Hierarchical Clustering Approach for Unsupervised Image Classification of Hyperspectral Data

Sanghoon Lee[1] and Melba M. Crawford[2]

[1] Dept. of Industrial Engineering, Kyungwon University
San 65 Bokjeong-dong, Sujeong-gu, Seongnam-si, Kyunggi-do 461-701, Korea
Tel.: +8231-750-5367 Email: shl@ kyungwon.ac.kr
[2] Center for Space Research, University of Texas at Austin, Austin, TX, USA.
Tel.: 512- 471- 5573 Email: crawford@csr.utexas.edu

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 of CN-chain 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

Most approaches to the image classification require *a priori* class-dependent knowledge of parameterized models for the data. In many instances, however, the parameter values of the models are not known *a priori*, and the process of gathering training samples to estimate parameters is often infeasible or very expensive. In addition, the classification results much depend on the number of classes selected in the specific analyzed area, but it is very complicate to determine the class number, as known as "cluster validation," particularly for remotely sensed data. Therefore, it is necessary that classification procedures perform the unsupervised learning of the parameters including the number of classes and the image classification simultaneously. For the unsupervised analysis, agglomerative hierarchical clustering [1] technique is one of the most appropriate approaches.

Due to advances in sensor technology, it is now possible to acquire hyperspectral data simultaneously in hundreds of bands. The hyperspectral imagery possesses much richer spectral information, but one of challenging problems in processing the high-dimensional data is the computational complexity resulting from processing the vast amount of data volume. Especially, the unsupervised classification that makes use of hierarchical clustering requires enormous processing time for the hyperspectral image. This study proposed a multistage approach that is very computationally efficient for the unsupervised classification of hyperspectral data. The proposed algorithm includes two stages of segmentation. In the first stage, the "local" segmentor 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 stage employs the hierarchical clustering procedures that merge "mutual closest neighbor (MCN)" pair satisfying a given clustering criterion by using "closest neighbor chain (CN-chain)." The "global" segmentor of the second stage, which has not spatial constraints for merging, clusters the segments resulting from the local segmentor by an agglomerative hierarchical clustering scheme. The segmented regions are classified into a small number of distinct states by a sequential merging operation. The local 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.

The analysis of hyperspectral images, which are collected by as many as 224 spectral bands from the airborne visible/infrared imaging spectrometer (AVIRIS), may require serious computational complication even for a quite efficient method. This problem can be generally alleviated by band reduction, band selection and data compression. In this study, the typical schemes of these three approaches, PCA[2], elimination of strongly correlated bands[3], and discrete wavelet transform[4], were applied for the multistage hierarchical clustering method. The classification results were compared with them using full bands.

II. CN-CHAIN SPATIAL CLUSTERING

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.

The computational efficiency of hierarchical clustering segmentation is mainly dependent on how to find the best pair to be merged. Let $I_n = \{1,2,\cdots,n\}$ be an index set of pixels of a sample image, $J_M = \{1,2,\cdots,M\}$ be an index set of regions associated with $\mathbf{G}_J = \{G_j \subseteq I_n \mid j \in J_M\}$ that is a partition of I_n, $\mathbf{R}_J = \{R_j \subseteq I_n \mid j \in J_M\}$ be a region neighborhood system such that R_j is the index set of neighborhood regions of region *j*. The closest neighbor of region *j* is defined as

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

where $d(j,k)$ is the dissimilarity measure between regions *j* and *k*, and 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 $j = \mathrm{CN}(k)$. if a cutting rule that

$$\mathrm{CR}(j,k) < \mathrm{CR}_{\max} \qquad (2)$$

is given as a merging condition of two regions, for any arbitrary region r_0 satisfying $\mathrm{CR}(r_0, r_1) < \mathrm{CR}_{\max}$, a CN-chain is established as the sequence of regions

$$r_0,\ r_1 = \mathrm{CN}(r_0),\ r_2 = \mathrm{CN}(r_1), \cdots, r_{h-1} = \mathrm{MCN}(r_h) \qquad (3)$$

such that the last two region constitutes an MCN pair, i.e.

$$r_{h-1} = \mathrm{CN}(r_h) \text{ and } r_h = \mathrm{CN}(r_{h-1}).$$

The sequence of (3) also satisfies

$$\mathrm{CR}(r_{k-1}, r_k) < \mathrm{CR}_{\max} \text{ for } k = 1,2,\cdots,h.$$

The CN-chain algorithm is outlined in the following:

1. Initialize that every pixel is defined as a region, i.e., $J_M \leftarrow I_n$.
2. Construct a CN-chain by starting region r_0 with the lowest index among J_M satisfying

   $$\mathrm{CR}(r_0, \mathrm{CN}(r_0)) < \mathrm{CR}_{\max}.$$

   If there exists no avaiable region, **STOP**.
3. Merge the last MCN pair, r_{h-1} and r_h by indexing the new region with the lower one of both region indices and update the partition by eliminating the region with the higher index, i.e.

   $$\begin{array}{l} r_{new} \leftarrow \min\{r_h, \mathrm{MCN}(r_h)\} \\ J_M \leftarrow J_M \setminus q \text{ where } q = \max\{r_h, \mathrm{MCN}(r_h)\}' \end{array}$$

   and update the regional parameters of the new region r_{new}.
4. For $h = 1$,
   if $\mathrm{CR}(r_0, \mathrm{CN}(r_0)) < \mathrm{CR}_{\max}$, reconstruct CN-chain from the starting region r_0.
   else **GOTO** Step 2.
   For $h > 1$,
   if $\mathrm{CR}(r_{h-2}, \mathrm{CN}(r_{h-2})) < \mathrm{CR}_{\max}$, reconstruct CN-chain from the starting region r_{h-2}.
   else **GOTO** Step 2.

5. **GOTO** Step 3.

This algorithm provides the unique optimal solution for spatial constraints of merging with a given cutting rule if the dissimilarity measure satisfies the following condition, referred to as "regional reducibility" :

$$d(p \cup q, k) > d(p, q),\ k \in \left(R_p \cup R_q\right) \tag{4}$$

where p = MCN(q).

One common objective in image segmentation involves minimizing the over-all intra-cluster sample variance. This results in the maximum likelihood solution for the case of unknown parameters in a Gaussian image field. Other statistical measures can also be employed to obtain well-posed solutions, but at a higher computational cost[5]. The advantages of intra-cluster sample variance are both simplicity and its ability to represent a basic important characteristic of clusters. In the local segmentation of spatial region growing, this study assumed a simple variance structure with no correlation between bands for the hyperspectral data processes. Under this assumption, the dissimilarity measure based on the intra-cluster sample variance is defined:

$$d(r,s) = \frac{n_r n_s \sum_{k=1}^{b} (\hat{\mu}_{rk} - \hat{\mu}_{sk})^2}{n_r + n_s} \tag{5}$$

where b is the number of bands, n_j and $\hat{\mu}_{jk}$ are the number of pixels and the kth band's average of G_j respectively. The reducibility of (5) is easily verified using the Lance-William's formula[6].

It is difficult to define an appropriate measure of homogeneity to establish rules for cutting the hierarchical tree. The model fitting approaches using information criteria which measure the trade-off between the likelihood and the penalty for increasing the model's order have been applied for cluster validation in image analysis[7]. One of these approaches involves selecting the optimal state that maximizes Schwarz's information criterion [8]:

$$\text{SIC} = -2\log L(h) + K(h)\log n \tag{6}$$

where h is the number of distinct states, $L(h)$ and $K(h)$ are the maximum value of the likelihood function and the number of independent parameters estimated when using h states respectively. The SIC statistic was derived assuming a nonzero prior based on an asymptotic approximation to Bayes' loss. Although the Schwarz's approximation may fail for small samples because of its asymptotic nature, the use of SIC is generally appropriate for selecting the number of classes in image analysis. According to the SIC, merging of two regions can be considered if the decrease in the log likelihood is less than (0.5log$n$) multiplied by the change in the number of parameters associated with a merged class.

For the intensity process of an additive Gaussian field with diagonal covariance matrix, the cutting rule is designed on the basis of the SIC :

$$\begin{aligned} \text{CR}(\text{r},\text{s}) &= n_{r\cup s}\sum_{k=1}^{b}\ln\hat{\sigma}^2_{r\cup s,k} - \left(n_r\sum_{k=1}^{b}\ln\hat{\sigma}^2_{r,k} + n_s\sum_{k=1}^{b}\ln\hat{\sigma}^2_{s,k}\right) \\ \text{CR}_{\max} &= \frac{K(1)\log n}{2} \end{aligned} \tag{7}$$

where $\hat{\sigma}^2_{j,k}$ is the kth band's sample variance of G_j and $K(1) = 2b$ when using (6) and (7).

III. Hierarchical Clustering Classification

Hierarchical clustering is an approach for step-by-step merging of small clusters into larger ones. In the classification, the image partition resulting from the local segmentor are classified into a small number of distinct states by sequentially merging two regions at each iteration . This study used a dissimilarity coefficient based on the Mahalanobis distance. The coefficient for two regions, r and s, is defined as

$$\lambda(r,s) = M_{r\cup s} - (M_r + M_s) \tag{8}$$

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

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

$$\hat{\boldsymbol{\Sigma}}_j = \frac{\sum_{k\in G_j}(\mathbf{x}_k - \bar{\mathbf{x}}_j)(\mathbf{x}_k - \bar{\mathbf{x}}_j)'}{n_j}$$

A. where $\mathbf{x}_k$ is the observation vector of the kth pixel.

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. However, a relatively homogeneous region can be separated into a number of smaller nonmeaningful regions, if the estimated number is too large. In most studies of cluster validation, the validation criteria are based on statistical rules or model fitting approaches. However, these conventional approaches usually fail to find a parsimonious model, which is typically desirable, especially for large images with a small number of classes. 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.

IV. EXPERIMENTS

In this study, the typical schemes of dimensionality reduction for hyperspectral data, PCA, elimination of strongly correlated bands, and discrete wavelet transform, were applied for the multistage hierarchical clustering method. The classification results were compared with them using full bands.

References

[1] Anderberg, M.R, *Cluster Analysis for Application*, Academic Press, NY, 1973.

[2] C.-I Chang and Q. Du, "Interference and noise-adjusted principal components analysis," *IEEE Trans. Geosci Remote Sensing*, vol. 37, pp. 2387-2396, 1999.

[3] J. A. Richards, *Remote Sensing Digital Image Analysis*, 2nd ed. New York: Springer-Verlag, 1993.

[4] S. Kaewpijit, J. Le Moigne, and T. El-Ghazawi, "Automatic reduction of hyperspectral imagery using Wavelet spectral analysis," *IEEE Trans. Geosci. Remote Sensing.*, Vol. 41, pp.863-871, 2003.

[5] S. Lee and M. M. Crawford, "Unsupervised Multistage Image Classification Using Hierarchical Clustering with a Bayesian Similarity Measure", *IEEE Trans. image processing*, accepted(April, 2004).

[6] G. N. lance and W. T. Williams, "A general theory of classificatory sorting strategies. I. Hierarchical Systems," *The Computer Journal*, vol. 9, pp.373-380, 1967.

[7] C. S. Won and H. Derin, "Unsupervised segmentation of noisy and textured images using Markov random fields," *Comp. Vision, Graphics, Image Processing*, vol. 54, pp. 308-328, 1992.

[8] G. Schwarz, "Estimation of the Dimension of a Model," *Annal., Math., Statist.*, vol. 6, pp. 461-464, 1978.