CHAPTER - 4

FEATURE BASED EFFECTIVE K-MEANS CLUSTERING METHOD IN RSI CLASSIFICATION

4.1. Introduction

The feature based classification of remotely sensed image is used to assign corresponding levels with respect to groups with homogeneous characteristics, with the aim of discriminating multiple objects from each other within the image. Every level of an image is called class. This will be executed on the basis of spectral or spectrally defined features such as density, texture and many other things in the feature space. This chapter focuses remote sensing image classification of color feature based using K-Means clustering method. K-Means is one of the unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters [1]. 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 centroid. When no point is pending, the first step is completed and we need to re-calculate k new centroids of the clusters resulting from the previous step. After identifying these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. Inside the loop, the k centroids change their location step by step repeatedly until no more changes are done. Here we introduce feature based K-Means clustering algorithms that consolidate data by clustering or grouping in remote sense application [1].
4.2. Common Procedure for Classification and Clustering

Remote sensing image classification refers to the task of extracting information classes from a multiband raster image. The resulting raster from image classification can be used to create thematic maps. Depending on the interaction between the analyst and the computer during classification, there are two types of classification: supervised and unsupervised. The objective of image classification is to identify and portray, as a unique gray level (or color), the features occurring in an image in terms of the object or type of land cover these features actually represent on the ground [2]. A broad group of digital image processing techniques of remote sensing application explains the image classification techniques which are most generally applied to the spectral data of a image or to the varying spectral data of a series of multidimensional images. Clustering involves a set of point into non-overlapping groups or points; where points in a cluster are more similar to one another than points in other clusters. The term more similar, when applied to clustered points, usually means closer by some measure of proximity. When a dataset is clustered, every point is assigned to some cluster and every cluster can be characterized by a single reference point, usually take an average of the points in the cluster. Any particular division of all points in a dataset into clusters is called a partitioning. Clustered data require considerably less storage space and can be manipulated more quickly than the original data. The value of particular clustering method will depend on how relatively the reference points represent the data as well as how fast the program runs [2].

Clustering is a vital element of remote sensing model identification field means distinguishing and classifying things that are provided with similar properties. Clustering technique classifies the pixels with the same characteristics into one cluster, thus forming different clusters according to coherence between the pixels in a cluster. The flow diagram of a general classification procedure of remotely sensed image is shown in the following Figure 4.1.
4.3. **Different Perspectives of Clustering Method**

Clustering is a grouping of data with similar characteristics. This similarity in a given set may vary according to data, because clustering is used in various fields such as numerical taxonomy, morphometric and systematics. Thus, a clustering algorithm that fits the numerical measure of optimization in a data may not optimize another set of data. There are many algorithms to solve a clustering problem. The algorithms used in our applet concentrate on joining, splitting, and switching search methods (also called bottom up, top down, and interchange, respectively). The clustering method uses a three-dimensional model for...
demonstration purposes. These algorithms can calculate clusters in n-dimensions also. There are many methods of clustering developed for wide variety of purposes. Clustering algorithms used for unsupervised classification of remote sensing data vary according to the efficiency with which clustering takes place. Clustering is divided into hierarchical clustering and non-hierarchical clustering as mentioned as follows. The distances to evaluate the similarity are selected from the following methods [2].

4.3.1. Hierarchical Clustering Method

Hierarchical clustering techniques proceed by either a series of successive mergers or a series of successive divisions. Hierarchical clustering is a widely used data analysis tool. This idea is to build a binary tree of the data that successively merges similar groups of points. This tree visualizing to provides a useful summary of the data [2]. The different hierarchical clustering method is shown below.

1. Nearest Neighbour Method – Nearest neighbor with minimum distance will from a new merged cluster.
2. Furthest Neighbour Method – Furthest neighbor with minimum distance will from a new merged cluster.
3. Centroid Method – Distance between the gravity centers of two clusters is evaluated for merging a new merged cluster.
4. Group Average Method – Root mean square distance between all pairs of data within two different clusters are used for clustering.
5. Ward Method – Root mean square distance between the gravity center and each member is minimized.

4.3.2. Non-hierarchical Clustering Method

At the initial stage, an arbitrary number of clusters should be temporarily chosen. The members belonging to each cluster will be checked by selected parameters or distance and relocated into the more appropriate clusters with higher
separability. The K-Means method is a best example of non-hierarchical clustering. This method is composed of the following procedure [3].

### Non-hierarchical Clustering Algorithm

**Step 1:** All members are relocated into the closest clusters by computing the distance between the member and the clusters.

**Step 2:** The center of gravity of all clusters is recalculated and the above procedure is repeated until convergence.

**Step 3:** If the number of clusters is within a certain specified number and the distances between the clusters meet a prescribed threshold, the clustering is considered complete.

| Table 4.1. Algorithm For Non-hierarchical Clustering |

4.3.3. **Threshold-based Clustering Algorithm**

In the threshold-based clustering algorithm, the number of clusters is unknown. However, two elements are classified to the same cluster if the distance between them is below a specified threshold. The algorithm proceeds as follows:

### Threshold-based Clustering Algorithm

**Step 1:** Select an element from the given data set. This element is assigned as the seed of a cluster by itself.

**Step 2:** For every unclassified element, find its distance from the centroid of the existing clusters.

**Step 3:** If the distance is less than the threshold, assign the element to this cluster. Recompute the centroid of the cluster as the average of all properties of all elements in the cluster.

**Step 2:** If no such cluster can be found after examining all current clusters, assign the element as the seed for a new cluster.

**Step 3:** If, as a result of the above step, the distance of new cluster to another cluster is smaller than the threshold, merge the two close clusters together and recompute the cluster distances.

**Step 3:** The algorithm stops after all the elements have been assigned to one or the other cluster.

| Table 4.2. Algorithm For Threshold-based Clustering |
Thus, this algorithm is sensitive to the specification of threshold. The threshold can be reduced or increased to get an appropriate number of clusters by repeated iterations [3].

4.4. Clustering Terminologies

Clustering is a process which separate a given dataset into homogeneous groups based on specific requirements. The similar objects are kept in a group whereas dissimilar objects are in different groups. Clustering plays an important role in various fields including image processing, mobile communication, computational biology, medicine and economics [4]. The main objective of this chapter is to provide a classification algorithm which will be used in image classification tool in an effective manner. K-Means algorithm is a well-known clustering algorithm popularly known as Hard C-Means algorithm [4]. This algorithm splits the given image into different clusters of pixels in the feature space, each of them defined by its center. Initially each pixel in the image is allocated to the nearest cluster. Then the new centers are computed with the new clusters. These steps are repeated until convergence. Basically we need to determine the number of clusters K first. Then the centroid will be assumed for these clusters. We could assume random objects as the initial centroids or the first K objects in sequence could also serve as the initial centroids [4]. The K-Means result relies on the data set to satisfy the assumptions made by the clustering algorithms is shown in Figure 4.2.
The algorithm looks for pixels with maximum probability to be boundary in order to determine the exact place of the edge. We decide the level to amended by computing the average of the link distances. As larger the distance, more separated are the clusters among them. In this way, the problem of finding the optimum number of classes is equivalent to minimize the following energy function.

\[ E = \frac{1}{D_{\text{avg}}} \]  

(4.1)

In this equation, \( D_{\text{avg}} \) is the average of the distances of the links. If an estimation of the possible number of regions is available, this information can be taken into account by including a new term on the decision criterion. Specifically, Eqn. 4.2 can be generalized as
In this equation, $D_{avg}$ is the average of the distances of the links to cut, $N_k$ is the previously estimated number of clusters given by a supervisor or extracted from the previous experimentation and $N_c$ is the number of clusters that the amended will produce.

The clustering algorithms assume that pixels belonging to the same cluster have a similar behavior, which is explained by uniform features like grey levels, texture or colour. The popular K-Means algorithm is an error minimization algorithm where the function to minimize is the sum of squared error.

$$E^2(K) = \sum_{k=1}^{K} \sum_{i \in C_k} (x_i - C_k)^2$$ (4.3)

Here, $C_k$ is the centroid of cluster of $K$ and $K$ is the number of clusters. It has two important factors such as linear time complexity and easy implementation.

### 4.5. Standard K-Means Clustering Method

The most common algorithm uses an iterative refinement technique. Due to its ubiquity it is often called the K-Means algorithm, it is also referred to as Lloyd's algorithm, particularly in the computer science community. Given an initial set of $k$ means $m_1, \ldots, m_k$, the algorithm proceeds by alternating between two steps in [5].

**Assignment Step**: Assign each observation to the cluster whose mean yields the least Within Cluster Sum of Squares (WCSS). Since the sum of squares is the squared Euclidean distance, this is intuitively the "nearest" mean.

$$S_i^{(t)} = \{ x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2 \text{ For all } j \neq i, 1 \leq j \leq k \}$$ (4.4)

where each $x_p$ is assigned to exactly one $S_i^{(t)}$, even if it could be assigned to two or more of them.

**Update Step**: Calculate the new means to be the centroids of the observations in the new clusters.
\[
M_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j
\] (4.5)

The algorithm has converged when the assignments no longer change. Since both steps optimize the WCSS objective, and there only exists a finite number of such partitioning, the algorithm must converge to a (local) optimum. There is no guarantee that the global optimum is found using this algorithm in [5]. The algorithm is often presented as assigning objects to the nearest cluster by distance. Using a different distance function other than Euclidean distance may stop the algorithm from converging.

4.6. Broad Spectrum of K-Means Clustering

The K-Means clustering algorithm is faster than the standard version and extends the size of the datasets that can be clustered. It differs from the standard version of the cluster algorithm in how the initial reference points are chosen and how data points are selected for the updating process. In the standard algorithm the initial reference points are chosen more or less arbitrarily. The proposed K-Means cluster algorithm reference points are chosen at random sampling from the whole population of data points. If the sample is sufficiently large, the distribution of these initial reference points should reflect the distribution of points in the entire data set [6]. Another difference between the standard and the new K-Means clustering algorithms is the way the data points are treated as random sampling. The each complete iteration, the standard algorithm examines all the data points in sequence. In contrast, the proposed k-means clustering algorithm examines only a random sample of data points. If the data set is very large and the sample is representative of the dataset, the algorithm should coverage much more quickly than the standard algorithm that examines every point in sequencing of updating the centroids during the initial partitioning, when the data point are first assigned to clusters [6].

Clustering algorithm uses an interchange method to partition a graph into clusters. An initial partition is given, and new partitions are obtained by switching
an object from one cluster to another. This method randomly picks $K$ points initially, where each point stands for a cluster to be made. A set of points is taken from the graph, and each point is added to the closest cluster. The closeness to the cluster is determined by calculating the distance between a point and the centroid of a cluster. Then each point is visited to recalculate the distance to the updated clusters. If the closest cluster of the point is not the one it currently belongs, the point will switch to the new cluster. When switching occurs, centroids of both modified clusters have to be recalculated. This procedure is repeated until no more switching takes place. The logical representation of K-Means clustering algorithm is given [6].

- Determine the centroid coordinate (Random assignment).
- Determine the distance of each Object pixel to the centroids.
- Group the object based on minimum distance with the centroid.

4.6.1. Developments

K-Means clustering in particular when using heuristics such as Lloyd's algorithm is rather easy to implement and apply even on large data sets. As such, it has been successfully used in various topics, including remote sensing image classification, computer vision, geostatistics, astronomy and agriculture. It often is used as a preprocessing step for other algorithms, for example to find a starting configuration [7].

4.6.2. Vector Quantization

K-means originates from signal processing, and still finds use in this domain. For example in computer graphics, color quantization is the task of reducing the color palette of an image to a fixed number of colors $k$. The K-Means algorithm can easily be used for this task and produces competitive results. Other uses of vector quantization include non-random sampling, as K-Means can easily
be used to choose $k$ different but prototypical objects from a large data set for further analysis [7].

4.6.3. Cluster Analysis

In cluster analysis, the $k$-means algorithm can be used to partition the input data set into $k$ partitions (clusters). However, the pure $k$-means algorithm is not very flexible, and as such of limited use. In particular, the parameter $k$ is known to be hard to choose, when not given by external constraints. Another limitation of the algorithm is that it cannot be used with arbitrary distance functions or on non-numerical data in [7].

4.6.4. Feature Learning

K-means clustering has been used as a feature learning step, in either semi-supervised learning or unsupervised learning. The basic approach is first to train a K-Means clustering representation, using the input training data (which need not be labeled). Then, to project any input datum into the new feature space, we have a choice of encoding functions, but we can use for example the thresholded matrix-product of the datum with the centroid locations, the distance from the datum to each centroid, or simply an indicator function for the nearest centroid or some smooth transformation of the distance. This use of K-Means has been successfully combined with simple, linear classifiers for semi-supervised learning in NLP and in computer vision. On an object recognition task, it was found to exhibit comparable performance with more sophisticated feature learning approaches such as auto encoders and restricted Boltzmann machines [8]. However, it generally requires more data than the sophisticated methods, for equivalent performance, because each data point only contributes to one feature rather than multiple.

4.6.5. Merits

K-Means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. It is possible to reduce the computational cost and gives a high discriminative power of regions present in the image. The K-
Means clustering algorithm is faster than the standard version and extends the size of the datasets that can be clustered.

The proposed K-Means cluster algorithm reference points are chosen at random sampling from the whole population of data points. Another difference between the standard and the new K-Means clustering algorithms is the way the data points are treated as random sampling.

4.6.6. Demerits

Thus, a clustering algorithm that fits the numerical measure of optimization in a data may not optimize another set of data. If the data set is very large then it should coverage much comparable time than the standard algorithm.

we did not use any training data and the work is divided into two stages. First enhancing color separation of satellite image using decorrelation stretching is carried out and then the regions are grouped into a set of five classes.

It avoids feature calculation for every pixel in the remote sensing image.

4.7. Proposed Modified K-Means Clustering Algorithm for RSI

The basic aim is to classify colors in an automated fashion using the color space and K-Means clustering. The entire process can be summarized in the following steps.

<table>
<thead>
<tr>
<th>Proposed Modified K-Means Clustering Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1:</strong> Read the remote sensing image</td>
</tr>
<tr>
<td><strong>Step 2:</strong> For color separation of remote sensing image applying the correlation stretching method</td>
</tr>
<tr>
<td><strong>Step 3:</strong> Convert image from RGB color space to L<em>a</em>b* color space</td>
</tr>
<tr>
<td><strong>Step 4:</strong> Classify the colors in a particular image of a<em>b</em> color space using K-Means Clustering method</td>
</tr>
<tr>
<td><strong>Step 5:</strong> Label every pixel of the image using the results from K-means algorithm</td>
</tr>
<tr>
<td><strong>Step 6:</strong> Create images that classify the image by color</td>
</tr>
<tr>
<td><strong>Step 7:</strong> Classify the nuclei of the image into a separate image window</td>
</tr>
</tbody>
</table>

Table 4.3. Algorithm for Proposed Modified K-Means Clustering
First step define to read the image from the remote sensing source is in jpeg format and processing of data from LSU Earth Scan laboratory. Second step express to the purpose of decorrelation stretch is visual enhancement. The decorrelation stretching is a way to enhance the color differences in an image. Third step state that the L*a*b* color space enables you to quantify these visual differences. The L*a*b* space consists of a luminosity L* or brightness layer, chromaticity layer a* indicating where color falls along the red-green axis, and chromaticity layer b* indicating where the color falls along the blue-yellow axis.

Figure 4.3. Flow Diagram of Feature Based K-Means Clustering Method
We can measure the difference between two colors using the Euclidean distance metric. Convert the image to L*a*b* color space. Fourth step deals clustering as a way to separate groups of objects. The color based K-means clustering treats each object as having a location in space. It requires specifying the number of clusters to be partitioned and distances metric to quantify how close two objects are to each other. Since the color information exists in the a*b* space. Use K-Means to cluster objects into five clusters using the Euclidean distance metric. The Euclidean distance is used as a metric and variance is used as a measure of cluster scatter in [9]. Fifth step deals for every object in the above input, K-Means returns an index corresponding to a cluster. Label every pixel in the image with its cluster index. Sixth step stated that image classification is the process of partitioning a digital image into multiple classifying objects. The goal of classification is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Remote sensing image classification is typically used to locate objects and boundaries in images. More precisely, image classification is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. Last step to deal with this method uses examples selected by a user for building a statistical model that captures the texture and shape variations of the nuclei structures from a given image dataset to be classified. Then programmatically determine the index of the cluster containing the blue object because k-means will not return the same cluster index value every time in [9]. We can do this using the cluster center value, which contains the mean a* and b* value of each cluster.

The result of remote sensing image classification is a set of segments that collectively cover the entire image or a set of contours extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property such as color, intensity or texture. Using pixel label, we have to separate objects in image by color which will to produce the result in five images.
4.8. Relative Error Measure Factor

In fact, the clustering algorithm adopts mathematical method of updating centroids during the initial partitioning, when the data points are first assigned to clusters. The following is a suitable error measurable factors for the K-Means clustering method is given in [9].

4.8.1. Cluster Error Measurable Factor

From a mathematical perspective, random sampling represents a return to original concept of the algorithm as a method of clustering data over a sequence space. The error measure $EMF_i$ for each region $R_i$ is given by

$$EMF_i = \int_{x \in R_i} \rho(x) \| x - z_i \|^2 \, dx$$  \hspace{1cm} (4.6)

Where $\rho(x)$ is the probability density function, a sequence function defined over the space and the total error measure $EMF$. In this mathematical concept of the algorithm, very large set of discrete data points can be thought of as a large sample and is a good estimate of the continuous probability density $\rho(x)$. It becomes a random sample of the dataset can also be a good estimate of $\rho(x)$. Such a sample yields a representative set of cluster centroids and a responsible estimate of the error measure without using all the points in the original dataset.

These modifications to the standard algorithm greatly accelerate the clustering process. Since both the reference points and the data points for the updates are chosen by random sampling, more reference points will be found in the densest regions of the dataset and the reference points will be updated by data points in the most critical regions. This clustering algorithm is about ten times faster than the standard algorithm. The computation time can be further reduced by making the individual steps in the algorithm more efficient. A substantial fraction of the computation time required by any of these clustering algorithms is typically spent in finding the reference point closest to a particular data point. It is a more elegant method of point location avoiding much of this time consuming process by reducing the number of reference points that must be spent to create data.
structures. Such structures range from particular orderings of reference points to trees in which reference points are organized into categories. The K-Means algorithm uses a tree method to cluster three-dimensional data [9].

4.8.2. Cluster Error Determination Factor

As k-means approach is iterative, it is computationally intensive and hence applied only to images subareas rather than to full scenes and can be treated as unsupervised training areas. Consider a single cluster of points along with its centroid or mean. If the data points are tightly clustered around the centroid, the centroid will be representative of all the points in that cluster. The standard measure of the spread of a group of points about its mean is the variance or sum of the squares of the distance between each point and the mean. If the data points are close to the mean, the variance will be small. A generalization of the variance, in which the centroid is replaced by a reference point that may or may not be a centroid, is used in cluster analysis to indicate the overall quality of partitioning. Specifically, the error determination factor EDF is the sum of all the variances [9].

\[
EDF = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \| x_{ij} - z_i \|^2
\]  

(4.7)

Where \( x_{ij} \) is the jth point in the ith cluster, \( z_i \) is the reference point of the ith cluster and \( n_i \) is the number of points in that cluster. The notation \( \| x_{ij} - z_i \| \) stands for the distance between \( x_{ij} \) and \( z_i \). Hence, the error determination factor EDF indicates the overall spread of data points about their reference points. The error determination factor provides an objective method for comparing partitioning as well as a test for eliminating unsuitable partitioning. At present, finding the best partitioning requires generating all possible combinations of clusters and comparing their error measures. When clustering is done for the purpose of data reduction, as in the case of the remote sensing image dataset, the goal is to find the best partitioning. We merely want a reasonable consolidation of \( N \) data points into \( k \) clusters and if necessary, in this efficient way to improve the quality of the initial
partitioning. All the data points are partitioned into $k$ clusters by assigning each point to the cluster of the closest reference point. Adjustments are made by calculating the centroid for each of those clusters and then using those centroids as reference points for the next partitioning of all the data points. It can be proved that a local minimum of the error measure $EDF$ corresponds to the reference point of its cluster than to any other reference point.

### 4.9. Experimental Result

Traditional clustering algorithms have been considered to compare the performance of classical versus the new proposal. Moreover, we have also compared our algorithms with those clustering versions that use pixel coordinates as artificial features. Note that this study is to provide significant improvements in classical K-Means clustering algorithms. The various experiment carried out in the above remote sense imagery data set in MATLAB 7.6. The complete process and the standard are summarized in subsequent Figure 4.4.

4.10. Performance Evaluation Matrices

The results of the proposed feature based K-Means clustering method is also compared with different cluster level of classified images from the remote sensing images and the resultant matrices is depicted in Table 4.4.
## Input Image
Remote Sensing Forestry based Original Satellite Image

<table>
<thead>
<tr>
<th>Output Image</th>
<th>Performance Matrices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
</tr>
<tr>
<td>DCS Image with Luminosity 50</td>
<td>53.87</td>
</tr>
<tr>
<td>DCS Image with Luminosity 60</td>
<td>53.89</td>
</tr>
<tr>
<td>DCS Image with Luminosity 70</td>
<td>53.89</td>
</tr>
<tr>
<td>DCS Image with Luminosity 80</td>
<td>52.90</td>
</tr>
<tr>
<td>Labeled Image Cluster Index 3</td>
<td>52.95</td>
</tr>
<tr>
<td>Labeled Image Cluster Index 4</td>
<td>52.97</td>
</tr>
<tr>
<td>Labeled Image Cluster Index 5</td>
<td>46.81</td>
</tr>
<tr>
<td>Image Cluster 5</td>
<td>46.89</td>
</tr>
<tr>
<td>DCS Image with Luminosity 90</td>
<td>32.98</td>
</tr>
<tr>
<td>DCS Image with Luminosity 100</td>
<td>28.91</td>
</tr>
<tr>
<td>Image Cluster 2</td>
<td>29.84</td>
</tr>
<tr>
<td>Image Cluster 3</td>
<td>29.87</td>
</tr>
<tr>
<td>Nuclei of Image 1</td>
<td>29.91</td>
</tr>
<tr>
<td>Nuclei of Image 2</td>
<td>28.90</td>
</tr>
<tr>
<td>Nuclei of Image 3</td>
<td>25.91</td>
</tr>
<tr>
<td>Nuclei of Image 4</td>
<td>25.93</td>
</tr>
</tbody>
</table>

### Finest Level of Performance

### Average Level of Performance

### Worst Level of Performance
### Table 4.4. Performance Evaluation Matrices of Feature Based Proposed Modified K-Means Clustering Algorithm

<table>
<thead>
<tr>
<th>Output Image</th>
<th>Performance Matrices</th>
<th>Overall Efficiency</th>
<th>Computational Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BSS</td>
<td>PBSS</td>
<td>MPS</td>
</tr>
<tr>
<td><strong>Finest Level of Performance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCS Image with Luminosity 50</td>
<td>0.97</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>DCS Image with Luminosity 60</td>
<td>0.96</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>DCS Image with Luminosity 70</td>
<td>0.99</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>DCS Image with Luminosity 80</td>
<td>0.99</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Labeled Image Cluster Index 3</td>
<td>0.91</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>Labeled Image Cluster Index 4</td>
<td>0.98</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>Labeled Image Cluster Index 5</td>
<td>0.99</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Image Cluster 5</td>
<td>0.97</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td><strong>Average Level of Performance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labeled Image Cluster Index 1</td>
<td>0.85</td>
<td>0.80</td>
<td>84.40 %</td>
</tr>
<tr>
<td>Labeled Image Cluster Index 2</td>
<td>0.85</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>Image Cluster 1</td>
<td>0.87</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>Image Cluster 4</td>
<td>0.80</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>Nuclei of Image 5</td>
<td>0.85</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td><strong>Worst Level of Performance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCS Image with Luminosity 90</td>
<td>0.78</td>
<td>0.75</td>
<td>76.75 %</td>
</tr>
<tr>
<td>DCS Image with Luminosity 100</td>
<td>0.79</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Image Cluster 2</td>
<td>0.79</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>Image Cluster 3</td>
<td>0.77</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Nuclei of Image 1</td>
<td>0.78</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Nuclei of Image 2</td>
<td>0.78</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>Nuclei of Image 3</td>
<td>0.76</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>Nuclei of Image 4</td>
<td>0.69</td>
<td>0.66</td>
<td></td>
</tr>
</tbody>
</table>
4.11. Comparison Graph for Test Scores Clustering

There are many possible approaches to clustering, in this work we use one of the most common advanced feature based effective K-Means clustering algorithm. There are many possible approaches to clustering, in this work the number of clusters is denoted in this paper by $N_c$ and must be selected carefully as the size of the final feature vector increases quadratically with $N_c$. The different level of comparison graph is shown in the following Figure 4.5.
4.12. Discussions

To investigate this method, we first run the K-Means algorithm and the input image with region of interest is divided into various clusters and the resultant image is shown in Figure 4.4. The outputs from remote sensing image compared to various performance levels (finest, average and worst). From Figure 4.4. a2, a3, b1 and b2 have to produce finest level of performance and Figure 4.4. c2, c3, e1, f1 and h1 have to produce the average level of performance and Figure 4.4. b3, c1, e2, e3, f3, g1, g2 and g3 have to produce worst level of performance in K-Means clustering method. The performance of the proposed feature based K-Means method has been evaluated in terms of sensitivity, specificity and accuracy. The Table 4.4 represents the performance comparison for K-Means classifier with different matric functions. Here total 50 satellite images are taken for testing but only one to concentrate to the study of our dissertation. The experimental results have shown that the proposed method achieves good classification accuracy and less standard error while compared to various levels of thresholding methods. The overall execution times of the proposed method are around 12.651204 seconds for the remote sensing data sets and acceptable times in the usual applications to
remote based satellite mapping of classification methods. The overall efficiency of this algorithm is 85.49% is significantly improved over the other threshold methods. The schemes proposed in this work can be further improved by introducing fuzzy logic concepts into the clustering process.

4.13. Conclusion

Remote Sensing image classification has been the basis of image processing, comprehension and model identification and a hot research subject of image processing technologies. This method presents theoretical and practical knowledge to image classification and clustering. The application scope of image classification is quite extensive. The experimental data include that our algorithm produces effective results. The clustering process is performed by using a classical algorithm such as K-Means, but including one important restriction is pixels with a high gradient are not classified. The reason of this restriction is that, in images rich on surface, the clustering will not be effective in boundary pixels. This is due to the fact that surface features are extracted using the local neighboring information, and in these positions the surface value will be a mixture of two or more regions. These algorithms are robust and very effective in producing desired classifications especially in the field of pattern recognition as per the region of interest as demonstrated by the experimental results. This work presents a novel image classification based on colour features from the images. In this we did not use any training data and the work is divided into two stages. First enhancing color separation of satellite image using decorrelation stretching is carried out and then the regions are grouped into a set of five classes using K-Means clustering algorithm. Using this two-step process, it is possible to reduce the computational cost avoiding feature calculation for every pixel in the image.
References


