N O T I C E

THIS DOCUMENT HAS BEEN REPRODUCED FROM MICROFICHE. ALTHOUGH IT IS RECOGNIZED THAT CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED IN THE INTEREST OF MAKING AVAILABLE AS MUCH INFORMATION AS POSSIBLE

(E80-10273) SOME APPROACHES TO OPTIMAL CLUSTER LABELING OF AEROSPACE IMAGERY (Lockheed Engineering and Management) 50 p HC A04/MF A01 CSCL 20C

N80-23520

Unclas
G3/43 00273

TECHNICAL REPORT

SOME APPROACHES TO OPTIMAL CLUSTER LABELING OF AEROSPACE IMAGERY

Job Order 73-302

This report describes Classification activities of the Supporting Research project of the AgRISTARS program.

PREPARED BY

C. B. Chittineni

APPROVED BY

T. C. Minter

T. C. Minter, Supervisor
Techniques Development Section

J. E. Wainwright

J. E. Wainwright, Manager
Development and Evaluation Department

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.
Under Contract NAS 9-15800

For

Earth Observations Division
Space and Life Sciences Directorate
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

May 1980

LEMSCO-14597

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
JSC-16355; SR-LO-00440		

4. Title and Subtitle
Some Approaches to Optimal Cluster Labeling of Aerospace Imagery

5. Report Date
May 1980

6. Performing Organization Code

7. Author(s)
C. B. Chittineni
Lockheed Engineering and Management Services Company, Inc.

8. Performing Organization Report No.
LEMSCO-14597

9. Performing Organization Name and Address
Lockheed Engineering and Management Services Company, Inc.
1830 NASA Road 1
Houston, Texas 77058

10. Work Unit No.

11. Contract or Grant No.
NAS 9-15800

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Lyndon B. Johnson Space Center
Houston, Texas 77058 Technical Monitor: J. D. Erickson

13. Type of Report and Period Covered
Technical Report

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This paper presents some approaches for the problem of labeling clusters using information from a given set of labeled and unlabeled patterns. Assigning class labels to the clusters is formulated as that of finding the best label assignment over all possible label assignments with respect to a criterion. Labeling clusters is also viewed as that of obtaining probabilities of class labels to the clusters with the maximization of likelihood function and probability of correct labeling as criteria. Closed form solutions are obtained for the probabilities of class labels to the clusters by maximizing a lower bound on the likelihood criterion. Expressions are derived for the asymptotic mean and variance of the resulting class proportion estimates. The problem of obtaining class labels to the clusters is further formulated as that of minimizing the variance of the proportion estimates of the classes that use both the given labeled and unlabeled patterns. Imperfections in the labels of the given labeled set are incorporated into the criteria. Probability of error is proposed as a criterion for grouping the modes into their natural classes using unlabeled patterns. Furthermore, the results of application of these techniques in the processing of remotely sensed multispectral scanner imagery data are presented.

17. Key Words (Suggested by Author(s))
Aerospace imagery, Asymptotic variance, Cluster labeling, Label imperfections, Maximum likelihood, Mode, Probabilities of class labels, Probability of correct labeling, Probability of error, Remote sensing, Unlabeled patterns, variance of proportion estimate

18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	52	

*For sale by the National Technical Information Service, Springfield, Virginia 22161

PREFACE

The techniques which are the subject of this report were developed to support the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing program. Under Contract NAS 9-15800, Dr. C. B. Chittineni, a principal scientist for Lockheed Engineering and Management Services Company, Inc., performed this research for the Earth Observations Division, Space and Life Science Directorate, National Aeronautics and Space Administration, at the Lyndon B. Johnson Space Center, Houston, Texas.

CONTENTS

TABLES

1. INTRODUCTION

Recently, there has been considerable interest in the development of systems for the classification of remotely sensed imagery data for inventorying natural resources, monitoring crop conditions, etc. Usually, the inherent classes in the data are multimodal, and nonsupervised classification or clustering techniques (refs. 1, 2) have been found to be effective (ref. 3). These usually break up the image into its inherent modes or clusters. One of the crucial problems in the application of clustering techniques for the classification of imagery data is to label the clusters.

There is considerable interest in the statistical literature in labeling the clusters (ref. 4). This problem is also common in labeling the regions, obtained by using segmentation algorithms, in the development of scene understanding systems. In the recent literature, relaxation labeling algorithms (refs. 5, 6, and 7) have been proposed for labeling the segmented regions, but these use relational or spatial properties of the regions through compatability coefficients. However, in cluster labeling, the relational properties of the clusters are either nonavailable or not meaningful. For example, in aerospace imagery, the regions of interest are crops, nonagricultural areas, etc. and can be anywhere in the image; hence, it is difficult to define relational properties.

It is the purpose of this paper to address the problem of labeling the clusters using the information from a given set of labeled patterns. It is assumed that the probability density functions and a priori probabilities of the clusters or modes are given. Let these respectively be $p(\Omega = i|X)$, δ_i, $i = 1, 2, \cdots, m$ where m is the number of modes or clusters. It is also assumed that a set of labeled patterns, $X_i(j)$ with labels $\omega_i(j) = i$; $j = 1, 2, \cdots, N_i$; $i = 1, 2, \cdots, C$ are given, where C is the number of classes. In remote sensing, the labels for these patterns are provided by analyst interpreter (AI) by examining imagery films and using some other information such as historic information, crop calender models, etc. Very often, the AI labels are imperfect. It is relatively expensive to acquire labels, and a large number of unlabeled patterns is usually available. Some approaches that use all the given information are presented in this paper for optimum cluster labeling.

The paper is organized as follows. In section 2, the problem of obtaining optimum class labels to the modes is formulated as the one that maximizes likelihood criterion by exhaustive search over all possible label assignments. Section 3 considers the problem of obtaining probabilities of class labels to the clusters using maximum likelihood criterion. A closed-form solution that maximizes a lowerbound on the criterion is presented in section 3. Also, expressions for the asymptotic mean and variance of resulting proportions are presented. In section 4, probability of correct labeling is used as a criterion for obtaining probabilities of class labels for the modes. In section 5, variance of the class proportion estimates is proposed as a criterion that uses both the given labeled and unlabeled pattern sets for obtaining the probabilities of class labels to the modes. Imperfections in the labels of the given, labeled set are considered in section 6. Section 7 contains the experimental results in the processing of remotely sensed imagery data and a concluding summary is given in section 8. The problem of grouping modes into their natural classes using unlabeled patterns is considered in appendix A. Appendix B considers a two-class problem in which the labeled patterns from a single class and a set of unlabeled patterns are given. Fixed point iteration schemes for the probability of correct labeling criterion are presented in appendix C. Appendix D addresses the problem of proportion estimation with impure clusters.

2. LABEL ASSIGNMENT TO CLUSTERS BY EXHAUSTIVE SEARCH

In general, the class-conditional density functions are multimodal. Let C_i be the number of modes of class i, where $\sum_{i=1}^{C} C_i = m$. By defining a criterion, the class label assignment to the modes that maximizes the criterion can be chosen as the optimal assignment. Let $p_i(X)$ be the density function of class i, $p_{ij}(X)$ be the density function of mode j of class i, q_{ij} be the a priori probability of mode j of class i, q_i be the a priori probability of class i, and p(X) be the mixture density function. Then we have the following relationships.

$$\left.\begin{aligned}
\sum_{i=1}^{C} q_i &= 1 \\
\sum_{j=1}^{C_i} q_{ij} &= 1 \\
p_i(X) &= \sum_{j=1}^{C_i} q_{ij} p_{ij}(X) \\
p(X) &= \sum_{i=1}^{C} q_i p_i(X)
\end{aligned}\right\} \tag{2-1}$$

Choosing the likelihood function as a criterion, the likelihood of occurrence of given patterns with their corresponding labels can be expressed by the quantity L', where

$$\begin{aligned}
L' &= \prod_{i=1}^{C} \left\{ \prod_{j=1}^{N_i} p[X_i(j), \omega_i(j) = i] \right\} \\
&= \prod_{i=1}^{C} \left\{ \prod_{j=1}^{N_i} p[X_i(j) | \omega_i(j) = i] P[\omega_i(j) = i] \right\} \\
&= \left(\prod_{i=1}^{C} \left\{ \prod_{j=1}^{N_i} \sum_{\ell=1}^{C_i} q_{i\ell} p_{i\ell}[X_i(j)] \right\} \right) \left(\prod_{i=1}^{C} q_i \right) .
\end{aligned} \tag{2-2}$$

Since logarithm is a monotonic function of its argument, taking logarithm of equation (2-2) results in

$$L = \log(L') = \sum_{i=1}^{C} \left(\sum_{j=1}^{N_i} \log \left\{ \sum_{\ell=1}^{C_i} q_{i\ell} \, p_{i\ell}[X_i(j)] \right\} \right) + \sum_{i=1}^{C} \log(q_i) \tag{2-3}$$

The a priori probability of mode j of class i, q_{ij}, may be estimated as follows. For a particular labeling assignment, let the modes 1, 2, ···, C_i belong to class i. Then

$$q_{ij} = \frac{\delta_j}{\sum_{r=1}^{C_i} \delta_r} \tag{2-4}$$

If the clustering algorithms generate a relatively fewer number of clusters (in remote sensing, typically around 12), optimal class label assignment for the clusters can easily be obtained by exhaustive search. By giving all possible class label assignments to clusters and computing the value of the criterion for each assignment, the optimal class label assignment can be chosen as the one that extremizes the criterion. If the density functions of the modes are Gaussion, the criterion takes a relatively simple form (for example, if a clustering algorithm based on the maximum likelihood equations (ref. 2) is used to fit the Gaussion density functions for the modes).

2.1 CASE IN WHICH THE NUMBER OF MODES IS EQUAL TO THE NUMBER OF CLASSES AND THE MODE-CONDITIONAL DENSITIES ARE GAUSSIAN

Consider a simple case in which the number of classes is equal to the number of modes and the mode-conditional densities are Gaussian. That is

$$p(X|\Omega = \ell) \sim N(M_\ell, \Sigma_\ell) \tag{2-5}$$

The problem is to assign class labels to the modes such that the criterion L of equation (2-3) is maximized. For a particular assignment of labels, let the mode ℓ be given a class label i. In a unimodal case, then

$$p(X|\omega = i) = p(X|\Omega = \ell) \tag{2-6}$$

Let

$$p(X|\omega = i) \sim N\left(M_i^d, \Sigma_i^d\right)$$

where

$$M_i^d = M_\ell,\ \Sigma_i^d = \Sigma_\ell,\ \text{and } q_i = \delta_\ell = \delta_i^d$$

Then, the log likelihood criterion becomes

$$L = \frac{-n}{2} C \log(2\pi) - \frac{1}{2} \sum_{i=1}^{C} \log(|\Sigma_i^d|) + \sum_{i=1}^{C} \log\left(\delta_i^d\right) - \frac{1}{2} L_1^d \tag{2-7}$$

where

$$L_1^d = \sum_{i=1}^{C} \left[\sum_{j=1}^{N_i} \left(X_i(j) - M_i^d\right)^T \Sigma_i^{d^{-1}} \left(X_i(j) - M_i^d\right) \right]$$

The maximization of L is equivalent to minimization of L_1^d. That is, when the number of modes is equal to the number of classes, the assignment of class labels to the modes by the maximization of the criterion L is based on the smallest Mahalanobis distance between the given labeled patterns of each class and the mode-conditional densities.

2.2 GENERAL CASE IN WHICH THE NUMBER OF MODES IS GREATER THAN THE NUMBER OF CLASSES

In this case, the criterion L becomes

$$L = \sum_{i=1}^{C} \left(\sum_{j=1}^{N_i} \log \left\{ \sum_{\ell=1}^{C_i} q_{i\ell}^d p_{i\ell}[X_i(j)] \right\} \right) + \sum_{i=1}^{C} \log\left(q_i^d\right) \tag{2-8}$$

where d refers to a particular assignment of class labels. The probabilities $q^d_{i\ell}$ and q^d_i are computed according to equation (2-4), for this label assignment. Equation (2-8) can be used as a criterion. However, for Gaussian densities, a simple criterion can be obtained using the fact that the logarithm is a convex upward function to derive a lower bound on L. Since logarithm is a convex upward function, we have the inequality

$$\log\left[\sum_{i=1}^{C} a_i g_i(X)\right] \geq \sum_{i=1}^{C} a_i \log[g_i(X)] \tag{2-9}$$

where

$$\left.\begin{aligned} &\sum_{i=1}^{C} a_i = 1 \\ &a_i \geq 0 \quad ; \quad i = 1,2,\cdots,C \end{aligned}\right\} \tag{2-10}$$

and

Let the density function of the ℓ^{th} mode of class i be

$$p_{i\ell}(X) \sim N\left(M^d_{i\ell}, \Sigma^d_{i\ell}\right) \tag{2-11}$$

using equations (2-9) to (2-11) in equation (2-8) yields

$$L \geq -C \frac{n}{2} \log(2\pi) - \frac{1}{2} L_2 \tag{2-12}$$

where

$$L_2 = \sum_{i=1}^{C} \sum_{\ell=1}^{C_i} q^d_{i\ell} \log\left(\left| \Sigma^d_{i\ell} \right|\right) + \left(\sum_{i=1}^{C} \sum_{\ell=1}^{C_i} q^d_{i\ell} \left\{ \sum_{j=1}^{N_i} \left(X_i(j) - M^d_{i\ell}\right)^T \Sigma^{d^{-1}}_{i\ell} \left[X_i(j) - M^d_{i\ell}\right]\right\}\right) - 2 \sum_{i=1}^{C} \log\left(q^d_i\right) \tag{2-13}$$

Thus, the optimal class label assignment can be chosen as the one that minimizes L_2. Combinatorial algorithms (ref. 8) can be used to efficiently generate all possible class label assignments for exhaustive search.

3. PROBABILISTIC CLUSTER LABELING BASED ON MAXIMUM LIKELIHOOD CRITERION

The last section addressed the problem of obtaining class labels to the clusters by exhaustive search. This section considers the problem of obtaining a probabilistic description for the class labels of the clusters. The criterion used is the likelihood function, but normalized as shown in equation (3-14).

$$L' = \frac{\prod_{i=1}^{C} \left\{ \prod_{j=1}^{N_i} p[X_i(j),\ \omega_i(j) = i] \right\}}{\prod_{i=1}^{C} \left\{ \prod_{j=1}^{N_i} p[X_i(j)] \right\}}$$

$$L' = \prod_{i=1}^{C} \left\{ \prod_{j=1}^{N_i} \frac{p[X_i(j),\ \omega_i(j) = i]}{p[X_i(j)]} \right\} \tag{3-1}$$

The mixture density p(X) can be written in terms of class-conditional densities as

$$p(X) = \sum_{i=1}^{C} P(\omega = i) p(X|\omega = i) \tag{3-2}$$

The mixture density p(X) can also be written in terms of mode-conditional densities as

$$\begin{aligned} p(X) &= \sum_{\ell=1}^{m} P(\Omega = \ell) p(X|\Omega = \ell) \\ &= \sum_{\ell=1}^{m} \sum_{i=1}^{C} P(\Omega = \ell, \omega = i) p(X|\Omega = \ell) \\ &= \sum_{i=1}^{C} P(\omega = i) \left[\sum_{\ell=1}^{m} P(\Omega = \ell|\omega = i) p(X|\Omega = \ell) \right] \end{aligned} \tag{3-3}$$

On comparing equations (3-2) and (3-3), the following assumption is made.

$$p(X|\omega = i) = \sum_{\ell=1}^{m} P(\Omega = \ell|\omega = i)p(X|\Omega = \ell) \tag{3-4}$$

Equation (3-4) can be rewritten as

$$p(\omega = i|X) = \sum_{\ell=1}^{m} P(\omega = i|\Omega = \ell)p(\Omega = \ell|X) \tag{3-5}$$

Since logarithm is a monotonic function of its argument, taking logarithm of L' of equation (3-1) and using equation (3-5) yields the following:

$$L = \log(L') = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \log\left\{\sum_{\ell=1}^{m} \alpha_{\ell i} p[\Omega = \ell|X_i(j)]\right\} \tag{3-6}$$

where $\alpha_{\ell i} = P(\omega = i|\Omega = \ell)$ is the probability that the label of model ℓ is class i. The probabilities $\alpha_{\ell i}$ satisfy the constraints given in equation (3-7).

$$\left.\begin{aligned} &\alpha_{\ell i} \geq 0 \quad ; \quad i = 1,2,\cdots,C \text{ and } \ell = 1,2,\cdots,m \\ &\sum_{i=1}^{C} \alpha_{\ell i} = 1 \quad ; \quad \ell = 1,2,\cdots,m \end{aligned}\right\} \tag{3-7}$$

Closed form solutions for $\alpha_{\ell i}$, by minimizing L of equation (3-6), subject to the constraints of equation (3-7), seem to be difficult. The probabilities $\alpha_{\ell i}$ can easily be obtained using optimization techniques such as Davidon-Fletcher-Powell (refs. 9, 10, 11).

3.1 FIXED-POINT ITERATION SCHEME FOR OPTIMAL $\alpha_{\ell i}$

The following fixed-point iteration equation [similar to maximum likelihood equations in parametric clustering (ref. 2)] for the solution of the above optimization problem can easily be obtained by introducing lagrangian multipliers.

$$\alpha_{\ell i} = \frac{\sum_{j=1}^{N_i} d_{\ell ij}}{\sum_{i=1}^{C} \sum_{j=1}^{N_i} d_{\ell ij}} \tag{3-8}$$

where
$$d_{\ell ij} = \frac{\alpha_{\ell i} p[\Omega = \ell | X_i(j)]}{\sum_{\ell=1}^{m} \alpha_{\ell i} p[\Omega = \ell | X_i(j)]} \tag{3-9}$$

Lennington (ref. 12) derived fixed-point iteration equations similar to equation (3-8) and Heydorn (ref. 13) developed a least squares solution for he probabilities $\alpha_{\ell i}$. However, closed form solutions for $\alpha_{\ell i}$ can be obtained with the criterion as the maximization of a lower bound on L. Using the inequality (2-9) in equation (3-6), a lower bound on the log likelihood function L can be obtained as

$$L \geq \sum_{i=1}^{C} \sum_{j=1}^{N_i} \sum_{\ell=1}^{m} \left\{ p[\Omega = \ell | X_i(j)] \log(\alpha_{\ell i}) \right\} \tag{3-10}$$

Introducing the lagrangian multipliers, the probabilities $\alpha_{\ell i}$ that maximize the lower bound of equation (3-10), subject to the constraints of equation (3-7), can easily be shown to be

$$\alpha_{\ell i} = \frac{N_i e_{i\ell}}{\sum_{r=1}^{C} N_r e_{r\ell}} \tag{3-11}$$

where
$$e_{i\ell} = \frac{1}{N_i} \sum_{j=1}^{N_i} p[\Omega = \ell | X_i(j)] \tag{3-12}$$

This solution simply states that the probability of the i^{th} class label for a given cluster ℓ is the ratio of the a posteriori probability of cluster ℓ given the labeled patterns from class i to the sum over all classes of the a posteriori probability of cluster ℓ given the labeled patterns from each class. Having obtained $\alpha_{\ell i}$; q_i, the proportion of class i, can be estimated as follows. Consider

$$q_i = \sum_{\ell=1}^{m} P(\omega = i, \Omega = \ell) = \sum_{\ell=1}^{m} \delta_\ell \alpha_{\ell i}$$

Hence, $\hat{q}_i$, the estimate of q_i, can be computed from the following.

$$\hat{q}_i = \sum_{\ell=1}^{m} \delta_\ell \hat{\alpha}_{\ell i} \tag{3-13}$$

3.2 EXPRESSIONS FOR THE ASYMPTOTIC MEAN AND VARIANCE OF PROPORTION ESTIMATE $\hat{q}_i$

One of the objectives in the processing of remotely sensed aerospace imagery data is the estimation of proportion of the crop of interest. Assuming δ_ℓ is constant, expressions are developed in this section for the asymptotic mean and variance of proportion estimate $\hat{q}_i$. The expected value of $\hat{q}_i$ can be written as

$$E(\hat{q}_i) = \sum_{\ell=1}^{m} \delta_\ell E(\hat{\alpha}_{\ell i}) \tag{3-14}$$

The delta method (ref. 14) can be used to compute the asymptotic variance of $\hat{q}_i$. This involves expanding $\hat{q}_i$ in a Taylor series around its true value $q_i = \sum_{\ell=1}^{m} \delta_\ell \alpha_{\ell i}$. The result of the expansion is

$$\mathrm{Var}(\hat{q}_i) = \sum_{u=1}^{m} \sum_{v=1}^{m} \mathrm{cov}(\hat{\alpha}_{ui}\ \hat{\alpha}_{vi}) \frac{\partial q_i}{\partial \alpha_{ui}} \frac{\partial q_i}{\partial \alpha_{vi}}$$

$$= \sum_{u=1}^{m} \sum_{v=1}^{m} \delta_u \delta_v \ \mathrm{cov}\ (\hat{\alpha}_{ui}\ \hat{\alpha}_{vi}) \tag{3-15}$$

The covariance of the estimates $\hat{\alpha}_{ui}$ and $\hat{\alpha}_{vi}$ can be expressed as

$$\mathrm{cov}(\hat{\alpha}_{ui}\ \hat{\alpha}_{vi}) = E\{[\hat{\alpha}_{ui} - E(\hat{\alpha}_{ui})][\hat{\alpha}_{vi} - E(\hat{\alpha}_{vi})]\} \tag{3-16}$$

The estimates $(\hat{\alpha}_{ui})$ are functions of the given labeled patterns $X_r(j)$; $j = 1, 2, \cdots, N_r$; $r = 1, 2, \cdots, C$. Let the mean of $X_r(j)$ be μ_r. Expanding $\hat{\alpha}_{ui}$ in Taylor's series around $X_r(j) = \mu_r$; $j = 1, 2, \cdots, N_r$; $r = 1, 2, \cdots, C$ and retaining only first order terms yields

$$\hat{\alpha}_{ui}[X_r(j)\ ;\ j = 1,2,\cdots,N_r\ ;\ r = 1,2,\cdots,C] = \hat{\alpha}_{ui}\Big|_{\substack{X_r(j)=\mu_r \\ j=1,2,\cdots,N_r \\ r=1,2,\cdots C}}$$

$$+ \sum_{r=1}^{C} \sum_{j=1}^{N_r} \left(\frac{\partial \hat{\alpha}_{ui}}{\partial X_r(j)} \Big|_{X_r(j)=\mu_r} \right)^T (X_r(j) - \mu_r) + \cdots \tag{3-17}$$

Thus, to a first order approximation

$$E(\hat{\alpha}_{ui}) = \hat{\alpha}_{ui}\Big|_{X_r(j)=\mu_r} = \frac{[p(\mu_i|\Omega = u)/p(\mu_i)]}{\sum_{r=1}^{C}[p(\mu_r|\Omega = u)/p(\mu_r)]} \tag{3-18}$$

Similar to equation (3-17), expanding $\hat{\alpha}_{vi}$ in Taylor's series around $X_r(j) = \mu_r$ and retaining only first order terms yields

$$\hat{\alpha}_{vi} = \hat{\alpha}_{vi}\Big|_{\substack{X_r(j)=\mu_r \\ j=1,2,\cdots,N_r \\ r=1,2,\cdots,C}} + \sum_{r=1}^{C} \sum_{j=1}^{N_r} \left(\frac{\partial \hat{\alpha}_{vi}}{\partial X_r(j)} \Big|_{X_r(j)=\mu_r} \right)^T [X_r(j) - \mu_r] + \cdots \tag{3-19}$$

Assuming the patterns are independent and using equations (3-17) and (3-19) in equation (3-16) results in

$$\mathrm{cov}(\hat{\alpha}_{ui}\ \hat{\alpha}_{vi}) = \sum_{r=1}^{C} \sum_{j=1}^{N_r} \sum_{s=1}^{C} \sum_{t=1}^{N_s} \left[\frac{\partial \hat{\alpha}_{ui}}{\partial X_r(j)} \Bigg|_{X_r(j)=\mu_r} \right]^T E\left\{[X_r(j) - \mu_r][X_r(j) - \mu_r]^T\right\} \left[\frac{\partial \hat{\alpha}_{vi}}{\partial X_s(t)} \Bigg|_{X_s(t)=\mu_s} \right]$$

$$= \sum_{r=1}^{C} \sum_{j=1}^{N_r} \left(\frac{\partial \hat{\alpha}_{ui}}{\partial X_r(j)} \Bigg|_{X_r(j)=\mu_r} \right)^T S_r \left(\frac{\partial \hat{\alpha}_{vi}}{\partial X_r(j)} \Bigg|_{X_r(j)=\mu_r} \right) \tag{3-20}$$

where $$S_r = E\left\{[X_r(j) - \mu_r][X_r(j) - \mu_r]^T\right\} \tag{3-21}$$

Differentiation of equation (3-11) yields

$$\left.\begin{aligned} \frac{\partial \hat{\alpha}_{ki}}{\partial X_i(j)} &= \frac{(1 - \hat{\alpha}_{ki})}{\left\{\sum_{r=1}^{C} \sum_{s=1}^{N_r} p[\Omega = k|X_r(s)]\right\}} \cdot \left(\frac{\partial \{p[\Omega = k|X_i(j)]\}}{\partial X_i(j)} \right) \\ \frac{\partial \hat{\alpha}_{ki}}{\partial X_r(s)} &= - \frac{\hat{\alpha}_{ki}}{\left\{\sum_{r=1}^{C} \sum_{s=1}^{N_r} p[\Omega = k|X_r(s)]\right\}} \cdot \left(\frac{\partial \{p[\Omega = k|X_r(s)]\}}{\partial X_r(s)} \right) \end{aligned}\right\} \quad k = u,v \tag{3-22}$$

$$\left.\begin{aligned} &\text{Let } \alpha^*_{ki} = \alpha_{ki} \ (X_r(s) = \mu_r \ ; \ s = 1,2,\cdots,N_r \ ; \ r = 1,2,\cdots,C) \\ &\nu^*_k = \sum_{r=1}^{C} \sum_{s=1}^{N_r} p(\Omega = k|X_r(s) = \mu_r) \\ &\text{and} \quad \theta^*_{ki} = \frac{\partial\ p[\Omega = k|X_i(j)]}{\partial X_i(j)} \Bigg|_{X_i(j)=\mu_i} \end{aligned}\right\} \quad k = u,v \tag{3-23}$$

If the mode-conditional densities are Gaussian, θ^*_{ki} can easily be computed from the following. Consider

$$p[\Omega = k|X_i(j)] = \frac{\delta_k p[X_i(j)|\Omega = k]}{\sum_{r=1}^{m} \delta_r p[X_i(j)|\Omega = r]} \tag{3-24}$$

Let
$$p(X|\Omega = k) \sim N(M_k, \Sigma_k)$$

Then we can write

$$\frac{\partial p(X_i(j)|\Omega = k)}{\partial X_i(j)} = \left\{-p[X_i(j)|\Omega = k]\Sigma_k^{-1}[X_i(j) - M_k]\right\} \tag{3-25}$$

Using equations (3-22) and (3-23) in equation (3-20) yields

$$\operatorname{cov}(\hat{\alpha}_{ui}\hat{\alpha}_{vi}) = \frac{\overset{*}{\alpha}_{ui}\overset{*}{\alpha}_{vi}}{\overset{*}{\nu}_u\overset{*}{\nu}_v} \sum_{r=1}^{C} \left(\theta^{*T}_{ur} S_r \theta^*_{vr}\right)$$

$$+ (1 - \overset{*}{\alpha}_{ui} - \overset{*}{\alpha}_{vi})\left(\theta^{*T}_{ui} S_i \theta^*_{vi}\right) \tag{3-26}$$

The variance of $\hat{q}_i$ can then be computed from equations (3-15) and (3-26), using class sample means and covariance matrices for μ_r and S_r.

4. CLUSTER LABELING BASED ON THE CRITERION OF PROBABILITY OF CORRECT LABELING

If the class-conditional densities are known, the a posteriori probabilities of the classes can be expressed as a function of pattern X and a priori probabilities. Since the label of the pattern $X_i(j)$ is i, for particular class-conditional densities and a priori probabilities, $p[\omega = i|X_i(j)]$ is the probability with which the pattern $X_i(j)$ is correctly recognized. Let p_{ii} be the probability that the pattern comes from class i and belongs to class i for particular class-conditional densities and a priori probabilities. Similarly, let $p_{i\ell}$ be the probability with which the pattern comes from class i and belongs to class ℓ. Then, these probabilities can be expressed as

$$
\begin{aligned}
p_{ii} &= P(\omega = i) \int p(\omega = i|X)p(X|\omega = i)dX \\
&= P(\omega = i)\underset{p(X|\omega=i)}{E}[p(\omega = i|X)]
\end{aligned} \tag{4-1}
$$

and

$$p_{i\ell} = P(\omega = i)\underset{p(X|\omega=i)}{E}[p(\omega = \ell|X)] \tag{4-2}$$

The probability of correct labeling or the probability with which a pattern comes from a particular class and belongs to the same class is

$$P_S = \sum_{i=1}^{C} p_{ii} \tag{4-3}$$

The error probability or the probability with which the pattern comes from a particular class and belongs to some other class is

$$P_e = \sum_{i=1}^{C} \sum_{\substack{\ell=1 \\ \ell \neq i}}^{C} p_{i\ell} \tag{4-4}$$

From equations (4-1) to (4-4), it is easily seen that

$$P_S + P_e = 1 \tag{4-5}$$

It is observed that equations (4-1) and (4-3) are based on treating the a posteriori probabilities of the classes as continuous variables and differ from the usual estimates based on the counts. The probability P_S can be estimated from the given, labeled pattern set as follows.

$$\hat{P}_S = \sum_{i=1}^{C} q_i \hat{u}_i \tag{4-6}$$

where

$$\left.\begin{aligned} \hat{u}_i &= \frac{1}{N_i} \sum_{j=1}^{N_i} r_i(j) \\ \text{and} \quad r_i(j) &= p[\omega = i \mid X_i(j)] \end{aligned}\right\} \tag{4-7}$$

The following analysis shows that the estimate for P_S of equation (4-6) has less variance than the estimate based on counting the classification errors. The estimate of equation (4-6) is unbiased. That is

$$E(\hat{P}_S) = \sum_{i=1}^{C} q_i E(\hat{u}_i) = \sum_{i=1}^{C} q_i u_i = P_S \tag{4-8}$$

Assuming the patterns are independent, an expression for the variance of P_S can be obtained as follows. Consider

$$\mathrm{Var}(\hat{P}_S) = E\left[(\hat{P}_S - P_S)^2\right] = E\left[\sum_{i=1}^{C} q_i \frac{1}{N_i}\left\{\sum_{j=1}^{N_i} [r_i(j) - u_i]\right\}\right]^2$$

$$= \sum_{i=1}^{C} \sum_{k=1}^{C} q_i q_k \frac{1}{N_i N_k} \sum_{j=1}^{N_i} \sum_{\ell=1}^{N_k} E\{[r_i(j) - u_i][r_k(\ell) - u_k]\}$$

$$= \sum_{i=1}^{C} q_i^2 \frac{1}{N_i^2} \sum_{j=1}^{N_i} \left(E\{[r_i(j)]^2\} - u_i^2 \right) \tag{4-9}$$

But, we have the relationship

$$0 \leq r_i(j) \leq 1 \tag{4-10}$$

Using equation (4-10) in equation (4-9) yields

$$\mathrm{Var}(\hat{P}_S) \lesssim \sum_{i=1}^{C} q_i^2 \frac{1}{N_i^2} \sum_{j=1}^{N_i} \left\{ E[r_i(j)] - u_i^2 \right\}$$

$$= \sum_{i=1}^{C} q_i^2 \frac{1}{N_i} u_i(1 - u_i) \tag{4-11}$$

Hence, the variance of $\hat{P}_S$ is less than the variance of the estimate based on counts of correctly classified patterns (ref. 15). The criterion of either the maximization of $\hat{P}_S$ or the minimization of $\hat{P}_e$ can be used to obtain probabilistic class label assignment for the clusters. Using the relationship of equation (3-4) between the class-conditional densities and the cluster-conditional densities in terms of probabilities $\alpha_{\ell i}$ in equation (4-6) results in

$$\hat{P}_S = \sum_{i=1}^{C} P(\omega = i) \frac{1}{N_i} \sum_{j=1}^{N_i} \sum_{\ell=1}^{m} \alpha_{\ell i} p[\Omega = \ell | X_i(j)]$$

$$= \sum_{i=1}^{C} q_i \sum_{\ell=1}^{m} \alpha_{\ell i} e_{i\ell} \tag{4-12}$$

where

$$e_{i\ell} = \frac{1}{N_i} \sum_{j=1}^{N_i} p[\Omega = \ell | X_i(j)] \tag{4-13}$$

The probabilities q_i and $\alpha_{\ell i}$ are related as follows.

$$q_i = \sum_{\ell=1}^{m} \alpha_{\ell i} \delta_\ell \quad ; \quad i = 1,2,\cdots,C \tag{4-14}$$

Now the problem of estimation of proportions and the probabilities of class labels for the clusters can be formulated as follows.

Find: q_i, $i = 1, 2, \cdots, C$ and $\alpha_{\ell i}$; $i = 1, 2, \cdots, C$; $\ell = 1, 2, \cdots, m$ such that $\hat{P}_S$ is maximized where

$$\hat{P}_S = \sum_{i=1}^{C} q_i \sum_{\ell=1}^{m} \alpha_{\ell i} e_{i\ell} \tag{4-15}$$

Subject to the constraints

$$\left.\begin{aligned} &\sum_{i=1}^{C} q_i = 1 \\ &\sum_{i=1}^{C} \alpha_{\ell i} = 1 \quad ; \quad \ell = 1,2,\cdots,m \\ &\sum_{\ell=1}^{m} \alpha_{\ell i}\delta_\ell = q_i \quad ; \quad i = 1,2,\cdots,C \\ &q_i \geq 0 \quad ; \quad i = 1,2,\cdots,C \\ \text{and} \quad &\alpha_{\ell i} \geq 0 \quad ; \quad i = 1,2,\cdots,C \quad ; \quad \ell = 1,2,\cdots,m \end{aligned}\right\} \tag{4-16}$$

Comparing equations (3-6) and (4-15), it is seen that q_i is now directly entered into the problem. Optimization techniques such as Davidon-Fletcher-Powell (refs. 9, 10, 11) can easily be used to solve the above problem.

5. USE OF LABELED AND UNLABELED PATTERNS FOR PROBABILISTIC CLUSTER LABELING

One of the important objectives in the processing of remotely sensed imagery data is to estimate the proportions of classes of interest. Ideally, these estimates should be unbiased and of minimum variance. It is the purpose of this section to develop a scheme that uses both the given labeled and unlabeled patterns for obtaining the probabilities of class labels for the clusters by minimizing the variance of the proportion estimates. It is assumed that we are given a set of labeled patterns $X_i(j) \varepsilon \omega_i(j) = i$; $j = 1, 2, \cdots, N_i$; $i = 1, 2, \cdots, C$ and a set of unlabeled patterns Z_i, $i = 1, 2, \cdots, N_u$. Let N_T be the total number of labeled and unlabeled patterns. That is

$$N_T = \sum_{i=1}^{C} N_i + N_u$$

Let Y_i, $i = 1, 2, \cdots, N_T$ be the given labeled and unlabeled patterns. Let the Bayes classifier be used to classify the given labeled and unlabeled pattern sets. For particular class-conditional densities and a priori probabilities, let the resulting confusion matrix of a given labeled pattern set and the classification matrix of an unlabeled pattern set be as shown in table 1.

Let ω be the given label and ω_C be the classifier label. Let $\lambda_{ij} = P(\omega = i|\omega_C = j)$ be the probability that the true label is i, given that the classifier label is j. Let $P_{ij} = P(\omega = i, \omega_C = j)$ be the probability that the true label of the pattern is i and the classifier label is j. Let $P_C(i) = P(\omega_C = i)$ be the probability that the classifier classifies a pattern into class i and $q_i = P(\omega = i)$ be the a priori probability of class i. Then we obtain

$$\begin{aligned} p_{ij} &= P(\omega = i, \omega_C = j) \\ &= P(\omega_C = j)P(\omega = i|\omega_C = j) \\ &= P_C(j)\lambda_{ij} \end{aligned} \qquad (5\text{-}1)$$

TABLE 1.- CLASSIFICATIONS OF LABELED AND UNLABELED PATTERN SETS

(a) Confusion matrix of labeled pattern set

True label	Classifier label				Number belonging to each class
	1	2	...	C	
1	n_{11}	n_{12}	...	n_{1C}	$n_{1.} = N_1$
2	n_{21}	n_{22}	...	n_{2C}	$n_{2.} = N_2$
⋮	⋮	⋮		⋮	⋮
C	n_{C1}	n_{C2}	...	n_{CC}	$n_{C.} = N_C$
Number classified into each class	$n_{.1}$	$n_{.2}$	...	$n_{.C}$	$n = n_{..}$

(b) Matrix of classifications of unlabeled set

Classifier label			
1	2	...	C
V_1	V_2	...	V_C

where

n_{ij} = number of labeled patterns for which the true or given label is i and the classifier label is j

C = number of classes

$n_{i.} = \sum_{j=1}^{C} n_{ij}$

$n_{.j} = \sum_{i=1}^{C} n_{ij}$

$n = n_{..} = \sum_{i=1}^{C} \sum_{j=1}^{C} n_{ij}$, the total number of labeled patterns

V_j = number of unlabeled patterns for which the classifier label is j

Since each classification is independent, the likelihood function of the observed n's and V's can be written as

$$L = K \prod_{i=1}^{C} \prod_{j=1}^{C} (\lambda_{ij})^{n_{ij}} \prod_{j=1}^{C} [P_C(j)]^{V_j+n_{.j}} \tag{5-2}$$

where K is a constant. The values of $P_C(j)$ and λ_{ij} which maximize L, subject to the probability constraints on $P_C(j)$ and λ_{ij}, can be shown to be (refs. 16, 17)

$$\hat{P}_C(j) = \frac{n_{.j} + V_j}{\sum_{\ell=1}^{C} (n_{.\ell} + V_\ell)} \tag{5-3}$$

and

$$\hat{\lambda}_{ij} = \frac{n_{ij}}{n_{.j}} \tag{5-4}$$

An estimate $\hat{q}_i$ for the proportion q_i may be obtained as follows.

$$q_i = P(\omega = i) = \sum_{j=1}^{C} P(\omega = i, \omega_C = j) = \sum_{j=1}^{C} P_C(j)\lambda_{ij} \tag{5-5}$$

From equations (5-3) through (5-5), the following is obtained.

$$\hat{q}_i = \frac{\sum_{j=1}^{C} \left[\frac{n_{ij}}{n_{.j}} (n_{.j} + V_j)\right]}{\sum_{\ell=1}^{C} (n_{.\ell} + V_\ell)} \tag{5-6}$$

The estimate of equation (5-6) can be interpreted as follows. The ratio $(n_{ij}/n_{.j})$ gives the proportion of the patterns truly belonging to class i of the patterns classified into class j. Multiplying this ratio by $(n_{.j} + V_j)$ and summing it from 1 to C gives an estimate of the patterns of class i in the patterns classified into all the classes. Dividing this by the total number of patterns gives an estimate for the proportion of class i.

It can be shown (refs. 16, 17) that the estimate of equation (5-6) is asymptotically unbiased and its asymptotic variance is given by the following.

$$\mathrm{Var}(\hat{q}_i) = \sum_{j=1}^{C} \frac{\lambda_{ij}(1 - \lambda_{ij})P_C(j)}{n} + \sum_{j=1}^{C} \frac{P_C(j)\lambda_{ij}^2 - q_i^2}{N_u} \tag{5-7}$$

$$= \frac{q_i}{n} - \frac{q_i^2}{N_u} + \left(\frac{1}{N_u} - \frac{1}{n}\right) \sum_{j=1}^{C} P_C(j)\lambda_{ij}^2 \tag{5-8}$$

For particular a priori probabilities and class-conditional densities, the probabilities $P_C(j)$ and λ_{ij} may be expressed as

$$P_C(j) = \int p(\omega = j|X)p(X)dX \tag{5-9}$$

and

$$\lambda_{ij} = P(\omega = i|\omega_C = j)$$

$$= \frac{P(\omega = i,\omega_C = j)}{\sum_{i=1}^{C} P(\omega = i,\omega_C = j)}$$

$$= \frac{q_i \int p(\omega = j|X)p(X|\omega = i)dX}{\sum_{i=1}^{C} q_i \int p(\omega = j|X)p(X|\omega = i)dX} \tag{5-10}$$

Using the given labeled and unlabeled patterns, estimates for $P_C(j)$ and λ_{ij} can be obtained from equations (3-5), (5-9), and (5-10) as follows:

$$\hat{P}_C(j) = \frac{1}{N_T} \sum_{\ell=1}^{N_T} p(\omega = j|Y_\ell) = \sum_{s=1}^{m} \alpha_{sj} e_{us} \tag{5-11}$$

where

$$e_{us} = \frac{1}{N_T} \sum_{\ell=1}^{N_T} p(\Omega = s|Y_\ell)$$

and

$$\hat{\lambda}_{ij} = \frac{q_i \frac{1}{N_i} \sum_{\ell=1}^{N_i} p[\omega = j | X_i(\ell)]}{\sum_{i=1}^{C} q_i \frac{1}{N_i} \sum_{\ell=1}^{N_i} p[\omega = j | X_i(\ell)]}$$

$$= \frac{q_i \sum_{s=1}^{m} \alpha_{sj} e_{is}}{\sum_{r=1}^{C} q_r \sum_{s=1}^{m} \alpha_{sj} e_{rs}} \tag{5-12}$$

Using equations (5-11) and (5-12) in equation (5-5) yields

$$q_i = \sum_{j=1}^{C} \hat{P}_C(j) \hat{\lambda}_{ij} \quad ; \quad i = 1,2,\cdots,C \tag{5-13}$$

Let S be the sum of asymptotic variances of proportion estimates. From equations (5-8), (5-11), and (5-12), the following estimate for S is obtained.

$$\hat{S} = \sum_{i=1}^{C} \mathrm{Var}\ (\hat{q}_i)$$

$$= \frac{1}{n} - \sum_{i=1}^{C} \frac{q_i^2}{N_u} + \left(\frac{1}{N_u} - \frac{1}{n}\right) \sum_{j=1}^{C} \left\{ \left(\sum_{s=1}^{m} \alpha_{sj} e_{us} \right) \left[\frac{\sum_{i=1}^{C} q_i^2 \left(\sum_{s=1}^{m} \alpha_{sj} e_{is} \right)^2}{\left(\sum_{s=1}^{m} \alpha_{sj} \sum_{r=1}^{C} q_r e_{rs} \right)^2} \right] \right\} \tag{5-14}$$

Now the problem of obtaining probabilistic class label assignment for the clusters may be formulated as follows.

Find: α_{sj}; $s = 1,2,\cdots,m$; $j = 1,2,\cdots,C$; and q_i, $i = 1,2,\cdots,C$ such that $\hat{S}$ of equation (5-14) is minimized, subject to the constraints of equations (4-16) and (5-13).

Optimization techniques such as Davidon-Fletcher-Powell (refs. 9, 10, 11) can easily be used to solve the above optimization problem.

6. FORMULATION WITH LABEL IMPERFECTIONS OF THE GIVEN LABELED SET

In practice, such as in the classification of remotely sensed, multispectral scanner imagery data, it is difficult and expensive to obtain labels for the pattern set. The labels for the patterns are usually provided by an analyst interpreter on examining imagery films and using some other information such as historic information, crop calendar models, etc. These labels are very often imperfect. Recently, Chittineni (refs. 16, 18, 19) investigated techniques for estimating the probabilities of label imperfections. Once the probabilities of label imperfections are known, these can be used in obtaining the class labels to the clusters through their densities and proportions. Let ω and ω' be the perfect and imperfect labels, respectively, each of which take values 1, 2, ···, C. The imperfections in the labels are described by the probabilities

$$\beta_{ji} = P(\omega' = i \mid \omega = j) \tag{6-1}$$

where

$$\sum_{i=1}^{C} \beta_{ji} = 1 \tag{6-2}$$

The a priori probabilities, the class-conditional densities, and the a posteriori probabilities of the classes with and without imperfections in the labels are related in terms of probabilities of label imperfections as (ref. 18)

$$P(\omega' = i) = \sum_{j=1}^{C} \beta_{ji} P(\omega = j) \tag{6-3}$$

$$P(\omega' = i) p(X \mid \omega' = i) = \sum_{j=1}^{C} \beta_{ji} P(\omega = j) p(X \mid \omega = j) \tag{6-4}$$

$$p(\omega' = i \mid X) = \sum_{j=1}^{C} \beta_{ji} p(\omega = j \mid X) \tag{6-5}$$

where it is assumed that

$$p(X|\omega = j) = p(X|\omega' = i, \omega = j) \tag{6-6}$$

Let β be the matrix of probabilities of label imperfections where

$$\beta = [\beta_{ij}] \tag{6-7}$$

Let

$$\nu = (\beta^T)^{-1} \tag{6-8}$$

Using equations (6-7) and (6-8) and inverting equation (6-4) results in

$$P(\omega = i)p(X|\omega = i) = \sum_{j=1}^{C} \nu_{ij} P(\omega' = j)p(X|\omega' = j) \tag{6-9}$$

Using these relationships, the criteria developed in the previous sections for labeling the clusters can be reformulated to take into account the imperfections in the labels, once β_{ji} are known or estimated. In the following, it is assumed that the probabilities of label imperfections β_{ji} are available.

6.1 MAXIMUM LIKELIHOOD CRITERION WITH IMPERFECTIONS IN THE LABELS

The log of likelihood function of equation (3-1) with imperfections in the labels can be written as

$$L = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \log \{p[\omega_i'(j) = i|X_i(j)]\} \tag{6-10}$$

Using equations (3-5) and (6-5) in equation (6-10) yields

$$L = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \log \left\{ \sum_{\ell=1}^{C} \beta_{\ell i} p[\omega = \ell | X_i(j)] \right\}$$

$$= \sum_{i=1}^{C} \sum_{j=1}^{N_i} \log \left\{ \sum_{\ell=1}^{C} \sum_{r=1}^{m} \alpha_{r\ell} \beta_{\ell i} p[\Omega = r | X_i(j)] \right\} \qquad (6\text{-}11)$$

For a given $\beta_{\ell i}$, the problem of estimating $\alpha_{r\ell}$ can be formulated as follows. Find: $\alpha_{r\ell}$; $r = 1, 2, \cdots, m$; $\ell = 1, 2, \cdots, C$ such that L of equation (6-11) is maximized, subject to the constraints of equation (3-7).

Closed form solutions for the above optimization problem seem to be difficult. However, the following fixed point iteration scheme, similar to equation (3-8), can easily be obtained by introducing the lograngian multipliers.

$$\alpha_{r\ell} = \frac{\sum_{i=1}^{C} \sum_{j=1}^{N_i} d_{ij\ell r}}{\sum_{\ell=1}^{C} \sum_{i=1}^{C} \sum_{j=1}^{N_i} d_{ij\ell r}} \qquad (6\text{-}12)$$

where

$$d_{ij\ell r} = \frac{\alpha_{r\ell} \beta_{\ell i} p[\Omega = r | X_i(j)]}{\sum_{s=1}^{C} \sum_{k=1}^{m} \alpha_{ks} \beta_{si} p[\Omega = k | X_i(j)]} \qquad (6\text{-}13)$$

Also, optimization methods such as Davidon-Fletcher-Powell (refs. 9, 10, 11) can easily be used to solve the above optimization problem.

6.2 THE CRITERION OF PROBABILITY OF CORRECT LABELING WITH LABEL IMPERFECTIONS

In section 4, the probability of correct labeling is proposed as a criterion for obtaining the probabilities of class labels for the clusters. From equations (4-1), (4-3), and (6-9), we obtain

$$P_S = \sum_{i=1}^{C} \int [p(\omega = i|X)][P(\omega = i)p(X|\omega = i)]dx$$

$$= \sum_{i=1}^{C} \sum_{j=1}^{C} \nu_{ij} P(\omega' = j) \int p(\omega = i|X)p(X|\omega' = j)dx \qquad (6\text{-}14)$$

An estimate for P_S can be obtained in terms of given, imperfectly labeled patterns as

$$\hat{P}_S = \sum_{i=1}^{C} \sum_{j=1}^{C} \nu_{ij} P(\omega' = j) \left[\frac{1}{N_j} \sum_{\ell=1}^{N_j} p[\omega = i|X_j(\ell)] \right] \qquad (6\text{-}15)$$

Using equations (3-5) and (6-3) in equation (6-15) yields

$$\hat{P}_S = \sum_{i=1}^{C} \sum_{j=1}^{C} \nu_{ij} P(\omega' = j) \left\{ \frac{1}{N_j} \sum_{\ell=1}^{N_j} \sum_{r=1}^{m} \alpha_{ri} p[\Omega = r|X_j(\ell)] \right\}$$

$$= \sum_{i=1}^{C} \sum_{j=1}^{C} \nu_{ij} \sum_{s=1}^{C} \beta_{sj} q_s \left\{ \frac{1}{N_j} \sum_{\ell=1}^{N_j} \sum_{r=1}^{m} \alpha_{ri} p[\Omega = r|X_j(\ell)] \right\}$$

$$= \sum_{s=1}^{C} q_s \sum_{r=1}^{m} \sum_{i=1}^{C} \alpha_{ri} b_{sri} \qquad (6\text{-}16)$$

where

$$b_{sri} = \sum_{j=1}^{C} \beta_{sj} e_{jr} \nu_{ij} \qquad (6\text{-}17)$$

Now the problem can be formulated as follows.

Find: α_{ri} ; i = 1, 2, ⋯, C; r = 1, 2, ⋯, m and q_s, s = 1, 2, ⋯, C such that $\hat{P}_S$ of equation (6-16) is maximized, subject to the constraints of equation (4-16).

Optimization techniques such as Davidon-Fletcher-Powell (refs. 9, 10, 11) can easily be used to solve the above optimization problem.

6.3 THE CRITERION OF VARIANCE OF PROPORTION ESTIMATE WITH IMPERFECTIONS IN THE LABELS

The probabilities of label imperfections can be taken into account in estimating the probabilities of class labels to the modes through the probabilities λ_{ij}. Using equation (6-9) in equation (5-10) yields

$$\lambda_{ij} = \frac{\sum_{\ell=1}^{C} \nu_{i\ell} P(\omega' = \ell) \int p(\omega = j|X) p(X|\omega' = \ell) dx}{\sum_{i=1}^{C} \sum_{\ell=1}^{C} \nu_{i\ell} P(\omega' = \ell) \int p(\omega = j|X) p(X|\omega' = \ell) dx} \tag{6-18}$$

An estimate for λ_{ij} can be obtained from the given, imperfectly labeled patterns as

$$\hat{\lambda}_{ij} = \frac{\sum_{\ell=1}^{C} \nu_{i\ell} P(\omega' = \ell) \left\{ \frac{1}{N_\ell} \sum_{s=1}^{N_\ell} p[\omega = j|X_\ell(s)] \right\}}{\sum_{i=1}^{C} \sum_{\ell=1}^{C} \nu_{i\ell} P(\omega' = \ell) \left\{ \frac{1}{N_\ell} \sum_{s=1}^{N_\ell} p[\omega = j|X_\ell(s)] \right\}} \tag{6-19}$$

Using equations (6-3), (3-5), (6-19), and (5-11) in equation (5-8) yields a criterion similar to equation (5-14).

7. EXPERIMENTAL RESULTS

This section presents some results obtained in the processing of remotely sensed Landsat multispectral scanner (MSS) imagery data. The objective of the processing is to estimate the proportion of class of interest in each image. Several segments were processed in the following manner. (A segment is a 9 by 11 kilometer or 5 by 6 nautical mile area for which the MSS image is divided into a rectangular array of pixels, 117 rows by 196 columns.) The image is overlaid with a rectangular grid of 209 grid intersections. Ground truth labels or true labels of the pixels or dots corresponding to each grid intersection are acquired. Also, for a subset of the pixels of 209 grid intersections, the labels are provided by analyst interpreter (AI) by examining the imagery films and using information such as historic information, crop calender models, etc. These are imperfect labels. There are two classes in the image. Class 1 is wheat and class 2 is nonwheat, designated "other." The class of interest is wheat.

Several acquisitions are used for each segment. The number of features or the number of channels used for each segment is listed in table 2. The Gaussian mode-conditional densities and a priori probabilities of the inherent modes in the data of each segment are obtained using maximum likelihood clustering algorithm (refs. 2, 20). The number of modes generated for each segment is listed in table 2. The probabilities of label imperfections of AI labels or β matrix are estimated for each segment and are listed in table 2. The theory developed in section 3 is applied in estimating the probabilities of class labels to the modes of each segment using AI labeled patterns and using ground truth labeled patterns. The proportion of class 1, the class of interest, is estimated for each segment using equation (3-13) and listed in table 2.

The theory developed in section 6.1 is used with the AI labeled patterns and the corresponding β matrix in estimating the probabilities of class labels to the modes. Equation (3-13) is then used in estimating the proportion of class 1 for each segment and the results are listed in table 2.

TABLE 2.- ESTIMATION OF PROPORTION OF CLASS 1 WITH MAXIMUM LIKELIHOOD CRITERION

Segment	Location	Number of labeled patterns	AI labels		GT labels		β-matrix computed comparing AI and GT labels $\begin{bmatrix}\beta_{11} & \beta_{12}\\ \beta_{21} & \beta_{22}\end{bmatrix}$	With imperfections initial α_{re} =		No. of features(d) and clusters(m)	GT proportion
			Closed form solution	Iterative scheme	Closed form solution	Iterative scheme		0.5	Closed form solution		
1005	Sherman, Texas	97	0.2311	0.2456	0.3621	0.3885	$\begin{bmatrix}0.5455 & 0.4545\\ 0.0308 & 0.9692\end{bmatrix}$	0.3044	0.3025	d = 8, m = 11	0.348
1060	Cheyenne, Colorado	106	0.1968	0.1975	0.3169	0.3251	$\begin{bmatrix}0.5667 & 0.4333\\ 0.0293 & 0.9737\end{bmatrix}$	0.2174	0.2172	d = 4, m = 12	0.231
1231	Jackson, Oklahoma	96	0.6378	0.6265	0.7395	0.7156	$\begin{bmatrix}0.9315 & 0.0685\\ 0.1304 & 0.8696\end{bmatrix}$	0.7141	0.7139	d = 6, m = 12	0.744
1520	Big Stone, Montana	91	0.2300	0.2109	0.2733	0.2696	$\begin{bmatrix}0.7917 & 0.2083\\ 0.0149 & 0.9851\end{bmatrix}$	0.3643	0.3647	d = 6, m = 10	0.301
1604	Renville, N. Dakota	101	0.3214	0.2963	0.5023	0.4982	$\begin{bmatrix}0.4600 & 0.5400\\ 0.1569 & 0.8431\end{bmatrix}$	0.4817	0.4814	d = 4, m = 7	0.524
1675	Mcpherson, S. Dakota	107	0.1019	0.1085	0.2977	0.3037	$\begin{bmatrix}0.2667 & 0.7333\\ 0.0390 & 0.9610\end{bmatrix}$	0.2193	0.2156	d = 8, m = 12	0.291
1805	Gregory, S. Dakota	144	0.1156	0.1181	0.1649	0.1896	$\begin{bmatrix}0.4211 & 0.5789\\ 0.0640 & 0.9350\end{bmatrix}$	0.1937	0.1932	d = 8, m = 12	0.154
1853	Ness, Kansas	91	0.3161	0.3246	0.3200	0.3379	$\begin{bmatrix}0.8077 & 0.1923\\ 0.0615 & 0.9385\end{bmatrix}$	0.3052	0.3052	d = 6, m = 12	0.306
1899	Walsh, N. Dakota	95	0.6273	0.6282	0.6287	0.6345	$\begin{bmatrix}0.8644 & 0.1356\\ 0.2500 & 0.7500\end{bmatrix}$	0.6269	0.6269	d = 8, m = 7	0.596
Bias			.80778E-01	.832E-01	-.11156E-01	-.17522E-01		.8667E-02	.93778E-02		
MSE			.12736F-01	.13575E-01	.11274E-02	.18077E-02		.1750E-02	.18369E-02		

From table 2, it is observed that the estimates obtained with the closed form solution of equation (3-11) for the probabilities of class labels to the nodes are in close agreement with the ones obtained using the fixed point iteration scheme. Better proportion estimates are obtained by taking the imperfections in the AI labels into account through the β matrix instead of estimating the proportions directly using AI labeled patterns.

The estimated proportions of class 1 by exhaustive search using maximum likelihood criterion and maximization of probability of correct labeling criterion with both the AI and ground truth labels are listed in table 3.

The estimated proportions of class 1 from the given, randomly labeled patterns are listed in table 4 for all the processed segments. On comparison of tables 2, 3, and 4, it is seen that there is improvement in the estimates through machine processing.

For all the segments, the method developed in reference 19 is used to estimate the probabilities of label imperfections of AI labels. The number of labeled patterns used for each segment is listed in table 5 and the number of unlabeled patterns used is 836. The estimated labeling accuracies and the proportion estimates obtained using maximum likelihood criterion to label the clusters with these probabilities of label imperfections are listed in table 5. From table 5, it is seen that there is considerable improvement in the proportion estimates with the use of estimated β-matrix over that obtained directly using imperfectly labeled patterns.

TABLE 3.- ESTIMATION OF PROPORTION OF CLASS 1 BY EXHAUSTIVE SEARCH

Segment	Location	No. of labeled patterns	Maximum likelihood criterion		Probability of correct labeling criterion		Ground truth proportion
			AI labels	GT labels	AI labels	GT labels	
1005	Sherman, Texas	97	0.2746	0.3220	0.2746	0.3746	0.348
1060	Cheyenne, Colorado	106	0.2171	0.2529	0.2171	0.2171	0.231
1231	Jackson, Oklahoma	96	0.7562	0.7245	0.7672	0.7491	0.744
1520	Big Stone, Montana	91	0.1837	0.3017	0.1986	0.2886	0.301
1604	Renville, N. Dakota	101	0.4114	0.5195	0.4378	0.5359	0.524
1675	Mcpherson, S. Dakota	107	0.1713	0.2843	0.3131	0.2809	0.2[illegible]
1805	Gregory, S. Dakota	144	0.1207	0.1981	0.1098	0.1914	0.164
1853	Ness, Kansas	91	0.3573	0.2719	0.2860	0.2960	0.306
1899	Walsh, N. Dakota	95	0.6313	0.6320	0.6312	0.6441	0.596
Bias			0.4238E-01	-0.2111E-03	0.2995E-01	-0.8077E-02	
MSE			0.5805E-02	0.5804E-03	0.3233E-02	0.4987E-03	

TABLE 4.- ESTIMATION OF PROPORTION OF CLASS 1 FROM RANDOMLY LABELED PATTERNS

Segment	Number of labeled patterns	Proportion of class 1 estimated from labeled patterns		GT proportion
		AI labels	GT labels	
[a]1005	97	0.2061	0.3368	0.348
[a]1060	106	0.1604	0.2830	0.231
[a]1231	96	0.7395	0.7604	0.744
[b]1520	91	0.2197	0.2637	0.301
[b]1604	101	0.3069	0.4950	0.524
[b]1675	107	0.0934	0.2897	0.291
[b]1805	144	0.1042	0.1389	0.164
[a]1853	91	0.2637	0.2857	0.306
[b]1899	95	0.6316	0.6484	0.596
Bias		0.86611E-01	0.37778E-03	
MSE		0.13840E-01	0.10134E-02	

[a]Segments in which class 1 is winter wheat.
[b]Segments in which class 1 is spring wheat.

TABLE 5.- ESTIMATED LABELING ERRORS AND PROPORTION ESTIMATES

Segment	Location	Number of AI labeled patterns		Estimated β-matrix using method developed in reference 19	Computed β-matrix comparing AI and GT labels	P_1 using β of column 3	P_1 directly with AI labels	GT proportion
		Wheat	Other					
1005	Sherman, Texas	20	77	$\begin{bmatrix} 0.8284 & 0.1716 \\ 0.0165 & 0.9835 \end{bmatrix}$	$\begin{bmatrix} 0.5455 & 0.4545 \\ 0.0308 & 0.9692 \end{bmatrix}$	0.2723	0.2456	0.348
1060	Cheyenne, Colorado	17	89	$\begin{bmatrix} 0.5732 & 0.4268 \\ 0.0431 & 0.9569 \end{bmatrix}$	$\begin{bmatrix} 0.5667 & 0.4333 \\ 0.0293 & 0.9737 \end{bmatrix}$	0.2173	0.1975	0.231
1231	Jackson, Oklahoma	71	25	$\begin{bmatrix} 0.9586 & 0.0414 \\ 0.1330 & 0.8670 \end{bmatrix}$	$\begin{bmatrix} 0.9315 & 0.0685 \\ 0.1304 & 0.8696 \end{bmatrix}$	0.7057	0.6265	0.744
1520	Big Stone, Montana	20	71	$\begin{bmatrix} 0.8629 & 0.1371 \\ 0.0363 & 0.9637 \end{bmatrix}$	$\begin{bmatrix} 0.7917 & 0.2083 \\ 0.0149 & 0.9851 \end{bmatrix}$	0.2154	0.2109	0.301
1604	Renville, N. Dakota	31	70	$\begin{bmatrix} 0.6155 & 0.3845 \\ 0.0000 & 1.0000 \end{bmatrix}$	$\begin{bmatrix} 0.4600 & 0.5400 \\ 0.1569 & 0.8431 \end{bmatrix}$	0.3496	0.2963	0.524
1675	Mcpherson, S. Dakota	10	97	$\begin{bmatrix} 0.5481 & 0.4519 \\ 0.0005 & 0.9995 \end{bmatrix}$	$\begin{bmatrix} 0.2667 & 0.7333 \\ 0.0390 & 0.9610 \end{bmatrix}$	0.1932	0.1085	0.291
1805	Gregory, S. Dakota	15	129	$\begin{bmatrix} 0.5227 & 0.4773 \\ 0.0626 & 0.9374 \end{bmatrix}$	$\begin{bmatrix} 0.4211 & 0.5789 \\ 0.0640 & 0.9360 \end{bmatrix}$	0.1757	0.1181	0.164
1853	Ness, Kansas	24	67	$\begin{bmatrix} 0.6342 & 0.3658 \\ 0.0027 & 0.9973 \end{bmatrix}$	$\begin{bmatrix} 0.8077 & 0.1923 \\ 0.0615 & 0.9385 \end{bmatrix}$	0.3563	0.3246	0.306
1899	Walsh, N. Dakota	60	35	$\begin{bmatrix} 0.9972 & 0.0028 \\ 0.0356 & 0.9644 \end{bmatrix}$	$\begin{bmatrix} 0.8644 & 0.1356 \\ 0.2500 & 0.7500 \end{bmatrix}$	0.6216	0.6282	0.596
Bias						.4421E-01	.832E-01	
MSE						.6446E-02	.13575E-01	

8. CONCLUDING SUMMARY

In the classification of imagery data such as in the machine processing of remotely sensed multispectral scanner data, unsupervised classification techniques have been found to be effective. Clustering techniques break up the image into its inherent modes. One of the crucial problems in the machine classification of imagery data is to label these clusters.

This paper addressed the problem of labeling the modes and proposed various techniques. It is assumed that the a priori probabilities of the modes and the mode-conditional probability densities are available. It is also assumed that a set of labeled patterns from the classes of the data and a set of unlabeled patterns are given. The labels of these patterns might be imperfect.

Using the given, labeled pattern set, the problem of assigning the class labels to the modes is formulated as a combinatorial problem. If the number of modes is small, best assignment of class labels to the modes can easily be obtained by exhaustive search, using criterion such as maximum likelihood.

The problem is also formulated as that of obtaining probabilistic class label assignment to the modes using maximization of either the likelihood function or the probability of correct labeling as a criterion. Closed form solution is obtained for the probabilities of class labels to the modes with the maximization of lower bound on the likelihood function as a criterion. With this solution and using the Taylor series expansion, expressions are developed for the asymptotic mean and variance of the proportion estimates of the classes. In the processing of remotely sensed data, one of the important objectives is to estimate the proportion of class of interest. Using the given, labeled and unlabeled patterns, the problem of obtaining class labels to the modes is formulated as that of minimizing the variance of the proportion estimates of the classes.

The criteria of the maximum likelihood, maximization of probability of correct labeling, and minimization of variance of proportion estimates are reformulated to take into account label imperfections in the given, labeled set for known probabilities of label imperfections.

In practice, it is often of interest to group the modes into their natural classes. A procedure that uses unlabeled patterns based on probability of error as a criterion is proposed for grouping the modes into their natural classes. Also, the problem of proportion estimation through cluster labeling with impure clusters is addressed.

Furthermore, experimental results in the processing of remotely sensed multispectral scanner imagery data are presented.

9. REFERENCES

1. Everitt, B.: Cluster Analysis. Wiley, New York, 1974.

2. Duda, R. O.; and Hart, P. E.: Pattern Classification and Scene Analysis. Wiley, New York, 1973.

3. Heydorn, R. P.: Methods for Segment Wheat-Area Estimation. Proceedings of the LACIE Symposium, NASA/JSC, Houston, Texas, 1978.

4. Cochran, W. G.: Sampling Techniques. Wiley, New York, 1953.

5. Rosenfeld, A.; Hummel, R.; and Tucker, S. W.: Scene Labeling by Relaxation Operations. IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-6, June 1976, pp. 420-433.

6. Rosenfeld, A.: Iterative Methods in Image Analysis. Proceedings of the IEEE Conference on Pattern Recognition and Image Processing, Troy, New York, June 1977, pp. 14-18.

7. Tucker, S. W.; Krishna Murty, E. V.; and Haar, R. L.: Relaxation Processes for Scene Labeling: Convergence, Speed, and Stability. IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-8, January 1978, pp. 41-48.

8. Beckenbach, E. F.: Applied Combinatorial Mathematics. Wiley, New York, 1964.

9. Fletcher, R.; and Powell, M. J. D.: A Rapidly Covergent Descent Method for Minimization. Computer J., vol. 6, April 1963, pp. 163-168.

10. Johnson, I. L.: The Davidon-Fletcher-Powell Penalty Function Method: A Generalized Iterative Technique for Solving Parameter Optimization Problems. Technical note TN D-8251, NASA/JSC, Houston, Texas, May 1976.

11. Cooper, L.; and Steinberg, D.: Introduction to Methods of Optimization. W. B. Saunders Co., Philadelphia, 1970.

12. Lennington, R. K.: Proportion Estimation Using Distributions. Lockheed Internal Note, Houston, Texas.

13. Haydorn, R. P.: Private Communication.

14. Rao, C. R.: Linear Statistical Inference and Its Applications. Wiley, New York, 1965.

15. Highleyman, W. H.: The Design and Analysis of Pattern Recognition Experiments. Bell System Tech. J., vol. 41, March 1962, pp. 723-744.

16. Chittineni, C. B.: Maximum Likelihood Estimation of Label Imperfection Probabilities and Its Use in the Identification of Mislabeled Patterns. Technical Memorandum, LEC-13678, JSC-16079, Houston, Texas, September 1979. Also to be presented at 1980 Machine Processing of Remotely Sensed Data Symposium, LARS, Purdue University.

17. Tenenbein, A.: Estimation From Data Subject to Measurement Error. Ph.D. Dissertation, Statistics Department, Harvard University, Cambridge, Mass., 1969.

18. Chittineni, C. B.: Learning With Imperfectly Labeled Patterns. Technical Memorandum, LEC-13068, JSC-14867, April 1979. Also in Proc. IEEE Converence on Pattern Recognition and Imagery Processing, Chicago, Illinois, 1979, pp. 52-62.

19. Chittineni, C. B.: Estimation of Probabilities of Label Imperfections and Correction of Mislabels. Technical Memorandum, LEMSCO-14356, JSC-16342, March 1980.

20. Lennington, R. K.; and Rassbach, M. E.: Mathematical Description and Program Documentation for Classy: An Adaptive Maximum Likelihood Clustering Method. Technical Memorandum, LEC-12177, Lockheed Electronics Company, Inc., Houston, Texas, 1979.

21. Fukunaga, K.; and Kessel, D. L.: Nonparametric Error Estimation using Unclassified Samples. IEEE Trans. on Information Theory, vol. IT-19, pp. 434-440, July 1973.

22. Minter, T. C.: Single Class Classification. Proc. of the Machine Processing of Remotely Sensed Data, LARS, Purdue University, June 3-5, 1975, pp. 2A-12 to 2A-15. .

APPENDIX A

USE OF UNLABELED PATTERNS FOR ASSIGNING MODES TO THEIR NATURAL CLASSES

In practice, it is often of interest to group the modes generated by the clustering algorithm into their natural classes. An approach is proposed in this appendix for grouping the clusters into their respective classes based on Bayes probability of error as a criterion and using unlabeled patterns. It is assumed that a set of unlabeled patterns X_i, $i = 1, 2, \cdots, N$ is given, and the number of classes C is given.

The Bayes classifier classifies a pattern X into a class whose a posteriori probability is largest. The conditional probability of error when X is classified according to the Bayes decision rule is

$$r(X) = 1 - \max_i [p(\omega = i|X)] \tag{A-1}$$

The Bayes probability of error is then given by

$$P_e = E[r(X)] = \int r(X)p(X)dx \tag{A-2}$$

Thus, if r(X) is known as a function of X, the Bayes probability of error P_e can be estimated by the sample mean $r(X_i)$ of N unlabeled patterns as

$$\hat{P}_e = \frac{1}{N} \sum_{i=1}^{N} r(X_i) \tag{A-3}$$

The estimate of equation (A-3) is unbiased and its variance is given by (ref. 21)

$$\mathrm{Var}(\hat{P}_e) = \frac{E[r^2(X)] - P_e^2}{N} \leq \frac{P_e(1 - P_e)}{N} - \frac{P_e}{CN} \tag{A-4}$$

The variance of $\hat{P}_e$ is at least $\frac{P_e}{CN}$ less than the variance of the error estimate based on counting misclassified labeled test patterns, $\frac{P_e(1 - P_e)}{N}$.

Using equations (3-5) and (A-1) in equation (A-3), an estimate of Bayes error probability can be written as

$$\hat{P}_e = 1 - \frac{1}{N} \sum_{i=1}^{N} \left\{ \max_{k} \left[\sum_{\ell=1}^{m} \alpha_{\ell k} p(\Omega = \ell | X_i) \right] \right\} \tag{A-5}$$

Now the problem can be formulated as follows.

Find: $\alpha_{\ell k}$; $k = 1,2,\cdots,C$ and $\ell = 1,2,\cdots,m$ such that $\hat{P}_e$ of equation (A-5) is minimized, subject to the constraints of equation (3-7).

Optimization techniques such as Davidon-Fletcher-Powell (refs. 9, 10, 11) can easily be used to solve the above optimization problem.

APPENDIX B
CLUSTER LABELING WITH A SET OF LABELED PATTERNS FROM A SINGLE CLASS AND A SET OF UNLABELED PATTERNS

In practice, the situation often arises in which the patterns of one class are easily and accurately labeled compared to the patterns of another class. For example, in remote sensing, it is easier to label the pixels of another class (ref. 22) and accuracy of labeling is higher for these pixels compared to the pixels of the wheat class. This appendix formulates the problem of obtaining class labels to the clusters using information from a given set of labeled patterns from a single class and a set of unlabeled patterns. It is assumed that there are only two classes. The probability of correct labeling is used as a criterion. Let $X_1(\ell)$, $\ell = 1, 2, \cdots, N_1$ be the given patterns from class 1 and Y_ℓ, $\ell = 1, 2, \cdots, N_u$ be the given set of unlabeled patterns. From equations (4-1) and (4-2), we get

$$P_S = \sum_{i=1}^{2} P(\omega = i) \int p(\omega = i|X)p(X|\omega = i)dX$$

$$= q_1 \int [p(\omega = 1|X) - p(\omega = 2|X)]p(X|\omega = 1)dX$$

$$+ \int p(\omega = 2|X)p(X)dX \qquad \text{(B-1)}$$

The probability P_S can be estimated using the given set of labeled and unlabeled patterns is as follows.

$$\hat{P}_S = q_1 \frac{1}{N_1} \sum_{\ell=1}^{N_1} \{p[\omega = 1|X_1(\ell)] - p[\omega = 2|X_1(\ell)]\}$$

$$+ \frac{1}{N_u}\left[\sum_{\ell=1}^{N_u} p(\omega = 2|Y_\ell)\right] \qquad \text{(B-2)}$$

Using equation (3-5) in equation (B-2) yields

$$\hat{P}_S = q_1 \sum_{r=1}^{m} (\alpha_{r1} - \alpha_{r2}) e_{1r} + \sum_{r=1}^{m} \alpha_{r2} e_{ur} \tag{B-3}$$

where e_{1r} is given by equation (4-13) and e_{ur} is given by the following.

$$e_{ur} = \frac{1}{N_u} \sum_{\ell=1}^{N_u} p(\Omega = r | Y_\ell) \tag{B-4}$$

Now the problem of estimating the probabilities of class labels to the clusters can be formulated as follows.

Find q_1 and α_{ri} where $r = 1,2, \cdots, m$ and $i = 1,2$ such that $\hat{P}_S$ is maximized, subject to the constraints

$$\left.\begin{array}{c} \sum_{i=1}^{2} \alpha_{ri} = 1 \quad ; \quad r = 1,2,\cdots,m \\ \sum_{r=1}^{m} \alpha_{r1}\, \delta_r = q_1 \\ 0 < q_1 < 1 \\ \alpha_{ri} > 0 \quad ; \quad i = 1, 2 \text{ and } r = 1,2,\cdots,m \end{array}\right\} \tag{B-5}$$

Optimization techniques such as Davidon-Fletcher-Powell (refs, 9, 10, 11) can easily be used to solve the above optimization problem.

APPENDIX C

FIXED POINT INTERATION SCHEMES FOR PROBABILISTIC CLUSTER LABELING WITH THE CRITERION OF PROBABILITY OF CORRECT LABELING

Fixed point iteration schemes are presented in this appendix for obtaining the probabilities of class labels to the clusters, assuming that the a priori probabilities q_i of the classes can be approximately estimated from the given labeled patterns.

C.1 THE LABELS OF THE GIVEN PATTERN SET ARE PERFECT

Since logorithm is a monotonic function of its argument, taking log of equation (4-15) and using inequality (2-9) results in

$$\begin{aligned} \hat{P}_S' &= \log (\hat{P}_S) \\ &= \log \left(\sum_{i=1}^{C} q_i \sum_{\ell=1}^{m} \alpha_{\ell i} e_{i\ell} \right) \\ &> \sum_{i=1}^{C} q_i \log \left(\sum_{\ell=1}^{m} \alpha_{\ell i} e_{i\ell} \right) \end{aligned} \tag{C-1}$$

A fixed point interation scheme for computing the probabilities of class labels to the clusters $\alpha_{\ell i}$, that maximize the lower bound of equation (C-1), subject to the constraints of equation (3-7), can easily be shown to be the following.

$$\alpha_{\ell i} = \frac{d_{\ell i}}{\sum_{i=1}^{C} d_{\ell i}} \quad ; \quad \ell = 1,2,\cdots,m \text{ and } i = 1,2,\cdots,C \tag{C-2}$$

where

$$d_{\ell i} = \frac{q_i \alpha_{\ell i} e_{i\ell}}{\sum_{r=1}^{m} \alpha_{ri} e_{ir}} \tag{C-3}$$

C.3 THE LABELS OF THE GIVEN PATTERN SET ARE IMPERFECT

The equation (6-16) can be written in terms of a priori probabilities of the imperfectly labeled classes as

$$\hat{P}_S = \sum_{j=1}^{C} q_j' \sum_{r=1}^{m} \sum_{i=1}^{C} \alpha_{ri} \nu_{ij} e_{jr} \tag{C-4}$$

Taking log of equation (C-4) and using in equality (2-9) yields

$$\begin{aligned} \hat{P}_S' &= \log (\hat{P}_S) \\ &> \sum_{j=1}^{C} q_i' \log \left(\sum_{r=1}^{m} \sum_{i=1}^{C} \alpha_{ri} \nu_{ij} e_{jr} \right) \end{aligned} \tag{C-5}$$

The probabilities α_{ri} that maximize the lower bound of equation (C-5) subject to the constraints of equation (3-7) can easily be shown to satisfy the following.

$$\alpha_{\ell i} = \frac{\alpha_{\ell i} d_{\ell i}}{\sum_{i=1}^{C} \alpha_{\ell i} d_{\ell i}} \quad ; \quad \ell = 1,2,\cdots,m \text{ and } i = 1,2,\cdots,C \tag{C-6}$$

where

$$d_{\ell i} = \sum_{j=1}^{C} q_j' \frac{\nu_{ij} e_{j\ell}}{\sum_{r=1}^{m} \sum_{i=1}^{C} \alpha_{ri} \nu_{ij} e_{jr}} \tag{C-7}$$

C.3 EXPERIMENTAL RESULTS

Some simulation results are presented in this section in estimating the proportion of the class of interest using the schemes of sections C.1 and C.2. The a priori probabilities q_i in equation (C-3) are estimated from the given labeled patterns for use in the fixed point iteration scheme of equation (C-2). The same labeled patterns and the cluster statistics of section 7 are used. The proportion of the class of interest, class 1, is estimated using equation (3-13). The fixed point iteration scheme of equation (C-6) is used to obtain probabilities of class labels to the clusters by taking into account the imperfections in the labels. The proportion of the class of interest, class 1, is estimated using equation (3-13). The results are listed in table C-1. From table C-1, it is seen that the better estimates are obtained by taking the imperfections in the labels into account.

TABLE C-1.- ESTIMATION OF PROPORTION OF CLASS 1 WITH THE CRITERION OF PROBABILITY OF CORRECT LABELING

Segment	Location	Number of labeled patterns	Fixed point iteration scheme of equation (C-2)		ν-matrix $\nu = (\beta^T)^{-1}$ computed comparing AI and GT labels $\begin{bmatrix} \nu_{11} & \nu_{12} \\ \nu_{21} & \nu_{22} \end{bmatrix}$	With imperfections initial $\alpha_{r\ell} = 0.5$ estimation with the scheme of equation (C-6)	No. of features(d) and clusters(m)	GT proportion
			AI labels	GT labels				
1005	Sherman, Texas	97	0.2759	0.3764	$\begin{bmatrix} 1.802 & -0.0487 \\ -0.8018 & 1.0490 \end{bmatrix}$	0.3129	d = 8, m = 11	0.348
1960	Cheyenne, Colorado	106	0.2171	0.2171	$\begin{bmatrix} 4.220 & -0.1711 \\ -3.220 & 1.1710 \end{bmatrix}$	0.2171	d = 4, m = 11	0.231
1231	Jackson, Oklahoma	96	0.6672	0.7325	$\begin{bmatrix} 1.086 & -0.1628 \\ -0.08551 & 1.1163 \end{bmatrix}$	0.6976	d = 6, m = 12	0.744
1520	Big Stone, Montana	91	0.2528	0.3020	$\begin{bmatrix} 1.268 & -0.01918 \\ -0.2682 & 1.0180 \end{bmatrix}$	0.2892	d = 6, m = 10	0.301
1604	Renville, N. Dakota	101	0.3448	0.5124	$\begin{bmatrix} 2.782 & -0.5177 \\ -1.782 & 1.518 \end{bmatrix}$	0.4218	d = 4, m = 7	0.524
1675	Mcpherson, S. Dakota	107	0.1970	0.2411	$\begin{bmatrix} 4.220 & -0.1711 \\ -3.220 & 1.1710 \end{bmatrix}$	0.2340	d = 8, m = 12	0.291
1805	Gregory, S. Dakota	144	0.1167	0.1918	$\begin{bmatrix} 2.622 & -0.1793 \\ -1.624 & 1.179 \end{bmatrix}$	0.2527	d = 8, m = 12	0.164
1853	Ness, Kansas	91	0.2860	0.2860	$\begin{bmatrix} 1.258 & -0.08247 \\ -0.2577 & 1.0820 \end{bmatrix}$	0.2917	d = 6, m = 12	0.306
1899	Walsh, N. Dakota	95	0.6319	0.7437	$\begin{bmatrix} 1.221 & -0.4069 \\ -0.2207 & 1.4070 \end{bmatrix}$	0.6314	d = 8, m = 7	0.596
Bias			.57289E-01	-.10889E-01		.28511E-01		
MSE			.64986E-02	.29717E-02		.43996E-02		

APPENDIX D

PROPORTION ESTIMATION WITH IMPURE CLUSTERS

Unsupervised classification or clustering algorithms are frequently used in the estimation of proportion of classes of interest in the remotely sensed imagery data. Based on extensive practical studies, it is observed that the clusters produced by many clustering algorithms are impure. That is, they contain more than one class. In the previous sections, approaches were presented for proportion estimation through cluster labeling assuming the clusters are pure.

This appendix considers the problem of estimating the proportion of classes of interest with impure clusters. Let ω, Ω, and Φ denote class, cluster, and mode, respectively. It is assumed that a set of labeled patterns $X_i(j) \in \omega = i$ where $j = 1,2,\cdots,N_i$ and $i = 1,2,\cdots,C$ are given. It is also assumed that the cluster proportions and the cluster densities are given. Probability of correct labeling is used as a criterion. An estimate of the probability of correct labeling for particular class-conditional densities can be written as

$$\hat{P}_S = \left(\sum_{i=1}^{C} q_i \left\{ \frac{1}{N_i} \sum_{\ell=1}^{N_i} p[\omega = i | X_i(\ell)] \right\} \right) \tag{D-1}$$

In the following, it is assumed that the number of inherent modes in the data is given, and for simplicity, it is assumed to be equal to the number of clusters. Equation (3-5) can be written with respect to the modes as

$$p(\omega = i|X) = \sum_{\ell=1}^{m} P(\omega = i|\Phi = \ell)\ p(\Phi = \ell|X) \tag{D-2}$$

Using equation (D-2) in equation (D-1) yields

$$\hat{P}_S = \sum_{i=1}^{C} q_i \sum_{\ell=1}^{m} n_{\ell i} e_{i\ell} \tag{D-3}$$

where $$\eta_{\ell i} = P(\omega = i | \Phi = \ell)$$

and $$e_{i\ell} = \frac{1}{N_i} \sum_{r=1}^{N_i} p[\Phi = \ell | X_i(r)] \qquad \text{(D-4)}$$

Similar to equation (3-4), a relationship between the cluster and class-conditional densities can be written as

$$p(X|\Omega = \ell) = \sum_{i=1}^{C} p(X, \omega = i | \Omega = \ell)$$

$$= \sum_{i=1}^{C} p(X|\omega = i, \Omega = \ell)\, P(\omega = i | \Omega = \ell)$$

$$= \sum_{i=1}^{C} p(X|\omega = i) P(\omega = i | \Omega = \ell) \qquad \text{(D-5)}$$

where it is assumed that

$$p(X|\omega = i) = p(X|\omega = i, \Omega = \ell) \qquad \text{(D-6)}$$

The probabilities $P(\Omega = \ell | \omega = i)$ can be estimated using the given labeled patterns as follows.

$$P(\Omega = \ell | \omega = i) = \int p(\Omega = \ell | X) p(X|\omega = i) dX$$

$$\approx \frac{1}{N_i} \sum_{r=1}^{N_i} p[\Omega = \ell | X_i(r)] \qquad \text{(D-7)}$$

From the Bayes rule we obtain

$$\alpha_{\ell i} = P(\omega = i | \Omega = \ell) = \frac{q_i}{\delta_\ell} P(\Omega = \ell | \omega = i) \qquad \text{(D-8)}$$

Let $$\zeta_{i\ell} = P(\Omega = \ell | \omega = i) \tag{D-9}$$

Using equations (D-8) and (D-9) in equation (D-5) yields

$$p(\Omega = \ell | X) = \sum_{r=1}^{m} p(\Phi = r|X) \left(\sum_{i=1}^{C} \zeta_{i\ell} \eta_{ri} \right) \; ; \quad \ell = 1,2,\cdots,m \tag{D-10}$$

For a given η_{ri} and estimated $\zeta_{i\ell}$, inverting equation (D-10) yields $p(\Phi = r|X)$. The mode proportions $\psi_\ell = P(\Phi = \ell)$ can be estimated from $p(\Phi = r|X)$ as follows.

$$\begin{aligned} \psi_\ell &= P(\Phi = \ell) \\ &= \int p(\Phi = \ell | X) p(X) dX \end{aligned} \tag{D-11}$$

Using the given patterns, ψ_ℓ can be estimated from equation (D-11). Now the problem of estimating the proportions may be formulated as follows.

Find: q_i and $\eta_{\ell i}$; $\ell = 1,2,\cdots,m$ and $i = 1,2,\cdots,C$ such that $\hat{P}_S$ of equation (D-3) is maximized subject to the constraints

$$\left.\begin{aligned} & q_i > 0 \; ; \quad i = 1,2,\cdots,C \\ & \eta_{\ell i} > 0 \; ; \quad \ell = 1,2,\cdots,m \\ & \qquad\qquad\;\; i = 1,2,\cdots,C \\ & \sum_{i=1}^{C} q_i = 1 \\ & \sum_{i=1}^{C} \eta_{\ell i} = 1 \; ; \quad \ell = 1,2,\cdots,m \\ & \sum_{\ell=1}^{m} \eta_{\ell i} \psi_\ell = q_i \; ; \quad \ell = 1,2,\cdots,m \end{aligned}\right\} \tag{D-12}$$

Optimization techniques such as Davidon-Fletcher-Powell (refs. 9, 10, 11) can easily be used to solve the above optimization problem.

NASA-JSC