Efficient Multistage Approach for Unsupervised Image Classification

Sanghoon Lee

Dept. of Industrial Engineering, Kyungwon University
San 65 Bokjeong-dong, Sujeong-gu, Seongnam-si, Kyunggi-do 461-701, Korea
Tel.: +82-31-750-5367 Email: shl@kyungwon.ac.kr

Abstract – A multi-stage hierarchical clustering technique, which is an unsupervised technique, has been proposed in this paper for classifying the hyperspectral data.. The multistage algorithm consists of two stages. The "local" segmentor of the first stage performs region-growing segmentation by employing the hierarchical clustering procedure with the restriction that pixels in a cluster must be spatially contiguous. The "global" segmentor of the second stage, which has not spatial constraints for merging, clusters the segments resulting from the previous stage, using a context-free similarity measure. This study applied the multistage hierarchical clustering method to the data generated by band reduction, band selection and data compression. The classification results were compared with them using full bands.

I. INTRODUCTION

In general, most of the statistical techniques for image analysis are tailored to supervised learning that requires the prior knowledge of the number of classes and parameter values related to the class characteristics. The statistical characteristics are obtained from the training samples representing the regions associated with each class. However, estimating the statistical properties from the training data is an expensive process and the parameters are often unavailable because they continuously change over time and it is sometimes extremely difficult to locate adequate training field due to political or geographical problems. In addition, supervised classification results depend on the number of classes selected in the specific analyzed area, which is not easy to determine exactly. Therefore, unsupervised analysis is a very important approach including statistical inference of the parameters characterizing regional properties without the prior knowledge about the contents of the observed scene.

An image processing system involves both: 1) low-level processing in which general purpose techniques are applied to improve the original content of the scene, and 2) high-level processing that involves specific tasks designed for interpretation of the scene. Image segmentation is a typical component of the low-level processing stage of an image analysis system. High-level processing tasks such as classification are then devoted to analysis of the resulting regions. A system of unsupervised analysis, which integrates both levels of image processing system, has been proposed in this study.

In order to regularize the ill-posed problems in image processing, various approaches have been proposed by many authors, which involve introducing generic constraints related to image data structure or characteristics. Although it is not obvious how to constrain a segmentation algorithm to obtain a well-posed solution, one essential structural characteristics involves hierarchy of scene information[1]. Under the constraint of the hierarchical structure, it is possible to determine natural image segments by combining hierarchical clustering in the measurement space with spatial region growing. Hierarchical clustering of step-by-step merging based on similarity measures[4] is one of the most plausible, well-grounded tools for the unsupervised procedure. The hierarchical clustering procedure assumes very little a prior knowledge. However, in the analysis of large data acquired over large geographical area or by high spatial resolution radiometer, conventional hierarchical clustering approaches performing a successive merge of sub-regions is not practicable because of intensive computational and memorial requirements. This study used a multistage approach that is computationally efficient for the unsupervised classification. The multistage algorithm includes two stages of segmentation. The first stage performs region-growing segmentation that confines merging to spatially adjacent clusters and then generates an image partition such that no union of any neighboring segments is uniform. This regional segmentor employs a hierarchical clustering procedure that merge "mutual closest neighbor (MCN)" pair satisfying a given clustering criterion. In the second stage of global segmentation, the image partition resulting from the regional segmentation is classified into a small number of distinct states by a sequential merging operation. The global segmentor uses the conventional agglomerative hierarchical clustering scheme which merges step-by-step small two regions into a large one.

In statistical image segmentation, Markov random field (MRF) models [2] have frequently been used to characterize geophysical connectedness. The MRF represents the local characteristics of image structure such that neighboring pixels have a higher probability of being members of the same class. Unfortunately, conventional "distribution-free" or "context-free" similarity measures between regions in typical clustering approaches used for image segmentation fail to include important information about image spatial structure that should be exploited when the scene is segmented. Lee and Crawford [3] utilize the idea of MRF in the context of hierarchical clustering for the region-growing segmentation to incorporate spatial contextual information for unsupervised image classification. However, incorporating spatial contextual information into the hierarchical clustering of segmentation requires computational complexity. This study proposed to apply the segmentation to the data restored by a modified anisotropic diffusion restoration. The restoration uses a probabilistic model based on the MRF, which represents geographical connectedness, for iterative diffusion processing. In every iteration, the "bonding-strength" coefficient associated with the spatial connectedness is adaptively estimated as a function of brightness gradient. The gradient function involves a constant called "temperature," which determines the amount of discontinuity and is continuously decreased in the iterations.

In this study, the proposed method has been extensively evaluated using simulated images that were generated from various patterns. These patterns represent the types of natural and artificial land-use. The simulated images are classified by a multistage classification including the modified anisotropic diffusion technique. This study makes experiments on the satellite images remotely sensed on the Korean peninsula..

II. MULTISTAGE HIERARCHICAL CLUSTERING CLASSIFICATION

One essential structural characteristic involves hierarchy of scene information. Under the constraint of the hierarchical structure, it is then possible to determine natural image segments by combining hierarchical clustering with spatial region growing. Hierarchical clustering is an approach for step-by-step merging of small clusters into larger ones. Clustering algorithm utilize a (di)similarity measure that is computed between all pairs of candidates being considered for merging, a rule for selecting the pairs to be merged, and a rule for "cutting" the hierarchical tree.

The computational efficiency of hierarchical clustering segmentation is mainly dependent on how to find the best pair to be merged. The closest neighbor of region j is defined as

$$\mathrm{CN}(j) = \arg\min_{k \in \mathbf{R}_j} d(j,k) \qquad (1)$$

where $d(j,k)$ is the dissimilarity measure between regions j and k, and $\mathbf{R}_j$ is the index set of regions considered to be merged with region j. The pair of regions is then defined as MCN iff $k = \mathrm{CN}(j)$ and $j = \mathrm{CN}(k)$. It is easily shown that the best pair is one of the MCN's. Thus, the search is limited in the set of MCN's in the hierarchical clustering procedure.

For the region growing segmentation, the clustering procedure successively merges a pair among them of two regions which neighbor in image space, that is, $\mathbf{R}_j$ is the collection of the regions which are spatially adjacent with region. If all the pixels in the sample image are initially considered to be individual clusters, the algorithm requires exorbitant memory for the values associated with the merging process for large multi-channel imagery. According to the merger, the neighbor configuration and the set of MCN's must be updated, and the computational time for the update of the configurations and the search of the best pairs exponentially increases as the number of clusters in the initial state increases. To alleviate the memory problem and improve the computational performance of the algorithm, a multi-window strategy of boundary blocking operation has been used by constructing a pyramid-like hierarchy system.

The multistage algorithm consists of two stages, as shown in Fig.1. The regional segmentation can be considered as a relaxation stage to reduce the obscurity in the image pattern, whereas the global segmentation is a classification stage in which the image is grouped into a number of physically meaningful regions.

This study used a disimilarity coefficient based on the Mahalanobis distance. Under the assumption that all the regions have an identical covariance matrix, the coefficient for for two regions, r and s, is defined as

$$\lambda(r,s) = M_{r+s} - (M_r + M_s) \tag{2}$$

$$M_j = \sum_{k \in G_j} (\mathbf{x}_k - \bar{\mathbf{x}}_j)' \hat{\Sigma}_A^{-1} (\mathbf{x}_k - \bar{\mathbf{x}}_j)$$

$$\bar{\mathbf{x}}_j = \frac{\sum_{k \in G_j} \mathbf{x}_k}{n_j} \quad \text{for } j = r, s, r+s$$

where $\mathbf{x}_k$ is the observation vector of the kth pixel, $\hat{\Sigma}_A$ is an estimation of the common covariance matrix, n_j and G_j are the number of pixels and the index set of the jth region respectively.

For the intensity process of an additive Gaussian field, if all the regions have an identical covariance matrix, the decision rule to determine the level corresponding to the optimal state in the hierarchy is based on the schwarz information criterion(SIC)[4] such that the clustering stops when, for all possible candidate cluster-pairs to be merged,

$$\lambda(r,s) \geq \frac{K(1)\log n}{2}. \tag{3}$$

A. Statistical rules including the SIC, however, usually fail to find a parsimonious model that is typically desired in most applications. The rule of (3) was applied for the regional segmentor to terminate merging in this study.

No clue to the optimal number of classes is generally provided in unsupervised analyses. If the estimated number of classes is too small, regions with different characteristics will not be properly partitioned in the classified image. A simple heuristic rule was used an alternative for the global segmentation such that the optimal number of classes is determined at the hierarchical level where the value of the similarity coefficient changes quite markedly in consecutive iterations of the clustering procedure.

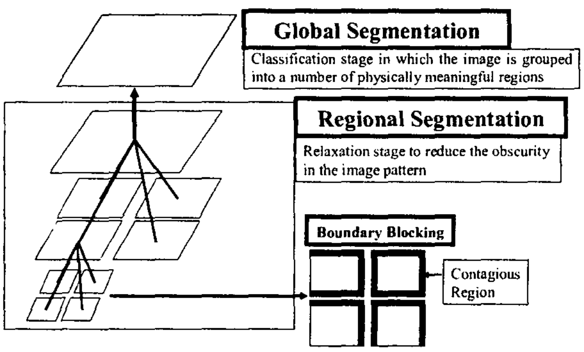


Fig. 1. Multistage hierarchical clustering segmentation.

III. MODIFIED ANISOTROPIC DIFFUSION (MAID)

For a probability structure of the mean process μ, a class of Gibbs measures is specified with the energy function $E_p(\mu)$ associated with pair potentials:

$$E_p(\mu) = \mu' \mathbf{A} \mu$$

where $\mathbf{A}=\{A_{ij}\}$ is the matrix related to the bonding strength coefficients such that

$$A_{ij} = \begin{cases} \sum_{k \in R_i} \alpha_{ik}, & \text{if } j = i \\ -\alpha_{ij}, & \text{if } j \in R_i \\ 0 & \text{otherwise} \end{cases}$$

where R_i is the index set of neighbors of the ith pixel and $\alpha_{jk} = \alpha_{kj}$ is the nonnegative coefficient which represents "bonding strength" between the jth and kth pixels.

Applying an iterative approach similar to the point-Jacobian iteration method, the mean intensity image is iteratively obtained based on *maximum a posteriori* criterion: at the hth iteration,

$$\hat{x}_i^h = D_{ii}^{-1}\left(\sigma_i^{-2} - \sum_{j \in R_i} S_{ij} \hat{x}_j^{h-1}\right)$$

where

$$D_{ii} = \sigma_i^{-2} + A_{ii}$$
$$S_{ii} = 0 \text{ and } S_{ij} = A_{ij} \text{ for } i \neq j.$$

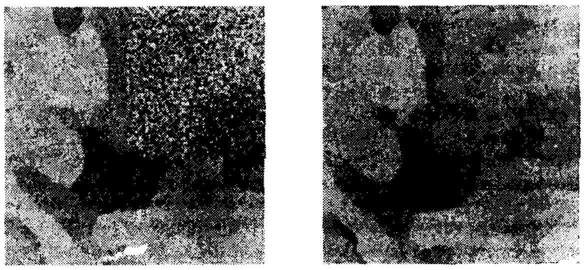
This iteration converges to a unique solution since

$$D_{ii}^{-1} \sum_{j \in R_i} S_{ij} < 1, \forall i.$$

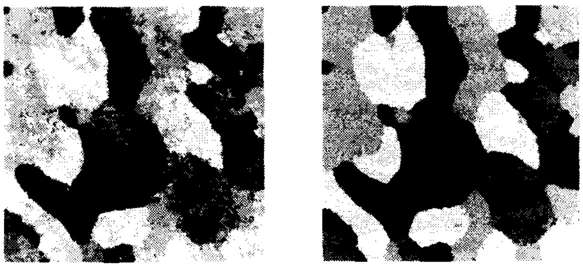
Unfortunately the bonding strength coefficients are unknown in most practice. This study employs an approach of anisotropic diffusion[5] which adaptively chooses the coefficients at each iteration. The coefficients are updated at every iteration as a function of the brightness gradient:

$$\alpha_{ij}^h = g\left(\left|\nabla_{ij} X_{ij}^h\right|\right) = g\left(\left|x_i^h - x_j^h\right|\right).$$

III. EXPERIMENTS



Simulated Image MAID-recovered Image

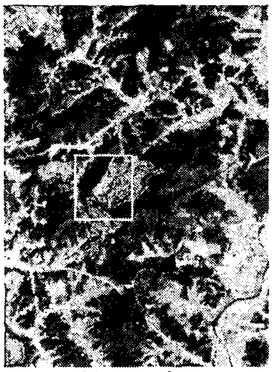


No Diffusion(Error: 25.9%) MAID(Error: 3.87%)

Fig. 2. Classification results of 3 band simulated images.

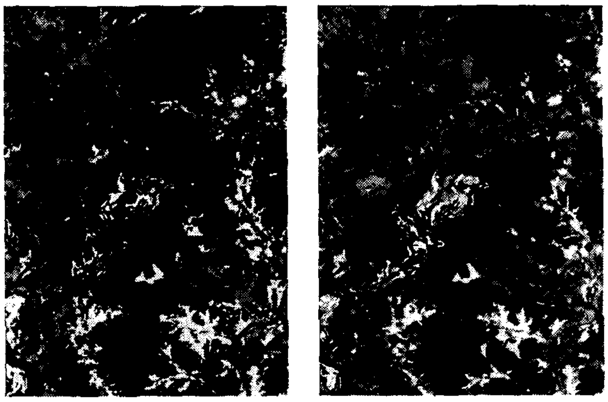
References

[1] S. Tanimoto and A. Klinger, *Structured Computer Vision*, Academic, New York, 1980.
[2] Anderberg, M.R, *Cluster Analysis for Application*, Academic Press, NY, 1973.
[3] S. Lee and M. M. Crawford, "Unsupervised Multistage Image Classification Using Hierarchical Clustering with a Bayesian Similarity Measure", *IEEE Trans. image



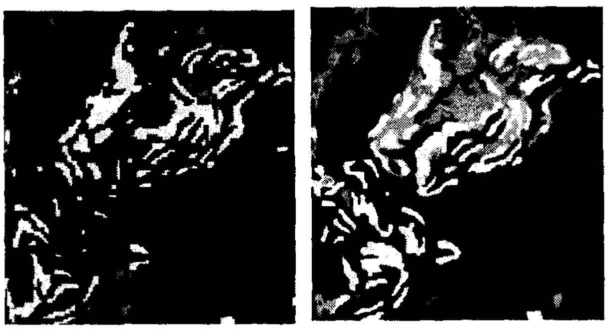
Observation

Fig. 2. Classification results of LANDSAT ETM+ (3 band: green, red, NIR).



No Diffusion MAID

processing, accepted(April, 2004).
[4] G. Schwarz, "Estimation of the Dimension of a Model," *Annal., Math., Statist.*, vol. 6, pp. 461-464, 1978.
[5] P.Perona and J.Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 629-639, 1990.



No Diffusion MAID

Fig. 3. Classification results of the rectangular area in the observation image of Fig. 2.