CHAPTER II

REVIEW OF LITERATURE

Remote sensing involves the acquisition of information about objects through the analysis of data collected by remote sensors that are not in physical contact with the objects of investigation. Often, these sensors are mounted on aircrafts, satellites and hips capture images of inaccessible or dangerous areas on the Earth. These sensors are either passive or active. Active sensors transmit artificially produced energy to a target and record its reflection on the target. Passive sensors, however, don’t transmit energy, but detect only energy emanating naturally from an object. Examples of passive sensors include traditional cameras, and radiometers and those of active sensors include Radio Detection and Ranging (RADAR) and Light Detection and Ranging (LIDAR) sensors.

Hyperspectral remote sensing is a new technology that has provided a variety of applications such as geology (mineral exploration), vegetation studies (species identification) and soil science (type mapping). Recent development of remote sensing has led the way for the development of hyperspectral sensors such as NASA’s AVIRIS, Hyperion Instrument and Analytical Spectral Devises (ASD) handheld spectrometer. These hyperspectral remote sensors measure reflected radiation as continuous and narrow wavelength bands and hence produce images with hundreds or thousands of spectral bands which can provide unique spectral signatures for each image pixel. In hyperspectral image, each species has a distinct spectral signature, making it unique and identifiable by that spectral signature. Such advantages make hyperspectral images extremely suitable for statistical pattern recognition. Before discussing about the hyperspectral image analysis, just recall about the basic steps of image segmentation.
2.1. Image segmentation

Image segmentation [14] is one of the prime and most significant tasks in image analysis and computer vision. Segmentation can be described as an intense partitioning of the input image into regions, each of which is examined to be homogeneous with respect to some criterion of interest. The objective of segmentation [75] is to find regions that exemplify objects or consequent parts of objects. Image segmentation is useful in many applications. It can describe the regions of interest in a scene or annotate the data. The existing segmentation algorithms are categorized into region based segmentation, data clustering and edge based segmentation. In this analysis, a new quantization technique for HSV (Hue, Saturation, Value) colour space is implemented to develop a colour histogram for K-Means clustering which operates across different dimensions in HSV colour space.

2.2. Multispectral image segmentation

Geographic Object Based Image Analysis (GEOBIA) [30] is a sub-discipline of Geographic Information Science (GI Science) devoted to developing automated methods to partition remote sensing imagery into meaningful image objects, this research investigated the use of a GEOBIA approach with the incorporation of object specific Gray Level Co-occurrence Matrix (GLCM). The GEOBIA approach has the potential to overcome inherent problems of high spectral variability within the same land-cover classes in VHR (Very High Resolution) imagery. Object Based Image Analysis (OBIA) is gaining rapid popularity in remote sensing science as a means of bridging Very High Spatial Resolution (VHSR) imagery [19]. Classification algorithms based on single-pixel analysis often do not give the aimed result when enforced to high spatial resolution remote sensing data [9]. The pixel oriented approach is not appropriate to use geometric and contextual attributes. An object
oriented approach is needed. Thus, a multispectral image segmentation method should be used to generate objects.

The Multispectral Local Difference Texem (MLDT) is an affordable procedure to be used in multispectral images that may contain extensive number of bands [58]. Multi and hyperspectral sensors acquire information in several spectral bands, which develop hyperspectral data in high dimensional spaces. Multispectral data are used in order to evaluate and interpret the types of vegetation, land, water and other man-made objects. Standard multispectral image interpretation approaches scarcely exploit the spectral, spatial relationships in the image. The multispectral image data are inherently treated as asset of independent spectral measurements at each pixel location without taking into account their spatial relations.

2.3. Hyperspectral image processing methods

The systematic approach to analyze the hyperspectral imaging has first introduced in [40]. This work serves as a cornerstone in the remote sensing system modelling research based on input statistical data. In this paper, the author explained the remote sensing system and several typical instruments which cover the optical spectrum ranging from 0.4 μm to 2.4 μm. Then the system and its working process were separated into different unit blocks. The surface reflectance statistics were assumed to be spectrally multivariate Gaussian with a spatial correlation. The scene was spatially modelled as having cells of diffuse reflectance (Lambertian assumption) with spatial correlation from cell to cell. Hyperspectral imaging is a new remote sensing scheme that develops hundreds of images, comparable to disparate wavelength channels, in the same area on the surface of the earth [59]. Supervised classification of hyperspectral image dataset is a demanding obstacle due to the limited availability of training samples. The high number of spectral bands acquired by hyperspectral
sensors enhances the competence to discriminate physical materials and object presenting new challenges to image analysis and classification [66]. Hyperspectral imaging sensors measure the energy of the received light in tens or hundreds of narrow spectral bands in each position in the image. Markov Random Fields (MRFs) are probabilistic models that are commonly used to integrate spatial context into image classification problems. The following figures 2.1 shows a general hyperspectral image classification system.

![Diagram](image_url)

Figure 2.1. Block diagram of a typical hyperspectral image analysis system.

Imaging spectroscopy, also known as hyperspectral imaging [6], is concerned with the measurement, analysis and interpretation of spectral acquired from a given scene (or specific object) at a short, medium or long distance by the satellite sensor. The special characteristics of hyperspectral datasets pose various processing obstacles, which must be automatically tackled under specific mathematical formalisms, such as classification and segmentation or spectral mixture analysis. The covariance matrix is a key component in a wide array of statistical signal processing tasks applied to remote sensing imagery from multispectral and hyperspectral sensors [67]. It paves to evaluate performance metrics that are observed in real hyperspectral imagery. Hyperspectral sensors simultaneously capture hundreds of narrow and contiguous spectral images from a wide range of the electromagnetic spectrum [28]. Hyperspectral sensors capture signals in a wide spectrum and it can be expected that different
parts of the spectrum will have differing representative capabilities for distinguishing the objects of interest.

With the recent developments in remote sensing instruments, hyperspectral images are now widely used in disparate application domains [45]. The special characteristics of hyperspectral data sets bring difficult processing problems. A well-known difficulty in supervised hyperspectral image classification is the limited availability of training data, which are difficult to obtain in practice as a matter of cost and time. A new method for segmentation and classification of hyperspectral images is proposed in this analysis [65]. The method is based on the construction of a Minimum Spanning Forest (MSF) from region markers. The construction of MSF belongs to graph based approaches for image segmentation. Furthermore, the segmentation and classification of the image were performed by constructing MSF based on the selected markers. Watershed transformation is one of the most powerful tools for image segmentation.

The segmentation by the watershed of hyperspectral images has shown to improve the result of classification in HSI. An imperative aspect of spectral image analysis is description of materials present in the object or scene being imaged. Band selection is a frequent approach for dimensionality reduction [73]. When the desired object information is unknown, an unsupervised band selection technique is employed to select the most distinctive and informative bands. Compared to supervised band selection techniques, unsupervised methods need no priori information about objects and classes.

In [5], a correlation matrix feature extraction based on spectral clustering (CMFESC) is proposed. A visual correlation matrix pseudo colour map of bands is proposed to emphasize the importance of second-order statistics in HSI. Moreover, the correlation matrix is also a similarity matrix of bands. Hence, the spectral clustering is applied to the correlation matrix and the membership values of the bands could be used for determining the
transformation matrix. CMFESC can solve the problem of selecting the thresholds in the greedy modular Eigen-space. In this study [51], unsupervised classification of hyperspectral imagery is carried out by using correlation clustering. As hyperspectral imagery is high dimensional data and suffers from “curse of dimensionality”, correlation clustering can be used to address these issues. The main advantage of the correlation clustering algorithm lies in its ability to find a subset of points (clusters) within a projected subspace and further, this projected subspace may differ from each cluster. For correlation clustering feature reduction method is tightly knitted with the clustering procedure. Instead of PCA, the SPCA is interlaced with the ORCLUS algorithm. Experiments are conducted on three real hyperspectral images. For all the dataset, performance of ORCLUS is acceptable. Major drawback lies in finding the appropriate values of the parameters. Even though the correlation clustering has appealing features for treatment of high dimensionality, but still more efforts and investigations are required to make it suitable for HSI. In this work [23], the pixel-based and the object-oriented image classification was utilized to perform land-cover mapping in a coal fire area. Twelve land-cover classes were intended to be classified as coal, coal dust, limestone, sandstone, metamorphic stone, mixed sandstone with shale, pure Gobi desert sand, Gobi desert sand with vegetation, river, agriculture, settlement, and bare land. The aim was to evaluate the performance of these two methods in (potential) surface coal fire area mapping. This work [7] has developed a new classification, strategy that integrates sparse representations and EMAPs for spatial-spectral classification of remote sensing data. Experiments reveal that the proposed approach, which combines the advantages of sparse representation and the rich structural information provided by EMAPs, can appropriately exploit the inherent sparsity present in EMAPs in order to provide state-of-the art classification results. This is mainly due to the fact that the samples in EMAP space can be approximately represented by a few numbers of atoms in the training dictionary after solving
the optimization problem, whereas the same samples could not be represented in the original spectral space with the same level of sparsity. The proposed strategy was tested on both simulated and real multi/hyperspectral data sets. A comparison with state-of-the-art classifiers shows very promising results for the proposed approach, particularly when a very limited number of training samples are available. This work [50], is introducing a novel spectral-spatial classifier for hyperspectral image data. This method is based on the consideration of both global posterior probability distribution and local probabilities which result from the whole image and set of previously derived class combination maps, respectively.

Even though hyperspectral data can provide finely resolved details about the spectral properties of features to be identified, it also has some limitations. When dealing with such high-dimensional data, one is faced with the “curse of dimensionality” problem [36]. One popular way [42] to tackle the curse of dimensionality is to employ a feature extraction technique.

### 2.4. Inter and intra band clustering method

Clustering targets to identify and create structure in data sets by determining and evaluating similarities among specific data patterns. The basic idea in most clustering algorithms is to identify a set of points and then update pattern membership to clusters iteratively, so as to achieve a better partition. Vector quantization provide a means of decomposition of the signal in approaches which takes the improvement of intra and inter band correlation as a more lithe partition for higher dimension vector spaces [8]. The vector quantization is a classical quantization approach from signal processing and image compression, which concedes the modelling of probability density functions of the distribution of prototype vectors. The accurate classification of remote sensing is an imperative task for many practical applications, such as precision agriculture, monitoring,
management of the environment, security and defence issues [64]. Besides that, a new spectral-spatial classification pattern for hyperspectral images is proposed. The method incorporates the results of a pixel wise support vector machine classification and the segmentation map attained by partitioned clustering using majority voting. The ISODATA algorithm and Gaussian mixture resolving schemes are used for image clustering. The cluster based segmentation methods desires in finding distinct structures in the spectral feature space. This clustering is a comprehensive partitioning of a set of pixels from the input image into homogenous groups of pixels. The cluster based segmentation of hyperspectral images has examined. Clustering is a search for hidden pattern that may exist in data sets. It is a technique of grouping data objects into disjoined clusters so that the data in each cluster are similar, yet disparate to the others.

A new unsupervised multispectral image segmentation algorithm had proposed in [29]. This method is based on a combination of the watershed transform, proposed size-weighted fuzzy clustering method, and MC method. The evaluation of this method has shown that it detects small regions (that often appear in remote-sensing images) better than other reported methods. Also, it produces homogeneous segments by accounting for spatial relations among image pixels by performing a cluster-based method. This method is based on fuzzy logic and has detected the overlapped regions observed in mist satellite images. The fuzzy clustering step requires prior information on the number of clusters, but the integration of this method with a hierarchical fuzzy clustering method that does not require one to know the number of objects may result in a fully automatic segmentation algorithm. The work [47] is addressing high dimensional hyperspectral data using diffusion maps, which consider the Eigen-function of Markov matrices as a system of coordinates on the original data set in order to obtain efficient representation of data geometric descriptions. The major difference between diffusion maps and methods like principle components analysis (PCA) is that in
diffusion maps, a kernel is chosen before the procedure. This kernel is chosen by our prior definition of the geometry of the data. In PCA, all correlations between values are considered, while only high correlation values are considered in diffusion maps. Diffusion maps have already been applied in the analysis of protein data, gene expression data, video sequences, and so on, and have achieved attractive performances. In this work [1] a hardware/software co-design approach to implementation of K-Means algorithm is addressed. A comparison with a software version was also presented in order to show the advantages of the former approach for this kind of problem, where high computation and parallelism processing are important issues. Even though the limitation of the FPGA and its low operating frequency, 40 MHz, 12.5 times lesser than software solution, the co-design approach has 2 times faster.

2.5. Techniques for hyperspectral image segmentation and multiband reduction

The following table illustrates the different author proposed different technique for hyperspectral image segmentation and multiband reduction which published in various year.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year and reference</th>
<th>Techniques</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen, et al.</td>
<td>2008 [14]</td>
<td>Quantization technique for HSV (HUE, Saturation, Value) colour space</td>
<td>HSV colour space is implemented to generate a colour.</td>
</tr>
<tr>
<td>Tarabalka, et al.</td>
<td>2009 [64]</td>
<td>ISODATA algorithm and Gaussian mixture resoling technique</td>
<td>This method is well for classification of images with large spatial structures</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Method Description</td>
<td>Notes</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Li, et al.</td>
<td>2011</td>
<td>Supervised Bayesian segmentation approach</td>
<td>Bayesian segmentation approach addressing ill-posed hyperspectral classification and segmentation problems</td>
</tr>
<tr>
<td>Yang et al.</td>
<td>2011</td>
<td>Similarity based unsupervised band selection</td>
<td>This algorithm is applied in the spatial domain for band selection</td>
</tr>
<tr>
<td>Fan, et al.</td>
<td>2009</td>
<td>Weighted Fuzzy C-Means clustering algorithm</td>
<td>This method can solve the FCM’s problem</td>
</tr>
<tr>
<td>Ghosh, et al.</td>
<td>2011</td>
<td>Context sensitive technique</td>
<td>Two fuzzy clustering algorithms, namely FCM and Gustafson-Kessel Clustering (GKC) algorithms have been used</td>
</tr>
<tr>
<td>Zalik</td>
<td>2008</td>
<td>K-Means clustering</td>
<td>It performance correct clustering without pre-assigning the exact number of clusters</td>
</tr>
<tr>
<td>Paoli, et al.</td>
<td>2009</td>
<td>Multi Objective Particle Swarm Optimization (MOPSO)</td>
<td>It solves the problem of initialization using a built-in boosting function</td>
</tr>
<tr>
<td>Louisis du Plessis, et al.</td>
<td>2011</td>
<td>Working with diffusion maps, which consider the Eigen-function of Markov matrices</td>
<td>It addressing the complicated hyperspectral data and identifying the minerals in core samples</td>
</tr>
<tr>
<td>Bar-Chen Kua, et al.</td>
<td>2012</td>
<td>The spectral clustering is applied to the correlation matrix of bands and the corresponding membership values determine the transformation matrix</td>
<td>The corresponding membership values determine the transformation matrix. It can solve the problem of selecting the thresholds in the greedy modular Eigen-space. In addition improve the clustering performance.</td>
</tr>
<tr>
<td>A.Mehta</td>
<td>2014</td>
<td>ORCLUS, a correlation clustering algorithm, is implemented and enhanced by making use of segmented principle component analysis (SPCA)</td>
<td>Preliminary analysis of algorithms on real hyperspectral imagery shows ORCLUS is able to produce acceptable results.</td>
</tr>
</tbody>
</table>

Various approaches for hyperspectral image segmentation and disparate clustering approaches for multiband reduction are depicted in the above table. A clustering and fusion method improves the classification performance of hyperspectral imagery. The PSOKHM
algorithm searches robustly the data cluster centers using the sum over all data points to all
the centers as a metric. To diminish computation complexity, band selection can be
conducted on automatically selected pixels from the N-FINDR algorithm. A remote sensing
image segmentation procedure that utilizes a single point iterative weighted FCM clustering
algorithm based upon the prior information. A large number of spectral channels in a
hyperspectral image increase the potential of discriminative physical materials and structures
in a scene.

2.6. Introduction to cluster analysis.

Cluster analysis is an important exploratory tool widely used in many areas such as
sociology, medicine and business. For example, in computational biology, the cluster has
been successfully implemented in inferring function of unknown genes and detecting classes
or sub-classes of diseases. The goal of cluster analysis [26] is to assign objects in a data set
into meaningful classes such that objects in the same class are more similar to each other than
to those in other classes. The reasonable way of summarizing the observed data into classes is
determined only based on the information provided by the data since there is no prior
knowledge about the classes at the beginning of an investigation. Cluster analysis involves
several procedures such as selecting clustering objects and clustering variables, variable
standardization, choosing the measure of association, selecting the clustering method,
determining the number of clusters and interpretation, validation and replication. Clustering
methods are applied in many application areas such as pattern recognition, data analyses,
information retrieval and image processing.
2.6.1. K-Means clustering method

K-Means is a typical algorithm. It is attractive in practice, because it is elementary and it is typically very fast. It segregates the input dataset into k-clusters. Each cluster is described by an adaptively changing centroid, starting from some initial values named seed points. K-Means enumerates the squared distances between the inputs and centroids, and assigns inputs to the nearest centroid. K.R.Zalik [78] recommends K-Means algorithm that implements the correct clustering without pre-assigning the exact number of clusters. As a traditional clustering algorithm, K-Means is suitable for its simplicity for implementation and it is generally applied to grouping pixels in images or video sequences.

The performance of the K-Means algorithm depends on initial cluster centers. Furthermore, the final partition depends on the initial configuration. Solving the selection of a correct cluster number has been tried in two ways. The first one invokes some heuristic approaches [33]. The clustering algorithm is run many times with the number of clusters gradually increasing from a certain initial value to some threshold value that is difficult to set. The second is to formulate cluster number selection by choosing a component number in a finite mixture model. As K-Means approach is iterative, it is computationally and hence applied only to image subareas rather than to full scenes and can be treated as unsupervised training areas. K-Means clustering aims is to partition the $n$ pixels into $k$ sets ($k \leq n$)

The main drawback of the K-Means algorithm is that the number of clusters is fixed: once $K$ is chosen it always returns $K$-cluster centers. Despite being widely used in data analysis, pattern recognition and image processing. K-Means has three major constraints:

- The number of clusters must be previously known and fixed.
- The results of K-Means algorithm depend on initial cluster centers.
- The algorithm contains the dead-unit problem.
2.6.2. Fuzzy C-Means (FCM) clustering method

Clustering technique is applied in many application areas such as pattern recognition, data mining, machine learning, etc. A.Sophark, et.al, [63] suggests a variation of the Fuzzy C-Means (FCM) algorithm that contributes image clustering. The modified algorithm is called Fuzzy Local Information C-Means can overcome the drawbacks of the known Fuzzy C-Means algorithms in addition it improves the clustering performance. The major FLICM is the use of a fuzzy local similarity measure, focusing to guarantee noise insensitiveness and image detail preservation.

In [63], a remote sensing image segmentation procedure that utilizes a single point iterative weighted FCM clustering is proposed based upon the prior information. This method can solve the FCM algorithm’s problem that the clustering quality is greatly affected by the data, distributing and the stochastic initializing the centrals of clustering. Clustering analysis based methods can provide a non-parametric, supervised approach to the analysis of each kind of images. Many modified classifiers based on FCM algorithms have been applied in image segmentation. FCM algorithm has proved its efficacy for image segmentation. However, still it lacks in getting robustness to noise and outliers, specifically in the absence of prior knowledge of the noise. To overcome this dilemma, a speculated Novel Multiple Kernel Fuzzy C-Means (NMKFCM) technique with spatial information is imported as a framework for image segmentation obstacle [24].

The FCM algorithm is one of the most popular clustering methods based on minimization of a generalized least-squared error function. Given a data set, \( X = \{ x_i, x_2, \ldots, x_n \} \subseteq R^{n \times q} \), where \( n \) is the number of samples, \( q \) is the dimension of the sample \( x_j \) (\( j = 1, 2, \ldots, N \)). The FCM algorithm is based on minimizing the criterion with respect to the membership value \( U_{ij} \) and the distance \( d_{ij} \). Here \( n \) is the number of objects and \( C \) is the
number of clusters, where $U_{ij}$ is the degree of membership that the object $x_j$ pertains to the cluster center $w_i$

$U = \{ U_{ij}, i = 1,2,...C, j = 1,2,...n \}$ which is the membership matrix.

$V = \{ v_i, i = 1,2,...C \}$ is the cluster prototype matrix and is the prototype of the center of cluster $i$. $m \in [1, \infty)$ is the fuzzy factor. According to many studies, $m \in [2, 2.5]$ is practical [39].

The FCM algorithm can be summarized by the following steps:

Step 1: Initialize matrix $U = [ U_{ij} ]$ with the initial value $U^{(0)}$

Step 2: At $k$-step: Calculate the cluster prototype matrix $V^{(k)} = [v_i]$ with $U^{(k)}$

Step 3: Update $U^{(k)}$, $U^{(k+1)}$

Step 4: If $|| U^{(k+1)} - U^{(k)} || < \epsilon$ then stop, otherwise go to step 2.

To sum up, the basic idea of the FCM algorithm is that use iterative method for solving equation (2) and (3), until a termination condition is met.

K-Means is one of the hard clustering methods. The conventional hard clustering methods classify each point of the data set to one cluster. Among the Fuzzy Clustering methods, FCM algorithm [21] is one of the most prominent techniques used in image segmentation because it has robust characteristics for ambiguity and can retain much more information than hard segmentation methods. FCM algorithm is a generalization of the Hard C-Means algorithm yields extremely good results in an image region clustering and object classification. Despite the starting centers selected arbitrarily, K-Means is fully deterministic, given the stating centers. A bad choice of initial centers can have a great impact on both performance and distortion.
2.6.3. Fast K-Means methods

A group of researchers worked on choosing the best centers to avoid the problems of K-Means of either obtaining the non-optimal solutions or empty clusters generations. [53] Worked on modifying the K-Means to avoid the empty clusters. They moved the center of every cluster into new locations to ensure that there will be no empty clusters. The comparison between their modified K-Means and the original K-Means show that the number of iterations is higher with the modified K-Means method. In case of the numerical examples which produce empty clusters, the proposed method cannot be compared with any other method since there is no modified K-Means algorithm available to avoid the empty clusters. On the other hand developed, [10] a procedure in which the centers have to pass a refinement stage to generate good starting points. [71] Used genetically guided K-Means where the possibility of empty clusters will be treated in the mutation stage. Another method of center initializing based on the values of attributes of the dataset is proposed by [41]. The later proposed method creates a complex procedure which leads to be computationally expensive. In normal K-Means algorithm, if the initial centers are located exactly at the means of the clusters of the data, then the algorithm requires only one step to assign the individual clusters to each data point. This work [31], address for trying to get the stage of moving any initial centers to a location which is either that of the means or near them. The big gap between these locations will decide how many times the normal K-Means is required to run to assign all data to their clusters. This algorithm [60] will quickly move the centers to locations which are near its means.

Even though K-Means offers no accuracy guaranteed, its simplicity and speed are very appealing in practice. By augmenting K-Means with a simple, randomized seeding technique, obtain an algorithm [17] that is $O (\log k)$ –competitive with the optimal clustering.
Experiments show this augmentation improves both the speed and the accuracy of K-Means, often quite dramatically.

### 2.6.4. Robust K-Means (RKM) clustering method

Prof. Karsin introduces an extension of K-Means algorithm that removes the outlier is entitled as Robust K-Means [39]. It is an improvement in accuracy and efficiency over previously used geometric clustering algorithms. It accomplishes this by taking an Information theory approach to geometric clustering using the Information Bottleneck method. The main goal is to retain as must relevant information about the location of the data points while compressing the data points into the clusters. The resulting algorithm is more robust to initial centroid placement and has a faster execution time than previous methods. In [49] study generalized K-Means and generalized trimmed K-Means performance from the viewpoint of Hampel’s robustness criteria, that is, investigate the influence function, breakdown point, and qualitative robustness, confirming the superiority provides by the trimming. It includes the study of two real data sets to make clear the robustness of generalized trimmed K-means.

Despite the wide use of clustering algorithms for hyperspectral image segmentation, their performance suffers from a Hughes phenomenon [34], a kind of the curse of dimensionality problem. When the dimension of the space grows and the size of the training set is fixed, the classification accuracy reaches a maximum for a given size, and then decreases. It has been shown that using the feature extraction technique, e.g., principal component analysis (PCA) [79], in pre-processing can improve the performance of clustering. Recently, the correlation matrix is used for dimension reduction by combining the greedy modular eigen-space and the positive Boolean function [12]. A visual correlation
matrix pseudo-colour map of bands is proposed to emphasize the importance of second-order
statistics in hyperspectral image [43]. However, it is hard to determine the threshold values
for the greedy modular eigen-space. Spectral clustering has become a popular clustering
algorithm [11], especially on nonlinear problems.

2.7. Particle swarm optimization (PSO)

Particle Swarm Optimization (PSO) algorithms represent a new approach for
optimization. A clustering based PSO finds the centroids of a user specified number of
clusters, where each cluster group together similar patterns. PSO is easy to implement and
has been successfully applied to solve a wide range of optimization problems such as linear
and discrete optimization problems. The basic concept of the algorithm is to create a swarm
of particles which move in the space around them searching for their goal or the place which
best suits their needs given by a fitness i.e. objective function. It is initialized with a group of
random particles and then searches for optimum by updating generations. For every iteration,
each particle is updated by two best values. The first one is the best solution, it has achieved
so far has entitled as pbest. Another best value that has tracked based on the particle swarm
optimizer is the best value, obtained so far by any particle in the population is named as
gbest.

In [54], a new methodology for clustering hyperspectral images is conferred. It targets at
synchronously solving the following three different issues:

- Estimation of the class statistical parameter.
- Detection of the best discriminating bands without requiring the apriority setting of
  their number by the user.
- Estimation of the number of data classes describing the examined image.
It is formulated within a Multi Objective Particle Swarm Optimization (MOPSO) framework. The objective of this research is to propose a novel methodology for hyperspectral images capable of simultaneously solving the above problems, i.e., Clustering, feature detection and class number estimation. The objective is to classify the image in an unsupervised way. Given its hyperspectral nature, it is preferable to perform beforehand a feature detection operation. PSO [72] is a population based search method, which exploits the concept of social sharing of information. This means that each individual (called particle) of a given population (called swarm) can profit from the previous experiences of all other individuals from the same population. PSO is a prominent and robust method for optimization problems [74]. But the main obstacle in resolving PSO to real world application is that PSO typically needs an extensive number of fitness evaluations before a gratifying result can be attained. In this analysis, the improved algorithm, Fuzzy C-Means based on Picard iteration and PSO (PPSO – FCM) is proposed. A new segmentation method [20] produces through combining PSO algorithm with one of region-based image segmentation methods, which are names Seeded Region Growing (SRG). This algorithm randomly initializes each particle in the swarm to contain K-seed point and then SRG algorithm is applied to each particle.

A plethora clustering methods are available for cluster analysis. Here we discussed only few and most used one. However, a fundamental problem in applying most of the existing clustering approaches is that the number of clusters needs to be pre-specified before the clustering is conducted. The clustering results may heavily depend on the number of clusters specified. It is necessary to provide educated guidance for determining the number of clusters in order to achieve appropriate clustering results.
2.8. Determining the number of clusters

A fundamental problem in cluster analysis is to determine the number of clusters, which is usually taken at a priori in most clustering algorithms. Clustering solutions may vary as different numbers of clusters are specified. The result of a clustering algorithm can be very different from each other on the same data set as the other input parameters of an algorithm can extremely modify the behaviour and execution of the algorithm. The aim of the cluster validity is to find the partitioning that best fits the underlying data. Usually 2D data sets are used for evaluating clustering algorithms as the reader easily can verify the result. But in case of high dimensional data the visualization and visual validation is not a trivial task therefore some formal methods are needed.

The process of evaluating the results of a clustering algorithm is called cluster validity assessment. Two measurement criteria have been proposed for evaluating and selecting an optimal clustering scheme [22]:

• Compactness: The member of each cluster should be as close to each other as possible. A common measure of compactness is the variance.

• Separation: The clusters themselves should be widely separated. There are three common approaches measuring the distance between two different clusters: distance between the closest member of the clusters, distance between the most distant members and the distance between the centers of the clusters.

There are three different techniques for evaluating the result of the clustering algorithms namely, External Criteria, Internal Criteria and Relative Criteria.

Both internal and external criteria are based on statistical methods and they have high computation demand. The external validity methods evaluate the clustering based on some user specific intuition. The internal criteria are based on some metrics which are based on the data set and the clustering schema.
2.8.1. Davies–Bouldin index

The Davies–Bouldin index (DBI) (introduced by David L. Davies and Donald W. Bouldin in 1979) is a metric for evaluating clustering algorithms [16]. This is an internal evaluation scheme, where the validation of how well the clustering has been done is made using quantities and features inherent to the dataset. The Preliminaries of the Davies–Bouldin Index is as follows,

Let \( A_i \) be a cluster of vectors. Let \( X_j \) be an \( n \) dimensional feature vector assigned to cluster \( A_i \).

\[
S_i = \frac{1}{T_i} \sum_{j=1}^{T_i} \| X_j - A_i \|_p 
\]

...(2.1)

Here \( A_i \) is the centroid of \( A_i \) and \( T_i \) is the size of the cluster \( i \). \( S_i \) is a measure of scatter within the cluster. Usually the value of \( p \) is 2, which makes this a Euclidean distance function between the centroid of the cluster, and the individual feature vectors. Many other distance metrics can be used, in the case of manifolds and higher dimensional data, where the Euclidean distance may not be the best measure for determining the clusters. It is important to note that this distance metric has to match with the metric used in the clustering scheme itself for meaningful results.

\[
M_{ij} = \left\| A_i - A_j \right\|^p = \left( \sum_{k=1}^{n} \| a_{k,i} - a_{k,j} \|_p \right)^{1/p} 
\]

...(2.2)

\( M_{ij} \) is a measure of separation between cluster \( A_i \) and cluster \( A_j \).

\( a_{k,i} \) is the \( k \)th element of \( A_i \), and there are \( n \) such elements in \( A \) for it is an \( n \) dimensional centroid. Here \( k \) indexes the features of the data, and this is essentially the Euclidean distance between the centers of clusters \( i \) and \( j \) when \( p \) equals 2. The Davies–Bouldin index [16] is based on similarity measures of cluster \( (R_{i,j}) \) whose bases are the dispersion measures of a cluster \( (S_{i}) \) and the cluster dissimilarity measure \( (d_{i,j}) \).
The similarity measure of clusters ($R_{i,j}$) can be defined freely, but it has to satisfy the following condition [22]:

- $R_{i,j} \geq 0$
- $R_{i,j} = R_{j,i}$
- If $S_i = 0$ and $S_j = 0$ then $R_{i,j} = 0$
- If $S_j > S_k$ and $M_{i,k}$ then $R_{i,j} > R_{i,k}$
- If $S_j = S_k$ and $M_{i,j} < M_{i,k}$ then $R_{i,j} > R_{i,k}$

Usually $R_{i,j}$ is defined in the following way:

$$R_{i,j} = \frac{S_i + S_j}{M_{i,j}} \quad \ldots(2.3)$$

$$d_{i,j} = d(v_i, v_j), \quad S_i = \frac{1}{\|C_i\|} \sum_{x \in C_i} d(x, v_j) \quad \ldots(2.4)$$

Then the Davies- Bouldin index is defined as

$$DB = \frac{1}{N} \sum_{j=1}^{N} D_j \quad \text{where} \quad \ldots(2.5)$$

$$D_j = \max_{j \neq i} R_{i,j} \quad \ldots(2.6)$$

DB is called the Davies-Bouldin index.

This topic comprises the review about hyperspectral image segmentation and multiband reduction. Furthermore, this review describes various methods and techniques such as K-Means, Fuzzy C-Means Clustering, Fast K-Means, Robust K-Means, and Particle Swarm optimization. As a whole there is a need for developing a new method for hyperspectral image segmentation and multiband reduction, which can be performed with minimum time and computational complexity.