecture 27-32 B. Krishna Mohan

## Interpretation of Kappa Coefficient

- Suppose 150 samples are independently verified and the following determinations are made:

| | | Inspector | | |
|---|---|---|---|---|
| | | **Accept** | **Reject** | **Total** |
| System | **Accept** | 20 | 19 | 39 |
| | **Reject** | 1 | 110 | 111 |
| | **Total** | **21** | **129** | **150** |





## Chance Agreement

| | Inspector | | |
|---|---|---|---|
| | **Accept** | **Reject** | **Total** |
| **Accept** | 5.46 | 33.54 | 39 |
| Sys. **Reject** | 15.54 | 95.46 | 111 |
| **Total** | 21 | 129 | 150 |

The chance values are produced by computing:
(row total x col. total) / Total values

## Kappa Coefficient for this example

- κ = (130 – 100.92) / (150 – 100.92) = 0.593
- 130 comes from the first table
- 100.92 comes from the chance table
- 150 is the total number of samples
- Chance agreement is deducted from actual agreement to estimate the degree of agreement between the system (classifier) and the inspector (ground truth)





## Cross Validation

- In the cross-validation approach to accuracy estimation, the ability of the classifier to generalize is determined from the training data itself.
- All the training samples except one are used for training and the one remaining sample is used for testing.
- This process is repeated several times

## Cross Validation

- This method is also known as ***Leave one out*** method of accuracy estimation.
  - Over a large number of runs the average accuracy of correct classification of the sample left out is computed.
  - The one sample being left out is randomly selected.
  - The risk is that small classes are likely to be ignored
  - Classes with large number of training samples likely to be chosen more often.





## Cross Validation

- An extension of this method is to divide the training data into k subsets, and each time one subset is left out, and used to test the generalization capability of the classifier.
- This way, a maximum of k runs will be adequate to extensively test the classifier

## Cross Validation

- How do we select the subsets?
- The procedure is to randomly sample the training data to make the subsets, at the same time ensure that samples from each class are chosen
- The solution for this is to adopt the stratified sampling approach
- The stratification is performed on the basis of the classes.
- The size of the sample from each class is proportional to the size of the class itself.





## Comments on Accuracy Estimation

- Many sources of both conservative and optimistic bias in classification accuracy assessment
- Bias occurs when a classification estimate is optimistic or conservative

## Conservative Bias

- Three significant sources of conservative bias:
  - Errors in reference data,
  - Positional errors, and
  - Minimum mapping unit of reference grid





## Errors in Reference Data

- Incorrect class assignment,
- Change in cover type between the time of imaging and the time of field verification,
- Mistakes in recording or processing the reference data, etc.), some of our correctly classified pixels may be incorrectly assessed as being misclassified.
- For example, if ten percent of our ground-truth samples were incorrect, we would estimate our classification to be 90 percent accurate (while in this example it is actually 100 percent) or vice versa.

IIT Bombay Slide 125

## Positional Errors

- Due to registration error, we can not be absolutely sure of the location of any given pixel
- For example, if we rectified an image to 30-meter output pixels with an RMS error of 1.0 pixel, the average positional error of the rectification *model* is +/- 30 meters
- Even if we use sub-meter GPS technology to go to the center coordinates of a pixel to be "ground truthed", we cannot be certain that we are in that pixel!
- Due to the positional error inherent in rectified images, some correctly classified pixels may not be correctly located during field sampling

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 126

## Minimum Mapping Unit Area

- If map produced by visual interpretation of image is used for reference, smallest mapped area by visual means can influence accuracy estimation by digital means – some single pixel classes may be marked incorrect, even if they are correctly classified by computer

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 127

## Optimistic Bias

- Three significant sources of optimistic bias:
  - Use of training data for accuracy assessment
  - Restriction of reference data sampling to homogeneous areas, and
  - Sampling of reference data not independent of training data

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 128

## Use of training data for accuracy assessment

- We might estimate our overall classification accuracy to be 100 percent
  - If we selected our reference data from training fields that encompassed large, pure, level stands.
  - However, if much of our image is a heterogeneous mixture of vegetation types of variable canopy structure and variable topographic conditions, our classified image is probably not really 100 percent accurate.

GNR607 Lecture 27-32 B. Krishna Mohan

## Restriction of reference data sampling to homogeneous areas

- Training fields relatively spectrally pure
- Easy to classify.
- If reference areas are near training fields, they are likely to also be relatively easy to classify correctly.
- Reference data should not be close to training data since the results are likely to be similar due to spatial correlation between samples





## Sampling from homogeneous blocks of classified pixels

- A common approach for reference data is to sample the coordinates of the center pixel of a small neighborhood belonging to the same class.
- This can lead to an optimistic estimate of classification accuracy since homogeneous areas are selected
- Heterogeneous areas are more difficult to correctly classify but excluded from selection as reference data

## Reporting the error matrix and classification accuracy are insufficient

- Sampling methods used for reference data should be reported in detail so that potential users can judge whether there may be significant biases in the classification accuracy assessment





## General Comments

- Accuracy assessments in general are expensive
- Attention must be paid to limitations imposed by
  - Region
  - Sensor
  - Weather etc. versus the statistical requirements imposed on the study

IIT Bombay Slide 133

## General Comments

- Sampling schemes must include
  - Appropriate sample design
  - Should be in relation to the question being asked and the kind of imagery being analyzed.
- What kind of sample unit will be used and how big will it be?
  - Single pixels?
  - Polygons?
  - Multiple Polygons?
  - How many samples should be taken in order to be statistically valid?

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 134

## Fractional Classification

- When coarse resolution imagery are being classified, and reference data available at higher resolution, it is desirable to report the number of high resolution pixels within each thematic class inside the coarse resolution pixel – mixture estimates
- If only the dominant class is reported, then the remaining useful information is lost

GNR607 Lecture 27-32 B. Krishna Mohan

## Mixed Pixels

- When spatial resolution is coarse, one pixel may contain parts of many landuse classes
  - e.g., tree, bare soil, grass
- Classification is estimating proportions of different classes within a pixel
- The problem is called *mixture modeling*





## Mixed Pixels

- Important when spectra of different classes are compared as in hyperspectral remote sensing
- Reference spectra are drawn from single classes
- Most pixel spectra are mixtures of more than one pure class
- Mixture modeling estimates the relative proportion of each class assuming a particular model for mixing

## Mixed Pixels

- Most common mixing model is a linear mixture modeling

$$R_n = \sum_{k=1}^{L} w_k A_{k,n} + \xi_n$$

R is the refLecture ance recorded for a mixed pixel, $w_k$ is the proportion and $R_k$ is the refLecture ance belonging to pure class k. ξ is additive noise present in the data

$w_k$ are estimated from known pure pixels and given refLecture ances of mixed pixels, minimizing the effect of noise





## Classifiers v/s Features

- The definition and selection of land cover classes is crucial
- The capability of existing features is to be ascertained for reliable classification
- Choice of classifier can be critical in some cases, while ancillary data can help in other cases

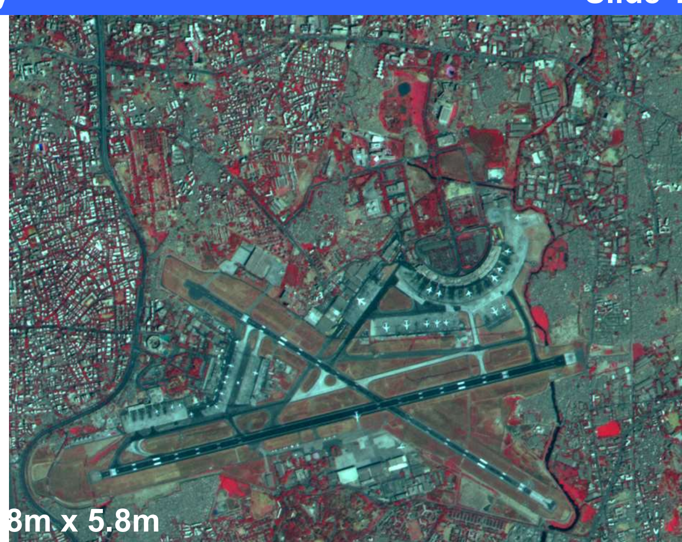
IIT Bombay Slide 136

**Type of features used:**
- Spectral
- Spatial
- Textural
- Contex-tual


GNR607 Lecture 27-32 B. Krishna Mohan

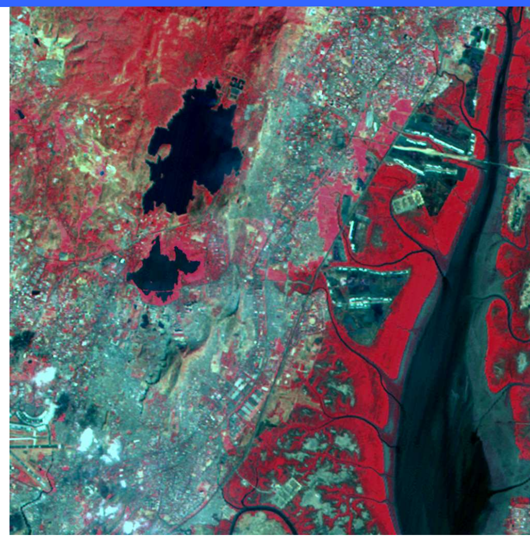
IIT Bombay Slide 137

**Type of features used:**
- Spectral
- Textural
- Contex-tual




GNR607 Lecture 27-32 B. Krishna Mohan

**Type of features used:** **Spectral**



23.25m x 23.25m



# Unsupervised Classification

## Unsupervised Classification

- When access to domain knowledge or the experience of an analyst is missing, the data can still be analyzed by numerical exploration, whereby the data are grouped into subsets or ***clusters*** based on statistical similarity





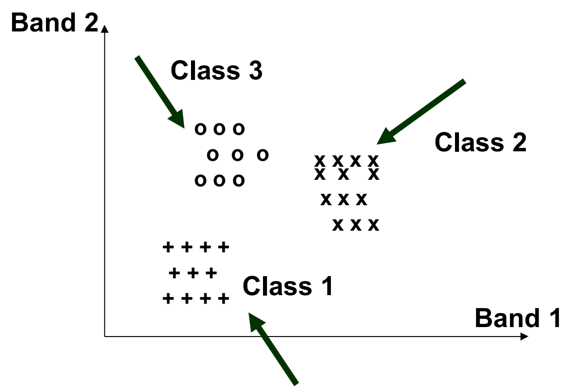
## Unsupervised Classification

- *Unsupervised classification* is also known as learning without teacher
- In the absence of reliable training data it is possible to understand the structure of the data using statistical methods such as *clustering algorithms*
- *Popular clustering algorithms are k-means and ISODATA.*

*The book **Pattern Recognition Principles by J.T. Tou and R.C. Gonzalez** contains discussion of several clustering algorithms with numerical examples*

# Concept of Clustering

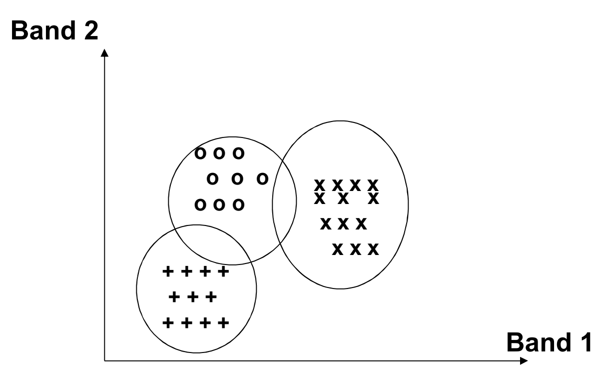
- 2-D Feature Space





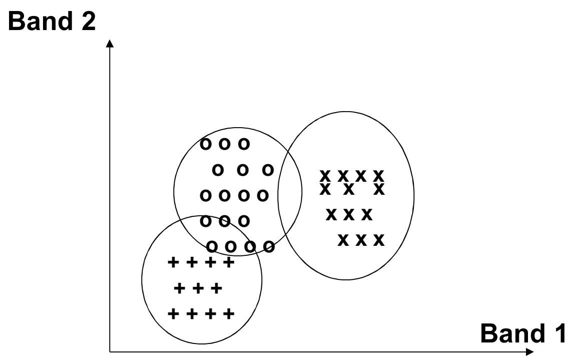



# Concept of Clustering



IIT Bombay Slide 143

# Overlapping Clusters




GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 144

# Clustering Algorithms

- All feature vectors are points in an L-dimensional space where L is the number of bands *(The letter K is reserved for the number of clusters!)*
- It is required to find which sets of feature vectors tend to form clusters
- ***Members of a cluster are more similar to each other than to members of another cluster – In other words, they possess low intra-cluster variability and high inter-cluster variability***

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 145

## A Simple Clustering Algorithm

- Let **p** be any element in the L-dimensional feature space. **p** can be selected randomly. Number of clusters = 1 initially.
- The Euclidean distance between **p** and any other vector **q** is given by

$$d(\mathbf{p},\mathbf{q}) = \sqrt{\sum_{i=1}^{L}(p_i - q_i)^2}$$

Let **r** be such that d(**p**,**r**) = $\max_q$(d(**p**,**q**))

Now **p** and **r** are representatives of two clusters, in other words, mean vectors. Number of clusters = 2.

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 146

## A Simple Clustering Algorithm

- Now compute d(**p,q**) and d(**r,q**) for each vector **q**.
- select vector **s** such that

  min{d(**p,s**), d(**r,s**)} $\geq$ t.d(**p,r**), t < 1
- **p, r** and **s** are the mean vectors of the three clusters
- Repeat till the desired number K of clusters are generated or till the minimum distance criterion fails to introduce a new cluster

GNR607 Lecture 27-32 B. Krishna Mohan

## A Simple Clustering Algorithm

- Each element **x** is now assigned to cluster j such that

$$d(\mathbf{m}_j, \mathbf{x}) \leq d(\mathbf{m}_i, \mathbf{x}),\ i=1,2,\ \ldots,\ K$$

$\mathbf{m}_j$ is the mean vector of cluster j.

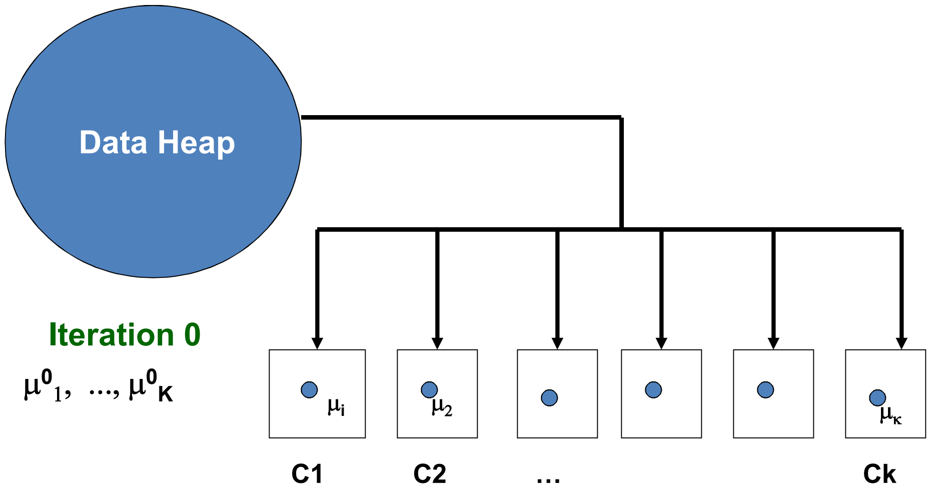
- The number of clusters formed depends on the starting point, that is the first vector **p** mentioned before. If the starting vector is different, the final clustering can be different.
- The other criterion that governs the formation of the clusters is the value t, the multiplying factor





## K-Means

- Iterative algorithm
- Number of clusters K is known by user
- Most popular clustering algorithm
- Initialize randomly K cluster mean vectors
- Assign each pixel to any of the K clusters based on minimum feature distance
- After all pixels are assigned to the K clusters, each cluster mean is recomputed.
- Iterate till cluster mean vectors stabilize



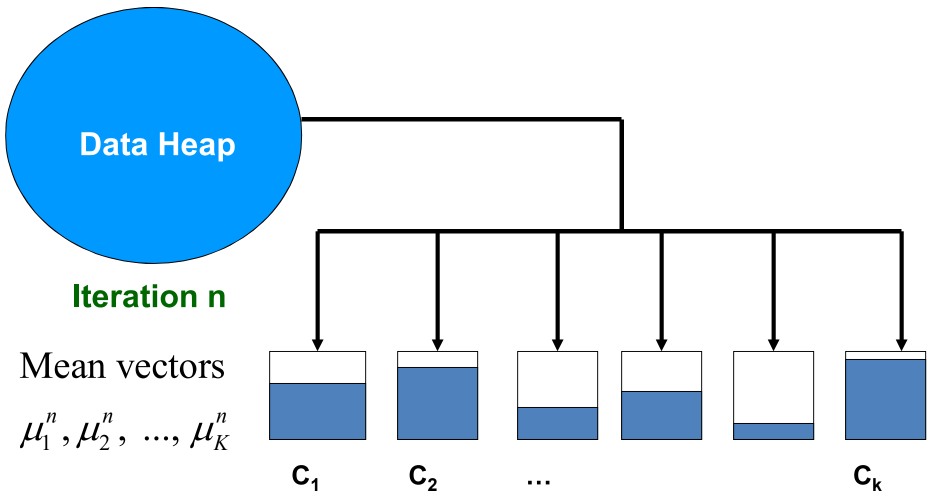





# K-Means algorithm

- Each cluster Ci is represented by its mean vector $\mu_i$
- $||\mathbf{x}-\mu_i||$ = $\sqrt{(X-\mu_i)^T(X-\mu_i)}$

Assign X to cluster $C_k$ where $||X-C_k|| = \text{Min}\,(||X-C_i||)$, i=1,2,…,K








# K-Means algorithm contd.

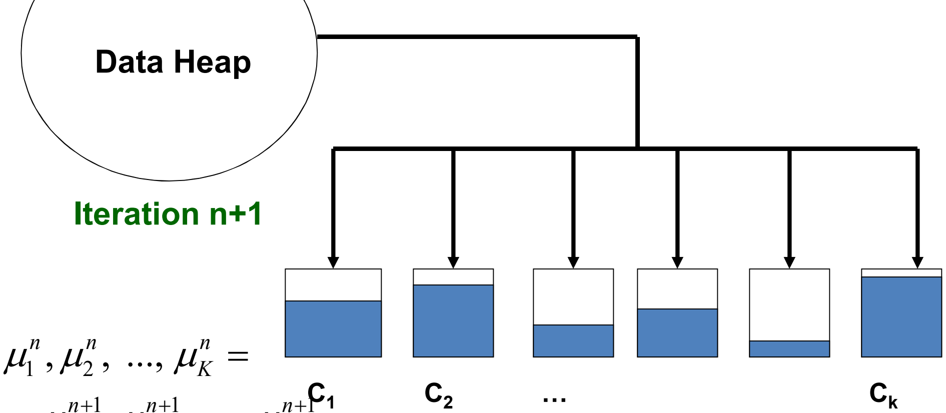
- After all the pixels are assigned to the K clusters, update $\mu_i$

$$\mu_i^{new} = \frac{1}{|C_i|} \sum_{X \in C_i} X$$

Replace existing cluster mean vectors by the updated mean vectors if their difference is more than a user-specified threshold

$$\mu_i^{old} \xleftarrow{i} \mu_i^{new} if\,(Max)_i \, \| \mu_i^{new} - \mu_i^{old} \| > d_{\max}$$

## Convergence!




Iteration n+1

$$\mu_1^n, \mu_2^n, \ldots, \mu_K^n = \mu_1^{n+1}, \mu_2^{n+1}, \ldots, \mu_K^{n+1}$$





## K-Means Algorithm

- Choice of K is important
- For normal remotely sensed images, about 15-20 iterations are maximum required for stabilization of the cluster centres
- Convergence can be based on average difference between cluster centres or maximum difference between any pair of cluster centres

## ISODATA ALGORITHM

- Iterative Self-Organizing Data Analysis Technique (the last A added to make the acronym sound better)
- Developed in biology in the 1960's by Ball and Hall
- See Tou and Gonzalez's classic "Pattern Recognition Principles" for an excellent exposition to clustering algorithms





## User specified parameters for ISODATA

- Generalization of K-Means algorithm
- Consists of many user-specified parameters
  - Minimum size of cluster
  - Maximum size of cluster
  - Maximum intra-cluster variance
  - Minimum separation between pairs of clusters
  - Maximum number of clusters
  - Minimum number of clusters
  - Maximum number of iterations

## User specified parameters for ISODATA

- If number of elements in cluster $c_k$ is less than $Size_{min}$, then $c_k$ is merged with another cluster $c_l$ such that
- $d(\mathbf{m}_k, \mathbf{m}_l) = \min_i\{d(\mathbf{m}_i, \mathbf{m}_k)\}$
- Number of clusters is reduced by 1
- All elements of cluster k are labeled i.





## User specified parameters for ISODATA

- Maximum size of cluster $Size_{max}$
- If $\#\{C_i\} > Size_{max}$, then it is split into two clusters
- Number of clusters is increased by 1
- Splitting criteria can be many.
- Cluster splitting is more complex compared to cluster merging

# User specified parameters for ISODATA

- A cluster should be comprised of homogeneous elements – similar feature vectors
- Intra-cluster variance

$$\sigma_i^2 = \frac{1}{N_i} \sum_{j=1, \mathbf{x} \in C_i}^{N_i} d(\mathbf{m_i} - \mathbf{x})^2$$

If $\sigma_i^2 > Var_{max}$, then split cluster i.





# User specified parameters for ISODATA

- The clusters should be separated by a minimum distance, else they could be merged into one cluster
- If $d(\mathbf{m_i}, \mathbf{m_j}) < Dist_{min}$, then merge clusters i and j.

## User specified parameters for ISODATA

- The clustering procedure should stop if the number of clusters reaches the upper limit $K_{max}$
- The clusters should not be allowed to merge if the number of clusters equals the minimum number $K_{min}$
- The K-means step should stop if the number of iterations reaches the limit $Iter_{max}$





## Cluster Evaluation

- Desired properties of clusters
  - High separation between different clusters
  - Low variability within each cluster

## High Separation Between Clusters

- Mahalanobis distance
- Compute mean vector and covariance matrix for each cluster
- High Mahalanobis distance between mean vectors desirable





## High Separation Between Clusters

- For small data, it is also possible to find minimum distance between samples of two clusters
- The minimum distance should be high

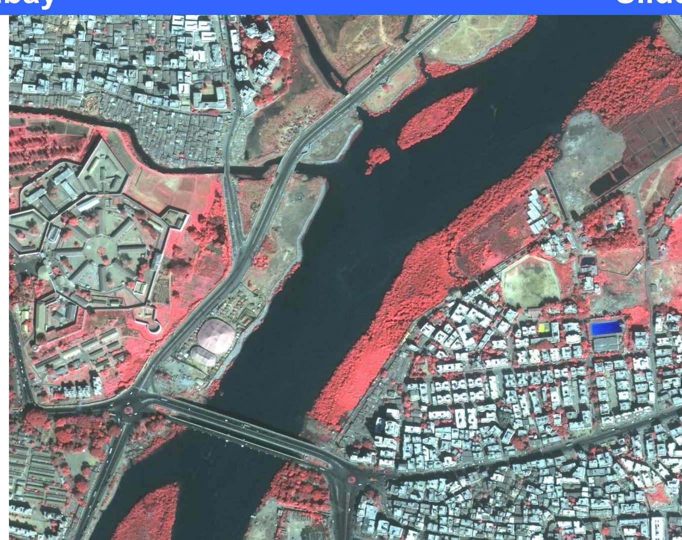
## Low Intra-Cluster Variability

- Trace of covariance matrix
- Within each cluster, find distance of an element from every other element
- Note maximum distance
- Max distance should be small





INPUT IMAGE



K-MEANS: K=7



# Change Detection

## Motivation for Change Detection

- Natural and manmade features are inventoried from time to time
- Some are static, some dynamic
- Changes should be monitored, in particular the landuse / landcover changes
- Change detection information – major reason to fund remote sensing programmes





## Change Detection - Definition

- Change detection is the process of identifying differences in the state of an object or phenomenon by observing it at different times
- Multitemporal images are analyzed to map changes
- Different applications require image at varying intervals of time

## Change Detection – ROI

- The geographic ROI (e.g., a block, village, or watershed) is especially important
- Must be completely covered by *n* dates of imagery.
- Failure to ensure → change detection maps with *data voids*
- Problematic when computing change statistics.





## LU/LC Classification Scheme

- It is desirable to use an established, standardized land-cover / land-use classification system for change detection
- Use of standardized classification systems allows change information to be compared with other studies.

## Concept of Change Detection

- To use remote sensing, the change must be detectable with our instruments
  - Spectrally
    - Distinguish use from cover
    - Allow sufficient time between images for changes to be noticeable
  - Spatially
    - Generally, grain size of change event >> pixel size





## Driving Force for Change Detection

- **Need for effective methods for deriving information on**
- **Land use change**
  - **Forest fragmentation**
  - **Urban growth**
  - **Loss of agricultural lands**
  - **Increase in impervious surface area**

IIT Bombay Slide 174

## Approaches for Change Detection

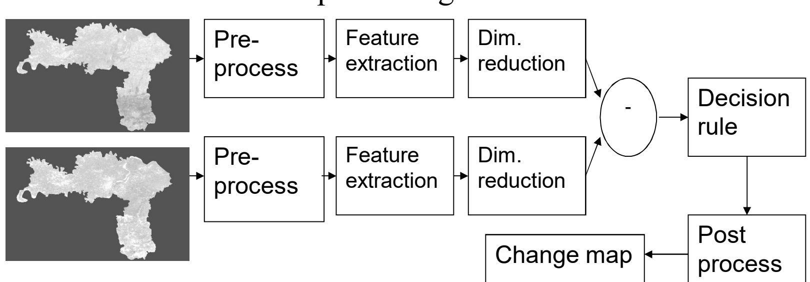
- **Compare the results of different land use and land cover change detection approaches**
  - Post-classification cross- tabulation
  - Cross-correlation analysis
  - Neural networks
  - Knowledge-based expert systems
  - Image segmentation & object-oriented classification

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 175

## Basic Steps in Change Detection

- Image pre-processing ➔ Feature Extraction
- Dimensionality reduction ➔ Image comparison
- Decision Rule ➔ Post-processing




GNR607 Lecture 27-32 B. Krishna Mohan

## Considerations before Image Comparison

- Precise registration of multi-temporal images.
- Precise radiometric and atmospheric calibration or normalization between multi-temporal images.
- Similar phenological states between multi-temporal images.
- selection of the same spatial and spectral resolution images if possible.





## Change Detection Logic

- Comparison of multiple-date *hard* land-cover classifications of remotely sensed data.
- Create a *hard* change detection map consisting of information about the change in discrete categories (e.g., change in forest, agriculture).
- This is still very important and practical in many instances, but *it is ideal to capture both* ***discrete and fuzzy changes*** *in the landscape.*

## Per-pixel or Object-oriented Change Detection

- Compare Date *n* and Date *n + 1* classification maps pixel by pixel. This is commonly referred to as *per pixel* change detection.
- *Object-oriented change detection* → comparison of two or more scenes having relatively homogenous image objects (patches or segments)
- The relatively homogeneous image objects in the two scenes are then subjected to various change detection techniques.



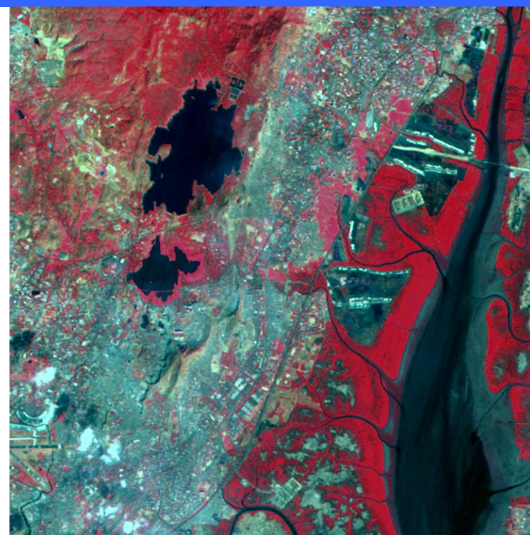


**Type of objects:**

- **Buildings**
- **Roads**
- **Water-bodies**
- **Shadows**
- **Clusters of trees**
- **Grasslands**

**…**

**Object characterization difficult in low resolution images**



23.25m x 23.25m





## Change Detection Techniques

- Algebra
- Transformation (e.g., PCA)
- Classification
- Advanced models
- Geographical Information System (GIS) approaches
- Visual analysis

## Remote Sensing System Considerations

- Attention should be paid to:
  - remote sensor system considerations, and
  - environmental characteristics.
- Failure to understand the impact of the various parameters on the change detection process can lead to inaccurate results.
- Ideally, hold the following resolutions constant: temporal, spatial (and look angle), spectral, and radiometric.





## System Parameter – Time of Acquisition

- First, use a sensor system that acquires data at approximately the *same time of day.* This eliminates diurnal Sun angle effects that can cause anomalous differences in the refLecture ance properties of the remote sensor data.

- Second, acquire remote sensor data on *anniversary dates*, e.g., Feb 1, 2004, and Feb 1, 2006. Anniversary date imagery minimizes the influence of seasonal Sun-angle and plant phenological differences that can negatively impact a change detection project.

## Spatial Resolution

- If same sensor data are not available for change detection, decide on a representative minimum mapping unit (e.g., 20 × 20 m) and then resample both datasets to this uniform pixel size.
- Note that the information content of the resampled data can never be greater than the IFOV of the original sensor system.





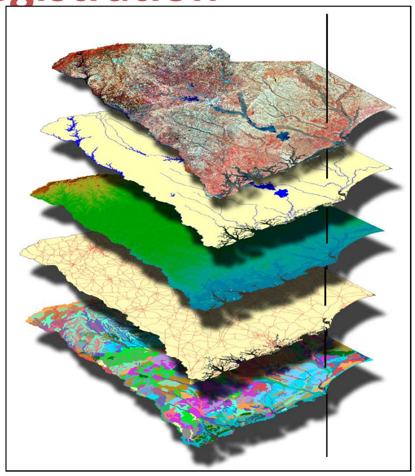
## Geo-registration

- Accurate spatial registration of at least two images is essential for digital change detection.
- Ideally, the remotely sensed data are acquired by a sensor system that colLecture s data with the same *instantaneous field of view* on each date.
- For example, Resourcesat LISS-IV colLecture ed at 5.8m m spatial resolution on two dates are relatively easy to register to one another.

IIT Bombay Slide 186

## Accuracy of registration

- Remotely sensed data used for change detection should be geometrically rectified to be within fraction of a pixel (0.5 pixel or less)



**Reproduced with permission from the Lecture notes of Prof. John Jensen, University of South Carolina**

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 187

## Geo-registration

- Geometric rectification algorithms are used to register the images to a standard map projection (such as UTM).
- *Rectification should result in the two images having a root mean square error (RMSE) of $\leq$ 0.5 pixel.*
- Mis-registration of the two images may result in the identification of spurious areas of change between the datasets. For example, just one pixel mis-registration may cause a stable road on the two dates to show up as a new road in the change image

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 188

## System Parameter: Look Angle

- Two images with significantly different look angles, colLecture ed by steerable sensors, can cause problems when used for change detection.
- Examples are IRS-1C, SPOT satellites. In nadir view, the sensors see the top of the vegetation canopy. In the oblique view (upto 26$^{o}$ angle), the sensor sees the side of the canopy. Differences in refLecture ance from the two datasets may cause spurious change detection results.
- Hence, choose two images with same viewing angle as far as possible

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 189

## System Parameter: Spectral Resolution

- Desirable to choose same sensor for two dates. Else, choose corresponding bands
- For example, Landsat MSS bands 4 (green), 5 (red), and 7 (near-infrared) and IRS bands 2 (green), 3 (red), and 4 (near-infrared), can be used successfully with Landsat ETM+ bands 2 (green), 3 (red), and 4 (near-infrared).
- Mismatched bands may result in poor accuracy of change detection

GNR607 Lecture 27-32 B. Krishna Mohan

## System Parameter: Radiometric Resolution

- Choose sensors with same radiometric resolution.
- Ideally, convert the brightness data to apparent surface refLecture ance, which eliminates the problem





## Radiometric Correction

- **Absolute Radiometric Correction**
  - Use of a model of atmosphere in conjunction with *in situ* atmospheric measurements (if possible) to correct for path radiance.
- **Relative Radiometric Correction**
  - Single image normalization using histogram adjustment
  - Multiple date image normalization using regression techniques

## Environmental Parameter: Soil Moisture

- If weather conditions are nearly same between the images in the time sequence, that will be ideal. If there is extreme heat or heavy precipitation, change detection techniques can result in misleading outputs. Rainfall data, if available, would be handy.
- At least, eliminate that part of the image (if possible to identify) where localized cloud burst or thunder storm would have happened, and limit the initial change detection exercise to the remaining part of the image.





## Environmental Parameter: Phenological Cycle Characteristics

- Natural ecosystems go through repeatable, predictable cycles of development. Humans also modify the landscape in predictable stages. These cycles of predictable development are referred to as *phenological cycles*.
- Analysts use this information to identify when remotely sensed data should be colLecture ed. Therefore, analysts must be familiar with the *biophysical* characteristics of the vegetation, soils, and water constituents of ecosystems and their phenological cycles.

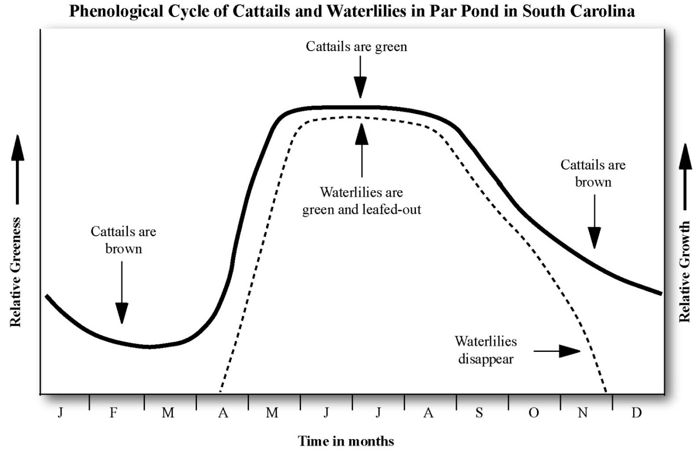
# Environmental Parameter: Vegetation Phenology

- Obtaining near-anniversary images greatly minimizes the effects of seasonal phenological differences
- When attempting to identify change in agricultural crops, the analyst must be aware of when the crops were planted. Large lags in planting date between fields of the same crop can cause serious change detection error.





Phenological Cycle of Cattails and Waterlilies in Par Pond in South Carolina




Reproduced with permission from the Lecture notes of Prof. John Jensen, University of South Carolina

**INPUT DATASETS**

Image of Varanasi town acquired on 06th December 1999.

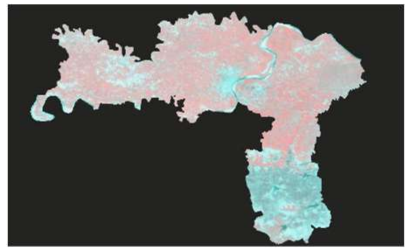


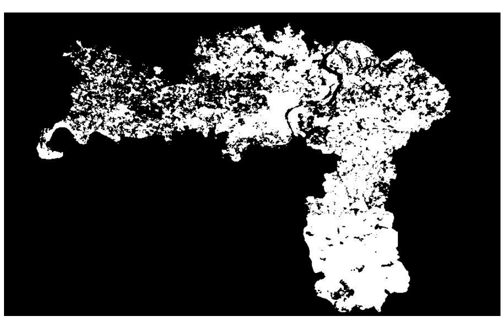
Image of Varanasi town acquired on 06th March 2000.

Number of bands = 2
Rows = 550
Columns = 900





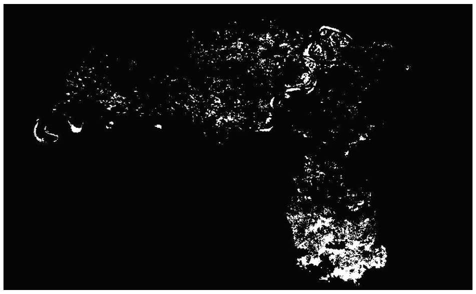
**INTEGRATION OF INTENSITY AND TEXTURE DIFFERENCES**



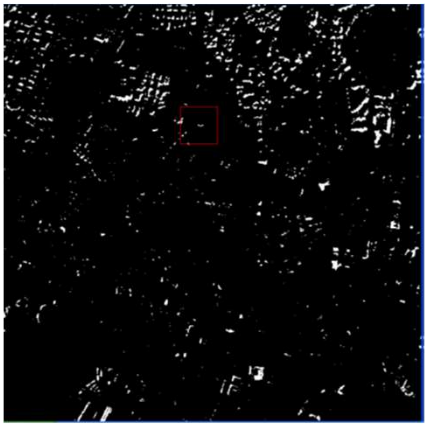
Change map of Varanasi area.

Change map of window-I of Powai area.

TEMPORAL PRINCIPAL COMPONENT ANALYSIS



Change map of Varanasi area.



Change map of window-II of Powai area.





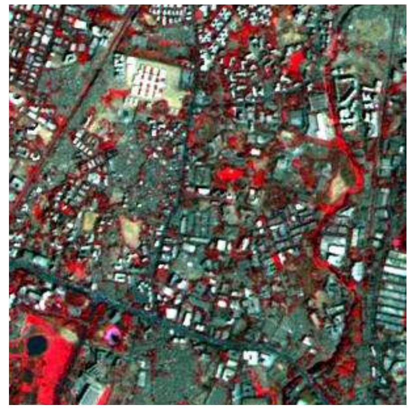
A window of the airborne image acquired over Southampton city in UK on 2001.

A window of the airborne image acquired over Southampton city in UK on 2002.

Number of bands = 3
Rows = 665
Columns = 739

Window-II of the Quickbird image acquired over Powai area on 2001 consisting of 2988 rows and 3009 columns with 0.6 metres resolution.

Window-II of the LISS-IV image acquired over Powai area on 2009 consisting of 305 rows and 303 columns with 5.8 metres resolution.

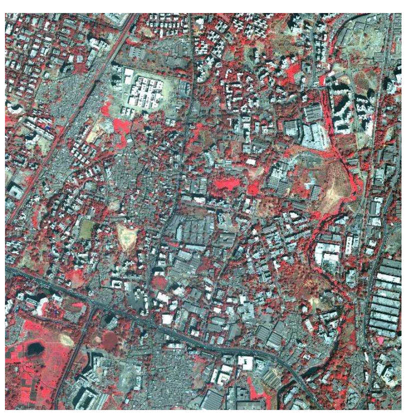


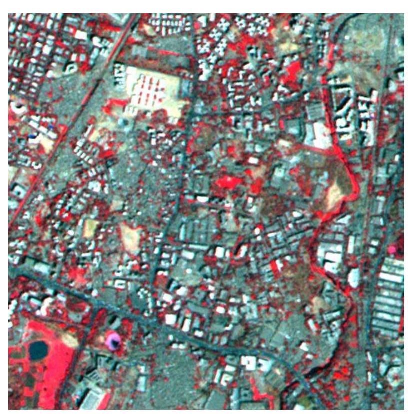




Window-II of the Quickbird image acquired over Powai area on 2001 consisting of 712 rows and 712 columns.

Window-II of the LISS-IV image acquired over Powai area on 2009 consisting of 712 rows and 712 columns.





The above images are resampled to 2.5 metres spatial resolution.

IIT Bombay Slide 202

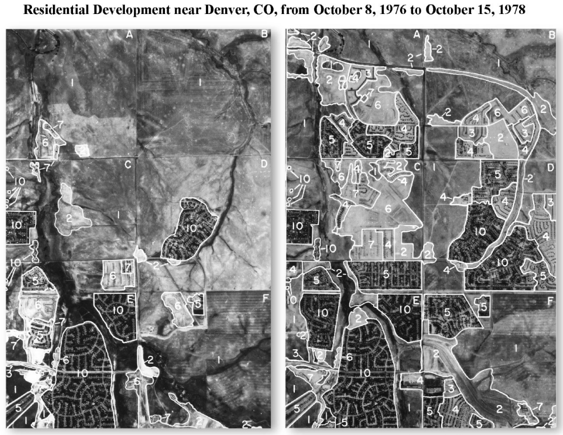
# Crop Changes

- Changes in row spacing and direction can have an impact.
- Analyst must know the crop's *biophysical* characteristics as well as the *cultural* land-tenure practices in the study area
- Choose the most appropriate remotely sensed data accordingly for change detection.
- Field visits during change detection should help.

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 203

**Residential Development near Denver, CO, from October 8, 1976 to October 15, 1978**



a. October 8, 1976. b. October 15, 1978.

**Reproduced with permission from the Lecture notes of Prof. John Jensen, University of South Carolina**

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 204

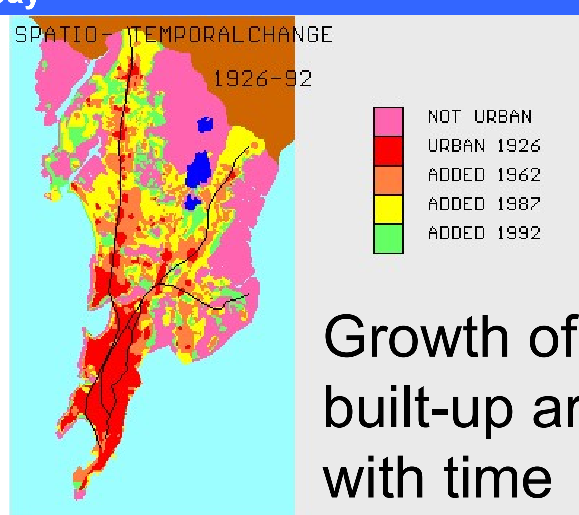
## selection of Change Detection Algorithm

- The selection of an appropriate change detection algorithm is very important.
- It will dictate whether "from–to" change information can be extracted.
- This is the need of many projects

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 205

## Visualization of Change Detection




Growth of built-up area with time

GNR607 Lecture 27-32 B. Krishna Mohan

## selection of the Change Detection Algorithm (from Jensen, 2004)

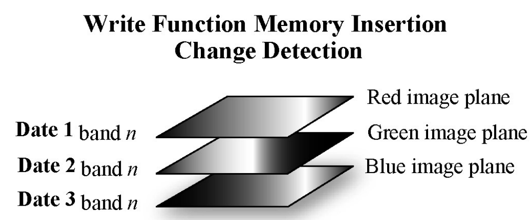
- ***Change detection algorithms*** **commonly used include:**
  - write function memory insertion
  - multi-date composite image
  - image algebra (e.g., band differencing, band ratioing)
  - post-classification comparison
  - binary mask applied to date 2
  - ancillary data source used as date 1
  - spectral change vector analysis
  - cross-correlation
  - visual on-screen digitization
  - knowledge-based vision systems.





## Write Function Memory Insertion Change Detection

- Load multidate imagery into memory, form image composites, identify changes happening on the test sites



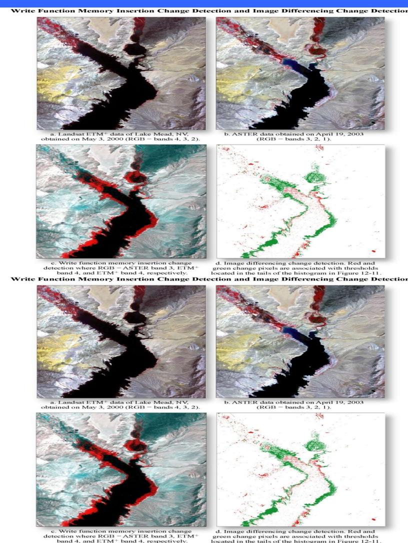

**Advantages**
* Visual examination of 2 or 3 years of nonspecific change
* Does not normally require atmospheric correction

**Disadvantages**
* Nonquantitative
* No "from–to" change class information

Reproduced with permission from the Lecture notes of Prof. John Jensen, University of South Carolina

IIT Bombay Slide 208

## Memory Insertion




Reproduced with permission from the Lecture notes of Prof. John Jensen, University of South Carolina

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 209

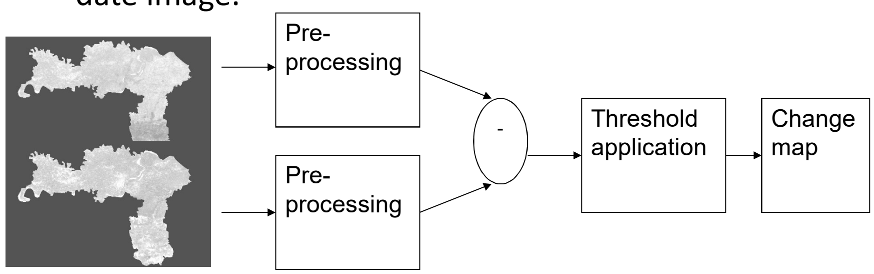
## Image Algebra Change Detection

- It is possible to identify the amount of change between two rectified images by *band ratioing* or *image differencing*.
- When 8-bit data are analyzed in this manner, the potential range of difference values
- 0 – 255 (ratioing) or
- –255 to 255 (differencing)

GNR607 Lecture 27-32 B. Krishna Mohan

## Image Differencing

- **Image differencing**
  This method is implemented by subtracting pixel-by-pixel the previous (first) date image from the recent (second) date image.








## Algebraic Techniques

- **Image regression**
  The pixel values of the second date image are calculated using the regression function producing the regressed image as the resultant. This regressed image is subtracted from the first date image. Large differences correspond to changes

- **Image ratioing**
  The ratio of the registered images of two dates is calculated band by band

IIT Bombay Slide 212

## Algebraic Change Detection

- **Vegetation index differencing**

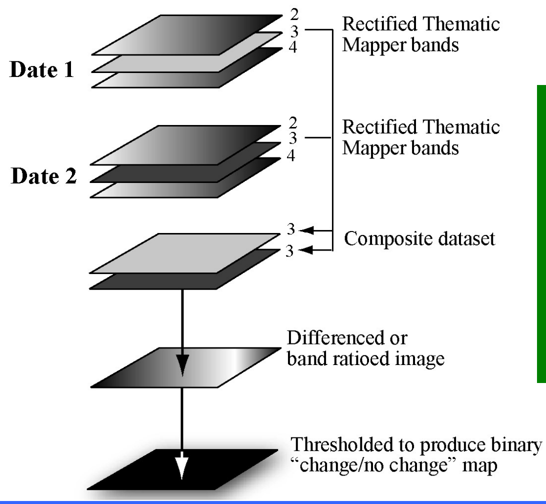
NDVI of data 1 – NDVI of date 2|

- Changes in vegetation cover correspond to larger differences of NDVI

GNR607 Lecture 27-32 B. Krishna Mohan

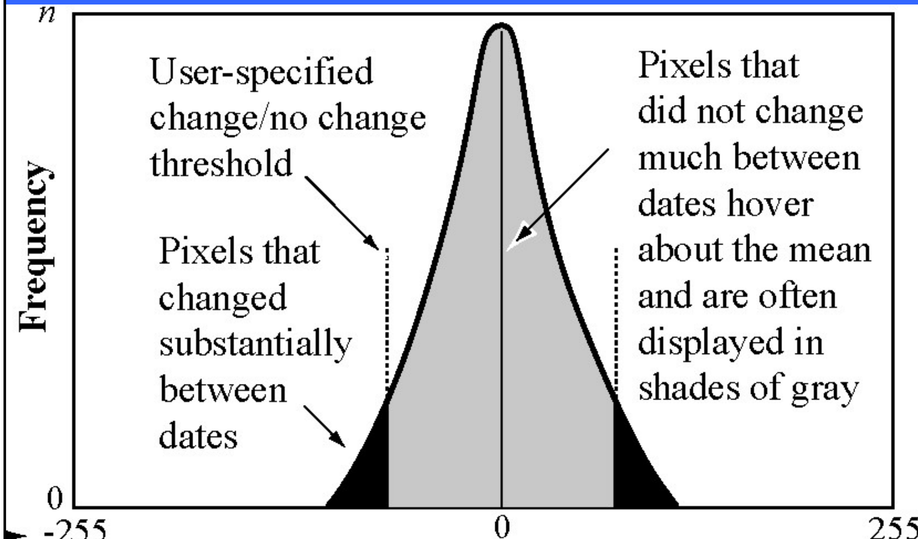
IIT Bombay Slide 213

**Image Algebra Change Detection**




**This figure and the figure in the next slide are reproduced with permission from the Lecture notes of Prof. John Jensen, University of South Carolina**

GNR607 Lecture 27-32 B. Krishna Mohan

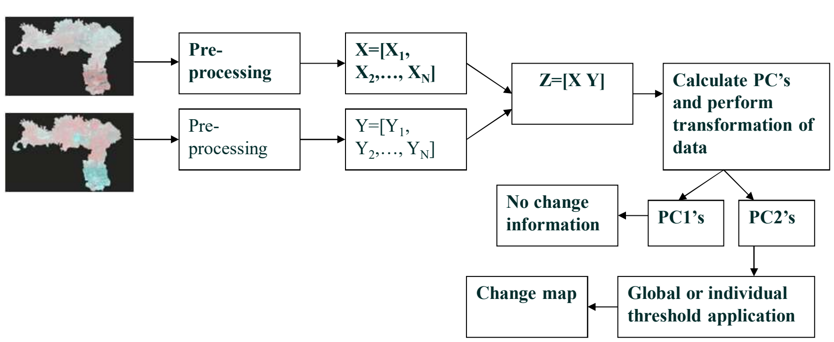







## TEMPORAL PRINCIPAL COMPONENT ANALYSIS

- In this method, a bi-temporal feature space is created. If there are N number of spectral bands in the images of two dates, then N change bands and N no change bands are produced.

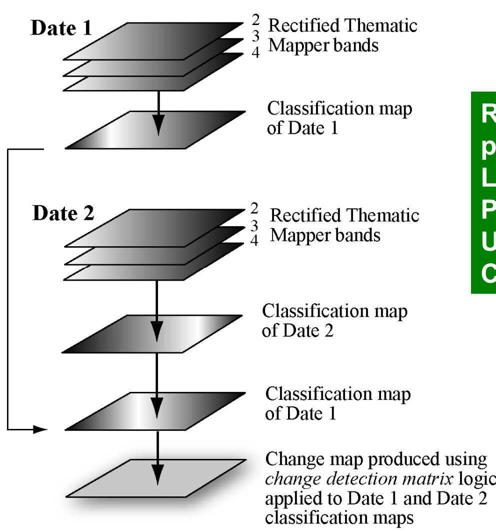


# Post-classification Comparison Change Detection

- *Post-classification comparison change detection* requires rectification and classification of each remotely sensed image.
- The two maps are then compared on a pixel-by-pixel basis using a *change detection matrix*.
- It is imperative that the individual classification maps used in the post-classification change detection method be as accurate as possible



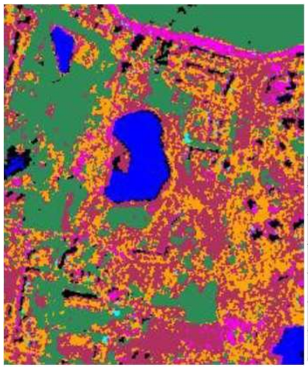




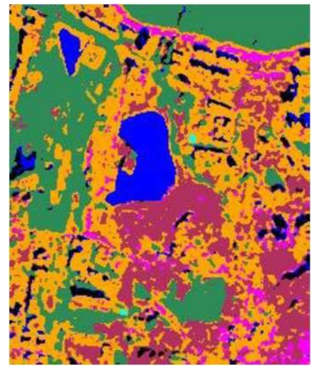

Reproduced with permission from the Lecture notes of Prof. John Jensen, University of South Carolina

## POST-CLASSIFICATION COMPARISON

SVM classification result of Quickbird window-I of Powai area.



SVM classification result of LISS-IV window-I of Powai area.



Legend for the classification result

| | |
|---|---|
| Lake | |
| Vegetation | |
| Roads | |
| Pool | |
| Shadows | |
| Open land | |
| Settlements | |

**Input parameters for SVM**
Kernel : Radial basis function
Gamma value in Kernel function : 0.333
Penalty parameter : 100
Number of Pyramid levels and classification probability threshold :0



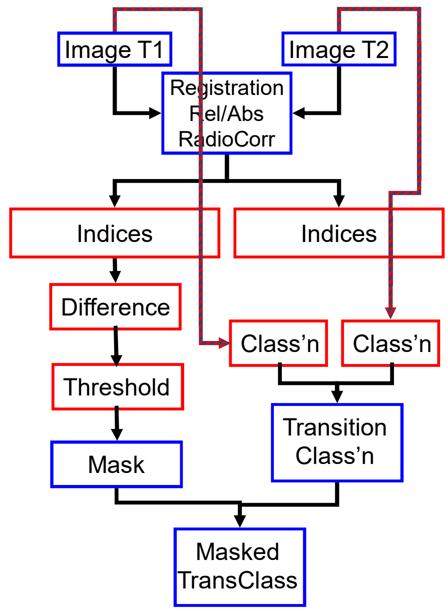


# Post-classification Change Detection

To: 2006

| From: 2000 | Water | Forest | Urban |
|---|---|---|---|
| Water | 1 | 2 | 3 |
| Forest | 4 | 5 | 6 |
| Urban | 7 | 8 | 9 |

**Reproduced with permission from the Lecture notes of Prof. John Jensen, University of South Carolina**




Reproduced with permission from the Lecture notes of Prof. John Jensen, University of South Carolina

## Change Detection Strategy





## When and How to Correct Atmospheric Effects?

- When is atmospheric correction necessary prior to image classification and change detection?
- *If* atmospheric correction is necessary, which method is best to use?

IIT Bombay Slide 222

## Types of Atmospheric Correction

- Absolute – Digital numbers converted to absolute refLecture ance
- Relative – The DNs in each image represent the same refLecture ance but are not themselves scaled to refLecture ance

GNR607 Lecture 27-32 B. Krishna Mohan

IIT Bombay Slide 223

## When is Atmospheric Correction not Necessary?

- When is radiometric correction *not* needed?
  - Single Data Classification
    - Radiometric correction only changes means, not covariance between bands
  - Post classification change detection
    - Make two classifications and compare-but this requires two training and two accuracy datasets
  - Classification of single multi-temporal dataset
    - Def: Combine multiple dates of imagery into single and classify total stack

GNR607 Lecture 27-32 B. Krishna Mohan

## When is atmospheric correction necessary?

- Image Differencing
  - To subtract un-corrected image of one date from image of another date
  - When computing and comparing NDVI
  - In practice, the DN threshold for change detection is often set empirically, but some algorithms assume that unchanging areas will have difference of zero





## Summary

- select ing a methodology for change detection requires a thorough understanding of the type of change you interested in.
- Methods that can largely bypass radiometric correction (e.g., Classification) are preferable

**Contd…**