CHAPTER 3

REMOTE SENSING IMAGE AND DATASET DESCRIPTION

3.1 INTRODUCTION

Remote sensing \cite{177} image processing is nowadays a mature research area. The techniques developed in the field allow many real-life applications with great societal value. For instance, urban monitoring, fire detection or flood prediction can have a great impact on economical and environmental issues \cite{110}. To attain such objectives, the remote sensing community has turned into a multidisciplinary field of science that embraces physics, signal theory, computer science, electronics, and communications. From a machine learning and signal/image processing point of view, all the applications are tackled under specific formalisms, such as classification and clustering, regression and function approximation, image coding, restoration and enhancement, source unmixing, data fusion or feature selection and extraction.

Remote sensing image processing \cite{96} is a mature research area allowing real-life applications with clear benefits for the Society. The main goals of remote sensing physical variables; and identifying materials on the land cover and analyzing the spectral signatures acquired by satellite or airborne sensors \cite{104}. Achievement of these objectives is possible because materials in a scene reflect, absorb, and emit are monitoring and modeling the processes on the Earth’s surface and their interaction with the atmosphere; measuring and estimating geographical, biological and electromagnetic radiation in a different way depending on their molecular composition and shape.
Remote sensing exploits this physical fact and deals with the acquisition of information [182] about a scene (or specific object) at a short, medium or long distance. Attending to the type of energy resources involved in the data acquisition, remote sensing imaging instruments can be passive or active [174]. Passive optical remote sensing relies on solar radiation as illumination source. The signal measured at the satellite by an imaging spectrometer is the emergent radiation from the Earth-atmosphere system in the observation. The radiation acquired by a (airborne or satellite) sensor is measured at different wavelengths and the resulting spectral signature (spectrum) is used to identify a given material [151], [156]. The field of spectroscopy is concerned with the measurement, analysis, and interpretation of such spectra. Some examples of passive sensors are infrared, charge-coupled devices, radiometers, or multispectral and Hyperspectral sensors. In active remote sensing, the energy is emitted by an antenna towards the Earth’s surface and the energy scattered back to the satellite is measured.

Radar systems, such as Synthetic Aperture Radar (SAR), are examples of systems for active remote sensing. In these systems, the time delay between emission and return is measured to establish the location, height, speed and direction of objects. The diversity of platforms and sensors implies a diversity and very articulated research area in which machine learning, signal and image processing [74] are very active. In fact, from a machine learning and signal/image processing viewpoint, all the applications are tackled under specific formalisms, such as classification and clustering or regression and function approximation. However, the statistical characterization of remote sensing images turns to be more difficult than in grayscale natural images because of pixel’s [128] higher dimensionality, particular noise and
uncertainty sources, the high spatial and spectral redundancy, and their inherently non-linear nature. It is worth to note that all these problems also depend on the sensor and the acquisition process. Consequently the developed methods for processing remote sensing images need to be carefully designed attending to these needs. Even if scientific production is high, the cross-fertilization between the remote sensing and the image processing communities is still far from being a reality.

3.2 REMOTE SENSING IMAGE SENSORS

Remote Sensing is a multifaceted technique and may vary based on the application and technologies development. Remote sensing [203] may be classified from many perspectives, for example based on platform, source of energy, number of bands and so on. Typically the remote sensors fall under the category based on energy source where there are two types such as:

ACTIVE SENSOR

On the other hand, the active sensor provides their energy source for illumination where the sensor emits radiation, which is directed towards the target to be investigated. The radiation reflected from that target is then detected and measured by the sensor. Active sensor [203] can be used for examining wavelengths that are not sufficiently provided by the sun, such as microwaves, or to better control the way a target is illuminated which is shown in the figure 3.1.

Figure 3.1   Active Sensors
PASSIVE SENSOR

Remote sensing systems which measure energy that are naturally available are called passive remote sensing [102]. Passive sensors can only be used to detect naturally occurring energy. Passive remote sensing within the optical region can only take place during the time when the sun is illuminating the earth, because the sun is the natural source of energy. Passive remote sensing is also possible in the microwave region. The figure 3.2 shows the passive sensors.

Figure 3.2 Passive Sensors

3.3 SPECTRAL RANGES AND INFORMATION

The electromagnetic spectrum [103], [105] is the distribution of electromagnetic radiation according to energy (or equivalently, by virtue of the relations in the previous section, according to frequency or wavelength).

Figure 3.3 Electro Magnetic Spectral Radiation
Thus one can see that visible light and gamma rays and microwaves are really the same things. They are all electromagnetic radiation; they just differ in their wavelengths. The figure 3.3 shows the electromagnetic spectral radiation.

3.4 REMOTE SENSING DATA

Remote sensing data [112] is possible to collect information about an object or geographic area using specialized instruments without direct contact with the object and also without going to the study area. In remote sensing information transfer is accomplished by the use of electromagnetic Radiation. Electro Magnetic Radiation is a form of energy that reveals its presence when it strikes the object. Receiving this energy for interpretation is called sensing [121]. The EMR emitted or reflected from an object is the common source of remote sensing data. The majority of remotely sensed data collected for earth resource are the results of sensors that record electromagnetic energy. This remote data collection was originally performed using aerial cameras[123]. But nowadays satellites are the main platforms for measuring the physical properties of earth digitally by using electronic sensors.

3.5 CHARACTERISTICS OF IMAGES

It is important to distinguish between the terms ‘images’ and ‘photographs’ in remote sensing. An image refers to any pictorial representation of what imaging media or wavelength or remote sensing devices have been used to detect and record the electromagnetic energy. Taking specifically about an image recorded photographically use the term ‘image’. Images that are captured on digital media are called digital image [115]. Many digital image processing techniques rely on the
digitally recorded pixel value. In a scanned photograph the pixel value represents the reflected amount of light from the photograph and not from the original object. Thus it differs from the originally captured digital image.

3.6 DATA USED IN REMOTE SENSING

DIGITAL DATA

Analog images are the images that have continuous color or gray tone. On the other hand, a group of divided petite cells, with numeral values of average intensity, the center representing the cell's value, is called a digital image [43]. It has coordinates of pixel number which is counted from left to right and when it is counted from top to bottom it is the line number. Sampling of pixel size [1], [9], [43] or sampling frequency is one of the most important factors. The optimum sampling is considered because when the sampling frequency is long or the pixel size is large the visualization of the image becomes worsened. While considering the reverse case the data volume in the image becomes very large.

GROUND DATA

Ground data [13] are otherwise called as ground “truth” which is the measurement, collection and observation of information about the real conditions on the ground in order to determine the relationship between remote sensing data and the object to be observed. Land truth called as the investigation of land. Generally ground data should be collected at the same time as data acquisition by the remote sensor, or at least within the time that the environmental condition does not change. It must not
be inferred that the use of the word "truth" implies that ground truth data are not without error. Analysis and data correction are the two applications which are for supplemental purpose. Analysis is for example investigation of a test area and collection of sample data for classification process. For the geometric correction a survey of ground control points is made which is called Data correction [65]. Ground data will mainly include identification of the object to be observed, and measurement is done using a spectrometer, as well as visual interpretation of aerial photographs and survey by existing maps, and a review of existing literature and statistics.

3.7 GROUND TRUTH DATA

Remote sensing is an efficient way to gather a large amount of information from large areas without actually going there. Ground data [40],[112],[116],[183] also called ground truth is defined as the observation, measurement and the collection of information about the actual conditions on the ground in order to determine the relationship between the remote sensing data and the object to be observed [150]. Ground truth refers to the gathering reference data on-site and deriving information then that property characterizes states, conditions, and parameters associated with the surface. Classification is identifying pixels to categorize it such as vegetation, urban etc [66]. Other than the image classification there are several reasons for which the ground truth data are collected.

3.7.1 REQUIREMENTS OF GROUND TRUTH DATA

Ground truth [10] data are generally required for a full and accurate characterization of ground features from remotely sensed data. The purpose in
accruing ground truth is ultimately to aid in calibrating and interpreting remotely recorded images by checking ground realities within the scene. Generally ground data should be collected at the same time as data acquisition by the remote sensor, or at least within the time that the environmental conditions do not change. It should not be inferred that the use of the word ‘truth’ implies that ground truth data are without error. Calibration and validation [40], [112], [146], [177] involve calibration of sensor as well as captured data. Calibration of sensor is performed to obtain a standard and desired reflectance in the output image for a given material. Calibration of data is performed to match the pixel value to the original reflectance of the object and validate the data for analysis. There are two applications, namely analysis and data correction. The former one is the ground investigation at a test area to collect the training sample data for classification and post-classification accuracy assessment [204]. The latter case is the survey of ground control points for geometric correction.

The terms to be investigated by ground truth data are as follows:

a. Information about the object type (e.g. vegetation), spectral characteristics (reflectance at different bands for a given material), circumstances, surface temperature, etc.

b. Information about the environment such as the elevation, irradiance, atmospheric clarity, air temperature, humidity, wind direction, ground surface condition, etc.

3.7.2 PARAMETERS OF GROUND TRUTH DATA

In order to determine specific conditions of land cover types such as vegetation, subsurface contamination, water quality criteria ground-truth [10],[77]
data collection is a necessity. Once it has been determined that ground-truth or vegetation data [78], [84], [92] are required with the following details.

- Atmospheric conditions
- Surface water
- Vegetation
- Soil
- Bare ground
- Rock
- Dark and light calibration targets

### 3.8 INFORMATION EXTRACTION IN REMOTE SENSING

In remote sensing, information extraction is mainly categorized into five types, namely Classification, Change Detection, Extraction of Physical Quantities, Extraction of Indices, and Identification of Specific Features [78],[119], [121]. Classification is a categorization of image data which uses spectral, temporal and spatial information. Change detection [23], [79], [91] is mainly associated with the extraction of change between multi-data images. Extraction of physical qualities is the measurement of elevation, atmospheric constituents and elevation from stereo information or spectral information. Extraction of indices is the calculation of a recently defined index, for example, one can consider the vegetation index from satellite data. Identification of specific features is the identification of disaster, archaeological, lineaments etc. Information extraction can be done by human or computer methods.
3.9 IMAGE ENHANCEMENT AND FEATURE EXTRACTION

Image enhancement [104], [114], [143], [147] is a technique of improvisation of an image into a better quality image for better feature extraction or image interpretation. The physically calibrated value is reconstructed which is termed as radiometric correction from the observed data. Feature extraction [12], [13],[160] can be defined as the process in which various parameters or functions are being applied to original image to quantify the image quality. These procedures can be considered as conversion of the image data. Image enhancement is mainly applied for image interpretation in the form of an image output and feature extraction is usually used for automated classification or analysis in a quantitative form.

3.10 ELEMENTS OF VISUAL IMAGE INTERPRETATION

There are a number of characteristics that enable the viewer to detect, recognize or even identify objects from the imagery [6], [36], [38], [46], [179]. These recognition elements are shape, size, pattern, shadow, color, texture and association. The common adjectives are given below.

<table>
<thead>
<tr>
<th>Elements</th>
<th>Common adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Latitude , Longitude and altitude</td>
</tr>
<tr>
<td>Size</td>
<td>Length, width, perimeter and area</td>
</tr>
<tr>
<td>Shape</td>
<td>Circular, elliptical, square, triangle</td>
</tr>
<tr>
<td>Shadow</td>
<td>Silhouette by solar illumination</td>
</tr>
<tr>
<td>Color</td>
<td>Grey, tone: Light (Bright), Intermediate (grey), Dark (black)</td>
</tr>
</tbody>
</table>
Color: IHS = Intensity, Hue (color), saturation;
RGB = Red, Green, Blue

6. Texture           Tone
7. Pattern           Objects on the Ground.

3.11 COLOR COMPOSITE OF IMAGES

A color image can be generated by compositing three selected bands of multi-band images with the use of three primary colors [58], [62], [75], [77]. Different color images may be obtained depending on the selection of three-band images and the assignment of the three primary colors. There are two methods of color composite; an additive color composite and a subtractive color composite. Additive color composite uses three light sources of three primary colors (RGB) for instance. The subtractive color composite uses the pigments of three primary colors (cyan, magenta, and yellow). Multi-spectral images generally contain more than three spectral bands. But one can view a maximum of three bands of any multi-spectral image at a time. This implies that several bands of a multi-spectral [80], [93], [135] image are invisible to human beings. But one can view those commonly invisible bands using the color combination technique. True color also called as natural color is a combination when image captured in blue band is passed through the blue color gun of a display device, green band through the green color gun and red band through the red color gun. In some cases especially where at least three bands in the visible region are not available, some bands which are out of visible region are mathematically combined in such a way that the appearance of the image resembles a visible normal color
appearance. Reasonable good natural color composite can be produced by the following combinations.

RED color gun = Red
GREEN color gun = (3 x Green + NIR) / 4 = 0.75 x Green + 0.25 x NIR
BLUE color gun = (3 x Green – NIR) / 4 = 0.75 x Green – 0.25 x NIR

3.11.1 COLOR DISPLAY OF IMAGE DATA

In effective visual interpretation of remote sensing data color display plays an eminent role. Color composite and Pseudo-color [58], [62], [75], [77],[149] are the two important color display methods where color composite is used to generate color with multi-band data and the latter to assign different colors to the gray scale single image.

COLOR COMPOSITE

A color image can be produced by composing three chosen multi-band images with the use of three primary colors [75], [88], [149]. Different color images are acquired based on the selection of three band images and the assignment of the three primary colors. The two color composite methods are Additive and Subtractive color composite. While the three primary colors light sources are being used like the color graphic display or multispectral viewer which is also called the Additive color composite. When the three pigments of three primary colors Cyan, Magenta and Yellow are used, for example, in color printing it is called subtractive color composite. When three filters of B, G and R are allocated to the similar spectral
regions of blue, green and red, about the same color as the natural scale is reproduced which is called a natural color composite. However in remote sensing multi-band images are not always separated in to the same spectral regions [170] as in the three primary color filters. In addition invisible regions, such as infrared, are often used, which are required to be displayed in color. As a color composite with an infrared band is not a natural color it can be called a false color composite.

3.11.2 COLOR REPRESENTATION-COLOR MIXING SYSTEM

Color stimulus is the light perceived as color by human eye which corresponds to the visible region of electromagnetic spectrum which has a specific spectral curve of radiance [154]. Psychological representation is considered more practical because the physical value like the spectral curve is not suitable for symbolizing color in daily life. Color mixing system and color appearance system are the two types of color representation. Color mixing system uses quantitative and physical approach while color appearance system uses the qualitative approach by color sample or color code.

Color combinations generated in such a manner is called as false natural color composites. Combinations other than true color combinations are known as false combinations. Particularly the color composite with the assignment of blue color gun to the green band, green gun to the red band, and red gun to the NIR band is very popular and is called Infrared color composite [58] which is the same as that found in color infrared films. Color representation can be classified into two types; a color mixing system using a quantitative and physical approach, and a color appearance system using a qualitative approach, by color code or color sample. The color mixing
system can generate any color by mixing of the three primary colors. The RGB color system specified by CIE uses three primary color stimuli; blue, green and red which all spectral colors ranging can be generated with the mixing combinations which is termed as color matching function.

3.11.3 FEATURE EXTRACTION

In the recent trends there are varied applications that are able to extort the precise information from the colored image database [14], [49], [51], [58]. There are various kinds of images which have their own semantics. Based on the content of images, different feature extraction techniques are available for information extraction. Commonly three properties are being considered – intensity, texture and color. Feature extraction is a method of capturing visual content of images for indexing and retrieval. Feature extraction is used to denote a piece of information which is relevant for solving the computational task related to certain systems where it is applied. Feature extraction [64], [80] includes structures and facilities (buildings, bridges, airports, etc.), water bodies, land cover, vegetation classification (agriculture, forestry, environmental), well locations, land use, pipelines, soil types and more.

3.11.4 DESCRIPTION OF FEATURE EXTRACTION

Feature extraction, in the context of remote sensing, can be defined as image processing techniques to identify and to classify mutual relationships or mutual meaning between image regions [12], [13], [38], [49], [51], [58]. The aggregation of image pixels forming image regions and their relationship to other image regions are interpreted and used as cues in the information retrieval process. A common approach
is to create hierarchical structures of image regions in which fine-scale image regions constitute portions of other coarse-scale image regions. Feature extraction differs from traditional pixel-based remote sensing [85],[97], [173] image classification algorithms in which each, individual pixel (or pixel vector in the case of images with more than one channel) is individually evaluated and assigned to one class. The difference between low-level information extraction techniques using traditional pixel-based classification methods and high-level information extracted by a human analyst is often referred to as the “semantic gap”.

Human analysts use a complex combination of different image cues such as colour, texture, shape, and context [27], [62],[124], [138], [173]. However, human analysis of large areas and multiple images is costly and time-consuming. As the volume of available remotely sensed imagery increases by many orders of magnitude, one of the challenges faced by many organizations and institutions is converting large quantities of images into actionable information and intelligence. Because human analysis of large areas and sometimes over multiple periods of time is costly and time consuming, scientists have recognized the importance of developing more sophisticated semi-automated or automated feature extraction techniques to improve the information extraction process.

The challenge resides in multifaceted problems where the relationship between image’s regions is too complex to be solved by explicit programming and these problems require the system to adapt and evolve when image conditions change. This provision is particularly important in remote sensing applications due to changing factors such as variation in sensor spatial and spectral resolutions, change in
environmental conditions between images, and specificity of the feature of interest. The use of stochastic algorithms to address these complex feature extraction problems are now being investigated as a possible alternative; due to their properties of deriving solutions from a small set of positive and negative examples through an optimized combinatorial search rather than being explicitly programmed. Evolutionary algorithms have been used to solve problems in different domains, including remote sensing applications [158], [176], [178]. The use of existing indices in new environments or the development of new spectral indices constitutes a time-consuming and complex problem. Different features with similar spectral signatures add to the complexity of creating such indices. Spectral indices level of complexity varies according to the relationship between feature’s spectral responses to different parts of the electromagnetic spectrum.

3.12 FUZZY LOGIC IN REMOTE SENSING IMAGE CLASSIFICATION

Fuzzy set theory was proposed by Zadeh in 1965 to treat the fuzziness in data [1], [14],[34]. In fuzzy set theory the membership grade the value considered is an intermediate value between 0 and 1 though in normal case the value considered is 0 or 1 only. The function of the membership grade is also called its "membership function" in Fuzzy theory [35], [148]. The membership function is defined by the user in deliberation of the fuzziness.

In remote sensing it is not easy to define the boundary between two different classes. For example one can consider that there is transitive vegetation or mixed vegetation between forest and grass land. In such cases fuzzy set theory [154] can be used in a qualitative sense to define class boundaries. Let the membership function be
Mf (k) of class k (k=1, n), the likelihood Lt of fuzzy class’ f can be defined as follows.

\[ Lt = \sum_{i=1}^{n} \{Mf(k)^* P(X | k)^* P(k)\} \cdot (P(i)^* P(X | i)) \]

In this classification, the concrete structure with clearly defined characteristics was first classified using the ordinary maximum likelihood [22], [148] method, while the loosely defined urban classes were classified by the fuzzy-based maximum likelihood method.

### 3.13 K-MEANS CLUSTERING IN REMOTE SENSING IMAGE

Starting from some limitations of the state-of-the-art algorithms, an unsupervised non-linear feature extraction K-Means [5], [8], [24], [83] Clustering algorithm is used in the research work. Hyperspectral data and Multispectral data have been taken for spectral feature extraction [101], [168]. Experiments on real hyperspectral [24], [25], [66] data are conducted for the purpose of classification. K-Means Clustering is used to classify the data using the extracted features. Experimental results have shown the effectiveness of the non-linear approach, which makes it possible to extract more informative features.

#### 3.13.1 CLASSIFICATION BASED ON K-MEANS ALGORITHM

A classification unit could be a pixel [43], a group of neighboring pixels or the whole remote sensing image. It refers to the use of spatial, temporal, and other related information, in addition to the spectral information of a classification unit in the classification of an image. In order to illustrate the differences between the supervised
and unsupervised classification, it is introduced with two concepts: information class and spectral class. Spectral class is a class which includes similar grey-level vectors in the multispectral space [80], [106], [116], [123]. One of the differences between a supervised classification and an unsupervised one is the ways of associating each spectral class to an information class. For supervised classification, first start with specifying an information class on the image. An algorithm is then used to summarize multispectral information from the specified areas on the image to form class signatures. This process is called supervised training. For the unsupervised case, however, an algorithm is first applied to the image and some spectral classes (also called clusters) are formed. The image analysts then try to assign a spectral class to the desirable information class. The following diagram 3.4 represents the system flow architecture of the K-Means Clustering Algorithm.

![Diagram of K-Means Clustering Algorithm]

**Figure 3.4 System Architecture of K-Means Clustering Algorithm**
3.13.2 K-MEANS CLUSTERING ALGORITHM

The K-Means algorithm [5], [8], [24], [83], [101], [168] is one of a group of algorithms called partitioning methods. The K-Means algorithm is very simple and can be easily implemented in solving many practical problems. The K-Means algorithm is the best-known squared error-based clustering algorithm. Consider the data set with ‘n’ objects, i.e. \( S = \{x_i, 1 \leq i \leq n\} \).

(a) Initialize a k-partition randomly or based on some prior knowledge. i.e. \( \{C_1, C_2, \ldots, C_k\} \)

(b) Calculate the cluster prototype matrix \( M \) (distance matrix of distances between k-clusters and data objects). \( M = \{m_1, m_2, m_3, \ldots, m_k\} \) where \( m_i \) is a column matrix \( 1 \times n \).

(c) Assign each object in the data set to the nearest cluster - \( C_m \)

i.e. \( X_j \rightarrow C \)

If \( \| x_j - C_m \| \leq \| x_j - C_i \| \) for \( 1 \leq j \leq k, j \neq m \)

Where \( j=1, 2, 3 \ldots n \).

(d) Calculate the average of each cluster and change the k-cluster centers by their averages.

(e) Again calculate the cluster prototype matrix \( M \).

(f) Repeat steps c, d, and e until there is no change for each cluster, and have the closest centroid;
Calculate new mean for each cluster until convergence criteria are met. As a result of this loop, the k- centroids may change their position in step-by-step manner. K-Means algorithms solve clustering problem. The procedure follows to classify a given data set through a certain number of clusters assumed as K-Clusters. The main idea is to define K-Centroids, one for each cluster.

These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroids. When no point is pending, the first step is completed and an early group age is done. At this point there is a need to re-calculate ‘k’ new centroids as bar centers of the clusters resulting from the previous step. After having these ‘k’ new centroids, a new binding has to be done between the same data set points and the nearest new centroids. A loop has been generated. As a result of this loop the K-Centroids change their location step by step until no more changes are done. In other words centroids do not move anymore. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function.

The objective function:

\[ \text{Objective function:} \]

Where \( d \) is a chosen distance measure between a data point \( x^i_j \) and the cluster centre \( c_j \) is an indicator of the distance of the ‘n’ data points from their respective cluster centres.
1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.

2. Assign each object to the group that has the closest centroid.

3. When all objects have been assigned, recalculate the positions of the K-Centroids.

4. Repeat second and third points until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

### 3.13.3 PSEUDO-CODE FOR K-MEANS ALGORITHM

Take ‘k’ number of clusters and ‘n’ number of samples in a multi-dimensional space. Here, ‘k’ number of iterations is followed for K-Centroids to obtain optimal clusters, at each iteration, and the solution is constructed. Finally, the best optimal solution from iteration and average time is calculated at the end. The rule mentioned below is followed for algorithm construction.

Step 1: Initialize all ‘k’ number of clusters and ‘n’ number of samples.

Step 2: Iteration (i) <= k

Step 3: Randomly select one centroid c_j, where j=1<j<k.

Step 4: Method1<=n

Step 5: Randomly choose one object o_i, where i=1<i<n.

Step 6: Object i on cluster j represent as o_{ij}. 

95
Step 7: Result of $o_{ij}$ is either 0 or 1

  If $o_{ij}=1$ means, $i$ belongs to cluster $j$.

  If $o_{ij}=0$ means, $i$ belongs to some other clusters.

Step 8: Calculate mean algorithm by $K_j=1/n_i(o_{ij})c_{ij}$

Step 9: Repeat Step 4 to Step 8 until method reaches in ‘n’ sample.

Step 10: Recalculate the position of centroid by $C_j=1/n_i(K_j*c_{ij})$.

Step 11: Repeat Step 2 to Step 10, upto centroids.

Step 12: Find the optimal clusters from the result of iteration by $\min_j o_{ij}|K_j - C_j|^2$.

3.13.4 EXPERIMENTAL RESULTS

Clustering is an important task with applications in many fields. In K-Means algorithm at different runs it produces some desired results when the initial centroids are chosen randomly. It is important to realize that the choice of the initial centroid has a huge effect on the final result. K-Means algorithm for multiple runs on large data sets is not work and it takes more time to complete [101], [168]. K-Means algorithm is applied to this data sets and clustering group of data by different colors. The data sets contain six numerical attributes with values between 0 and 255. Each data set contains ‘n’ points with the points centered around ‘k’ cluster centers. The ‘k’ cluster centers are first allocated randomly and each of the six attributes values from a range of 0 to 255 is uniformly allocated in K-Means algorithm. To calculate the minimum distance (D) between two cluster points is defined by $D/r$, where ‘r’ is a variable used to define the cluster [19], [37]. This process is repeated until it reaches
data points and clusters. The results of the experiment have been shown with Kappa analysis, Execution time and confusion matrix is being used in analyzing and testing the results acquired from K-Means Clustering Algorithm and the classified remote sensing image is also being given.

(i) **LANDSAT 7 ETM+**

The first dataset is a portion of Landsat-7 ETM+ image [27], [92] (bands 1, 2, 3, 4, 5 and 7) acquired over the west of haerbin, Heilongjiang, China on August 11, 2001. This site mainly contains two land cover types, which are vegetation and exposed land. In this result analysis, vegetation is represented by green, dark areas represent dry land, and slightly darker green area represents grass land and deep darker area represents paddy fields. A brown area mainly represents exposed land; slightly brown area mainly symbolizes the dry salt flats which are blocked by forest land and dry land that are highly mixed too. The figure 3.5 is the image of Landsat-7 ETM+.

![LANDSAT-7 ETM+ Image](image)

**Figure 3.5** LANDSAT-7 ETM+ Image
The Landsat-7 ETM+ image is classified with K-Means clustering algorithm. The Landsat-7 ETM+ image dataset consists of 670 instances which are group for classification. The result of the classified images is shown in figure 3.6.

Figure 3.6 Results of Landsat-7 ETM+ Image Using K-Means Algorithm
The K-Means clustering algorithm classifies the Landsat-7 ETM+ image into five classes. The five classes are labeled as Paddy field, forest, grass land, dry flat salts and dry land. These classes are classified according to the feature extraction based on color. The class labels are measured in terms of Overall Accuracy, Producer Accuracy and User Accuracy. The Landsat-7 ETM+ image is classified with K-Means clustering algorithm and the output of the classified image is shown in figure 3.7.

The K-Means Clustering algorithm produces the producer accuracy for the class Paddy field is 80.91%, forest is 56.16%, grass land is 63.87%, Dry salt flats is 57.93%, and dry land is 70.81% and the User accuracy for the Paddy field is 69.28%, forest is 43.15%, grass land is 56.90%, Dry salt flats is 91.25% and the Dry land is 77.98%. These classes values can be accurately measured using the statistical method like Kappa analysis and the performance of the algorithm is measured with Execution time. The remaining part leading to classification is termed as misclassification, which is measured using Error matrices.
The portion of remote sensing image is Quick Bird [72] image which covers a small area of the south part of the city of Trento, Italy acquired on July 17, 2006. In this image vegetation is represented as green, dark area represents dry land, slightly darker green area represents grass land and deep darker area represents paddy fields. The brown area mainly represents exposed land; slightly brown area symbolizes the dry salt flats which are blocked by forest land and dry land that are highly mixed too. The figure 3.8 shows the Quick Bird Image.

Figure 3.8  QUICK BIRD Image
The above image represents the Quick Bird image which is to be classified using K-Means clustering Algorithm. The Quick Bird image is classified with K-Means clustering algorithm and produces different spectral classes. This image dataset consists of 670 instances which are categorized for classification. The result of the classified image is shown in the figure 3.9.

Figure 3.9  Results of Quick Bird Image Using K-Means Algorithm
Figure 3.10   Results of the Classified Quick Bird Image Using K-Means Algorithm

The K-Means clustering algorithm classifies the Quick Bird image into seven classes. The seven classes are labeled as agricultural field, road, tree, soil, roof, shadow and grass. These classes are classified according to the feature extraction based on color. The class labels are measured in terms of Error matrices. The Error matrices for the classification of the images shows the Overall Accuracy, Producer Accuracy and User Accuracy which shows the percentage of classification done by the K-Means algorithm. The Quick Bird Image is classified with K-Means clustering algorithm and the output of the classified image is shown in figure 3.10.

The producer accuracy of K-Means for the class Agricultural field is 64.00%, road is 92.52%, Tree is 71.11%, Soil is 69.77%, Roof is 53.73%, Shadow is 80.18%, Grass is 81.08% and the K-Means produces the user accuracy for the class Agricultural field is 79.34%, road is 84.62%, Tree is 44.45%, soil is 58.25%, roof is 82.76%, shadow is 72.36% and grass is 63.83%. These classes values can be accurately measured using the statistical method like Kappa analysis and the performance of the algorithm is measured with the help of Execution time. The remaining part leading to classification is termed as misclassification, which is measured using Error matrices.
(iii) MODIS IMAGE

The Modis [131], [171] (bands 1, 2,3,4,5 and 7) is acquired over the west of Chang Chum, Jilin, China on June 12, 2008. This site mainly contains two land-cover types which are vegetation and exposed land. In this analysis, vegetation is represented by green, dark area is represented as dry land, and slightly darker green area represents grass land and deep darker area represents paddy fields. The brown area mainly represents exposed land; slightly brown area mainly symbolizes the dry salt flats which are blocked by forest land and dry land that are mixed too. The figure 3.11 shows the Modis Image.

Figure 3.11 MODIS Image
The Modis image is classified with K-Means clustering algorithm. This image dataset consists of 670 instances which are categorized for classification. The result of the classified Modis image using K-Means clustering algorithm is shown in figure 3.12.

Figure 3.12 Results of Modis Image Using K-Means Algorithm
The K-Means clustering algorithm classifies the Modis image into seven classes. The seven classes are labeled as wet land, river, cultivated land, open grass land, bare soil, dry salt flats, and grass land. These classes are classified according to the feature extraction based on color. The class labels are measured in terms of Error matrices. The Error matrices for the classification of the images shows the Overall Accuracy, Producer Accuracy and User Accuracy which shows the percentage of classification done by the K-Means algorithm. The Modis Image is classified with K-Means clustering algorithm and the output of the classified image is shown in figure 3.13.

The producer accuracy of K-Means for the class agricultural field is 72.10%, road is 73.86%, tree is 70.72%, soil is 74.55%, roof is 69.64%, shadow is 90.20% and grass is 93.15%. The User accuracy produced by the K-Means clustering algorithm for the class agricultural field is 59.70%, road is 79.27%, tree is 73.99%, soil is 95.35%, roof is 76.47%, shadow is 77.97% and grass is 80.00%. These classes values can be accurately measured using the statistical method like Kappa analysis and the performance of the algorithm is measured with the help of Execution time. The remaining part leading to classification is termed as misclassification, which is measured using Error matrices.
(iv) ASTER IMAGE

The portion of this Aster image (bands 1, 2, 3, 4, 5 and 7) acquired over the west of Haerbin, Heilongjiang, China on August 14, 2005. This site mainly contains two land-cover types which are vegetation and exposed land. The figure 3.14 shows the Aster Image.

Figure 3.14 ASTER Image
The Aster image is to be classified with K-Means clustering algorithm. This image dataset consists of 670 instances which are categorized for classification. The result of the classified Aster Image is shown in figure 3.15.

Figure 3.15 Results of Aster Image Using K-Means Algorithm
The K-Means clustering algorithm classifies the Aster image into five classes. The five classes are labeled as paddy field, forest, grass land, dry salt flats and dry land. The class labels are measured in terms of Error matrices. The Error matrices for the classification of the images shows the Overall Accuracy, Producer Accuracy and User Accuracy which shows the percentage of classification done by the K-Means algorithm. The Aster image is classified with K-Means clustering algorithm and the output of the classified image is shown in figure 3.16.

The producer accuracy of K-Means for the class Paddy Field is 80.91%, Forest is 56.75%, Grass Land is 62.66%, Dry Salt Flats is 58.54% and Dry Land is 70.81% and User accuracy for the class Paddy Field is 69.28%, Forest is 43.75%, Grass Land is 55.62%, Dry Salt Flats is 91.46% and Dry Land is 77.97%. These classes values can be accurately measured using the statistical method like Kappa analysis and the performance of the algorithm is measured with the help of Execution time. The remaining part leading to classification is termed as misclassification, which is measured using Error matrices.
SPOT-5 IMAGE

The portion of Spot-5 (bands 1, 2,3,4,5 and 7) image dataset is acquired over the oil field, Carthage, East Texas in the year 2004. This site mainly contains two land cover types, which are vegetation and exposed land. In this analysis, vegetation is represented by green, dark areas represent dry land, and slightly darker green area represents grass land and deep darker area represents paddy fields. A brown area mainly represents exposed land; slightly brown area mainly symbolizes the dry salt flats which are blocked by forest land and dry land that are mixed too. The figure 3.17 shows the image of Spot-5.

Figure 3.17 SPOT-5 Image
The Spot-5 image is to be classified with K-Means clustering algorithm. This image dataset consists of 670 instances which are categorized for classification. The result of the classified Spot-5 image is shown in the figure 3.18.

Figure 3.18 Results of Spot-5 Image Using K-Means Algorithm
Figure 3.19 Results of Classified Spot-5 Image Using K-Means Algorithm

The K-Means clustering algorithm classifies the SPOT image with five classes. The five classes are labeled as paddy field, forest, grass land, dry salt flats and dry land. These classes are classified according to the feature extraction based on color. The class labels are measured in terms of Error matrices. The Error matrices for the classification of the images show the Overall Accuracy, Producer Accuracy and User Accuracy which shows the percentage of classification done by the K-Means algorithm. The SPOT-5 image is classified with K-Means clustering algorithm and the output of the classified image is shown in figure 3.19.

The producer accuracy of K-Means for the class Paddy Field is 64.00%, Forest is 62.52%, Grass Land is 59.25%, Dry Salt Flats is 70.58% and Dry Land is 68.24% and User accuracy for the class Paddy Field is 84.95%, Forest is 76.74%, Grass Land is 43.83%, Dry Salt Flats is 53.93% and Dry Land is 82.11%. The remaining part leading to classification is termed as misclassification, which is measured using Error matrices.
3.13.5 PERFORMANCE ANALYSIS

KAPPA ANALYSIS

Kappa coefficients are widely used as Classification accuracy assessment for remote sensing data.

\[
\text{Accuracy} \% = \frac{\text{Kappa coefficients}}{\text{Error Matrix}}
\]

Kappa coefficients are widely used as Classification accuracy assessment for remote sensing Image dataset. The result of performing kappa analysis is a KHAT statistic (an estimate of Kappa), which is computed as

\[
KHAT = \frac{m^2 - 1}{m^2 + \sum_{i=1}^{r} (x_{ii} - \frac{N^2}{r^2})}
\]

Where \( r \) is the number of rows in the error matrix (also called the confusion matrix), \( x_{ii} \) is the number of observations in row \( i \) and column \( i \), \( X_{ii} \) and \( X_{ii}^* \) are the marginal totals of row \( i \) and column \( i \), respectively, and \( N \) is the total number of observations. Similarly, Cohen’s kappa is utilized as measure of agreement between the original images. Each Pixel is classified by its luminance, whose range is 0 to 255. Consequently, the error matrix is of the size \( r \times r \), where \( r = 256 \). Pixels of the original image are defined as reference data, where pixels of the image as classified data. The overall performance measures shown in the following table 3.1.
Table 3.1 Performance Analysis of K-Means Clustering Algorithm

<table>
<thead>
<tr>
<th>IMAGE DATASET/PERFORMANCE</th>
<th>KAPPAVALUE (%)</th>
<th>OVERALL ACCURACY (%)</th>
<th>EXECUTION TIME (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LANDSAT-7 ETM+</td>
<td>75.21</td>
<td>67.16</td>
<td>61.00</td>
</tr>
<tr>
<td>QUICK BIRD</td>
<td>73.00</td>
<td>71.30</td>
<td>60.00</td>
</tr>
<tr>
<td>MODIS</td>
<td>75.00</td>
<td>75.80</td>
<td>56.00</td>
</tr>
<tr>
<td>ASTER</td>
<td>72.00</td>
<td>66.80</td>
<td>60.00</td>
</tr>
<tr>
<td>SPOT-5</td>
<td>57.00</td>
<td>56.10</td>
<td>64.00</td>
</tr>
</tbody>
</table>

The table 3.1 shows the performance analysis of K-Means algorithm. The performance analysis is measured with the Kappa value, Overall accuracy and Execution time and with different image datasets like Landsat-7 ETM+, Quick Bird, Modis, Aster and Spot-5 Images. The table shows that the Kappa value is high for Landsat 7ETM+ which shows 75.21 % whereas it is low for Spot-5 Image and the Overall accuracy is high for Modis Image which shows 75.80 % whereas it is low for Spot-5 Image and Execution time is less for Modis Image which shows 56.00 ms and high for Spot-5 Image.
Figure 3.20  Graphical Representation of Performance Analysis Using K-Means Clustering Algorithm

The figure 3.20 shows the graphical representation of the performance analysis of K-Means clustering algorithm for the given Image datasets mentioned in the table 3.1. The graph shown that in x axis Image datasets have been taken and in y axis performance (Kappa analysis, Overall accuracy and Execution time) is taken where performance is clearly shown in the graph.
3.14 KERNEL INDUCED FUZZY POSSIBLISTIC C-MEANS CLUSTERING ALGORITHM

Classification Techniques are used on large databases to develop models describing different data classes. Numerous algorithms have been developed and tested to classify the image. The main purpose of these algorithms is to lessen the human efforts and errors in minimum time. This chapter also discusses the Kernel Induced Fuzzy Possiblistic C-Means Clustering Algorithm (KIFPCM) in detail. Kernel Induced Fuzzy Possiblistic C-Means Clustering Algorithm [95], [118] is a classification algorithm which works using remote sensing Image datasets. The algorithm is applied on testing images to get confusion matrix and analysis of accuracy assessment with the help of classified pixels.

Kernel [23], [165], [178] methods provide a machine learning paradigm for building nonlinear methods from linear ones. Kernel methods intrinsically cope with nonlinearities in a very flexible way and are effective when dealing with low numbers of high-dimensional samples. Many types of kernels like linear, polynomial; Radial Basis Function (RBF), sigmoid etc. are used in the Kernel Induced Fuzzy Possiblistic C-Means Clustering Algorithm (KIFPCM).

The concept of Fuzzy logic [54] is deployed to get the feasible solution in Kernel Induced Fuzzy Possiblistic C-Means Clustering Algorithm (KIFPCM). Fuzzy partitions are more flexible than crisp partitions in that each object can have membership in more than one cluster. Note, if $U$ is probabilistic [, the partition values are interpreted as a probability $p(i|ok)$ that $ok$ is in the $i$-th class. Fuzzy and probabilistic partitions are essentially equivalent from the point of view of clustering
algorithm development. The set of all fuzzy $c$-partitions is described. Each column of the fuzzy partition $U$ must sum to $1$, thus ensuring that every object is involved in a partition. The original Fuzzy C-Means (FCM) [1], [68] uses the probabilistic constraint that the memberships of a data point across classes sum to $1$. While this is useful in creating partitions, the memberships resulting from FCM [17], [95] and its derivatives, however, do not always correspond to the intuitive concept of degree of belonging or compatibility. The figure 3.21 shows the system architecture of the Kernel Induced Fuzzy Possiblistic C-Means Clustering Algorithm.

![System Architecture of KIFPCM Clustering Algorithm](image)

**Figure 3.21** System Architecture of KIFPCM Clustering Algorithm
The Possiblistic approach to clustering (PCM) by minimizing the following object function.

$$J_m(U,V) = \sum_{i=1}^{C} \sum_{k=1}^{n} u_{ik}^m \| x_k - v_i \|^2 + \sum_{i=1}^{C} \eta_i \sum_{k=1}^{n} (1 - u_{ik})^m$$

Where $\eta_i$ are suitable positive numbers. The first term demands that the distances from data points to the prototypes to be as low as possible, whereas the second term forces the $u_{ik}$ to be as large as possible, thus avoiding the trivial solution. It is recommended to select $\eta_i$ as

$$\eta_i = K \frac{\sum_{k=1}^{n} u_{ik}^m \| x_k - v_i \|^2}{\sum_{k=1}^{n} u_{ik}^m}$$

Typically, $K$ is chosen to be 1. The updating of prototypes is the same as that in FCM, but the memberships of PCM are updated as follows

$$U_{ik} = \frac{1}{1 + (\| x_k - v_i \|^2 / \eta_i)^{1/(m-1)}}$$

By following similar steps in KFCM [31], [95] construct the Kernel Possiblistic C-Means (KPCM) algorithm [125] by minimizing the following object function

$$J_m(U,V) = \sum_{i=1}^{C} \sum_{k=1}^{n} u_{ik}^m \| \phi(x_k) - \phi(v_i) \|^2 + \sum_{i=1}^{C} \eta_i \sum_{k=1}^{n} (1 - u_{ik})^m$$

As in KFCM, adopt the Gaussian kernel function [4], [134]. Then the updating of memberships is

117
Here use the Gaussian function as the kernel function [134], and \( \eta_i \) are estimated using

\[
\eta_i = k \frac{\sum_{j=1}^{n} u_{ik}^m 2(1 - k(x_i, v_j))}{\sum_{k=1}^{n} u_{ik}^m}
\]

Typically, \( K \) is chosen to be 1, and the updating of prototypes is the same as Equation.

\[
V_i = \frac{\sum_{k=1}^{n} u_{ik}^m k(x_k, v_i)x_k}{\sum_{k=1}^{n} u_{ik}^m k(x_k, v_i)}
\]

Kernel function is used to transform patterns into a higher dimensional feature space [23], [25], [68]. The transformation of the feature space into higher dimensional space can allow the naturally distributed groupings of data to be partitioned more effectively. The key idea in kernel-based clustering is that the transformation function need not be explicitly specified. The kernel function is defined as the dot product of two values obtained by the transforming function into input space. The summary of the algorithm is listed below

### 3.14.1 STEPS OF KIFPCM CLUSTERING ALGORITHM

**Step:** Initialization of memberships

Initialize fuzzier \( m \), stopping criterion \( \xi \), \( k=0 \);

Set initial \( u_{ij}(0) \) with memberships resulting from FKCM
Set initial $\sigma^2$ resulting from FKCM;

**Step:2 Minimization of objective function**

**REPEAT** $k \leftarrow k+1$ Compute $d_2(x_i,v_j)$ using the equation

$$k(x_k,x_k) = \frac{\sum_{j=1}^{N} u_{ij}^m}{\sum_{j=1}^{N} u_{ij}^m} + \frac{\sum_{j=1}^{N} u_{ij}^m}{\sum_{j=1}^{N} u_{ij}^m}$$

Compute $n_{ij}$ and $u_{ij}^{(k)}$

**UNTIL**

$$\|u_{ij}^{(k)} - u_{ij}^{(k-1)}\| < \epsilon$$

### 3.14.2 EXPERIMENTAL RESULTS

In Kernel Induced Fuzzy Possiblistic C-Means Clustering algorithm runs and it produces the desired results. The algorithm is tested with five different types of image datasets like Landsat-7 ETM+, Quick Bird, Modis, Aster, Spot-5 images. The image dataset descriptions are discussed in the section 3.28. The result obtained shows the classification of the images. These images are classified with different classes which show the importance of measuring the accuracy of classification.

**(i) LANDSAT-7 ETM+**

The Landsat-7 ETM+ image [27], [92] is classified with Kernel Induced Fuzzy Possiblistic C-Means clustering Algorithm. This image dataset consists of 670 instances which are categorized for classification. The result of the classified image is shown in the figure 3.22.
Figure 3.22 Results of Landsat-7 ETM + Image Using KIFPCM Algorithm
The Results of the Classified Landsat-7 ETM+ Image [92] Using Kernel Induced Fuzzy Possiblistic C-Means Clustering Algorithm is shown in the figure 3.23. The Kernel Induced Fuzzy Possiblistic C-Means Clustering algorithm produces producer accuracy for the class Paddy field is 84.73%, forest is 58.90%, grass land is 65.16%, Dry salt flats is 66.67%, and dry land is 74.60% and the User accuracy exhibits by the algorithm for the class Paddy Field is 72.55%, forest is 48.86%, grass land is 61.21%, Dry salt flats is 88.42% and dry land is 81.66%. These class values are measured using the Error Matrices [150] table which shows the classification values of each class. Other performance can be measured using statistical method like Kappa analysis and the performance of the algorithm is measured with Execution time.
(ii) QUICK BIRD IMAGE

The Quick Bird image [72] is classified with Kernel Induced Fuzzy Possibilistic C-Means Clustering Algorithm and produces various spectral classes. This image dataset consists of 670 instances which are categorized for classification. The result of the classified images is shown in the figure 3.24.

Figure 3.24 Results of Quick Bird Image Using KIFPCM Algorithm
Figure 3.25 Results of the Classified Quick Bird Image Using KIFPCM Algorithm

The result of the classified Quick Bird Image [72] Using Kernel Induced Fuzzy Possiblistic C-Means Clustering Algorithm is shown in the figure 3.25. The Kernel Induced Fuzzy Possiblistic C-Means Clustering exhibits producer accuracy for the class Agricultural field is 85.33%, road is 94.39%, Tree is 57.77%, Soil is 52.32%, Roof is 47.76%, Shadow is 90.99% and Grass is 94.59%. The user accuracy produced by the Kernel Induced Fuzzy Possiblistic C-Means Clustering exhibits producer accuracy for the class agricultural field is 84.76%, road is 82.78%, tree is 46.42%, soil is 77.58%, roof is 91.42%, shadow is 67.33% and grass is 55.55%. These class values can be measured with the help of Error matrices for classification accuracy.
(iii) MODIS IMAGE

The Modis image [171] is classified with Kernel Induced Fuzzy Possiblistic C-Means Clustering Algorithm. This image dataset consists of 670 instances which are categorized for classification. The result of the classified images is shown in the figure 3.26.

Figure 3.26 Results of Modis Image Using KIFPCM Algorithm
The result of the classified Modis Image Using Kernel Induced Fuzzy Possibilistic C-Means Clustering Algorithm is shown in the figure 3.27. The Kernel Induced Possibilistic C-Means Clustering produces the producer accuracy for the class agricultural field is 74.77%, road is 70.45%, tree is 75.69%, soil is 71.81%, roof is 76.78%, shadow is 94.11% and grass is 78.82%. The User accuracy produced by the Kernel Induced Possibilistic C-Means Clustering produces the producer accuracy for the class agricultural field is 65.35%, road is 78.48%, tree is 83.53%, soil is 84.04%, roof is 64.17%, shadow is 88.88% and grass is 89.14%.

These class values can be measured using the Error matrices [39], [150] for classification accuracy and the performance of the algorithm is measured with the help of Execution time.
(iv) ASTER IMAGE

The Aster image is to be classified with Kernel Induced Fuzzy Possibilistic C-Means Clustering Algorithm. This image dataset consists of 670 instances which are categorized for classification. The result of the classified image is shown in the figure 3.28.

Figure 3.28 Results of Aster Image for KIFPCM Algorithm
The results of the classified Aster image using Kernel Induced Fuzzy Possiblistic C-Means Clustering Algorithm is shown in the figure 3.29. The Kernel Induced Fuzzy Possiblistic C-Means Clustering algorithm produces the producer accuracy for the class Paddy Field is 82.44%, Forest is 60.81%, Grass Land is 64.00%, Dry Salt Flats is 58.59%, and Dry Land is 74.05%. The Kernel Induced Fuzzy Possiblistic C-Means Clustering algorithm produces the User accuracy for the class Paddy Field is 71.05%, Forest is 46.87%, Grass Land is 58.53%, Dry Salt Flats is 86.20%, and Dry Land is 81.05%.
(v) SPOT-5 IMAGE

The SPOT-5 image [64] is to be classified with Kernel Induced Fuzzy Possiblistic C-Means Clustering Algorithm. This image dataset consists of 670 instances which are categorized for classification. The result of the classified image is shown in the figure 3.30.

Figure 3.30 Results of Spot-5 Image Using KIFPCM Algorithm
The result of the classified Spot-5 Image Using Kernel Induced Fuzzy Possibilistic C-Means Clustering Algorithm is shown in the figure 3.31. The producer accuracy of KIFPCM for the class Paddy Field is 65.33%, Forest is 54.39%, Grass Land is 59.25%, Dry Salt Flats is 70.58% and Dry Land is 68.24% and User accuracy for the class Paddy Field is 89.90%, Forest is 77.09%, Grass Land is 46.37%, Dry Salt Flats is 53.33% and Dry Land is 79.52%.

The Error matrices show the classified value for the class in the Spot-5 Images. The performance of the algorithm is measured with the help of Execution time. The remaining part leading to classification is termed as misclassification, which is measured using Error matrices.
3.14.3 PERFORMANCE ANALYSIS

Table 3.2 Performance Analysis of KIFPCM Clustering Algorithm

<table>
<thead>
<tr>
<th>IMAGE DATASET/PERFORMANCE</th>
<th>KAPPA VALUE (%)</th>
<th>OVERALL ACCURACY (%)</th>
<th>EXECUTION TIME (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LANDSAT-7 ETM+</td>
<td>76.89</td>
<td>71.19</td>
<td>52.00</td>
</tr>
<tr>
<td>QUICK BIRD</td>
<td>76.00</td>
<td>74.60</td>
<td>51.00</td>
</tr>
<tr>
<td>MODIS</td>
<td>78.00</td>
<td>77.40</td>
<td>50.00</td>
</tr>
<tr>
<td>ASTER</td>
<td>78.00</td>
<td>68.80</td>
<td>56.00</td>
</tr>
<tr>
<td>SPOT-5</td>
<td>68.00</td>
<td>58.20</td>
<td>58.00</td>
</tr>
</tbody>
</table>

The table 3.2 shows the performance Analysis of Kernel Induced Fuzzy Possiblistic C-Means Clustering Algorithm. The performance analysis is measured with the Kappa value, overall accuracy and Execution time with different image datasets like Landsat-7 ETM+, Quick Bird, Modis, Aster and Spot-5 Images. The table shows that the Kappa value is high for both Modis and Aster Image which shows 78.00 % whereas it is low for Spot-5 Image and the Overall accuracy is high for Modis Image which shows 77.40 % whereas it is low for Spot-5 Image which shows 58.20 and Execution time is less for Modis Image which shows 56.00 ms and high for Spot-5 Image.
3.15 SUMMARY

This chapter discusses the remote sensing image with spectral ranges of remote sensing image. In this chapter remote sensing image data with ground truth information has also been discussed. The implementation of K-Means and KIFPCM algorithm with different image datasets and the results of the algorithm have been discussed. The next chapter discusses the implementation of proposed Fuzzy Cylindrical Partial Relations Clustering (FCPR) Algorithm for five different image datasets.