Hierarchical Segmentation and Classification of Satellite Images and Merging Using Minimum Spanning Tree

Ch. Lakshmi Narayana 1, Dr. K.Venkata Ramana2
1 M.Tech, Dept.of Computer Science and Systems Engineering, Andhra University(A), Visakhapatnam, Andhra Pradesh, India.
2 M.Tech, Ph.D, Assistant Professor, Department of CS&SE, Andhra University College Of Engineering(A), Visakhapatnam, India.

Abstract: This paper presents various leveled grouping calculations for land cover mapping issue utilizing multi-ghostly satellite pictures. In unsupervised methods, the programmed era of number of groups and its habitats for an enormous database is not misused to their maximum capacity. Thus, a various leveled clustering calculation that utilizes part and combining systems is proposed. At first, the part strategy is utilized to hunt down the most ideal number of groups and its focuses utilizing Mean Shift Clustering (MSC), Niche Particle Swarm Optimization (NPSO) furthermore, Glowworm Swarm Optimization (GSO). These 3 techniques are individually combined with MST technique. Utilizing these clusters furthermore, its focuses, the combining technique is utilized to aggregate the information focuses in light of a parametric. A execution correlation of the proposed various leveled clustering calculations (MSC, NPSO and GSO) combined with Minimal Spanning Tree(MST) is introduced for Worldview 3 Satellite image.

1. Introduction.

Land is the basic building block of human civilization. By nature, this precious gift cannot be expanded. To make best use of land and its natural resource, we need good factual knowledge of the land and its features. Accurate knowledge on land-use is very vital for planning and efficient operation. The satellite image is one of the sources which can capture the temporal nature of this knowledge for land utilization. As computer science has raised a stadium where computers are able to perform some “intelligent” tasks, a wide research area has been established in solving the problem of automatic image classification. Land cover mapping information can be used to audit land usage, in the context of city planning and land-usage [1]. For a given satellite image, if there is a lack of ground truth information then unsupervised technique can be applied for automatically classifying a satellite image into distinct land cover regions [2]. In unsupervised technique without prior knowledge of labels, data sets are sub-divided into groups or clusters, based on some attributes. The main aim is to ensure that the distance between intra-cluster is minimum and inter-cluster is maximum. The clustering problems can be developed and analysed hierarchically [3]. Hierarchical clustering constructs a hierarchy of clusters by splitting a large cluster into smaller ones and merging smaller cluster into their nearest centroid [4]. There are two main approaches: (i) The divisive approach, which splits a larger cluster into two or more smaller ones; (ii) The agglomerative approach, which builds a larger cluster by merging two or more smaller clusters. The hierarchical cluster provides a comprehensive description for any data set. There are many clustering methods to split and merge the data set. They are broadly classified into parametric and non-parametric [3]. In parametric methods prior assumptions are made regarding the shape or number of clusters - K-means clustering [5]. In non-parametric methods no prior assumptions need to be made - Mean Shift Clustering (MSC) [6].

A popular parametric method K-means clustering, is essentially a function minimization technique, where the objective function is the squared error. Non-parametric technique such as MSC is a procedure for locating the maxima of a mapped function given a set of discrete data points sampled from that function. It is useful for detecting the modes of density given a density function. In this paper, we present a new nature inspired technique Glowworm Swarm Optimization (GSO) [7] clustering algorithm for hierarchical splitting and merging of automatic multi-spectral satellite image classification (land cover mapping problem). To compare GSO hierarchical clustering and classification model with Mean Shift Clustering (MSC). The main challenge here is how best the clusters can be splitted and merged to classify the data set to their respective group. The performance of the hierarchical clustering models – GSO and MSC depends on automatic generation of number of cluster centers to classify
efficiently. The performance of clustering is evaluated using classification accuracy - individual, average and overall accuracy. In the proposed model, clustering is analyzed for worldview3 [9]. The MST clustering algorithm is known to be capable of detecting clusters with irregular boundaries. Unlike traditional clustering algorithms, the MST clustering algorithm does not assume a spherical shaped clustering structure of the underlying data.

A minimum spanning tree (MST) or minimum weight spanning tree is a subset of the edges of a connected, edge-weighted undirected graph that connects all the vertices together, without any cycles and with the minimum possible total edge weight. That is, it is a spanning tree whose sum of edge weights is as small as possible. More generally, any undirected graph (not necessarily connected) has a minimum spanning forest, which is a union of the minimum spanning trees for its connected components.

There are quite a few use cases for minimum spanning trees. One example would be a telecommunications company which is trying to lay out cables in new neighborhood. If it is constrained to bury the cable only along certain paths (e.g. along roads), then there would be a graph representing which points are connected by those paths. Some of those paths might be more expensive, because they are longer, or require the cable to be buried deeper; these paths would be represented by edges with larger weights. Currency is an acceptable unit for the underlying data. A spanning tree for that graph would be a subset of those paths that has no cycles but still connects to every house; there might be several spanning trees possible. A minimum spanning tree would be one with the lowest total cost, thus would represent the least expensive path for laying the cable.

2. Related Work.

Unsupervised techniques can be used for grouping distinct land cover regions, provided there is a lack of ground truth information [10]. Based on certain similarity metric, the data is sub-divided into clusters [11], [12], using unsupervised methods where the number of clusters is not known a priori [13]. The objective is to maximize the inter-cluster distances while the intra-cluster distances are minimized. The clustering problems can be studied using hierarchical approach [14], by breaking a large cluster and merging smaller groups into their closest centroid [15]. Two approaches are used in this hierarchical clustering method: (i) divisive methods, where a large cluster is split into several small clusters; (ii) agglomerative methods, where many small clusters are merged to form a large cluster.

The grouping of the same clusters is regarded as a fundamental task in land cover mapping problem, which transforms the remotely sensed images to generate thematic land-use/land-cover maps [13]. Several methods to compute a single-band gradient function from satellite images have been studied previously by Tarabalka et al. including pixel-wise classification methods [12], [16]. Studies show that hierarchical step-wise optimization and spectral clustering have given good results for analyses of satellite images [17]. And these results have been improved by a combination of probabilistic classification and hierarchical step-wise optimization algorithm [18].

In the literature different methods have been developed to cluster data sets by splitting and merging [14]. Broadly, they can be classified into parametric and non-parametric methods. In parametric methods such as -means clustering [19], prior assumptions of the number of clusters are made. This is essentially a function minimization technique, where the objective function is the squared error distance measure. In non-parametric methods such as Mean Shift Clustering (MSC) [20], [21], no prior assumptions are made on the number of clusters. This is a procedure for locating the maxima of a mapped function given a set of discrete data points sampled from that function. It is useful for detecting the modes of density given a density function.

2. Splitting.

Splitting and merging attempts to divide an image into uniform regions. The basic representational structure is pyramidal, i.e. a square region of size $m$ by $m$ at one level of a pyramid has 4 sub-regions of size $m/2$ by $m/2$ below it in the pyramid. Usually the algorithm starts from the initial assumption that the entire image is a single region, then computes the homogeneity criterion to see if it is TRUE. If FALSE, then the square region is split into the four smaller regions. This process is then repeated on each of the sub-regions until no further splitting is necessary.

A. The Mean Shift Algorithm.

The Mean Shift algorithm is a non-parametric, iterative mode seeking algorithm. It was originally presented in 1975 by Fukunaga and Hostetler [12]. It was significantly improved by Cheng in 1995 [6] which rekindled the interest for it. In recent years it has become popular in computer vision applications [7] and it will also be used for the
first stage of the two-stage Mean Shift Spectral Clustering approach.

**Mean Shift Clustering Algorithm.**

Let \( x_1, x_2, \ldots, x_n \) where \( n \) is the number of data points in \( d \)-dimensional space. In the mean shift clustering, each data point is shifted to the average of the other data points in its neighbourhood. This is done by using a Gaussian kernel, based on Euclidean distance between two data points \( (r) \) which is given by

\[
K(r) = e^{-\|r\|^2}
\]

The main idea behind mean shift clustering is to treat the points in the \( d \)-dimensional feature space as a probability density function. The dense regions in the feature space correspond to the local maxima of the underlying distribution. We perform a gradient ascent procedure on the local estimated density until convergence is reached. The stationary points of this procedure represent the modes of the distribution.

The mean shift for point \( x_i \) is

\[
m(x_i) = \frac{\sum_{j=1}^{n} x_j g(||x_i - x_j||^2)}{\sum_{j=1}^{n} g(||x_i - x_j||^2)}
\]

where \( h \) is a bandwidth to limit the density within the kernel function \( g = -k'(x), x \) is the mean of all \( n \) data points. A complete discussion about various aspects of mean shift clustering and its applications are given in [12], [13].

**B. Niche PSO.**

The NichePSO algorithm developed by Brits et al., is a niche particle swarm optimizer where many solutions are located in parallel. Many swarms are improve from an initial particle population. As niches are found, subswarms are formed. Finally, each subswarm denotes one of the potential solutions to the problem.

**Niche PSO Algorithm:**

Create and Initialize a n-dimensional main swarm, \( S \);  
repeat  
Train the main swarm, \( S \), for one iteration using the cognition-only model.  
Update the fitness of each main swarm particle \( x_{S_i} \)  
for each subswarm \( S_j \) do  
Train subswarm particles, \( x_{S_{j,k}} \) using a full model PSO which incorporates the social component;  
Update each particles’s fitness;  
Update the swarm radius \( R_{S_j} \);  
end  
if a merging condition is satisfied then  
merge the corresponding subswarms;  
end  
Allow subswarms to absorb any particles from the swarm that moved into the subswarm;  
if a particle in main swarm converged then  
Create a new subswarm with the particle and its closest neighbour;  
end  
Until Stopping condition is true;

**C. Glowworm Swarm Optimization.**

Most kinds of glowworms can locate its position and exchange information by sending out rhythmed short beam. The idea of GSO is glowworm individual finding flaring neighbors in its searching scope. Move from initial position to a better one and at last assemble into one or more extreme value point.

In GSO algorithm, Glowworm individuals’ attraction is only related to its brightness. Attraction of individual is proportional to brightness and inversely proportional to the distance between the two individuals. The position of individuals account for objective function value. Define dynamic decision domain as individual searching scope. When updating position, individuals move by step.

\[
l_j(t+1) = \max\{0, (1-\rho)l_j(t) + \gamma l_j'(t+1)\}
\]

\[
x_i(t+1) = x_i(t) + \frac{x_i(t) - x_i(0)}{\|x_i(t) - x_i(0)\|}
\]

\[
\tau_j(t+1) = \min[\tau_j, \max\{0, \tau_j(t) + \lambda (\|x_j(t)\| - |x_j(0)|)\}]
\]

**Splitting the Data Set Using GSO.**

Step 1. Initialization – Randomly choose \( N \) points as agents assign luciferin value \( l_{i_0} \) and sensor range \( r_s \) for each agent.  
Step 2. Luciferin update phase using (1).  
Step 3. Move the glowworms using (2).  
Step 4. Vary the local decision range using (3).  
Step 5. Repeat Steps 2 to 5 till convergence is achieved.

**3. Merging:**

The square regions which are split in splitting are now merged if they are similar to give larger irregular regions. The problem (at least from a programming point of view) is that any two regions may be merged if adjacent and if the larger region
satisfies the homogeneity criteria, but regions which are adjacent in image space may have different parents or be at different levels (i.e. different in size) in the pyramidal structure. The process terminates when no further merges are possible.

A. K-Means Algorithm:
The k-means algorithm is one of the best known clustering algorithms [16]. Since it has relatively low computational complexity it is widely used, particularly for large data sets. K-means is an iterative clustering algorithm where the number of clusters, k, is an input parameter. It uses a crisp (hard) membership function, which means that each data sample is assigned to one and only one cluster. This leads to a non-differentiable cost function. Hence optimization is divided into two steps, where first the cluster representatives and then the memberships are kept constant while the other is optimized. Each cluster is represented by a point-representative, which causes the algorithm to favour dense clusters. The cost function is a function of the memberships and the dissimilarity measure.

The k-means algorithm can be derived from minimizing a cost function if the dissimilarity measure is the squared Euclidean distance. Then it can easily be shown that the point representatives that minimize the cost function are the k cluster means. Hence the name k-means. The cluster memberships of the data samples are not known in advance and the initial cluster representatives are usually initialized at random. However, the number of clusters are assumed to be know. In the first step, each data sample is assigned to its closest cluster representative. Then new cluster representative for a particular cluster is set equal to the mean of the data samples assigned to that cluster. These steps are repeated to some convergence criteria are met. In clustering we do not have training data with known labels (cluster assignments) available. The membership vectors ui must then be estimated from the data. A meaningful group assignment of data points is the goal of all clustering algorithms.

Given a set of points \( x_1, \ldots, x_N \in \mathbb{R}^P \) with unknown underlying probability distribution, a graph \( G(V,E) \) is constructed with vertex set \( V \) and edge set \( E \) where each of the \( x_i \) is represented by a node in the graph. The edge between a given pair of vertices \( x_i \) and \( x_j \) is weighted by the Euclidean distance:

\[
w(x_i, x_j) = 1 - \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)
\]

\( \sigma \) has been fixed by grid search [15] technique using 50% of the total data for cross-validation. The value of \( w \) will be smaller if \( x_i \) and \( x_j \) are similar to each other in some sense whereas will get increased in case the data points differ. A minimum spanning tree \( T(V,E_T) \) of \( G \) is constructed henceforth using the Kruskal’s approach [15]. The proposed clustering method checks iteratively whether a given edge of the tree can be deleted to form two compact sub-clusters. The edges are considered in the descending order of the corresponding edge weights because it has been assumed that edges with larger weights are the ones which span different clusters. Algorithm 2 [1] describes the proposed clustering process. A is a measure of the cluster compactness. If a given edge \( E_{w(t)} \) has drastically larger weight value than the component edge weights of its two adjoining connected components (CC’s), then it can be presumed that \( E_{w(t)} \) spans two different clusters and hence can be deleted. Given a CC, the corresponding \( \alpha \) controls the gap allowed in the feature space between this CC and any other CC.

Merging:
Put all regions on ProcessList
Repeat
Extract each region from ProcessList
Traverse remainder of list to find similar region (homogeneity criterion)
If they are neighbours then
merge the regions and
recalculate property values;
until (no merges are possible)

B. MST Based Clustering Algorithm:
5. Results:
The simulation results of the satellite image are shown below.

Figure 3. Original Image

Figure 4. Mean shift output

Figure 5: MS+ MST output

Figure 6. GSO output

Figure 7. GSO+MST

Figure 8. NSPO output
7. Conclusion.

In this paper, we have introduced the progressive clustering strategies for land cover mapping issue. The satellite picture utilized for this review is worldview3. The various leveled strategy receives MSC. NPSO and GSO have been performed on the image. The outputs of each technique is merged with MST technique to prove the best recognition of the image. Hence it is shown from simulation result that the proposed clustering method is feasible for satellite image classification.

References.