US Army Corps
of Engineers
Construction Engineering
Research Laboratories

USACERL Technical Report EC-94/21
June 1994

A Multiscale Random Field Model for Bayesian Image Segmentation

by
Michael Shapiro, Charles Bouman, and Calvin F. Bagley

The Army's Land Condition Trend Analysis (LCTA) program collects both space-based remotely-sensed data and ground-level data for natural resource inventory and evaluation. Coupling remotely sensed digital data with traditional ecological ground data could help Army land managers inventory and monitor natural resources. This study used LCTA data sets to test image segmentation algorithms that may be used to interpret remotely sensed digital data.

Many approaches to Bayesian image segmentation have used maximum a posteriori (MAP) estimation in conjunction with Markov random fields (MRF). This study developed a new approach to Bayesian image segmentation that replaces the MRF model with a novel multiscale random field (MSRF), and replaces the MAP estimator with a sequential MAP (SMAP) estimator derived from a novel estimation criteria. Together, the proposed estimator and model result in a noniterative segmentation algorithm that can be computed in time proportional to MN, where M is the number of classes and N is the number of pixels.

Results show that the SMAP algorithm improves classification accuracy when applied to the segmentation of multispectral remotely sensed images with ground-truthed data, especially with the introduction of mixed signatures to represent the spectral information about the classes.

The contents of this report are not to be used for advertising, publication, or promotional purposes. Citation of trade names does not constitute an official endorsement or approval of the use of such commercial products. The findings of this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

DESTROY THIS REPORT WHEN IT IS NO LONGER NEEDED

DO NOT RETURN IT TO THE ORIGINATOR

USER EVALUATION OF REPORT

REFERENCE: USACERL Technical Report EC-94/21, *A Multiscale Random Field Model for Bayesian Image Segmentation*

Please take a few minutes to answer the questions below, tear out this sheet, and return it to USACERL. As user of this report, your customer comments will provide USACERL with information essential for improving future reports.

1. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which report will be used.)

_____

_____

_____

2. How, specifically, is the report being used? (Information source, design data or procedure, management procedure, source of ideas, etc.)

_____

_____

3. Has the information in this report led to any quantitative savings as far as manhours/contract dollars saved, operating costs avoided, efficiencies achieved, etc.? If so, please elaborate.

_____

_____

4. What is your evaluation of this report in the following areas?

   a. Presentation:_____

   b. Completeness:_____

   c. Easy to Understand:_____

   d. Easy to Implement:_____

   e. Adequate Reference Material:_____

   f. Relates to Area of Interest:_____

   g. Did the report meet your expectations?_____

   h. Does the report raise unanswered questions?_____

i. General Comments. (:e what you think should be changed to make this report and future reports of this type more resp .ve to your needs, more usable, improve readability, etc.)

_____

_____

_____

_____

_____

5. If you would like to be contacted by the personnel who prepared this report to raise specific questions or discuss the topic, please fill in the following information.

Name: _____

Telephone Number: _____

Organization Address: _____

_____

_____

6. Please mail the completed form to:

Department of the Army
CONSTRUCTION ENGINEERING RESEARCH LABORATORIES
ATTN: CECER-IMT
P.O. Box 9005
Champaign, IL 61826-9005

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1994	Final

4. TITLE AND SUBTITLE
A Multiscale Random Field Model for Bayesian Image Segmentation

5. FUNDING NUMBERS
4A162720
A896
NN-TS2
DACA
8890-D0029

6. AUTHOR(S)
Michael Shapiro, Charles Bouman, and Calvin F. Bagley

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Construction Engineering Research Laboratories (USACERL)
P.O. Box 9005
Champaign, IL 61826-9005

8. PERFORMING ORGANIZATION REPORT NUMBER
TR EC-94/21

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Assistant Chief of Staff for Installation Management (ACSIM)
ATTN: DAIM-ED-N
600 Army Pentagon
Washington, DC 20310-0600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Copies are available from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The Army's Land Condition Trend Analysis (LCTA) program collects both space-based remotely-sensed data and ground-level data for natural resource inventory and evaluation. Coupling remotely sensed digital data with traditional ecological ground data could help Army land managers inventory and monitor natural resources. This study used LCTA data sets to test image segmentation algorithms that may be used to interpret remotely sensed digital data.

Many approaches to Bayesian image segmentation have used maximum a posteriori (MAP) estimation in conjunction with Markov random fields (MRF). This study developed a new approach to Bayesian image segmentation that replaces the MRF model with a novel multiscale random field (MSRF), and replaces the MAP estimator with a sequential MAP (SMAP) estimator derived from a novel estimation criteria. Together, the proposed estimator and model result in a noniterative segmentation algorithm that can be computed in time proportional to MN, where M is the number of classes and N is the number of pixels.

Results show that the SMAP algorithm improves classification accuracy when applied to the segmentation of multispectral remotely sensed images with ground-truthed data, especially with the introduction of mixed signatures to represent the spectral information about the classes.

14. SUBJECT TERMS
Land Condition Trend Analysis (LCTA) Multiscale Random Field (MSRF)
Bayesian image segmentation
GRASS

15. NUMBER OF PAGES
52

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

FOREWORD

This study was conducted for the Office of the Assistant Chief of Staff for Installation Management (OACSIM) under Project 4A162720A896, "Environmental Quality Technology"; Work Unit NN-TS2, "Imagery Data for Training Area Management." The technical monitor was Dr. Victor Diersing, DAIM-ED-N.

The work was performed by Environmental Compliance and Modeling Division (EC), Environmental Sustainment Laboratory (EL), U.S. Army Construction Engineering Research Laboratories (USACERL). Part of the work for this project was conducted by the School of Electrical Engineering at Purdue University, Lafayette, IN, under contract DACA-8890-D0029, "Perform Design and Development of Image Clustering and Segmentation of Remotely-Sensed Multispectral Images Using Contextual Information." The USACERL principal investigator was Michael Shapiro. Calvin Bagley, CECER-ENR, of the USACERL Environmental Natural Resources Division (USACERL-EN), and Mark Johnson, CECER-ECS, were associate investigators. Special gratitude is owed to Dr. David Tazik and Gary Senseman, CECER-EN, for specification of requirements and testing of the product described in this report. John Bandy is Chief, CECER-EC and David Joncich is Acting Chief, CECER-FL. The USACERL technical editor was William J. Wolfe, Information Management Office.

LTC David J. Rehbein is Commander and Acting Director, USACERL. Dr. Michael J. O'Connor is Technical Director.

CONTENTS

A MULTISCALE RANDOM FIELD MODEL FOR BAYESIAN IMAGE SEGMENTATION

1 INTRODUCTION

Background

Many approaches to Bayesian image segmentation have used maximum a posteriori (MAP) estimation in conjunction with Markov random fields (MRF). This approach performs well, but has a number of disadvantages: exact MAP estimates cannot be computed; approximate MAP estimates are computationally expensive to calculate; and unsupervised parameter estimation of the MRF is difficult.

This study developed a new approach to Bayesian image segmentation that directly addresses these problems. The new method replaces the MRF model with a novel multiscale random field (MSRF), and replaces the MAP estimator with a sequential MAP (SMAP) estimator derived from a novel estimation criteria. Together, the proposed estimator and model result in a noniterative segmentation algorithm that can be computed in time proportional to MN where M is the number of classes and N is the number of pixels. The classes are modeled as a Gaussian mixture model. Given sample pixels for each class, an unsupervised, computationally efficient method was developed both for estimating the number of subclasses in each class as well as the parameters for each subclass.

This multispectral SMAP algorithm and the multimodal mixed signature model described in this report have been implemented in version 4.1 of the Geographic Resources Analysis Support System (GRASS). GRASS is a public domain geographic information system (GIS) and image processing system originally developed by researchers at the U.S. Army Construction Engineering Research Laboratories (USACERL), in Champaign, IL. GRASS is used to input, display, analyze, and output geographic data by users in both military and nonmilitary, and public and private agencies, based all over the world. GRASS is also a component of the Integrated Training Area Management (ITAM) program implemented by USACERL to assist in the management of Army training lands.

The ITAM program elements include inventory and monitoring of Army installation natural resources, integration of land management and training missions, training area rehabilitation, and

environmental awareness training for military personnel. The collection of ground-level data for natural resource inventory and evaluation is conducted through the Land Condition Trend Analysis (LCTA) element of the ITAM program (Tazik et al. 1992). Both space-based remotely-sensed data and ground-level data are collected for Army installations where the ITAM program is implemented. A responsibility to inventory and monitor installation natural resources has led to the development of the LCTA program. Research to support the LCTA program helps develop and apply available technologies to enhance natural resource stewardship on U.S. Army lands. Coupling remotely-sensed digital data with traditional ecological ground data has been identified as a technology with a potential to improve land managers' capabilities to inventory and monitor Army installation natural resources. LCTA data sets were used to test the image segmentation algorithms described in this report.

Objective

The objective of this study was to develop a new approach for processing remotely-sensed imagery using an advanced image segmentation algorithm that contextually smooths or filters digital data so that the classified polygons better model the ecological units on the ground. This algorithm was designed as an analysis tool to improve the ability to use ground-truthed LCTA data to verify remotely sensed electronic data.

Approach

Literature on existing approaches to Bayesian image segmentation was reviewed. A new contextual-based multispectral image processing procedure was defined, developed, and implemented in the GRASS using a multimodal mixed Gaussian distribution to represent spectral signatures for classes. The accuracy of the procedure was tested with data for areas where both ground-truth data and multispectral remotely-sensed imagery were already available.

Scope

The algorithms developed in this study and programs that implement these algorithms are part of the 4.1 release of GRASS, and may not be accurate for post-4.1 releases of GRASS.

Mode of Technology Transfer

The information in this report will be transferred to the field through its implementation in GRASS. GRASS technology is disseminated through training programs, Cooperative Research and Development Agreements, User Conferences, and the *GRASSClippings* quarterly newsletter, the GRASS Support Center, extensive documentation, institutional structures at Army and Interagency levels, communication networks, among other forums.

User feedback on the algorithms described in this report and their implementation in GRASS is desirable and important to the development of GRASS. Users are encouraged to provide such feedback to the GRASS development staff at USACERL via electronic mail grassbug@zorro.cecer.army.mil, and through the GRASS Information Center, U.S. Army Construction Engineering Research Laboratories (USACERL), P.O. Box 9005, Champaign, IL 61826-9005, Phone 217/373-7220, FAX 217/373-7222.

2 TECHNICAL OVERVIEW OF BAYESIAN IMAGE SEGMENTATION

Haralick and Shapiro (1985) have suggested that a good segmentation of an image should separate the image into simple regions with homogeneous behavior. In recent years, many authors have used Bayesian estimation techniques as a framework for computing segmentations that best compromise between these two opposing objectives (Therrien 1983; Besag 1986; Derin and Elliott 1987). These methods model the shape of segmented regions in addition to the behavior of pixels in each homogeneous region. The segmentation is then computed by estimating the best label for each pixel.

A number of estimation techniques and region models have been used for the Bayesian segmentation problem. Typically, the labels of image pixels are modeled as a Markov random field (MRF) or equivalent Gibbs distributions (Besag 1974). These models are used because they only require the specification of spatially local interactions using a set of local parameters. This is important since spatially local interactions result in segmentation algorithms that only require local computations. Most often, the image is then segmented by approximately computing the maximum a posteriori (MAP) estimate of the pixel labels.

These statistical approaches to segmentation provide an important framework, and have improved results in the application of segmentation to natural scenes (Pappas 1972), tomographic cross sections (Sauer and Bouman 1993), texture images (Therrien 1983), and multispectral remotely sensed images (Zhang, Haralick, and Campbell 1990; Jeon and Landgrebe 1991; Kato, Zerubia, and Berthod 1992; and Jeon and Landgrebe 1993). However, the approach has a number of important disadvantages.

Computing the MAP estimate requires the minimization of a discrete functional with many local minima. Exact minimization is intractable, so the true MAP estimate must be approximately minimized. Methods for approximation include simulated annealing (Geman and Geman 1984), greedy minimization (Besag 1986), dynamic programming (Derin and Elliott 1987), and multiresolution minimization (Gidas 1989; Bouman and Liu 1988; Bouman and Liu 1989; and Perez and Heitz 1992). However, all of these approaches require approximations in the two dimensional case, and are either iterative or computationally expensive.

The MRF model has a limited ability to describe large-scale behaviors. For example, even though segmented regions are likely to be at least 50 pixels wide, it is difficult to accurately incorporate this

information by specifying the interactions of adjacent pixels. The model can be improved by using a larger neighborhood for each pixel, but this rapidly increases the number of parameters of interaction, and the complexity of the segmentation algorithms. The fundamental limitation of local models is that they do not allow behavior to be directly controlled at different spatial scales. This is of critical importance since scale variation occurs naturally in images, and is important in quantifying image behavior (Pentland 1984; and Peleg et al. 1984).

The MAP estimate does not have desirable properties for the segmentation problem (Marroquin, Mitter, and Poggio 1987; Dubes et al. 1990). The MAP estimate minimizes the probability that any pixel in the image will be misclassified. This is an excessively conservative criteria since any segmentation algorithm is likely to result in some misclassified pixels. In practice, it has been noted that MAP estimation has some undesi ıle global properties that may actually make an approximate minimization more desirable (Besag 1986; Dubes et al. 1990). For example, in multiresolution segmentation, MRF correlations parameters were found to *increase* at coarser scales (Bouman and Liu, 1988, 1991). This is counter to the physical intuition that coarser sampling should produce less correlation.

The maximizer of the posterior marginals (MPM) estimator has been suggested as an alternative to MAP estimation, since it minimizes the probability of classification error (Marroquin, Mitter, and Poggio 1987). However, it may only be approximately calculated in a computationally expensive procedure similar to simulated annealing. Also, the MPM criteria does not consider the spatial placement of errors when distinguishing among the quality of segmentations.

Finally, parameter estimation of MRFs is difficult. When parameters are above the "critical temperature," there may be no consistent estimator as the image size grows to infinity (Pickard 1987). Methods have been developed to estimate MRF parameters from images being segmented (Lakshmanan and Derin (1989), but they are computationally expensive.

This study addressed these difficulties by introducing a new approach to Bayesian image segmentation. This method replaces the MRF model with a novel multiscale random field (MSRF), and replaces the MAP estimator with a sequential MAP (SMAP) estimator derived from a new estimation criteria. Together, the proposed estimator and model result in a noniterative segmentation algorithm that can be computed in time proportional to MN where M is the number of classes and N is the number of pixels. A method for estimating the parameters of the MSRF model directly from the image during the

segmentation process is also developed. This allows images with differing region sizes to be segmented accurately without specific prior knowledge of their behavior.

This MSRF model is composed of a series of random fields progressing from coarse to fine scale. Each field is assumed to only depend on the previous coarser field. Therefore, the series of fields form a Markov chain in scale or resolution. Further, this assumes that points in each field are conditionally independent given their coarser scale neighbors. This leads to a rich model with computationally tractable properties. In fact, Luettgen, Karl, Willsky and Tenney (1993) have shown in independent work that models similar to the MSRF actually include MRFs as a subclass.

Earlier work (Chou et al 1989; Chou, Golden, and Willsky 1991; Basseville et al. 1992; Basseville, Benveniste, and Willsky 1992a, 1992b) has shown that Markov Chains in scale can be used to model continuously valued Gaussian processes in one and two dimensions. This work has resulted in fast algorithms for problems such as optical flow estimation (Chou et al. 1991). Estimation for these models is performed using a generalization of Kalman filtering (Basseville et al. 1992; Basseville, Benveniste, and Willsky 1992a, 1992b). This approach is ideal for Gaussian models since the MAP, conditional mean, and minimum mean squared estimates coincide, and may be computed using only recursively computed first and second order statistics. However, since this model requires discrete values to represent pixel classes, these methods are not applicable.

The MSRF model has a number of advantages over fixed scale MRFs. The Markov chain structure facilitates straightforward methods for parameter estimation since it eliminates difficulties with intractable normalizing constants (partition functions) found in MRFs. Yet the model does not impose an unnatural spatial ordering to the pixels since the Markov chain is in scale. Also, since explicit parameters are available to control both coarse and fine scale behavior, the MSRF model can more accurately describe image behavior.

The SMAP estimation method results from minimizing the expected size of the largest misclassified region. This is accomplished by assigning progressively larger cost to errors at a coarser scale. Intuitively, the criteria accounts for the fact that an error at coarse scale is more serious since it causes the misclassification of many pixels. The SMAP criteria results in a series of optimization steps going from a coarse to fine scale. At each scale, the best segmentation is computed given the previous coarser segmentation and the observed data. Each maximization step is computationally simple and noniterative if the region parameters are known. The complete procedure is similar to pyramidal pixel linking

procedures (Burt, Hong, and Rosenfeld 1981; Antonisse 1982), but requires local computations much like those used in Bayesian networks (Pearl 1988).

Unknown region parameters may be estimated using an iterative procedure at each scale. This iterative procedure, based on the expectation maximization (EM) algorithm (Dempster, Laird, and Rubin 1977), is implemented by subsampling the image. Therefore, parameter estimation only increases the required computation by approximately a factor of two.

3 MULTISCALE SEGMENTATION APPROACH

The random field Y is the image that must be segmented into regions of distinct statistical behavior. (Upper case letters denote random quantities, while lower case letters denote the corresponding deterministic realizations.) Individual pixels in Y are denoted by Y_s where s is a member of a two dimensional lattice of points S.

The basis of this segmentation approach is a hierarchical or doubly stochastic model (Figure 1). This model assumes that the behavior of each observed pixel depends on a corresponding unobserved label pixel in X. Each label specifies one of M possible classes, each with its own statistical behavior. The dependence of observed pixels on their labels is specified through $p_{y|x}(y|x)$, the conditional distribution of Y given X. Prior knowledge about the size and shapes of regions will be modeled by the prior distribution $p(x)$.

Since a variety of features can be used with this approach, it is a general framework for the segmentation problem. For the texture segmentation problem, a stochastic texture model can be used for $p_{y|x}(y|x)$ (Derin and Elliott 1987; Manjunath, Simchony, and Chellappa 1990; Bouman and Liu 1991), or texture feature vectors can be extracted at each pixel (Laws 1980; Unser and Eden 1989, 1990) and modeled with a multivariate distribution. However, segmentation of multispectral remotely sensed images will serve as the target application for the examples. In this case, a multivariate Gaussian mixture distribution will be used to describe the behavior of the spectral components of the image. Appendix D describes the method used for estimating the parameters for this mixture model.

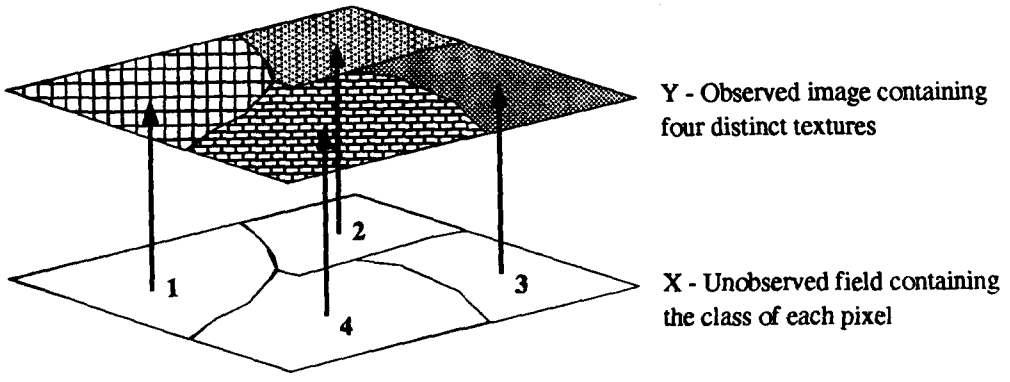


Figure 1. Structure of a Doubly Stochastic Random Field Used in Segmentation.

The following sections describe the general structure of a MSRF model for $p(x)$ and develop a sequential MAP estimation approach for computing the best segmentation. Detailed models and recursion formulas resulting from this framework are then derived in Chapter 4.

Multiscale Random Field Model

The multiscale Random Field (MSRF) model is composed of a series of random fields at varying scales or resolutions. Figure 2 shows the pyramid structure of the MSRF. At each scale, n, the segmentation or labeling is denoted by the random field $X^{(n)}$, and the set of lattice points is denoted by $S^{(n)}$. In particular, $X^{(0)}$ is assumed to be the finest scale random field with each point corresponding to a single image pixel. Each label at the next coarser scale, $X^{(1)}$, then corresponds to a group of four points in the original image. Therefore, the number of points in $S^{(1)}$ is one-fourth the number of points in $S^{(0)}$.

The fundamental assumption of the MSRF model is that the sequence of random fields from coarse to fine scale form a Markov chain. Therefore, the distribution of $X^{(n)}$ given all coarser scale fields only depends on $X^{(n+1)}$. This is a reasonable assumption since $X^{(n+1)}$ should contain all the relevant information from previous coarser scales. Formally, this Markov chain relation may be stated as:

$$P(X^{(n)} = x^{(n)} \mid X^{(l)} = x^{(l)}\, l > n) = P(X^{(n)} = x^{(n)} \mid X^{(n+1)} = x^{(n+1)}) = p_{x^{(n)} \mid x^{(n+1)}}(x^{(n)} \mid x^{(n+1)}) \qquad \text{[Eq 1]}$$

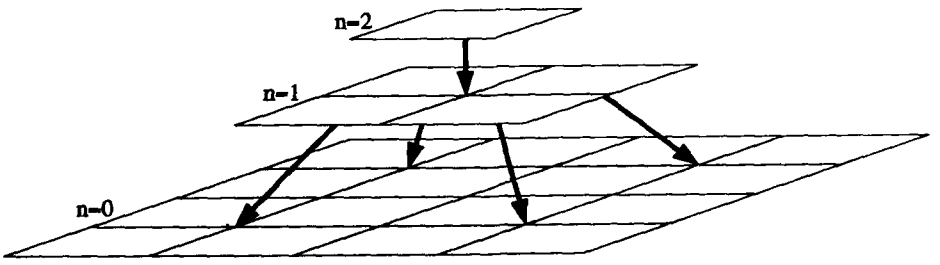


Figure 2. Pyramid Structure of the MSRF.

Correspondingly, the exclusive dependence of Y on $X^{(0)}$ implies that:

$$P(Y \in dy \mid X^{(n)} n > 0) = P(Y \in dy \mid X^{(0)}) = p_{y|x^{(0)}}(y \mid x^{(0)}) \qquad \text{[Eq 2]}$$

The joint distribution of X and Y may then be expressed as the product of these distributions:

$$P(Y \in dy, X = x) = p_{y|x^{(0)}}(y|x^{(0)}) \left\{ \prod_{n=0}^{L-1} p_{x^{(n)}|x^{(n+1)}}(x^{(n)}|x^{(n+1)}) \right\} p_{x^{(L)}}(x^{(L)}) \qquad \text{[Eq]}$$

where L is the coarsest scale in X. This Markov structure in scale has the isotropic behavior associated with MRFs, but in addition, the causal dependence in scale results in a noniterative segmentation algorithm and direct methods of parameter estimation.

Sequential MAP Estimation

To segment the image, Y, the pixel labels in X must be accurately estimated. Bayesian estimation techniques are the natural approach since the existence of a prior distribution $p(x)$ is assumed. Generally Bayesian estimators attempt to minimize the average cost of an erroneous segmentation by solving the optimization problem:

$$\hat{x} = \arg\min_{x} E[C(X, x) \mid Y = y] \qquad \text{[Eq 4]}$$

where $C(X,x)$ is the cost of estimating the true segmentation, X, by the approximate segmentation, x. Notice that X is a random quantity whereas x is a deterministic argument. Of course, the choice of the functional, C(- , -), is of critical importance since it determines the relative importance of errors.

A look at the assumptions of its derivation will help explain the deficiencies of the MAP estimate. The MAP estimate is the solution to [Eq 4] when the cost functional is given by:

$$C_{MAP}(X,x) = 1 - \delta(X-x) \qquad \text{[Eq 5]}$$

where $\delta(X-x)$ is 1 when X = x and 0 otherwise. Since $C_{MAP}(X,x) = 1$ whenever any pixel is incorrectly labeled, the MAP estimate maximizes the probability that *all* pixels will be correctly labeled. Of course, a segmentation need not be completely accurate at all pixels to be useful. Even good segmentations will normally have erroneously classified pixels along region boundaries. This is particularly true in high resolution images where the misclassification of a single pixel is not significant. Therefore, the MAP estimate can be excessively conservative (Marroquin, Mitter, and Poggio 1987; Dubes et al. 1990).

The implications of the MAP criteria appear even more inappropriate for the estimation of the MSRF introduced in the previous sections. The cost function used for MAP estimation of a MSRF is:

$$\begin{aligned} C_{MAP}(X,x) &= 1 - \delta(X-x) \\ &= 1 - \prod_{n=0}^{L} \delta(X^{(n)} - x^{(n)}) \end{aligned} \qquad \text{[Eq 6]}$$

This cost function is 1 if a labeling error occurs at any scale, n, of the segmentation. Consequently, this function assigns equal cost to a single mislabeled pixel at $n=0$ or the mislabeling of approximately 256 pixels at $n=4$. This cost assignment is clearly undesirable.

Ideally, a desirable cost function should assign progressively greater cost to segmentations with larger regions of misclassified pixels. To achieve this goal, the following alternative cost function is proposed:

$$C_{SMAP}(X,x) = \frac{1}{2} + \sum_{n=0}^{L} 2^{n-1} C_n(X,x) \qquad \text{[Eq 7]}$$

where

$$C_n(X,x) = 1 - \prod_{i=n}^{L} \delta\left(X^{(i)} - x^{(i)}\right) \qquad \text{[Eq 8]}$$

The behavior of C_{SMAP} is solely a function of the coarsest scale, K, that contains a misclassified pixel. More precisely, let K be the unique scale such that $X^{(K)} \neq x^{(K)}$, but $X^{(i)} = x^{(i)}$ for all $i>K$. Then the functions C_n are given by :

$$C_n(X,x) = \begin{cases} 1 \text{ if } n \leq K \\ 0 \text{ if } n > K \end{cases} \qquad \text{[Eq 9]}$$

and the total cost is given by $C_{SMAP}(X, x) = 2^K$. This error at scale K will generally lead to the misclassification of a group of pixels at the finest scale. The width of this misclassified group of pixels will be approximately $2^K = C_{SMAP}(X, x)$. Therefore, the SMAP cost function has the following intuitive interpretation.

$$C_{SMAP}(X,x) \approx \text{width of the largest grouping of misclassified pixels} \qquad \text{[Eq 10]}$$

The estimator that minimizes this proposed cost is determined by evaluating [Eq 4].

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} \mathrm{E}\left[C_{\mathrm{SMAP}}(\mathrm{X},\mathrm{x}) \,|\, \mathrm{Y} = \mathrm{y} \right] \\ &= \arg\min_{\mathbf{x}} \sum_{\mathrm{n}=0}^{\mathrm{L}} 2^{\mathrm{n}-1} \left\{ 1 - \mathrm{P}(\mathrm{X}^{(\mathrm{i})} = \mathrm{x}^{(\mathrm{i})} \; \mathrm{i} \geq \mathrm{n} \,|\, \mathrm{Y} = \mathrm{y}) \right\} \\ &= \arg\max_{\mathbf{x}} \sum_{\mathrm{n}=0}^{\mathrm{L}} 2^{\mathrm{n}} \, \mathrm{P}(\mathrm{X}^{(\mathrm{i})} = \mathrm{x}^{(\mathrm{i})} \mathrm{i} \geq \mathrm{n} \,|\, \mathrm{Y} = \mathrm{y}) \end{aligned} \qquad \text{[Eq 11]}$$

Since the random fields, $X^{(n)}$, form a Markov Chain, this estimate is computed recursively in the scale parameter n is done by assuming that $\hat{x}^{(i)}$ has been computed for $i>n$, and using this result to compute $\hat{x}^{(n)}$. Appendix A shows that this recursive approach yields the following expression for the solution:

$$\hat{x}^{(n)} = \arg\max_{x^{(n)}} \left\{ \log p_{x^{(n)}|x^{(n+1)},y}\left(x^{(n)} | \hat{x}^{(n+1)}, y\right) + \mathcal{E}\left(x^{(n)}\right) \right\} \qquad \text{[Eq 12]}$$

where $\mathscr{E}$ is a second order term which may be bounded by

$$0 \le \mathscr{E}(x^{(n)}) \le \max_{x^{(n-1)}} p_{x^{(n-1)}|x^{(n)},y}(x^{(n-1)}|\hat{x}^{(n)}, y) << 1 \qquad \text{[Eq 13]}$$

The approximation that $\mathscr{E} << 1$ is very good. To see this, notice that $x^{(n-1)}$ is an interpolation of the coarser segmentation, $\hat{x}^{(n)}$, given the image, y. Normally, there will be many pixels in the interpolation $x^{(n-1)}$ for which the correct labeling is uncertain. This is particularly true around the boundaries of objects. Since the number of unique labeling combinations for these pixels is enormous, the probability of any particular combination will be small. In fact for these models, this probability goes to 0 as the number of pixels, N, increases. Therefore:

$$\lim_{N \to \infty} \mathscr{E}(x^{(n)}) = 0 \qquad \text{[Eq 14]}$$

At very coarse scales, the number of labels becomes small, and often only one reasonable interpolation will exist (i.e., $\mathscr{E} \approx 1$). However, in this case the correct labeling of pixels at the coarser scale, n, must also be unambiguous, and any reasonable estimator should have good performance. Ignoring the contribution of $\mathscr{E}$ results in the following recursive equations:

$$\hat{\mathbf{x}}^{(L)} = \arg\max_{\mathbf{x}^{(L)}} \log \mathrm{p}_{\mathbf{x}^{(L)}|\mathbf{y}}(\mathbf{x}^{(L)}|\mathbf{y}) \qquad \text{[Eq 15]}$$

$$\hat{\mathbf{x}}^{(n)} = \arg\max_{\mathbf{x}^{(n)}} \log \mathrm{p}_{\mathbf{x}^{(n)}|\mathbf{x}^{(n+1)},\mathbf{y}}(\mathbf{x}^{(n)}|\hat{\mathbf{x}}^{(n+1)},\mathbf{y}) \qquad \text{[Eq 16]}$$

The recursion is initialized by determining the MAP estimate of the coarsest scale field given the observed data. The segmentation at each finer scale is then found by computing the MAP estimate of $X^{(n)}$ given $\hat{x}^{(n+1)}$ and the image, y. Due to this structure, this estimator is referred to as a sequential MAP (SMAP) estimator.

By using Bayes rule, the Markov properties of X, and assuming that $X^{(L)}$ is uniformly distributed, the SMAP recursion may be rewritten in a form that is more easily computed:

$$\hat{x}^{(L)} = \arg\max_{x^{(n)}} \log p_{y|x^{(L)}}(y|x^{(L)}) \qquad \text{[Eq 17]}$$

$$\hat{x}^{(n)} = \arg\max_{x^{(n)}} \left\{ \log p_{y|x^{(n)}}(y \mid x^{(n)}) + \log p_{x^{(n)}|x^{(n+1)}}(x^{(n)} \mid \hat{x}^{(n+1)}) \right\} \qquad \text{[Eq 18]}$$

The two terms in [Eq 18] play roles analogous to those of the likelihood function and prior distribution in conventional Bayesian estimation. The first term of the maximization gives the likelihood of the observed data y given the labeling at scale n. The second term of the maximization embodies the a priori information about the behavior of X. Therefore, this term biases the solution to favor segmentations with large regions and smooth boundaries.

The SMAP estimator has a number of additional advantages. Chapter 4 introduces specific models so that each optimization step of [Eq 18] may be computed with a single noniterative pass. This is in contrast to the MAP and MPM estimators, which require computationally expensive iterative optimization methods (Geman and Geman 1984; Besag 1983; Marroquin, Mitter, and Poggio 1987). Further, the SMAP estimator has a subtle advantage over MPM. The MPM method chooses each pixel individually to minimize the probability of error. However, it does not consider the spatial placement of errors. Since the SMAP method attempts to minimize the spatial size of errors, it will tend to generate a subjectively more desirable segmentation.

4 SEGMENTATION ALGORITHM

The specific models used for Y and X are defined by specifying the conditional density $p_{y|x^{(0)}}(y \mid x^{(0)})$, together with the coarse to fine scale transition densities $p_{x^{(n)}|x^{(n+1)}}(x^{(n)}|x^{(n+1)})$. An adaptive segmentation algorithm that estimates the parameters of the MSRF during the segmentation process can then be developed.

For the multispectral segmentation problem, it is best to use models with observed pixels that are conditionally independent given their class labels. This implies that the spatial texture of regions will not be used as a discriminating feature. Instead, the multispectral characteristics of classes help to discriminate distinct regions. This approach is supported by the fact that spatial correlation has been found to be weak within regions of multispectral images corresponding to a single ground cover (Landgrebe 1980; Kettig and Landgrebe 1976). Using this assumption, the conditional density function for the image has the form:

$$p_{y|x^{(0)}}(y|x^{(0)}) = \prod_{s \in S^{(0)}} p_{y_s|x_s^{(0)}}(y_s|x_s^{(0)}) \qquad \text{[Eq 19]}$$

where $p_{y_s|x_s^{(0)}}(\cdot \mid k)$ is the conditional density function for an individual pixel given the class label k. Since each pixel is composed of D multispectral components, $p_{y_s|x_s^{(0)}}(\cdot \mid k)$ is a multivariate density function. Here a multivariate Gaussian mixture density is used in numerical experiments, but other distributions can just as well be applied.

More generally, the methods used throughout this study are applicable to any model that can be expressed in the form:

$$\log p_{y|x^{(0)}}(y|x^{(0)}) = \sum_{s \in S^{(0)}} l_s(y|x_s^{(0)}) + c(y) \qquad \text{[Eq 20]}$$

where the functions l_s depend on all of y, and c is an arbitrary function of y. This type of model has been used extensively in texture segmentation applications (Derin and Elliott 1987; Derin and Cole 1986; Manjunath, Simchony, and Chellappa 1990; Bauman and Liu 1991).

The choice of models for X was restricted to have two important properties. First, the pixels in $X^{(n)}$ were conditionally independent given the pixels in $X^{(n+1)}$. Second, each pixel $X_s^{(n)}$ was only dependent on a local neighborhood of pixels at the next coarser scale. This set of neighboring locations to s is denoted by ∂s. Based on these properties, the transition distribution from coarse to fine scale takes the form:

$$p_{x^{(n)}|x^{(n+1)}}(x^{(n)} \mid x^{(n+1)}) = \prod_{s \in S^{(n)}} p_{x_s^{(n)}|x_{\partial s}^{(n+1)}}(x_s^{(n)} \mid x_{\partial s}^{(n+1)}) \qquad \text{[Eq 21]}$$

where $p_{x_s^{(n)}|x_{\partial s}^{(n+1)}}$ is the probability density for $x_s^{(n)}$ given its neighbors at the coarser scale, $x_{\partial s}^{(n+1)}$. These two assumptions assure that the pixels in $x^{(n)}$ will be conditionally independent given the pixels in $x^{(n+1)}$. However, note that nearby pixels in $x^{(n)}$ may still be highly dependent since they will share coarser scale neighbors.

The choice of neighborhood ∂s for the pixel s is important since it will define the structure of the multiscale pyramid. Two types of neighborhoods are used. The first corresponds to a quadtree structure and allows simple and exact calculation of the SMAP segmentation. The second neighborhood corresponds to a more general graph structure and can account for more complex interactions that occur across blocks of the quadtree. Therefore, this model produces smoother, less blocky segmentations. Unfortunately, the graph structured pyramid does not allow exact calculation of the SMAP estimator. Therefore, a hybrid model, which incorporates both the quadtree and graph structure, is used to approximate the exact SMAP estimate.

Quadtree Model

The first pyramid structure considered is based on a conventional quadtree. Figure 3a shows the structure of the quadtree, and the one dimensional analog is shown in Figure 3b. Since this is a tree structure, each point in the pyramid depends only on a single point at the coarser scale. This coarser scale neighbor of a point s is called the father of s and will be denoted by the function $d(s)$. In a similar vein, the n^{th} successive father of a point will be denoted $d^n(s)$, and $d^{-n}(s)$ will be the set of all points that occur n levels down from s in the tree structure.

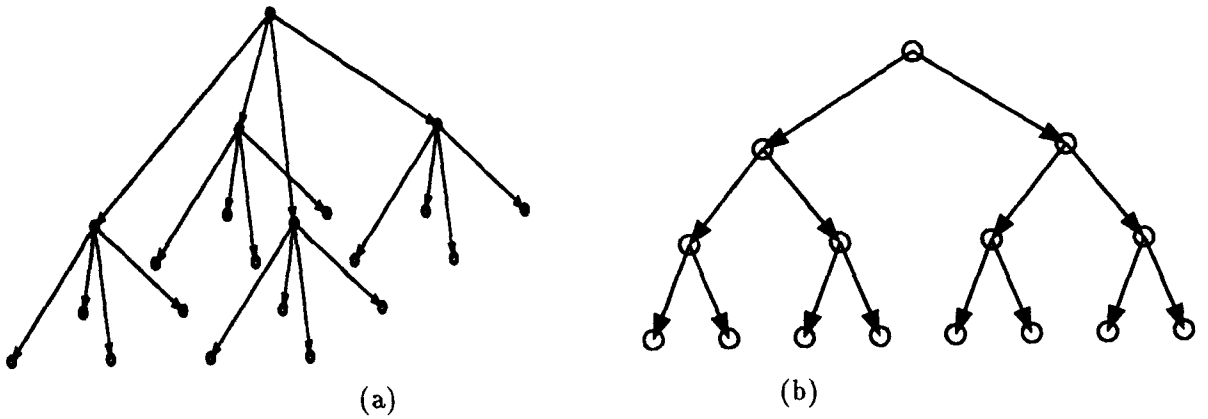


Figure 3. (a) Quadtree Structure Used for MSRF Model; (b) One-Dimensional Analog to Quadtree Structure.

The following transition function models the probability that $X_s^{(n)}$ has class m, given that its father is of class k:

$$p_{x_s^{(n)}|x_{\partial s}^{(n+1)}}(m \mid k) = \theta_{n,0}\delta_{m,k} + \frac{1-\theta_{n,0}}{M} \qquad \text{[Eq 22]}$$

where $\delta_{m,k}$ is the unit sample function and M is the number of possible classes. The parameter $\theta_{n,0} \in [0,1]$ is the probability that the labeling will remain the same from scale $n+1$ to n. If a class change does occur, it is equally likely to be any one of the remaining class types. At fine resolutions, the neighboring pixels are more likely to have the same class. Therefore, $\theta_{n,0}$ will generally be an increasing function of resolution (decreasing function of n). Notice that this distribution only depends on the the scale n through $\theta_{n,0}$ but does not depend on the particular pixel s.

An important property of the quadtree structure is that the conditional distribution of Y given $X^{(n)}$ has a product form that may be computed using a simple fine-to-coarse recursion. Let $y_s^{(n)}$ be the set of leaves of the quadtree that are on the branch starting at $s \in S^{(n)}$. Specifically, $y_s^{(n)} = \{y_r : d^{n}(r) = s\}$. Appendix B shows that the conditional density for Y given $X^{(n)}$ has the product form:

$$p_{y|x^{(n)}}(y \mid x^{(n)}) = \prod_{s \in S^{(n)}} p_{y_s^{(n)}|x_s^{(n)}}(y_s^{(n)} \mid x_s^{(n)}) \qquad \text{[Eq 23]}$$

Furthermore, the density functions $p_{y_s^{(n)}|x_s^{(n)}}$ may be computed using the following recursion where M is the number of class labels:

$$p_{y_s^{(n+1)}|x_s^{(n+1)}}(y_s^{(n+1)}|k) = \prod_{r\in d^{-1}(s)} \sum_{m=1}^{M} P_{y_r^{(n)}|x_r^{(n)}}(y_r^{(n)}|m)\ p_{x_r^{(n)}|x_{\partial r}^{(n+1)}}(m|k) \qquad \text{[Eq 24]}$$

In practice, dynamic range considerations mandate that the logarithm of these functions be computed and stored. Therefore, define the log likelihood function as:

$$l_s^{(n)}(k) \triangleq \log p_{y_s^{(n)}|x_s^{(n)}}(y_s^{(n)}|k) \qquad \text{[Eq 25]}$$

where the dependence on y is suppressed for clarity. Substituting the transition distribution of [Eq 22] into [Eq 24] and converting to log likelihood functions yields the new recursion:

$$l_s^{(0)}(k) = \log p_{y_s|x_s^{(0)}}(y_s|k) \qquad \text{[Eq 26]}$$

$$l_s^{(n+1)}(k) = \sum_{r\in d^{-1}(s)} \log\left\{\theta_{n,0}\exp\{l_r^{(n)}(k)\} + \frac{1-\theta_{n,0}}{M}\sum_{m=1}^{M}\exp l_r^{(n)}(m)\right\} \qquad \text{[Eq 27]}$$

To perform the SMAP segmentation, these log likelihood functions must be computed at every point in the quadtree. The number of points in the pyramid is approximately given by $\frac{4N}{3}$, where N is the number of pixels at the finest scale. Since each of the log likelihood functions may be stored in the form of M numbers, the total required storage is approximately $\frac{(4MN)}{3}$. Because the second sum of [Eq 27] is not a function of k, each log likelihood function can be computed in time proportional to M. Therefore, the total computation time for evaluating the log likelihood functions is $O(MN)$.

Note that in the limiting case of $\theta_{n,D}(k) = 1$, the recursion reduces to simple averaging:

$$l_s^{(n+1)}(k) = \sum_{r \in d^{-1}(s)} l_r^{(n)}(k) \qquad \text{[Eq 28]}$$

This is equivalent to assuming that all the pixels in the block of pixels, $y_s^{(n+1)}$, must have the same label. Of course, this is often not the case since some of the pixel blocks are likely to fall on region boundaries. These blocks then have a mixture of pixels from the two regions. Therefore, simple averaging can cause pixels on region boundaries to be misclassified. This is particularly true if the statistical average of the two regions appears to have characteristics of a third class. Another advantage of the more accurate recursion, [Eq 27], is that a small group of anomalous pixels will not adversely affect the classification of a large region. Simple linear averaging of the log likelihood function tends to be easily biased by a few pixels that are statistical outliers, whereas the more accurate recursion will tolerate such errors.

Once the likelihood functions are computed, the SMAP segmentation may be efficiently computed using [Eq 18]:

$$\hat{x}^{(n)} = \arg\max_{x^{(n)}} \sum_{s \in S^{(n)}} \left\{ l_s^{(n)}(x_s^{(n)}) + \log p_{x_s^{(n)} | \hat{x}_{\partial s}^{(n+1)}}(x_s^{(n)} \mid \hat{x}_{\partial s}^{(n+1)}) \right\} \qquad \text{[Eq 29]}$$

This expression is easily evaluated by minimizing with respect to each pixel individually:

$$\hat{x}_s^{(n)} = \arg\max_{1 \le k \le M} \left\{ l_s^{(n)}(k) + \log p_{x_s^{(n)} | \hat{x}_{\partial s}^{(n+1)}}(k \mid \hat{x}_{\partial s}^{(n+1)}) \right\} \qquad \text{[Eq 30]}$$

This coarse-to-fine segmentation operation requires order *O(MN)* computation time. Therefore, the complete segmentation process consists of a fine-to-coarse and a coarse-to-fine operation, each of which require *O(MN)* computation.

While the quadtree model results in an exact expression for computing the SMAP segmentation, it does not completely capture some aspects of image behavior. The following section introduces an augmented model that improves on the quadtree.

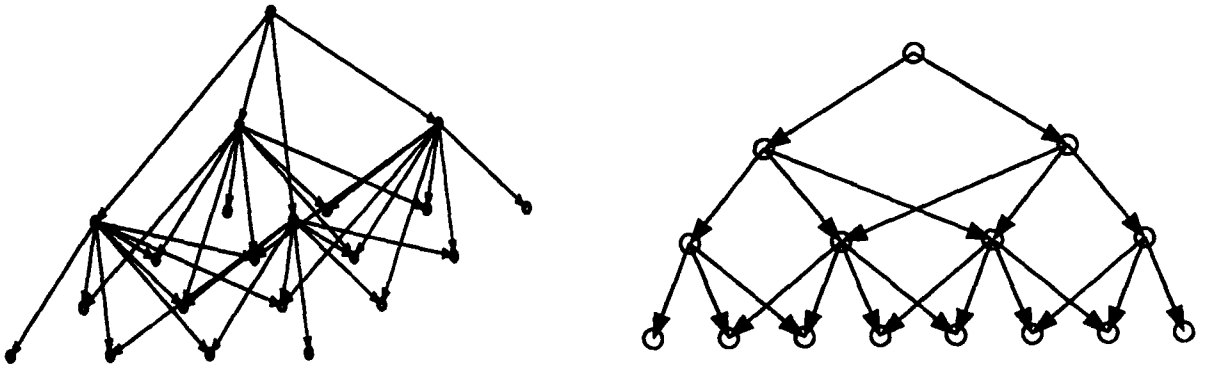
Pyramidal Graph Model

An important disadvantage of the quadtree model is that spatially adjacent pixels may not have common neighbors at the next coarser scale. Therefore, the model does not enforce continuity of region boundaries when they pass across branches of the quadtree. For example, if the image is broken into four quadrants at the second level of the quadtree, then a region boundary will not be constrained to be smooth as it passes from one of these quadrants to the next.

This weakness may be corrected by increasing the number of coarse scale neighbors for each pixel. Figure 4a shows such a pyramidal graph structure and Figure 4b shows an analogous one dimensional graph. Notice that each point has three neighbors at the next coarser scale.

To express the positions of these neighbors, a pixel was explicitly denoted at scale n as $s=(i, j)$, where i and j are the row and column indices starting at 0 and ranging through the width minus one, and the height minus one respectively. The three neighbors at scale $n+1$ may then be computed using the function $odd(i)$=1 if i is odd and -1 if i is even, and the greatest smaller integer function $\lfloor \cdot \rfloor$:

$$\begin{aligned} s_1 &= (\lfloor i/2 \rfloor, \lfloor j/2 \rfloor) \\ s_2 &= (\lfloor i/2 \rfloor, \lfloor j/2 \rfloor) + (odd(i), 0) \\ s_3 &= (\lfloor i/2 \rfloor, \lfloor j/2 \rfloor) + (0, odd(j)) \end{aligned} \qquad \text{[Eq 31]}$$



(a) (b)

Figure 4. (a) Augmented Pyramidal Graph Structure Used for MSRF Model; (b) One-Dimensional Analog of Pyramidal Graph Structure.

The transition function chosen for this pyramid graph is a natural extension of the transition function used for the quadtree based model:

$$\begin{aligned} & \tilde{p}_{x_s^{(n)} | x_{\partial s}^{(n+1)}}(m \mid i,j,k) \\ &= P\left(X_s^{(n)} = m \mid X_{s_1}^{(n+1)} = i, X_{s_2}^{(n+1)} = j, X_{s_3}^{(n+1)} = k\right) \\ &= \frac{\theta_{n,1}}{7}(3\delta_{m,i} + 2\delta_{m,j} + 2\delta_{m,k}) + \frac{1 - \theta_{n,1}}{M} \end{aligned} \qquad \text{[Eq 32]}$$

The notation $\tilde{p}$ is used to distinguish from the transition distribution used for the quadtree model. As with the quadtree model, the parameter $\theta_{n,1} \in [0,1]$ determines the probability that the label of the fine scale point will be the same as one of the coarser scale points. Conversely, $1 - \theta_{n,1}$ is the probability that a new label will be randomly chosen from the available labels.

The disadvantage of the pyramid graph structure is that the likelihood function for the labels does not have a product form as was the case for the quadtree in [Eq 23]. Therefore, there is no simple fine-to-coarse recursion with the form of [Eq 24] for computing the likelihood of the image y given the labels $x^{(n)}$.

For computations at a single scale, n, this problem may be circumvented by assuming that the pyramid has a quadtree structure for scales finer than n, and a graph structure for coarser scales. Figure 5 shows a one-dimensional analog to this hybrid pyramid structure for n=2. Notice that, for levels above n, the pyramid is a graph, but below n the pyramid has a simple tree structure. Using this hybrid pyramid, the conditional likelihood of [Eq 18] then has the computable form:

$$\log p_{y,x^{(n)} | x^{(n+1)}}(y, x^{(n)} \mid \hat{x}^{(n+1)}) = \sum_{s \in S^{(n)}} l_s^{(n)}(x_s^{(n)}) + \log \tilde{p}_{x_s^{(n)} | x_{\partial s}^{(n+1)}}(x_s^{(n)} \mid \hat{x}_{\partial s}^{(n+1)}) \qquad \text{[Eq 33]}$$

which results in the following formula for the SMAP estimate of $X^{(n)}$:

$$\hat{x}_s^{(L)} = \arg \max_{1 \le k \le M} l_s^{(L)}(k) \qquad \text{[Eq 34]}$$

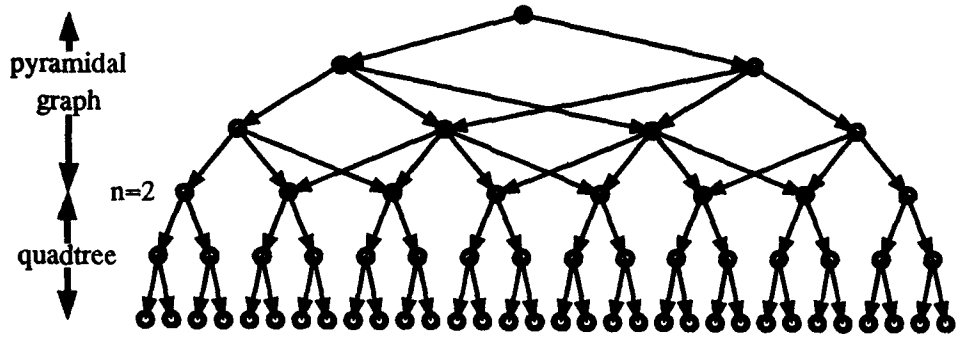


Figure 5. One-Dimensional Analog to Hybrid Graph Structure, Where Scales $n>2$ Use a Pyramidal Graph Structure, and Scales $n\leq 2$ Use a Quadtree Structure.

$$\hat{x}_s^{(n)} = \arg \max_{1 \leq k \leq M} \left\{ l_s^{(n)}(k) + \log \tilde{p}_{x_s^{(n)} | x_{\partial s}^{(n+1)}}(k \mid \hat{x}_{\partial s}^{(n+1)}) \right\} \qquad \text{[Eq 35]}$$

where $\tilde{p}$ is the transition function of [Eq 32], and $l_s^{(n)}(k)$ are computed using the recursion of [Eq 27].

Of course, the application of the above formula at all scales is an approximation to the true SMAP segmentation since the model may not legitimately be changed during the segmentation process. However, this approximation is reasonable since the likelihood functions are primarily dependent on the image data and have only a secondary dependence on the pyramid structure. Intuitively, if the pixels in $y_s^{(n)}$ appear to be principally from class k, then the likelihood function $l_s^{(n)}(k)$ should be relatively large regardless of the pyramid structure used.

All the following analysis and experimentation will assume the use of this hybrid quadtree-graph structure. As in the case of the quadtree, segmentation using the hybrid structure is done in two steps, each of which only requires order *MN* computation. The first step is a fine-to-coarse computation given by [Eq 27]. The second step is a coarse-to-fine computation given by [Eq 35]. Together the total computation is of order *O(MN)*.

Parameter Estimation

In typical applications, one does not have prior information about the exact behavior of the segmentation X. However, it is possible to determine this information directly from the image as it is being segmented by estimating the parameters $\theta_n = [\theta_{n,0}, \theta_{n,1}]$ during the segmentation process.

The parameters $\theta_{n,0}$ are required for fine-to-coarse operations used in computing the log likelihood functions, and the parameters $\theta_{n,1}$ are required for coarse-to-fine operations used in the SMAP segmentation. However, both of these parameter vectors will be estimated during the coarse-to-fine operations. This means that the segmentation process will require two full passes composed of the following steps:

1. Perform fine-to-coarse operations using initial parameter values $\theta_{n,0} = 1$.
2. Estimate parameters $\theta_{n,0}$ and $\theta_{n,1}$ during coarse-to-fine operations.
3. Perform fine-to-coarse operations using estimated parameters.
4. Re-estimate parameters $\theta_{n,1}$ during final coarse-to-fine segmentation.

The estimation procedure of 2 and 4 above is performed sequentially at each scale. Each transition parameter vector θ_n is estimated concurrently with the computation of the segmentation $\hat{x}^{(n)}$. The computational overhead of this estimation procedure is greatly reduced by subsampling the image at high resolutions. Since the number of pixels at high resolutions is great, this subsampling does not substantially impact on the accuracy of the estimated parameters.

The two-pass process implies that parameter estimation will increase computation by at least a factor of two. However, the additional computation required within each iteration is minimal due to subsampling. So the total increase in computation for parameter estimation is generally close to two.

Begin by deriving the sequential parameter estimation procedure. The transition parameter $\theta_{n,1}$ is estimated by finding the maximum likelihood value given the image, y, and the previous coarse scale segmentation, $\hat{x}^{(n+1)}$. Formally compute the solution to the optimization criteria:

$$\hat{\theta}_{n,1} \in \arg\max_{\theta_{n,1} \in \Omega} p_{y|x^{(n+1)}}(y|\hat{x}^{(n+1)}, \theta_{n,1}) \qquad \text{[Eq 36]}$$

where Ω is the set of valid parameter values. Using the hybrid pyramid structure of previous section, the log likelihood function, L, has the specific form:

$$\begin{aligned} L(\theta_{n,1}) &= \log p_{y|x^{(n+1)}}(y \mid \hat{x}^{(n+1)}, \theta_{n,1}) \\ &= \sum_{s \in S^{(n)}} \log\left(\sum_{k=1}^{M} \exp\left(l_s^{(n)}(k)\right) \tilde{p}_{x_0^{(n)}|x_{\partial 0}^{(n+1)}}(k \mid \hat{x}_{\partial s}^{(n+1)}, \theta_{n,1}) \right) \end{aligned} \qquad \text{[Eq 37]}$$

where the dependence of $\tilde{p}$ on n is through the parameter $\theta_{n,1}$. Notice that $\tilde{p}$ uses the subscript $x_0^{(n)} \mid x_{\partial 0}^{(n+1)}$ to emphasize that the conditional distribution assumed in [Eq 32] does not depend on s.

This likelihood, $L(\theta_{n,1})$, may be maximized as a function of $\theta_{n,1}$ by using the expectation-maximization (EM) algorithm (Dempster, Laird, and Rubin 1977; Bamm et al. 1970). From the form of the function $\tilde{p}\, L(\theta_{n,1})$ is known to be a convex function of $\theta_{n,1}$. This fact shows that the EM algorithm is guaranteed to converge to a value of $\theta_{n,1}$ which maximizes L.

To apply the EM algorithm, first compute the function:

$$\begin{aligned} Q(\theta_{n,1}, \theta'_{n,1}) = E\Big[&\log\left(p_{y|x^{(n)}}(y \mid X^{(n)})\, p_{x^{(n)}|x^{(n+1)}}(X^{(n)} \mid \hat{x}^{(n+1)}, \theta_{n,1})\right) \\ &\mid Y = y, X^{(n+1)} = \hat{x}^{(n+1)}, \theta'_{n,1} \Big] \end{aligned} \qquad \text{[Eq 38]}$$

where $X^{(n)}$ is a random object. The EM algorithm uses the following iterative procedure to find a sequence of parameters that converge to $\hat{x}_{n,1}$.

$$\theta_{n,1}^{p+1} \in \arg\max_{\theta_{n,1} \in \Omega} Q(\theta_{n,1}, \theta_{n,1}^{p}) \qquad \text{[Eq 39]}$$

To further simplify this expression a sufficient statistic for the transition distribution $\tilde{p}_{x_0^{(n)}|x_{\partial 0}^{(n+1)}}(m \mid i, j, k, \theta_{n,1})$ of [Eq 32] must be formulated. This statistic counts the number of distinct transitions that occur from

a particular set of coarse scale neighbors, $x_{\partial s}^{(n+1)}$ to a point $x_s^{(n)}$. Specifically, the sufficient statistic, T, is defined so that:

$$\log \tilde{p}_{x_0^{(n)}|x_{\partial 0}^{(n+1)}}(m \mid i,j,k,\theta_{n,1}) = \sum_{l=0}^{1} \sum_{h=0}^{2} T_{l,h}(m \mid i,j,k), V_{l,h}(\theta_{n,1}) \qquad \text{[Eq 42]}$$

where T and V have the functional forms:

$$T_{l,h}(m \mid i,j,k) = \begin{cases} 1 & \text{if } \delta_{m,i} = l \text{ and } \delta_{m,j} + \delta_{m,k} = h \\ 0 & \text{otherwise} \end{cases} \qquad \text{[Eq 40]}$$

and

$$V_{l,h}(\theta_{n,1}) = \log\left(\frac{\theta_{n,1}}{7}(3l + 2h) + \frac{1 - \theta_{n,1}}{M}\right) \qquad \text{[Eq 41]}$$

Substituting [Eq 40] into the expression for Q obtains the simpler expression:

$$\theta_{n,1}^{p+1} \in \arg\max_{\theta_{n,1} \in \Omega} \left\{ \sum_{l=0}^{1} \sum_{h=0}^{2} \bar{T}_{l,h}(\theta_{n,1}^{p}), V_{l,h}(\theta_{n,1}) \right\} \qquad \text{[Eq 42]}$$

where:

$$\begin{aligned} \bar{T}_{l,h}(\theta_{n,1}) &= \sum_{s \in S^{(n)}} E\left[T_{l,h}(X_s^{(n)} \mid \hat{x}_{\partial s}^{(n+1)}) \mid Y = y, X^{(n+1)} = \hat{x}^{(n+1)}, \theta_{n,1} \right] \\ &= \sum_{s \in S^{(n)}} \frac{\sum_{k=1}^{M} T_{l,h}(k \mid \hat{x}_{\partial s}^{(n+1)}) \exp(l_s^{(n)}(k)) \tilde{p}_{x_0^{(n)}|x_{\partial 0}^{(n+1)}}(k \mid \hat{x}_{\partial s}^{(n+1)}, \theta_{n,1})}{\sum_{k=1}^{M} \exp(l_s^{(n)}(k)) \tilde{p}_{x_0^{(n)}|x_{\partial 0}^{(n+1)}}(k \mid \hat{x}_{\partial s}^{(n+1)}, \theta_{n,1})} \end{aligned} \qquad \text{[Eq 43]}$$

Evaluation of $\bar{T}_{l,h}$ is computationally expensive since it requires a summation over all the points in $S^{(n)}$. However, accurate estimation of θ_n only requires a representative sampling of the image data. This is particularly true at high resolutions where the number of pixels far exceeds the number required to accurately estimate these parameters. Therefore, the points in $S^{(n)}$ at period $P^{(n)}$ in both the horizontal and vertical directions are subsampled, and $P^{(n)} \propto \sqrt{2^{-n}}$ is chosen so as to still have an increasing number of

samples at finer scales. Experimental results given in the following section will show that subsampling substantially reduces computation without adversely affecting the performance of parameter estimation.

The two steps of each iteration in the EM algorithm are then:

E - Compute $\bar{T}$ using the parameter, $\theta_{n,1}^{p}$ sampling period $P^{(n)}$, and [Eq 44].

M - Compute $\theta_{n,1}^{p+1} \in \arg\max_{\theta_{n,1} \in \Omega} Q(\theta_{n,1}, \theta_{n,1}^{p})$ using [Eq 43].

The M step can be efficiently computed since it requires the maximization of a convex function Q over an interval. This may be done with, for example, the golden section search method (Press et al 1988). The question remains as to whether the algorithm converges to the global maximum of L. Appendix C answers this question by adapting the convergence results of Wu (1983), and Redner and Walker (1984) to prove the following theorem:

Theorem 1: *If (i) Ω is a compact, convex set, (ii) $L(\theta)$ and $Q(\theta, \theta')$ are continuous and differentiable on an open set containing Ω, and (iii) $L(\theta)$ is convex, then any limit point, $\hat{\theta}$,* of the sequence $\{\theta^{p}\}_{1}^{\infty}$ *has the property that*:

$$\hat{\theta} \in \arg\max_{\theta \in \Omega} L(\theta) \qquad \text{[Eq 45]}$$

Notice that assumption (ii) does not strictly hold since Q becomes infinite at the points 0 and 1. In practice, this problem can be resolved by estimating $\theta_{n,1}$ over some smaller interval $\Omega = [0+\varepsilon_1, 1-\varepsilon_1]$. In any case, the introduction of ε_l is required for numerical stability.

Finally, the parameters $\theta_{n,0}$ are estimated using the statistics $\bar{T}$ resulting from convergence of the EM algorithm.

$$\hat{\theta}_{n,0} = \frac{\sum_{h=0}^{2} \bar{T}_{1,h}}{\sum_{l=0}^{1} \sum_{h=0}^{2} \bar{T}_{l,h}} \qquad \text{[Eq 46]}$$

These values are then used in the second pass of fine-to-coarse computations.

The complete SMAP segmentation algorithm with parameter estimation is summarized below:

1. Set the initial parameter values for all n, $\hat{\theta}_{n,0}=1$, and $\hat{\theta}_{L-1,1}=0.5$.

2. Compute the likelihood functions using [Eq 27] and the parameters $\hat{\theta}_{n,0}$.

3. Compute $\hat{x}^{(L)}$ using [Eq 34].

4. For scales $n=L-1$ to $n=0$

   (a) Use the EM algorithm to iteratively compute $\hat{\theta}_{n,1}$ and $\tilde{T}$. Subsample by $P^{(n)}$ when computing $\tilde{T}$, and stop when $|\theta_{n,1}^{p+1}-\theta_{n,1}^{p}|<\varepsilon_2$.

   (b) Compute $\hat{\theta}_{n,0}$ using [Eq 46].

   (c) Compute $\hat{x}^{(n)}$ using [Eq 35].

   (d) Set $\hat{\theta}_{n-1,1}=\hat{\theta}_{n,1}(1-10\varepsilon_2)$

5. Repeat steps 2 through 4.

Since global convergence of the EM algorithm is guaranteed, the choice of ε_1 and ε_2 only impacts on the accuracy of convergence. The values $\varepsilon_1=10^{-6}$ and $\varepsilon_2=10^{-4}$ were used in this study, and were never found to be problematic. Generally, smaller values of ε_1 and ε_2 will lead to better estimates but slower convergence. Also, ε_1 should be chosen so that $\varepsilon_1<<\varepsilon_2$. Also note that step 4d is used to accelerate convergence of the EM algorithm by starting the new parameter values near the previous parameter values.

5 EXPERIMENTAL RESULTS AND IMPLEMENTATION

Results

This chapter compares the performance of the SMAP algorithm with MAP segmentation using a MRF prior model for the pixel classes. All results of the SMAP algorithm used unsupervised estimation of the MSRF parameters. Unless otherwise stated, subsampling was used in all experiments, and the subsampling period was always chosen using the formula:

$$P(n) = \max\{\lfloor 2^{(L-n-3)/2} \rfloor, 1\} \qquad \text{[Eq 47]}$$

This generally resulted in sufficient sampling to accurately estimate parameters while keeping the computational overhead of parameter estimation to a minimum.

For this comparison to MAP estimation, a conventional 8pt neighborhood MRF model (Bouman and Liu 1991) was chosen for the class labels X. Specifically, the probability distribution has the form:

$$p_x(x) = \frac{1}{z} \exp\left\{-\lambda\left(\sqrt{2}-1\right)\left(t_1(x) + t_2(x)/\sqrt{2}\right)\right\} \qquad \text{[Eq 48]}$$

where λ=1.5, t_1 is the number of horizontal and vertical neighbors of different class, and t_2 is the number of diagonal neighbors of different class. This appeared to yield the best overall results.

Since the MAP estimate can not be exactly computed, an optimization method must also be chosen. Here the methods used were simulated annealing (SA) (Geman and Geman 1984), ICM (Besag 1986), and multiple resolution segmentation (MRS) (Bouman and Liu 1991). The ICM and SA methods were started with the maximum likelihood estimate of the segmentation. SA used a temperature schedule of the form:

$$\frac{1}{T_{n+1}} = \frac{1}{T_n} + \Delta \qquad \text{[Eq 49]}$$

where T_0=1, T_{final}=*0.2666* and Δ were chosen to achieve the desired number of iterations. Both 100 and 500 iterations were used to compare the performance. After the desired number of iterations, ICM was

used (equivalently T=0) to assure convergence to a local minima. These annealing parameters were chosen since they seemed to give the best performance over the range of test images used. In practice, the choice of annealing parameters must be a compromise since the optimal parameters depend on the specific image being processed.

The MRS algorithm differs from ICM and SA since it effectly uses different values of λ at each scale. The scale dependent λ is used because the MAP estimate would otherwise have unreasonable behavior at coarse scales (Bouman and Liu 1991). Intuitively, the MRS algorithm attempts to approximately correct the undesirable properties of the MAP estimator by varying the prior model.

To test accuracy with real data, a multispectral remotely sensed image was segmented and the results were compared to measured ground truth to determine classification accuracy. The image used was a 927 × 1097 three-band SPOT (Systeme Probatoire d'Observation de la Terre) image with a spatial resolution of 20m. Ground truth information was collected along 90 positions (transects) placed randomly throughout the full image (Tazik et al. 1992). Each transect is 100m long with a random orientation.* Along each transect, detailed measurements of ground cover were made, but for this experiment only five classes were considered based on the aggregate measure of percent bare ground for each transect: 0 percent; 1-10 percent; 11-21 percent; 21-30 percent; and 31-100 percent. Sixty of the transects were randomly chosen to use for training the class models, and the remaining 30 were used to test the segmentation algorithm performance.

Each of the five classes was modeled as a multivariate Gaussian mixture. The mixture model is important because it captures the multimodal spectral behavior that is typical of most classes. The method for estimating the parameters of the mixture model is described in Appendix D.

Only 100 iterations of SA were used due to the excessive computation time required to process the large image. The percent misclassification was also computed using the 30 testing transects for each class. The results are tabulated in Table 1. For each class and algorithm, the average region size was computed.

Classes were formed from ranges of percentage bare ground, and percent classification accuracy was tabulated for each class. The classification accuracy of the SMAP segmentation was substantially better than the maximum likelihood algorithm and slightly better than ICM. The SMAP and SA algorithms had

* Also, to increase the number of pixels associated with each transect, the transects were considered to be 60m wide, thus covering approximately 15 pixels on the image.

Table 1

Tabulated Results for the Segmentation of Multispectral SPOT Data With Ground Truth

		Class 1	Class 2	Class 3	Class 4	Class 5	Class Average
Percent Bare Ground		0%	1-10%	11-21%	21-30%	31-100%	
Accuracy	**SMAP**	88.1%	27.8%	17.5%	27.5%	93.5%	50.88%
	SA 100	88.1%	22.5%	17.5%	27.5%	96.0%	50.32%
	ICM	87.5%	23.2%	17.5%	27.5%	93.5%	49.84%
	ML	78.5%	26.5%	17.5%	20.0%	85.1%	45.52%
Average Region Area	**SMAP**	131.7	18.7	10.7	21.2	57.7	48.0
	SA 100	126.9	16.4	7.5	14.7	49.9	43.1
	ICM	112.2	13.2	7.2	12.9	44.2	37.9
	ML	32.1	3.9	3.9	4.7	16.1	12.1

approximately comparable accuracy with SMAP performing 5.3 percent better for class 2 and SA performing 2.5 percent better for class 5.

It is also interesting that the SMAP segmentation produces the largest average region sizes. This is because the SMAP segmentation produces fewer regions containing a small number of pixels.

Implementation

The multispectral SMAP segmentation algorithm and the multimodal mixed signature model have been implemented in the Geographical Resources Analysis Support System (GRASS) Version 4.1 (Westervelt, Shapiro, and Gerdes 1993). The mixed signature model is generated from ground-truth data by the GRASS *i.gensigset* module, and the SMAP segmentation is performed by the GRASS *i.smap* module.

6 CONCLUSIONS AND RECOMMENDATIONS

This study has derived a new criterion and model for statistical image segmentation. The SMAP estimator is proposed because it minimizes the expected size of the largest misclassified region, and it results in a computationally simple segmentation algorithm. The MSRF model uses a pyramid structure to capture the characteristics of image behavior at various scales. Because the MSRF has a Markov Chain structure, its parameters can be estimated efficiently from the image during segmentation.

Experiments indicate that the SMAP algorithm performs comparably to or substantially better than MAP estimation using a Markov random field model and simulated annealing. In addition, the SMAP algorithm requires less computation than ICM and much less then simulated annealing. The SMAP algorithm was tested on multispectral SPOT data using LCTA transects ground-level data and found to improve segmentation accuracy.

The single test run that was done used an image from only one date, a single sensor (SPOT), and a single LCTA ground-truth component, i.e., percent ground cover. Also, the linear nature of the LCTA transects required that the transects be artificially widened to include more pixels. This is necessary for statistical sampling reasons and for the spatial nature of the remotely-sensed imagery.

It is recommended that:

1. Further tests with LCTA ground-truth data sets using other components (e.g., plant community classes) should be used to test the efficacy of the segmentation model for other ground measures.

2. If LCTA transect information is to be used, rigorous methods should be found to "widen" the transects to more closely match the resolution of the imagery and to encompass more image pixels. This general recommendation is not isolated to the requirements of these algorithms, but is a general requirement for classification of remotely-sensed imagery using ground-truth data.

3. Other physiographic regions should be used to test the applicability of these algorithms to other regions of the earth.

4. Timing of the ground-truth data acquisition and the acquisition of the imagery may have impacts on the ability to correlate these data. At least one set of images should be acquired that relate very closely in time to the acquisition of the ground-truth to minimize the temporal effects due to differences in acquisition dates.

5. The efficacy of these algorithms should be tested with other sensors. SPOT has 3 spectral bands, each with 20m spatial resolution. Sensors with more spectral resolution (e.g., Thematic Mapper (TM) that has 6 spectral bands, with 30m resolution) could be used for testing. Also, since the segmentation algorithm is contextual, higher spatial resolution imagery (e.g., airborne multispectral scanners) could also be tested.

REFERENCES

Antonisse, H., "Image Segmentation in Pyramids," *Computer Vision Graphics and Image Processing*, Vol 19 (1982), pp 367-383.

Basseville, M., A. Benveniste, and A. Willsky, *Multiscale Autoregressive Processes, Part I: Schur-Levinson Parametrizations*, Vol 40, No. 8 (August 1992), pp 1915-1934.

Basseville, M., A. Benveniste, and A. Willsky, *Multiscale Autoregressive Processes, Part II: Lattice Structures for Whitening and Modeling*, Vol 40, No. 8 (August 1992), pp 1935-1954.

Basseville, M., A. Benveniste, K. Chou, S. Golden, R. Nikoukhah, and A. Willsky, "Modeling and Estimation of Multiresolution Stochastic Processes," *IEEE Transactions on Information Theory*, Vol 38, No. 2 (March 1992), pp 766-784.

Baum, L., T. Petrie, G. Soules, N. Weiss, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *Annals of Mathematical Statistics*, Vol 41, No. 1 (1970), pp 164-171.

Besag, J., "On the Statistical Analysis of Dirty Pictures," *Journal of the Royal Statistical Society B*, Vol 48, No. 3 (1986), pp 259-302.

Besag, J., "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society B*, Vol 36, No. 2 (1974), pp 192-236.

Bouman, C., and B. Liu, "Multiple Resolution Segmentation of Textured Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 13, No. 2 (February 1991), pp 99-113.

Bouman, C., and B. Liu, "Segmentation of Textured Images Using a Multiple Resolution Approach," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (New York, NY, 11-14 April 1988), pp 1124-1127.

Burt, P., T. Hong, and A. Rosenfeld, "Segmentation and Estimation of Image Region Properties Through Cooperative Hierarchical Computation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol SMC-11, No. 12 (December 1981), pp 802-809.

Chou, K., S. Golden, and A. Willsky, Modeling and Estimation of Multiscale Stochastic Processes, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (Toronto, Canada, 14-17 May 1991), pp 1709-1712.

Chou, K., A. Willsky, A. Benveniste, and M. Basseville, "Recursive and Iterative Estimation Algorithms for Multi-Resolution Stochastic Processes," *Proceedings of the 28th Conference on Decision and Control*, Vol 2 (Tampa, Florida, 13-15 December 1989), pp 1184-1189.

Dempster, A., N. Laird, and D. Rubin, "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society B*, Vol 39, No. 1 (1977), pp 1-38.

Derin, H., and H. Elliott, "Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol PAMI-9, No. 1 (January 1987), pp 39-55.

Derin, H., and W. Cole, "Segmentation of Textured Images Using Gibbs Random Fields," *Computer Vision Graphics and Image Processing*, Vol 35 (1986), pp 72-98.

Dubes, R., A. Jain, S. Nadabar, and C. Chen, "MRF Model-Based Algorithms for Image Segmentation," *Proceeedings of the 10th International Conference on Pattern Recognition* (Atlantic City, NJ, June 1990), pp 808-814.

Geman, S., and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol PAMI-6, No. 6 (November 1984), pp 721-741.

Gidas, B., "A Renormalization Group Approach to Image Processing Problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 11, No. 2 (February 1989), pp 164-180.

Haralick, R., and L. Shapiro, "Image Segmentation Techniques," *Computer Vision Graphics and Image Processing*, Vol 29 (1985), pp 100-132.

Jeon, B., and D.A. Landgrebe, "Classification With Spatio-Temporal Interpixel Class Dependency Contexts," *IEEE Transactions on Geoscience and Remote Sensing* (in publication).

Jeon, B., and D.A. Landgrebe, "Spatio-Temporal Contextual Classification Based on Markov Random Field Model," *Proceedings of 1991 International Geoscience and Remote Sensing Symposium* (Espoo, Finland, 1991), pp 1819-1822.

K. Chou, S. Golden, M. Luettgen, and A. Willsky, "Modeling and Estimation of Multiresolution Stochastic Processes and Random Fields," *Proceedings of the Seventh Workshop on Multidimensional Signal Processing* (Lake Placid, New York, 23-25 September 1991), p 3.8,.

Kato, Z., J. Zerubia, and M. Berthod, "Satellite Image Classification Using a Modified Metropolis Dynamics," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol 3 (San Francisco, CA, 23-26 March 1992), pp 573-576.

Kettig, R., and D. Landgrebe, "Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects," *IEEE Transactions on Geoscience and Electronics*, Vol GE-14, No. 1 (January 1976), pp 19-26.

Lakshmanan, S., and H. Derin, "Simultaneous Parameter Estimation and Segmentation of Gibbs Random Fields Using Simulated Annealing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 11, No. 8 (August 1989), pp 799-813.

Landgrebe, D., "The Development of a Spectral-Spatial Classifier for Earth Observational Data," *Pattern Recognition*, Vol 12 (1980), pp 165-175.

Laws, K.I., *Textured Image Segmentation*, Ph.D. Dissertation (Department. of English, University of Southern California, Los Angeles, CA, 1980).

Manjunath, B., T. Simchony, and R. Chellappa, "Stochastic and Deterministic Networks for Texture Segmentation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol 38, No. 6 (June 1990), pp 1039-1049.

Marroquin, J., S. Mitter, and T. Poggio, "Probabilistic Solution of Ill-Posed Problems in Computational Vision," *Journal of the American Statistical Association*, Vol 82 (March 1987), pp 76-89.

Pappas, T., "An Adaptive Clustering Algorithm for Image Segmentation," *IEEE Transactions on Signal Processing*, Vol 40, No. 4 (April 1992), pp 901-914.

Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann Publishers, San Mateo, CA, 1988).

Pentland, A., "Fractal-Based Description of Natural Scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol PAMI-6 (November 1984), pp 661-674.

Perez, P., and F. Heitz, "Multiscale Markov Random Fields and Constrained Relaxation in Low Level Image Analysis," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol 3 (San Francisco, CA, 23-26 March 1992), pp 61-64.

Pickard, D., "Inference for Discrete Markov Fields: The Simplest Nontrivial Case," *Journal of the American Statistical Association*, Vol 82 (March 1987), pp 90-96.

Press, W., B. Flannery, S. Teukolsky and Vetterling, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1988).

Rissanen,, J., "A Universal Prior for Integers and Estimation by Minimum Description Length," *Annals of Statistics*, Vol 11, No. 2 (1983), pp 417-431.

Sauer, K., and C. Bouman, "A Local Update Strategy for Iterative Reconstruction From Projections," *IEEE Transactions on Signal Processing* (in publication).

Tazik, D., S. Warren, V. Diersing, R. Shaw, R. Brozka, C. Bagley, and W. Whitworth, *U.S. Army Land Condition-Trend Analysis (LCTA) Plot Inventory Field Methods*, Technical Report N-92/03/ADA247931 (U.S. Army Construction Engineering Research Laboratories [USACERL], February 1992).

Unser, M., and M. Eden, "Multiresolution Feature Extraction and Selection for Texture Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 11, No. 7 (July 1989), pp 717-728.

Unser, M., and M. Eden, "Nonlinear Operators for Improving Texture Segmentation Based on Features Extracted by Spatial Filtering," *IEEE Transactions on Systems Man Cybernetics.*, Vol 20, No. 4 (July/August 1990), pp 804-815.

Westervelt, J., M. Shapiro, and D.P. Gerdes, *GRASS Version 4.1 User's Reference Manual* (Draft) (USACERL Office of GRASS Integration, Champaign, IL, 1993).

Wu, C., "On the Convergence Properties of the EM Algorithm," *Annals of Statistics*, Vol 11, No. 1 (1983), pp 95-103.

Zhang, M., R. Haralick, and J. Campbell, "Multispectral Image Context Classification Using Stochastic Relaxation," *IEEE Transactions on Systems Man Cybernetics.*, Vol 20, No. 1 (February 1990), pp 128-140.

APPENDIX A: Optimality Property for SMAP Iteration

Assume that $\hat{x}^{(i)}$ has been computed for $i>n$. Compute $\hat{x}^{(n)}$ using the Markov chain structure of the random fields $X^{(n)}$.

$$\hat{x}^{(n)} = \arg\max_{x^{(n)}} \max_{x^{(i)}\, i \neq n} \sum_{k=0}^{L} 2^k P(X^{(i)} = x^{(i)}\, i \geq k \mid Y = y)$$

$$= \arg\max_{x^{(n)}} \max_{x^{(i)} i<n\, x^{(i)} i>n} \max \sum_{k=0}^{L} 2^k P(Y \in dy, X^{(i)} = x^{(i)} i \geq k)$$

$$= \arg\max_{x^{(n)}} \max_{x^{(i)} i<n}$$

$$\left\{ \sum_{k=0}^{n} 2^k P(Y \in dy, X^{(i)} = x^{(i)}\, k \leq i \leq n \mid X^{(i)} = \hat{x}^{(i)} i>n)\, P(X^{(i)} = \hat{x}^{(i)}\, i>n) \right. \qquad \text{[Eq A1]}$$

$$\left. + \sum_{k=n+1}^{L} 2^k P(Y \in dy, X^{(i)} = \hat{x}^{(i)}\, i>k) \right\}$$

$$= \arg\max_{x^{(n)}} \max_{x^{(i)} i<n} \sum_{k=0}^{n} 2^k P(Y \in dy, X^{(i)} = x^{(i)}\, k \leq i \leq n \mid X^{(n+1)} = \hat{x}^{(n+1)})$$

$$= \arg\max_{x^{(n)}} \left\{ P(Y \in dy, X^{(n)} = x^{(n)} \mid X^{(n+1)} = \hat{x}^{(n+1)}) \right.$$

$$\left. + \max_{x^{(i)} i<n} \sum_{k=0}^{n-1} 2^{k-n} P(Y \in dy, X^{(i)} = x^{(i)} k \leq i \leq n \mid X^{(n+1)} = \hat{x}^{(n+1)}) \right\}$$

Next, define a residue term, $R(x^{(n)})$, so that the following equality holds:

$$\begin{aligned}\hat{x}^{(n)} &= \arg\max_{x^{(n)}} P(Y \in dy, X^{(n)} = x^{(n)} \mid X^{(n+1)} = \hat{x}^{(n+1)})\,(1 + R(x^{(n)})) \\ &= \arg\max_{x^{(n)}} p_{x^{(n)} \mid x^{(n+1)}, y}(x^{(n)} \mid \hat{x}^{(n+1)}, y)\,(1 + R(x^{(n)}))\end{aligned} \qquad \text{[Eq A2]}$$

Specifically, $R(x^{(n)})$ is given by

$$R(x^{(n)}) = \max_{x^{(i)}, i<n} \frac{\sum_{k=0}^{n-1} 2^{k-n} P(Y \in dy, X^{(i)} = x^{(i)}, k \le i \le n \mid X^{(n+1)} = \hat{x}^{(n+1)})}{P(Y \in dy, X^{(n)} = x^{(n)} \mid X^{(n+1)} = \hat{x}^{(n+1)})} \qquad \text{[Eq A3]}$$

Since this expression is the ratio of positive quantities, it follows that $R \ge 0$. Further, R may be bounded from above as follows:

$$\begin{aligned}R(x^{(n)}) &= \max_{x^{(i)}, i<n} \sum_{k=0}^{n-1} 2^{k-n} P(X^{(i)} = x^{(i)}, k \le i \le n-1 \mid X^{(n)} = x^{(n)}, Y = y) \\ &\le \max_{x^{(n-1)}} \sum_{k=0}^{n-1} 2^{k-n} P(X^{(n-1)} = x^{(n-1)} \mid X^{(n)} = x^{(n)}, Y = y) \\ &\le \max_{x^{(n-1)}} P(X^{(n-1)} = x^{(n-1)} \mid X^{(n)} = x^{(n)}, Y = y) \\ &= \max_{x^{(n-1)}} p_{x^{(n-1)} \mid x^{(n)}, y}(x^{(n-1)} \mid x^{(n)}, y)\end{aligned} \qquad \text{[Fq A4]}$$

Finally,

$$\begin{aligned}\hat{x}^{(n)} &= \arg\max_{x^{(n)}} \left\{\log p_{x^{(n)} \mid x^{(n+1)}, y}(x^{(n)} \mid \hat{x}^{(n+1)}, y) + \log(1 + R(x^{(n)})) - 1\right\} \\ &= \arg\max_{x^{(n)}} \left\{\log p_{x^{(n)} \mid x^{(n+1)}, y}(x^{(n)} \mid \hat{x}^{(n+1)}, y) + \mathscr{E}(x^{(n)})\right\}\end{aligned} \qquad \text{[Eq A5]}$$

where

$$0 \le \mathscr{E}(x^{(n)}) = \log(1 + R(x^{(n)})) - 1$$
$$\le \max_{x^{(n-1)}} p_{x^{(n-1)} \mid x^{(n)}, y}(x^{(n-1)} \mid x^{(n)}, y) \qquad \text{[Eq A6]}$$

APPENDIX B: Product Form for Quadtree

This appendix shows by induction that, for a quadtree based pyramid structure, the distribution of Y given $X^{(0)}$ has the form of [Eq 23] and that the terms in the product may be computed using the recursion of [Eq 24]. For $n=0$, these relations are true by assumption. So, assuming the result for scale n yields:

$$
\begin{aligned}
P(Y \in dy \mid X^{(n+1)} = x^{(n+1)}) &= \sum_{x^{(n)}} P(Y \in dy \mid X^{(n)} = x^{(n)}) p_{x^{(n)} \mid x^{(n+1)}}(x^{(n)} \mid x^{(n+1)}) \\
&= \sum_{x^{(n)}} \left\{ \prod_{r \in S^{(n)}} p_{y_r^{(n)} \mid x_r^{(n)}}(y_r^{(n)} \mid x_r^{(n)}) \right\} \left\{ \prod_{r \in S^{(n)}} p_{x_r^{(n)} \mid x_{\partial r}^{(n+1)}}(x_r^{(n)} \mid x_{\partial r}^{(n+1)}) \right\} \\
&= \prod_{r \in S^{(n)}} \sum_{x_r^{(n)}=1}^{M} p_{y_r^{(n)} \mid x_r^{(n)}}(y_r^{(n)} \mid x_r^{(n)}) p_{x_r^{(n)} \mid x_{\partial r}^{(n+1)}}(x_r^{(n)} \mid x_{\partial r}^{(n+1)}) \\
&= \prod_{s \in S^{(n+1)}} \prod_{r \in d^{-1}(s)} \sum_{x_r^{(n)}=1}^{M} p_{y_r^{(n)} \mid x_r^{(n)}}(y_r^{(n)} \mid x_r^{(n)}) p_{x_r^{(n)} \mid x_s^{(n+1)}}(x_r^{(n)} \mid x_s^{(n+1)}) \\
&= \prod_{s \in S^{(n+1)}} p_{y_s^{(n+1)} \mid x_s^{(n+1)}}(y_s^{(n+1)} \mid x_s^{(n+1)})
\end{aligned}
\qquad \text{[Eq B1]}
$$

where

$$
p_{y_s^{(n+1)} \mid x_s^{(n+1)}}(y_s^{(n+1)} \mid k) = \prod_{r \in d^{-1}(s)} \sum_{m=1}^{M} p_{y_r^{(n)} \mid x_r^{(n)}}(y_r^{(n)} \mid m) p_{x_r^{(n)} \mid x_s^{(n+1)}}(m \mid k) \qquad \text{[Eq B2]}
$$

APPENDIX C: Global Convergence of EM Algorithm

This appendix proves theorem 1 by extending basic results on the convergence of the EM algorithm (Wu 1983, Redner and Walker 1984).

By the stated assumptions and theorem 4.1(v) of Redner and Walker, any limit point, $\hat{\theta}$, of the sequence $\{\theta^p\}_1^\infty$ has the property that $\hat{\theta} \in arg\ \max_{\theta \in \Omega} Q(\theta, \hat{\theta})$. Since Ω is convex and Q is differentiable, this implies that for all $\theta \in \Omega$

$$D^{1,0} Q(\hat{\theta}, \hat{\theta})(\theta - \hat{\theta}) \leq 0 \qquad \text{[Eq C1]}$$

where $D^{1,0}$ computes the gradient of Q with respect to the first argument.

Let Ω_o be an open set containing Ω such that Q and L are continuous, differentiable on Ω_o. Then define the continuous, differentiable function $H(\theta,\hat{\theta}) = Q(\theta,\hat{\theta}) - L(\theta)$ on Ω_o. It has been shown (Dempster, Laird, and Rubin 1977) that H has the property $\hat{\theta} \in arg\ \max_{\theta \in \Omega} H(\theta, \hat{\theta})$, which implies that

$$\begin{aligned} DL(\hat{\theta}) &= D^{1,0} Q(\hat{\theta}, \hat{\theta}) + D^{1,0} H(\hat{\theta}, \hat{\theta}) \\ &= D^{1,0} Q(\hat{\theta}, \hat{\theta}) \end{aligned} \qquad \text{[Eq C2]}$$

Therefore, for any $\theta \in \Omega$,

$$DL(\hat{\theta})(\theta - \hat{\theta}) = D^{1,0} Q(\hat{\theta}, \hat{\theta})(\theta - \hat{\theta}) \leq 0 \qquad \text{[Eq C3]}$$

Since L is a convex function and Ω is a convex set, it follows that $\hat{\theta}$ is an (implied) global maximum of L.

APPENDIX D: Automatic Parameter Estimation for Mixture Signatures

It is often desirable to segment multispectral images into regions corresponding to various information classes. For example, forest, grasslands, or urban areas are examples of information classes that may be of interest. In practice, these information classes often contain subclasses, each with distinctive spectral characteristics. The objective of mixture classes is to form a probabilistic class model composed of a number of spectral subclasses. Each subclass is then characterized by a set of parameters describing the mean and variation of the spectral components.

For each mixture class, it is necessary to determine the number of subclasses and the parameters of each subclasses. This can be done by using a representative sample of training data from each class and estimating the number of subclasses and their parameters from this data. Specifically, for a given information class with K subclasses, the following parameters are required for the k^{th} subclass where $1 \le k \le K$.

π_k - the probability that a pixel has subclass k.

μ_k - the spectral mean vector for subclass k.

R_k - the spectral covariance matrix for subclass k.

Let $x_1, x_2, \cdots, x_N$ be the N samples used to train the class in question. Assuming that each subclass has a multivariate Gaussian distribution, the probability that the pixel x_n is from subclass k is given by:

$$p_{x|k}(x_n|k) = \frac{1}{(2\pi)^{M/2}} |R_k|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (x_n - \mu_k)' R_k^{-1} (x_n - \mu_k) \right\} \qquad \text{[Eq D1]}$$

where M is the number of spectral components.

The integer value K is known as an order parameter since it identifies the order of the model. Initially, assume that K is known, and then introduce a specific method for computing it. The parameters π_k, μ_k, and R_k may be computed for a fixed K by using the expectation maximization (EM) algorithm (Baum et al. 1970; Dumpster, Laird, and Rubin 1977; Redner and Walker 1984). The EM algorithm is an iterative method for computing maximum likelihood (ML) estimates. The ML estimates of π_k, μ_k, and

R_k are the values of these parameters that maximize the log probability of the data x. The log of this probability may be exactly computed as:

$$\log p_x(x) = \sum_{n=1}^{N} \log\left(\sum_{i=1}^{K} p_{x|k}(x_s|k)\,\pi_k\right) \qquad \text{[Eq D2]}$$

and is often referred to as the log likelihood.

The EM algorithm works by clustering the data into groups corresponding to each subclass. However, instead of the membership to each subclass being deterministic, the membership is represented using a "soft" probability. The probability that pixel x_n belongs to subclass k may be computed using Bayes rule as:

$$p_{k|x}(k|x_n) = \frac{p_{x|k}(x_n|k)\,\pi_k}{\sum_{k=1}^{K} p_{x|k}(x_n|k)\,\pi_k} \qquad \text{[Eq D3]}$$

Then using these "soft" subclass memberships compute new spectral mean and covariance estimates for each subclass. These new estimates are denoted by $\hat{\pi}_k$, $\hat{\mu}_k$, and $\hat{R}_k$ to distinguish them from the initial estimates values π_k, μ_k and R_k:

$$\begin{aligned}
\hat{\pi}_k &= \frac{1}{N}\sum_{n=1}^{N} p_{k|x}(k|x_s) \\
\hat{\mu}_k &= \frac{1}{N}\sum_{n=1}^{N} x_n\, p_{k|x}(k|x_s) \qquad \text{[Eq D4]} \\
\hat{R}_k &= \frac{1}{N}\sum_{n=1}^{N} (x_n-\hat{\mu}_k)(x_n-\hat{\mu}_k)^t\, p_{k|x}(k|x_s)
\end{aligned}$$

The application of [Eq D2] followed by [Eq D4] improves the parameter estimates by increasing the log likelihood of [Eq D2]. Therefore, iterative application of these two operations will converge to a (local) maximum of the log likelihood. Specifically, the procedure is repeated until the change in $p_x(x)$ is less than ε where:

$$\epsilon = \frac{1}{100}\left(1 + M + \frac{(M+1)M}{2}\right)\log N \qquad \text{[Eq D5]}$$

Estimation of K is difficult because ML estimation does not work for K. Intuitively, the log likelihood may always be increased by adding more subclasses since more subclasses may be used to more accurately fit the data. This problem is addressed in the work of Rissanen (1985). Rissanen suggested that the best model order is the one that minimizes the total number of bits required to store or transmit the data and parameters. Using a theoretical argument, Rissanen showed that this minimum description length (MDL) model can be approximately achieved by minimizing the criteria:

$$\text{MDL} = -\log p_x(x) + \frac{1}{2}(\text{\# of parameters})\log \text{N} \qquad \text{[Eq D6]}$$

where the number of parameters is the total number of degrees of freedom in the parameter sets of all subclasses:

$$(\text{\# of parameters}) = K\left(1 + M + \frac{(M+1)M}{2}\right) - 1 \qquad \text{[Eq D7]}$$

Notice that, for fixed K, minimizing the MDL criteria is equivalent to maximizing the log likelihood.

The overall strategy for computing the mixture class parameters may be outlined as follows:

1. Initialize the class with a large number of subclasses, $K=K_o$.

2. Apply the iterative EM algorithm to minimize the MDL criteria.

3. Record the value of the MDL criteria, K and the subclass parameters.

4. If the number of subclasses is greater than 1, combine the two subclasses that have minimum "distance," and go back to step 2.

5. Choose the subclass parameters corresponding to the minimum value of MDL.

It remains to define the measure of distance between subclasses, and the initialization procedure. The initial subclass parameters are chosen to be:

$$\hat{\pi}_k = \frac{1}{K_o}$$

$$\hat{\mu}_k = x_n \text{ where } n = \lfloor (k-1)(N-1)/(K_o-1) \rfloor + 1 \qquad \text{[Eq D8]}$$

$$\hat{R}_k = I$$

where I is the identity matrix and $\lfloor \cdot \rfloor$ is the greatest smaller integer function. The function used to measure distance between the subclasses is:

$$d(k,j) = \frac{N\pi_k}{2} \log\left(\frac{|R_{k,j}|}{|R_k|}\right) + \frac{N\pi_j}{2} \log\left(\frac{|R_{k,j}|}{|R_j|}\right) \qquad \text{[Eq D9]}$$

where $R_{k,j}$ may be computed using

$$\mu_{k,j} = \frac{\pi_k}{\pi_k + \pi_j} \mu_k + \frac{\pi_j}{\pi_k + \pi_j} \mu_j \qquad \text{[Eq D10]}$$

$$R_{k,j} = \frac{\pi_k}{\pi_k + \pi_j} \left(R_k + (\mu_k - \mu_{k,j})(\mu_k - \mu_{k,j})^t\right) + \frac{\pi_j}{\pi_k + \pi_j} \left(R_j + (\mu_j - \mu_{k,j})(\mu_j - \mu_{k,j})^t\right)$$

USACERL Distribution

Chief of Engineers
ATTN: CECPW-IM-LH (2)
ATTN: CECPW-IM-LP (2)
ATTN: CERD-L
ATTN: DAEN-ZCI-P (2)
ATTN: CECW-PF
ATTN: CECW-EP-S
ATTN: CECW-RE
ATTN: CEIM-P
ATTN: CECW-OR

US Army Engineer District
ATTN: Chief, Regulatory Functions
Buffalo 14207
Norfolk 23510
Huntington 25701
Wilmington 28402
Charleston 29402
St. Paul 55101
Chicago 60606
Little Rock 72203
Galveston 77550
Albuquerque 87103
Los Angeles 90053
San Francisco 94105
Sacramento 95814
Portland 97208
Seattle 99362
New England 02254
ATTN: CENED-OD-R
New York 10278
ATTN: CENAN
Philadelphia 19106
ATTN: CENAP
Baltimore 21203
ATTN: CENAB-OP-R
Savannah 31402
ATTN: CESAS-OP-FP
Jacksonville 32232
ATTN: CESAJ-R-RD
Mobile 36628
ATTN: CESAM-OP-S
Nashville 37202
ATTN: CEORN
Memphis 38103
ATTN: CELMM-CO-R
Vicksburg 39180
ATTN: CELMV
Louisville 40201
ATTN: CEORL
Detroit 48226
ATTN: CENCE-CO
Rock Island 61204
ATTN: CENCR
St. Louis 63101
ATTN: CELMS-OD-F
Kansas City 64106
ATTN: Chief, Permits Section
Omaha 68102
ATTN: CEMRO-OP
New Orleans 71060
ATTN: Chief, Permits Section
Tulsa 74121
ATTN: CEWST
Fort Worth 76102
ATTN: CESWF-OD-O

US Army Engr Divisions
North Central 60606
ATTN: CENCD-CO
Southwest 75242
ATTN: CESWD-CO-R

US Military Academy 10996
ATTN: Dept of Geo & Envr Engr (2)
ATTN: Natural Resources Branch

Aberdeen Proving Ground, MD 21005
ATTN: ISC-TECOM
ATTN: HSHB-CI

US EPA Research Lab 97333

Yakima Firing Center 98901

USATHAMA 21010
ATTN: CETHA-RM-I
ATTN: AMXTH-RM
ATTN: CETHA-IR-S

CECPW 22060
ATTN: CECPW-FN
ATTN: CECPW-SP
ATTN: CECPW-FM-A

Bureau of Land Management
WASH DC 20240
Fort Collins 80526
Denver 80225

US Dept of Commerce 20233

Army National Guard 20310
ATTN: NGB-AREC

US Army Concepts Analysis Agency 20814

FBI Academy 22135

USA Foreign Science Tech Ctr 22901

Naval Oceanographic Office 39522

CEWES
ATTN: CEWES-IM-DA 39181
ATTN: CADD Center 39180

Wright-Patterson AFB 45433
ATTN: NSBIT AL/OEBN

5th Inf, Fort Polk 71459
ATTN: Envr & Nat Rsrcs Mgmt Div (2)

US Army Materiel Cmd 61299
ATTN: AMXEN-U (2)

US Army Europe
HQ USAREUR 09403
ATTN: AEAEN-FE-E (2)

Dugway Proving Ground 84122
ATTN: DPG-EN-E (2)

Yuma Proving Ground 85365
ATTN: ATEYP-ES-E

White Sands Missile Range
ATTN: STEWS-ES-E 88002

Nat'l Geophysical Data Ctr
ATTN: Code E-GCI 80303

Hohenfels Training Area 09173
ATTN: AETTH-DEH
ATTN: AETTH-DEH-ENV-APO

US Army Forts
Fort Belvoir, VA 22060
ATTN: CEETL-CL-GC (2)
ATTN: CETEC-CA-D
ATTN: AMSEL-RD-NV-VMD-TST
ATTN: Envr & Nat Res Div
Fort Monroe, VA 23651
ATTN: ATBO-GE
ATTN: ATEN-FN
Fort Drum 13603
ATTN: AFZS-EH-E
Fort Jackson 29207
ATTN: ATCJ-EHN
Fort Gillem 30050
ATTN: FCEN-CED-E
Fort Gordon 30905
ATTN: ATZH-DIE (2)
Fort Stewart 31314
ATTN: AFZP-DEN-W
Fort Benning 31905
ATTN: Nat. Resource Mgmt Div (2)
Fort McClellan 36205
ATTN: ATZN-FEE
Fort Rucker 36362
ATTN: ATZQ-EH
Fort Knox 40121
ATTN: ATZK-EHE
Fort Campbell 42223
ATTN: AFZH-DEH
Fort Benjamin Harrison 46216
ATTN: ATZI-ISP (2)
Fort McCoy 54656
ATTN: AFZR-DEN
Fort Riley 66442
ATTN: AFZN-DE-N (2)
Fort Chaffee 72905
ATTN: ATZR-ZFE (2)
Fort Sill 73503
ATTN: Fish & Wildlife Br (2)
Fort Leonard Wood 65473
ATTN: ATZT-DEH-EE
Fort Dix 08640
ATTN: ATZD-EHN
Fort Eustis 23604
ATTN: Ranges & Targets Dir
Fort Worth 76115
ATTN: Cartographic Ctr (2)
Fort Hood 76544
ATTN: AFZF-DE-ENV
Fort Bliss 79916
ATTN: ATZC-DEH-E
Fort Carson 80913
ATTN: AFZC-ECM-NR
Fort Huachuca 85613
ATTN: ATZS-EHB
Fort Irwin 92310
ATTN: AFZJ-EH
Fort Lewis 98433
ATTN: AFZH-DEQ
ATTN: ATZH-EHQ
Fort Richardson 99505
ATTN: DEH
Fort Bragg 28307
ATTN: DEH

National Weather Service 20910

US Geological Survey 22092

Pine Bluff Arsenal 17602
ATTN: SMCPB-EMB

US Army Cold Regions Research & Engr Lab
ATTN: CECRL-IS 03755

NASA/SSC/STL 39529

Defense Technical Info Center
ATTN: DTIC-FAB (2)

134
+17
6/94

This publication was reproduced on recycled paper.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1994—3510-S/00039