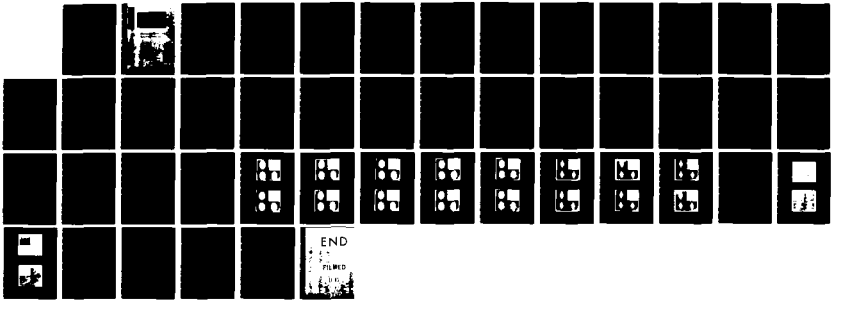
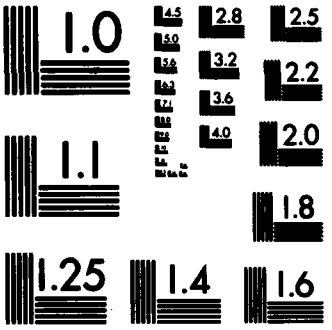
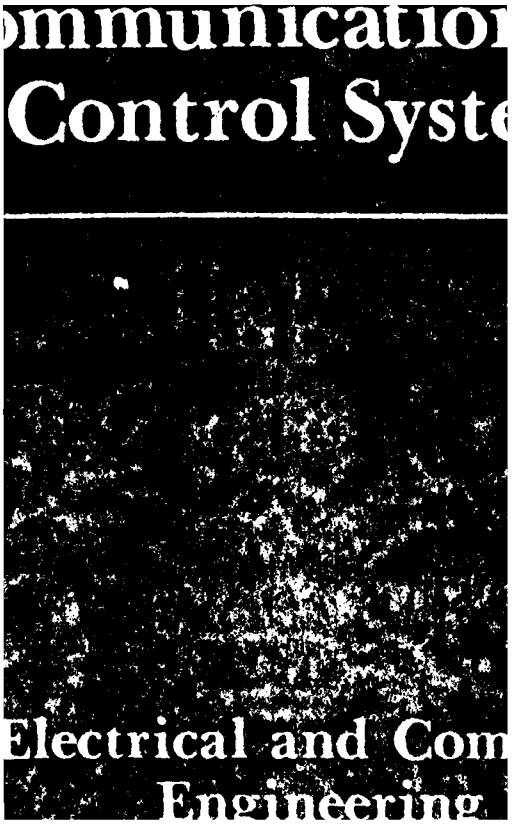
AD-A133 966 BAYES SMOOTHING ALGORITHMS FOR SEGMENTATION OF IMAGES MODELLED BY MARKOV. (U) MASSACHUSETTS UNIV AMHERST DEPT OF ELECTRICAL AND COMPUTER EN. H DERIN ET AL. AUG 83 1/1

UNCLASSIFIED UMASS-ECE-AUG83-1 N00014-83-K-0059 F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A


ommunicatio
Control Syste
Electrical and Com
Engineering

12

BAYES SMOOTHING ALGORITHMS FOR SEGMENTATION OF IMAGES MODELLED BY MARKOV RANDOM FIELDS

By

Haluk Derin[†], Howard Elliott[†],

Roberto Cristi[†] and Donald Geman[††]

[†]Department of Electrical and Computer Engineering

[††]Department of Mathematics and Statistics

University of Massachusetts

Amherst, Massachusetts 01003

Technical Report #UMASS-ECE-AU83-1

August 1983

This research has been supported by the Office of Naval Research under grant N00014-83-K-0059.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: UMASS-ECE-Aug83-1	2. GOVT ACCESSION NO.: AD-A133966
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): Bayes Smoothing Algorithms for Segmentation of Images Modelled by Markov Random fields	5. TYPE OF REPORT & PERIOD COVERED: Final
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): H. Derin, H. Elliott, R. Cristi, D. Geman	8. CONTRACT OR GRANT NUMBER(s): N00014-83-K-0059
9. PERFORMING ORGANIZATION NAME AND ADDRESS:	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS:	12. REPORT DATE: August 1983
	13. NUMBER OF PAGES:
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Markov random fields, Bayes smoothing, image segmentation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A new image segmentation algorithm is presented, based on recursive Bayes smoothing of images modelled by Markov random fields and corrupted by independent additive noise. The Bayes smoothing algorithm yields the a posteriori distribution of the scene value at each pixel, given the total noisy image, in a recursive way. The a posteriori distribution together with a criterion of optimality then determine a Bayes estimate of the scene. Examples are given where the algorithm is applied to test imagery and also SEASAT SAR imagery.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

A new image segmentation algorithm is presented, based on recursive Bayes smoothing of images modelled by Markov random fields and corrupted by independent additive noise. The Bayes smoothing algorithm yields the a posteriori distribution of the scene value at each pixel, given the total noisy image, in a recursive way. The a posteriori distribution together with a criterion of optimality then determine a Bayes estimate of the scene.

The algorithm presented is an extension of a 1-D Bayes smoothing algorithm to 2-D and it gives the optimum Bayes estimate for the scene value at each pixel. Computational concerns in 2-D, however, necessitate certain simplifying assumptions on the model and approximations on the implementation of the algorithm. In particular, the scene (noiseless image) is modelled as a Markov mesh random field, a special class of Markov random fields, and the Bayes smoothing algorithm is applied on overlapping strips (horizontal/vertical) of the image consisting of several rows (columns). It is assumed that the signal (scene values) vector sequence along the strip is a vector Markov chain. Since signal correlation in one of the dimensions is not fully used along the edges of the strip, estimates are generated only along the middle sections of the strips. The overlapping strips are chosen such that the union of the middle sections of the strips gives the whole image.

Different versions of the Bayes smoothing algorithm based on different assumptions on the Markov random field are implemented and applied to segmentation of some two level test images. The results are very good even for very low signal-to-noise ratios and compare favorably with existing segmentation algorithms. The algorithms are also applied to remotely sensed

SAR data obtained from SEASAT, yielding remarkably good segmentation of these images as well.

I. INTRODUCTION

Image segmentation and restoration have received enormous attention in the image processing literature. In this report, we present an image segmentation algorithm based on Markov random field (MRF) models for the image and a Bayesian smoothing approach for the actual processing. The image is assumed to be the sum of the realizations of two independent stochastic processes: the "scene", a MRF, and the noise field, consisting of iid (independent identically distributed) Gaussian random variables (r.v.). The Markov scene model provides a powerful mechanism for incorporating the spatial dependence of pixels in relative proximity of each other.

The Bayes smoothing approach used in the algorithm is an extension to 2-D of the 1-D algorithm presented by Askar and Derin [1]. The 1-D Bayes smoothing algorithm is based on a model where the signal is a Markov process corrupted by independent noise. The natural extension to 2-D is to model the scene as a MRF, although extension to 2-D results in severe computational complexity. It is well known that in 1-D the Bayes smoothing estimate performs better (i.e., has a smaller mean-squared error) than the linear smoothing or filtering estimates. In particular, it minimizes the expected cost of the estimate, using all available data and without any restriction of linearity. For a quadratic cost function the Bayes estimate is the a posteriori mean and for uniform cost function the Bayes estimate is the maximum a posteriori (MAP) estimate.

Some of the earlier work on image segmentation and estimation in 2-D is due to Nahi and Assefi [2], Nahi [3], Habibi [4], Nahi and Franco [5], Woods and Radewan [6]. Most of this work involves attempts to extend Kalman filtering ideas to 2-D by devising various 2-D scanning schemes and defining appropriate autoregressive models. Although Bayesian estimation is mentioned

in some of these studies, it is basically a linear filtering or linear smoothing that is being sought. To reduce the prohibitive computational load in 2-D certain reasonable suboptimality assumptions have been proposed by Woods and Radewan [6], namely, Kalman strip processor and reduced update Kalman filtering. Despite the Markovian characteristics of some of these models, the connection with MRF is incidental and undeveloped.

There is a vast body of literature on various spatial interaction models and image restoration and parameter estimation algorithms based on these models, e.g., [7]-[10]. Woods in [7] presents an autoregressive model characterization of causal, unilateral Markov random field models. Chellappa and Kashyap [8], [9] present an extensive exposition of simultaneous auto-regressive (SAR) and conditional Markov (CM) spatial interaction models, and some parameter estimation and neighborhood determination schemes for these models. They also report on linear image restoration algorithms for the SAR and CM models. For a detailed treatment on these and other models, we refer the reader to these references and to the paper by Jain [10], on image processing models, and to references therein.

MRF models, Bayesian approach and MAP formulations in the context of image segmentation have been used before by other researchers. MRF models and MAP formulation is combined in the recent work by Kaufman et al [11] with reduced update Kalman filtering techniques and in Therrien [12], [13] with two dimensional autoregressive texture models for texture based segmentation. Elliott et al [14], Elliott and Srinivasan [15], and Cooper et al [16] have applied MAP techniques to boundary estimation in noisy images. Hansen and Elliott [17] use MRF models, MAP formulation and some simplifying assumptions on the model to devise image segmentation

algorithms based on two stages of dynamic programming. In the recent work by Geman and Geman [18] and Elliott et al [19], the Gibbs distribution characterization of MRF is used to develop image segmentation algorithms.

The image segmentation algorithm in this report is based on a special conditional characterization of MRF's. The objective of the algorithm is to determine recursively the a posteriori distribution and subsequently an estimate of the scene value (pixel intensity) at each pixel given the whole (noise corrupted) image. The algorithm yields the optimal estimate of each pixel value in the scene given the observed image under quite general assumptions, namely, a MRF model for the scene, and independent noise. However, computational complexity of the algorithm in 2-D, necessitates restrictions on the class of MRF models, and suboptimal implementation of the algorithm. The model for the scene is restricted to the special class of MRF's called Markov mesh random fields (MMRF) which are characterized by causal transition distributions. Furthermore the processing is done over relatively narrow strips rather than over the full image as proposed by Woods and Radewan [6] in a Kalman filtering set up. Estimates are obtained at middle sections of the strips and these pieced together from overlapping strips yield an estimate of the full scene. Various versions of the algorithm are applied to some binary test images with varying noise levels and to some SAR images obtained from SEASAT. The segmentation results even for low signal-to-noise ratio images are remarkably good.

There is an important distinction between the a posteriori distribution of the scene value at each pixel based on the full image, which is sought here, and the scene configuration that will maximize the a posteriori distribution of the full scene given the full image, which is the goal of some earlier work. Although the latter may seem more desirable, the full scene

estimate is more constrained, and more susceptible to burst type errors. We have performed extensive simulations in 1-D comparing the performances of these approaches and observed that, in most cases, the individual pixel estimates were just as good as those of the full scene. Furthermore, the former approach provides the additional flexibility to compute other Bayes estimates corresponding to other cost functions, if desired.

This report is organized as follows. In Section II, we give a brief exposition of the Bayes smoothing algorithm in 1-D. Then we present in Section III the problem statement and the MRF model. Extension of the Bayes smoothing algorithm to 2-D and its implementation for Markov mesh random fields is presented in Section IV. In Section V the algorithm is applied to various test images and SAR data, followed by concluding remarks in Section VI.

II. BAYES SMOOTHING ALGORITHM IN 1-D

Let $\{X_k\}$ denote the signal to be estimated. It is assumed that $\{X_k\}$ is a Markov sequence with the known transition probability density function $f(x_k|x_{k-1})$ and initial distribution $f(x_1)$. Note that throughout this report random variables are denoted by capital letters and their realizations by respective lower case letters. The observation at each step is given as

$$Y_k = g_k(X_k, W_k) \qquad k = 1,2,\ldots,N \tag{1}$$

where W_k is the observation noise. It is also assumed that $\{W_k\}$ is an independent sequence, $\{X_k\}$ and $\{W_k\}$ are mutually independent and that g_k is an arbitrary measurable function.

Based on realizations from the observation sequence $y^N = \{y_1,y_2,\ldots,y_N\}$, it is desired to compute an estimate of the signal X_k for each k. A Bayes estimate $\hat{X}_k = \hat{X}_k(Y^N)$, where $Y^N = \{Y_1,Y_2,\ldots,Y_N\}$, is one that minimizes the

expected value of a cost function $C(X_k,\hat{X}_k)$ associated with the estimation procedure, namely

$$E\{C(X_k,\hat{X}_k)\} = \int C(x_k,\hat{x}_k)\ f(x_k,y^N)\ dx_k\ dy^N \tag{2}$$

Equivalently, we seek to minimize

$$E\{C(X_k,\hat{X}_k)|Y^N = y^N\} = \int C(x_k,\hat{x}_k)\ f(x_k|y^N)\ dx_k \tag{3}$$

for each y^N. So it is necessary to determine the a posteriori distribution $f(x_k|y^N)$.

Theoretically, the a posteriori distribution $f(x_k|y^N)$ is determined by the joint statistics of the random sequence $X^N = \{X_1,X_2,\ldots,X_N\}$, the statistics of the noise sequence W^N and the functions g_k. However, in order for this to be of any practical value, the a posteriori distribution $f(x_k|y^N)$ must be determined recursively for $1 \leq k \leq N$. The following are some results along these lines due to Askar and Derin [1].

Theorem 1. Under the assumptions stated above in (1)

$$f(x_k|x_j,y^N) = f(x_k|x_j,y^{j-1}) \qquad k < j \leq N. \tag{4}$$

where $y^k = \{y_1,y_2,\ldots,y_k\}$ is a realization of the random sequence $Y^k = \{Y_1,Y_2,\ldots,Y_k\}$.

Theorem 2. Under the assumptions stated above in (1)

$$f(x_k|y^N) = f(x_k|y^k) \int \frac{f(x_{k+1}|x_k)}{f(x_{k+1}|y^k)}\ f(x_{k+1}|y^N)\ dx_{k+1} \tag{5}$$

where recursive relationships for the conditional filtered and predicted densities $f(x_k|y^k)$ and $f(x_{k+1}|y^k)$ respectively were determined by Ho and Lee [20] as

$$f(x_k|y^k) = \frac{f(y_k|x_k)\ f(x_k|y^{k-1})}{\int f(y_k|x_k)\ f(x_k|y^{k-1})\ dx_k} \tag{6}$$

and

$$f(x_{k+1}|y^k) = \int f(x_{k+1}|x_k)\ f(x_k|y^k)\ dx_k \tag{7}$$

The proof of Theorem 1 and proofs of the recursive relationships (5) - (7) follow from Bayes' rule and model assumptions. The smoothing algorithm given in (5) is a backward algorithm where k is running from N-1 to 1. So the smoothing algorithm involves two passes over the data, one forward pass during which the conditional filtered and predicted densities are computed according to (6) and (7) and stored, and one backward pass during which the smoothing a posteriori densities are computed. Note that during the backward pass the observation sequence y^N is not explicitly used.

The general implementation of this smoothing algorithm could lead to storage and processing difficulties. However, under Gaussian assumptions the densities are characterized by a few parameters, and computation and storage is straightforward. Also, if the signal $\{X_k\}$ is a sequence of r.v.'s taking only finitely many values, then the integrals reduce to finite summations and the implementation of the algorithm is relatively straightforward.

In this smoothing algorithm, attention is concentrated on computing $f(x_k|y^N)$ the a posteriori distribution of individual signal values, given the full observation set y^N. An alternative Bayes estimation scheme would be to compute $f(x^N|y^N)$ (or equivalently $f(x^N,y^N)$) yielding a Bayes estimate of the whole signal sequence. Recursive expressions for these densities can easily be written down for models at hand. However, the actual implementation requires computations and memory in the order of K^N where K is the number of values r.v. X_1 takes and hence it is not feasible to compute

$f(x^N|y^N)$ unless for very short sequences, even for binary r.v.'s. On the other hand, the x^N that will maximize $f(x^N|y^N)$ can be computed using dynamic programming, thus obtaining a MAP estimate of X^N. Such a MAP formulation is presented by Scharf and Elliott [21] and used in image segmentation by Hansen and Elliott [17].

An extensive simulation in 1-D was carried out and the performance of the two approaches on different binary signal sequences were compared. It was observed that, taking the mode of $f(x_k|y^N)$ as the estimate for X_k for $k = 1,2,\ldots,N$ and then piecing them together to get the estimate for the full scene X^N performs just as well, if not better, in most cases as the MAP estimate obtained by maximizing $f(x^N|y^N)$. Of course, both estimates performed much better than a simple threshold comparison estimate which does not exploit the dependence in the signal sequence.

III. PROBLEM STATEMENT AND MARKOV RANDOM FIELD MODELS

Problem Statement

Suppose the image is a random field $Y = \{Y_{ij}\}$ defined over a finite $N_1 \times N_2$ lattice of points (pixels) defined as $L = \{(i,j): 1 \leq i \leq N_1, 1 \leq j \leq N_2\}$. Suppose further that the image random field Y is a function of a scene (true picture) random field X and a corruptive noise random field W, each defined over the same $N_1 \times N_2$ lattice. The functional relationship between the random fields is such that at each pixel the image r.v. is a function of the scene r.v. and the noise r.v. at that pixel, that is

$$Y_{ij} = g_{ij}(X_{ij}, W_{ij}) \qquad (k,j) \in L \tag{8}$$

Although, with assumptions that will be stated below, the model allows for an arbitrary measurable function, $g_{ij}(\cdot,\cdot)$ e.g., g_{ij} that can model satur-

ation, clipping or non-linearity effects, the interest is mainly in the case where the image is the scene plus noise. In other words,

$$Y_{ij} = X_{ij} + W_{ij} \qquad (i,j) \in L \tag{9}$$

or in matrix form

$$Y = X + W \tag{10}$$

It is assumed that X is a Markov random field (MRF), that W consists of independent r.v.'s at each pixel, and that W and X are mutually independent. The common distribution of the W_{ij} and the joint distribution of X is unspecified.

If X_{ij} represents the brightness level of the scene at pixel (i,j), it is reasonable to assume that X_{ij} will depend mostly on the values of neighboring pixels and that the correlation between the values of two pixels decreases with distance. This spatial continuity is inherent in the MRF model, in which the neighborhood structure can be specified according to the degree of spatial continuity desired.

Our objective is to determine $f(x_{ij}|y)$, the a posteriori distribution of the scene at pixel (i,j) given the full image y, for each pixel in the lattice. From these, an estimate of each X_{ij} and finally combining them an estimate for the full scene X is obtained. In this study the estimate of X_{ij} is taken as the mode of the a posteriori distribution $f(x_{ij}|y)$, which is a reasonable choice especially for binary scenes the algorithm is applied to.

For the model under consideration the a posteriori distribution $f(x_{ij}|y)$ can be determined recursively, as will be explained in Section IV. However, due to the amount of computation, certain approximations are made, such as

processing the image in sections (strips). These and other considerations, which are discussed in Section IV, lead naturally to certain subclasses of MRF's called Markov mesh random fields (MMRF).

Markov Random Fields (MRF)

The foundations of MRF lies in the physics literature on ferromagnetism, originating in the work of Ising [22] in 1925. We will give only the basic definitions; for a detailed treatment we refer the reader to Kinderman and Snell [23].

Definition 3.1. Let L be a finite lattice

$$L = \{(i,j): 1 \leq i \leq N_1, 1 \leq j \leq N_2\} \tag{11}$$

Neighborhood of a ϵ L, denoted by η_a, is defined as

$$\eta_a = \{b \epsilon L: \ b \neq a \text{ and } a \epsilon \eta_b\} \quad \forall a \epsilon L \tag{12}$$

And, a neighborhood system on L is defined as

$$\P = \{\eta_a : a \epsilon L \tag{13}$$

Definition 3.2. Given a finite lattice L and a neighborhood system ¶ on L, a random field $X = \{X_a, a \epsilon L\}$ is a Markov random field if and only if

$$P(X_a = x_a | X_b = x_b, b \epsilon L - \{a\}) = P(X_a = x_a | X_b = x_b, b \epsilon \eta_a) \tag{14}$$

for each a ϵ L.

Elaborate spatial dependence can be modelled by suitably large neighborhood systems. Indeed, any random field L is a MRF with respect to the neighborhood system ¶ for which $\eta_{ij} = L$, $\forall (i,j) \epsilon L$. However, relatively simple neighborhood systems are adequate in modelling most scenes of interest, and we are mainly interested in the two simplest neighborhood systems:

(i) the 4 neighbor system $\P^1$, and (ii) the 8 neighbor system $\P^2$, both depicted in Figure 1. Certain obvious adjustments need to be made for the neighborhoods of the pixels along the boundaries.

The probabilities on the left hand side of (14) are called the "local characteristics" of the field, and it can be shown that, for <u>any</u> random field, these local characteristics uniquely determine the joint distribution P(X). Now every MRF (with $P(X = x) > 0$ for all x) is also a so-called "Gibbs distribution" (and vice versa), and this equivalence allows us to write down explicitly the joint distribution $P(X = x)$.

It also follows from the Gibbs distribution equivalence that the joint distribution of a MRF with $\P^1$ or $\P^2$ neighborhood system can be expressed as

$$P(X = x) = \prod_{(i,j)\in L} \rho_{ij}(x_{ij}, x_{i-1,j}, x_{i,j-1}, x_{i-1,j-1}) \qquad (15)$$

with necessary adjustments for $i = 0$ or $j = 0$ or both. The expression in (15) for $P(X = x)$ can be obtained as an extension of the Hammersley and Clifford Theorem (see Besag [24]). Note that ρ_{ij}'s in (15) are not unique and calculation of a set of them may be extremely difficult due to the necessary normalization. Fortunately computation of ρ_{ij}'s is not necessary. The important point is that such a representation for $P(X = x)$ exists.

The following are implied by the representation in (15) and will be used in the smoothing algorithm.

<u>Theorem 3.1</u>. Let $X = \{X_a, a\in L\}$ be a MRF on lattice L with neighborhood system $\P^1$ or $\P^2$. The columns (and rows) of this MRF constitute a vector Markov chain.

This is an interesting and potentially useful result concerning finite lattice MRF's with $\P^1$ or $\P^2$ neighborhood systems.

Theorem 3.2. For $\{X_a, a\varepsilon L\}$ a MRF on lattice L with neighborhood system $\P^1$ or $\P^2$

$$P(x_a, a\varepsilon A | x_b, b\varepsilon B) = P(x_a, a\varepsilon A | x_b, b\varepsilon C) \tag{16}$$

where A, B, and C are subsets of L shown in Figure 2.

η^1 - 4 neighbor system η^2 - 8 neighbor system

Figure 1. η^1 and η^2 neighborhood systems

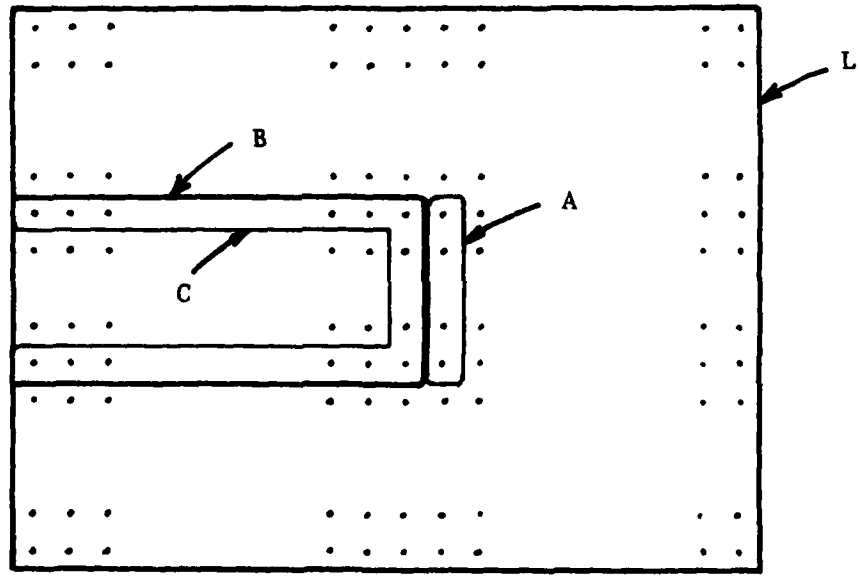


Figure 2. Subsets of L that are used in Theorem 3.2

Markov Mesh Random Field (MMRF)

Neither the conditional distribution nor the joint distribution characterizations of MRF's are immediately applicable in the Bayes smoothing algorithm that is being proposed. What is needed is a characterization more directly causal than (15), which brings us to the so-called Markov mesh random fields (MMRF) introduced by Abend et al [25] and Kanal [26]. Similar causal, unilateral characterizations have also been used by Kashyap and Chellappa [8], [9].

Let L be the lattice described above. We define

$$A_{i,j} = \{(k,\ell)\}L \mid k < i \text{ or } \ell < j\} \quad \text{for } (i,j)\epsilon L$$
$$B_{i,j} = \{(k,\ell)\epsilon L \mid k \leq i \text{ and } \ell \leq j\ ,\ (k,\ell)\neq(i,j)\} \text{ for } (i,j)\epsilon L \tag{17}$$

A MMRF is defined by the property

$$P(x_{ij} \mid x_a,\ a\epsilon A_{i,j}) = P(x_{ij} \mid x_a, a\epsilon C_{i,j}), \tag{18}$$

where $C_{i,j}$ is a subset of $B_{i,j}$ and determines the particular type of MMRF.

Such a random field has the following, very useful property:

$$P(X = x) = \prod_{(i,j)\epsilon L} P(x_{ij} \mid x_a,\ a\epsilon C_{i,j}) \tag{19}$$

This is the causal relationship we need for the smoothing algorithm. Corresponding to any choice of $\{C_{i,j}\}$, there is a family $\{D_{i,j}\}$ such that the MMRF satisfies

$$P(X_{ij} = x_{ij} \mid X_a = x_a,\ a\epsilon L - \{(i,j)\})$$
$$= P(X_{ij} = x_{ij} \mid X_a = x_a,\ a\epsilon D_{i,j}) \tag{20}$$

Moreover, the $\{D_{i,j}\}$ constitute a neighborhood system, and hence it follows from (20) that a MMRF is a MRF. The $\{D_{i,j}\}$ corresponding to a particular

choice of $\{C_{i,j}\}$ can be determined using (19).

Of special interest to us is the case in which $C_{i,j} = \{(i,j-1),(i-1,j-1),(i-1,j)\}$. The corresponding $\{D_{i,j}\}$ is the 8-neighbor system $\P^2$. For this choice of $C_{i,j}$, the joint distribution in (19) is equivalent to (15), a joint distribution expression for a MRF with $\P^1$ or $\P^2$ neighborhood system. The subsets of the lattice relevant to a MMRF and the particular $C_{i,j}$ are shown in Figure 3. Some of the interesting properties of this particular MMRF are given below:

Theorem 3.3. For a MMRF with $C_{i,j} = \{(i,j-1),(i-1,j-1),(i-1,j)\}$ the following are true:

$$\text{(i)} \quad P(X_a = x_a,\ a\varepsilon E_{imj} | X_b = x_b,\ b\varepsilon F_{i,j}) = P(X_a = x_a,\ a\varepsilon E_{i,j} | X_b = x_b,\ b\varepsilon G_{i,j}) = \prod_{a\varepsilon E_{i,j}} P(X_a = x_a | X_b = x_b,\ b\varepsilon C_a) \tag{21}$$

$$\text{(ii)} \quad P(X_a = x_a,\ a\varepsilon E_{i,j} | X_b = x_b,\ b\varepsilon H_{i,j}) = P(X_a = x_a,\ a\varepsilon E_{i,j} | X_b = x_b,\ b\varepsilon K_{i,j}) = P(x_{ij} | x_{i,j-1}, x_{i,j-2}, \ldots, x_{i,1}) \prod_{a\varepsilon E_{i,j}-\{i,j\}} P(x_a | x_b,\ b\varepsilon C_a) \tag{22}$$

The sets $E_{i,j}$, $F_{i,j}$, $G_{i,j}$, $H_{i,j}$ and $K_{i,j}$ are shown in Figure 4.

The utility of the MMRF model which stems from the causal relationships above (especially of course (19)), are already discussed in regard to strip processing. In addition, these models may be readily simulated and constitute a surprisingly rich subclass of MRF's.

Pickard Random Field

There is yet another interesting subclass of MRF's first noted by Pickard [27] as a "curious binary lattice process". Let us call it a Pickard random field (PRF). It is fully characterized by the joint distribution

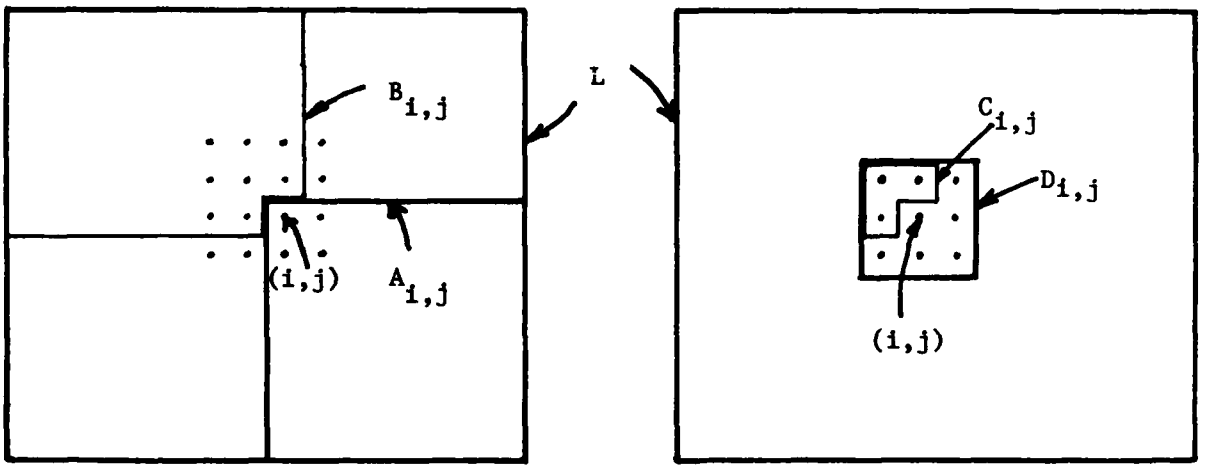


Figure 3. Subsets of L describing a MMRF

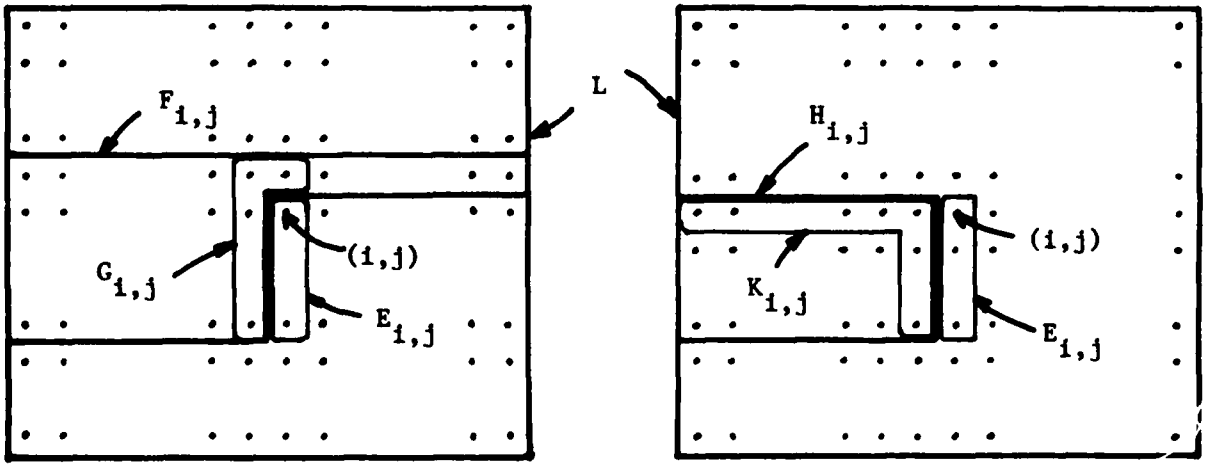


Figure 4. Subsets of L that are used in Theorem 3.3

of four r.v.'s on a 2 x 2 block, $\begin{pmatrix} X_a & X_b \\ X_c & X_d \end{pmatrix}$, with the property that

$$P(X_b = x_b | X_a = x_a, X_c = x_c) = P(X_b = x_b | X_a = x_a) \tag{23}$$

For the binary case, the joint distribution of the block is then completely specified by three parameters: $P(X_a = 1)$, $P(X_b = 1 | X_a = 1)$ and $P(X_d = 1 | X_c = 1, X_a = 1, X_b = 1)$, subject to certain constraints.

A PRF is then defined constructively. From the joint distribution of the 2 x 2 block, initial and transition distributions of a 2-dimensional vector Markov chain are generated, giving rise to a 2 x N_1 random field, namely this chain considered over the index set $[1,2,\ldots,N_1]$. Finally initial and transition distributions of an N_1 - dimensional vector Markov chain are generated, thus providing in the usual way the joint distribution of an $N_1 \times N_2$ random field. The random field so generated can easily be made to be stationary and isotropic.

Here are some of the relevant properties of a PRF. For a detailed treatment, including some of the proofs see Pickard [27].

Theorem 3.4. The PRF described above has the following properties:

(i) The rows (columns) in each set of k consecutive columns (rows) of the random field form a k-dimensional vector Markov chain.

(ii) $$P(X_a = x_a, a \varepsilon E_{i,j} | X_b = x_b, b \varepsilon M_{i,j}) = P(X_a = x_a, a \varepsilon E_{i,j} | X_b = x_b, b \varepsilon Q_{i,j}) \tag{24}$$

(iii) $$P(X_a = x_a, a \varepsilon E_{i,j} | X_b = x_b, b \varepsilon H_{i,j}) = P(X_a = x_a, a \varepsilon E_{i,j} | X_b = x_b, b \varepsilon Q_{i,j}) \tag{25}$$

Sets $E_{i,j}$, $M_{i,j}$, $Q_{i,j}$, and $H_{i,j}$ are shown in Figure 5.

Theorem 3.5. Any PRF is an MIRF, and hence an MRF.

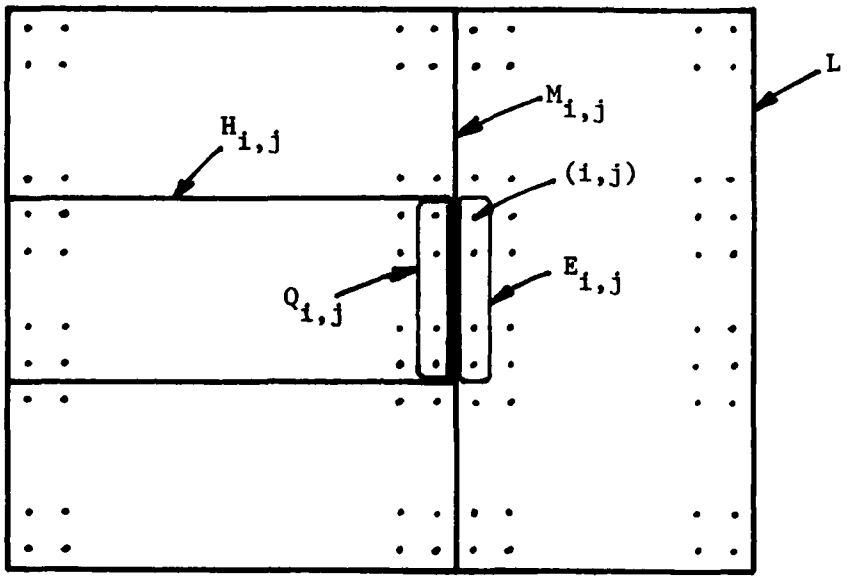


Figure 5. Subsets of L that are used in Theorem 3.4

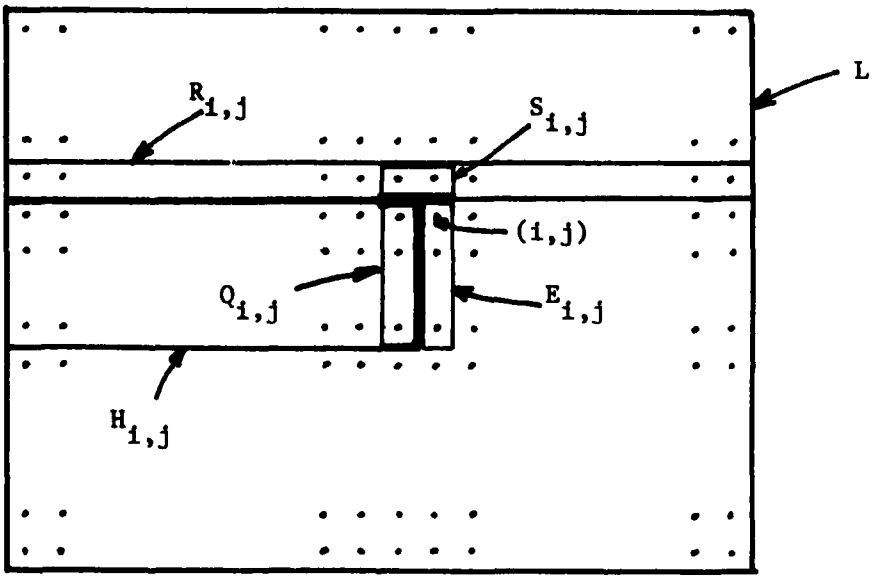


Figure 6. Subsets of L that are used in (28)

The proof of this interesting result is lengthy but straightforward and hence will not be included here.

Actually, the PRF is too special to model most of the scenes of interest to us, even after an extension past the binary case. We have included this brief discussion in order to illustrate, in an especially concrete way, various properties of PRF's that are relevant for the smoothing algorithm described in the next section.

IV. BAYES SMOOTHING ALGORITHM IN 2-D

Recall that the objective of the algorithm is to determine the a posteriori distribution $f(x_{ij}|y)$ recursively, where the image Y consists of a MRF X on the $N_1 \times N_2$ lattice corrupted by additive independent noise, see (8) - (10) in Section III.

The Bayes smoothing algorithm in 1-D (described in Section II) remains valid if all quantities involved, namely X_k, W_k and Y_k, are random vectors, so that $\{X_k\}$ is a vector Markov chain and $\{W_k\}$ is a sequence of iid random vectors and independent of $\{X_k\}$. The components of the random vectors $\{W_k\}$ need not be independent, but the computation of $f(y_k|x_k)$ is drastically simplified when they are. Thus, straightforward application of the 1-D Bayes smoothing algorithm on the vector processes recursively, yields $f(x_k|y)$, the a posteriori distribution of the random vector X_k (possibly a column of the scene) given the observation random field Y (possibly the noise corrupted image). The a posteriori distribution $f(x_{ki}|y)$ of the r.v. X_{ki}, the ith entry of the kth vector, can be simply obtained from $f(x_k|y)$ by integrating over the other components of X_k.

Thus, we have a recursive scheme to determine the a posteriori distribution of the individual pixel values given the noisy image, under the addi-

tional assumption that columns of the scene constitute a vector Markov chain. We have pointed out in Section III that the columns (or rows) of a MRF with $\P^1$ or $\P^2$ neighborhood system do constitute a vector Markov chain and thus for such MRF's the a posteriori distribution of individual pixel values given the full image can be obtained exactly. Specifying a cost function and determining the corresponding estimate for each pixel, this is the "best" that can be done in estimating pixel values using the observed image under the assumptions of the model.

Despite the optimistic tone of the previous paragraph there are overwhelming computional difficulties in implementing the smoothing algorithm on the columns of an image matrix. Even for a binary scene on a 128 x 128 lattice, at each step of the algorithm there are on the order of 2^{128} calculations necessary, because the column vector X_k has 2^{128} possible values. Obviously, drastic simplifications are necessary.

Inspired by the ideas of Woods and Radewan [6] on Kalman strip processor, we propose to process the image in relatively narrow strips, i.e., groups of rows. Implementing the algorithm on a strip will yield $f(x_k'|y')$, the a posteriori distribution of X_k', the kth column restricted to the strip, given $Y' = y'$, Y' being the restriction of the image to the strip. Clearly, $f(x_k'|y') \neq f(x_k'|y)$. However, due to the fact that the dependence between pixels usually decreases as the distance between pixels increases, we will assume that the a posteriori distribution of the pixels along the center section of the strip will not be influenced very much by the observed image values outside the strip. Denoting by X_k'' the restriction of X_k' to the center section of the strip, it is thus assumed that

$$f(x_k''|y) \doteq f(x_k''|y') \tag{26}$$

This is a reasonable assumption whenever pixels that make up X_k'' are "far away" from those outside the strip, as measured by the effective "correlation length" of the field.

The a posteriori distribution $f(x_k'|y')$ is simply obtained by applying the algorithm on a strip. Then integrating $f(x_k'|y')$ over the entries of X_k' that are outside the center section of the strip gives $f(x_k''|y')$, which in turn is approximately equal to $f(x_k''|y)$. The a posteriori distribution of the individual pixel values that are entries of X_k'' are simply obtained by integrating $f(x_k''|y')$ over all the other entries of X_k''. The estimates of the pixel values that are entries of X_k'' are then determined from the corresponding a posteriori distribution to be the mode or the mean or some other parameter of the a posteriori distribution depending on the specified cost function or the estimation strategy.

The strips are arranged such that the union of the center sections is equal to the full scene. Thus, the estimates along center sections combined provide an estimate of the full scene. The processing along the strips can be done simultaneously.

A drawback to this approach is the assumption that the column vectors $\{X_k'\}$ along the strip constitute a vector Markov chain, which is not necessarily true for an arbitrary MRF. Moreover, even if $\{X_k'\}$ is a vector Markov vector chain, the transition distribution $P(x_k'|x_{k-1}')$ of this vector Markov chain must be known in order to be able to use the algorithm.

To address both of these issues we concentrate on the Markov mesh random field models discussed in Section III. Looking at Theorem 3.3 (ii), we see that the column vectors along a strip are not exactly a vector Markov chain. But if we overlook the dependence of X_{ij} on pixels other than the adjacent one, in other words, if we assume that

$$P(x_{ij}|x_{i,j-1},x_{i,j-2},\cdots,x_{i,1}) \doteq P(x_{ij}|x_{i,j-1}) \tag{27}$$

then it follows from (22) that the columns along the strip do constitute a vector Markov chain and (22) gives a simple explicit expression to compute $P(x'_k|x'_{k-1})$ in terms of $P(x_{ij}|x_b,\ b\varepsilon C_{i,j})$'s, the causal transition distribution characterizing the MMRF, and the transition distribution $P(x_{ij}|x_{i,j-1})$.

Aside from the approximation expressed in (27) which is not an unreasonable one in a Markov setting, starting from a joint distribution on a 2 x 2 block, the building blocks of the MMRF, $P(x_{ij}|x_b,\ b\varepsilon C_{i,j})$ for $C_{i,j} = \{(i,j-1),(i-1,j-1),(i-1,j)\}$ and $P(x_{ij}|x_{i,j-1})$ needed in (22) are computed and hence $P(x'_k|x'_{k-1})$ is computed using (22). Thus, with a MMRF model and with the approximation of (27) the Bayes smoothing algorithm is readily implemented over strips of the image. Let us denote this version of the algorithm, in which strips can be processed simultaneously, by Algorithm A.

It should be noted that for Pickard random field (PRF) described in Section III, the desired properties of the strip processor, namely

(i) columns along a strip are a vector Markov chain,

(ii) $P(x'_k|x'_{k-1})$ is explicitly determined

are both valid <u>without</u> any approximation. But the PRF model is so restrictive that the segmentation so obtained for our scenes is not as good as that obtained by various MMRF models, making the approximation in (27).

The restrictive nature of the PRF is illustrated by the following example. Consider a binary scene with $P(X_{ij} = 1) = 0.5$ and $P(X_{ij} = 1|X_{i,j-1} = 1) = 0.8$. It follows from the model constraints that $P(X_{ij} = 1|X_{i,j-1} = 1,\ X_{i-1,j-1} = 1,\ X_{i-1,j} = 1) \geq 0.9375$ and that it has to be less than 0.94 for spatial continuity constraints such as $P(X_{i,j} = 1|X_{i,j-1} = 0,\ X_{i-1,j-1} = 1,\ X_{i-1,j} = 0)$ being relatively small (≤ 0.05). For these reasons, the PRF model will not be

discussed any further.

Alternate Version of Bayes Strip Processor Algorithm

For a MMRF with $C_{i,j} = \{(i,j-1), (i-1,j-1), (i-1,j)\}$ the following relationship holds:

$$P(X_a = x_a,\ a\varepsilon E_{i,j} | X_b = x_b,\ b\varepsilon H_{i,j} UR_{i,j})$$

$$= P(X_a = x_a,\ a\varepsilon E_{imj} | X_b = x_b,\ b\varepsilon Q_{i,j} US_{i,j})$$

$$= \prod_{a\varepsilon E_{i,j}} P(x_a | x_b,\ b\varepsilon C_a) \qquad (28)$$

The sets $E_{i,j}$, $H_{i,j}$, $R_{i,j}$, $Q_{i,j}$ and $S_{i,j}$ are shown in Figure 6. In simple terms, this means that given the row above the strip, the columns of the strip form a vector Markov chain, and that $P(x'_{k+1} | x'_k, \underline{x})$, where $\underline{x}$ denotes the row above the strip, is readily obtained in terms of the building blocks (causal transition distribution) $P(x_a | x_b, b\varepsilon C_a)$ of the MMRF.

It is also true that the Bayes smoothing algorithm (4) - (7) described in Section II is still valid for a MMRF model with vector processes (a strip) and with $\underline{x}$, the row above the strip, given, meaning that $\underline{x}$ is simply added to all the conditional distributions in (4) - (7). The transition distribution $P(x'_{k+1} | x'_k, \underline{x})$ is computed using (28). So the algorithm in this case yields $f(x'_k | y', \underline{x})$ recursively.

In order to apply this version of the algorithm, however, we need to know $\underline{x}$ the row above the strip. Clearly this information is not available. But an estimate $\hat{\underline{x}}$ of $\underline{x}$ is determined during the processing of a previous strip. So if the estimate $\hat{\underline{x}}$ is used instead of $\underline{x}$, $f(x'_k | y', \hat{\underline{x}})$ is obtained as an approximation to the a posteriori distribution $f(x'_k | y', \underline{x})$. Then integrating $f(x'_k | y', \hat{\underline{x}})$ over the variables that are not to be estimated, the a posteriori distribution of the variables to be estimated is obtained.

Using the estimate of the row above the strip is equivalent to adjoining a boundary condition to the segmentation problem. The variables to be estimated along the strip could again be along the center section of the strip or could be along the top of the strip adjacent to the "row above", whose estimate is used. Let us refer to these two new versions of the algorithm as Algorithm B and Algorithm B' respectively. Note that Algorithms B and B' have a sequential nature because of the use of the previous estimate $\hat{\underline{x}}$.

In Section V, we will report on the application of Algorithms A, B and B' to obtain binary segmentation of some noisy test images and some remotely sensed SAR images obtained by SEASAT.

At this point, it would be in order to summarize all the assumptions and approximations made to obtain an implementable 2-D Bayes smoothing algorithm. Aside from the general assumptions of the scene being a MRF on a finite lattice corrupted by additive independent noise the following assumptions and approximations are made:

i. Scene is a MRF with $\P^1$ or $\P^2$ neighborhood system.

ii. Scene is a MMRF with "support set" $C_{i,j} = \{(i,j-1), (i-1,j-1), (i-1,j)\}$ (a special case of a $\P^2$ neighborhood MRF).

iii. Pixels far apart tend to be independent and as a consequence of this $f(x''_k|y) \doteq f(x''_k|y')$.

iv. $P(x_{ij}|x_{i,j-1},x_{i,j-2},\ldots,x_{i,1}) \doteq P(x_{ij}|x_{i,j-1})$ (assumed in Algorithm A).

v. $\hat{\underline{x}}$, estimate of the row above the strip is used instead of $\underline{x}$, the row above itself, in all the conditional distributions (assumed in Algorithms B and B').

V. BINARY SEGMENTATION USING BAYES SMOOTHING STRIP PROCESSOR ALGORITHMS

The scene X is assumed to be a binary (equivalently two-level: ℓ_1, ℓ_2) MMRF with support set $C_{i,j} = \{(i,j-1), (i-1,j-1), (i-1,j)\}$. The joint distribution of a 2 x 2 block from the scene is specified as follows:

$$P\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = q_1, \quad P\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = q_2, \quad P\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = q_3$$

$$P\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = q_4, \quad P\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = q_5, \quad P\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = q_6 \tag{29}$$

The joint distribution of the 2 x 2 block is invariant under multiples of 90° rotations. Various transition probabilities necessary for the algorithms are determined in terms of the parameters $q_1, q_2, .., q_6$. Different degrees of spatial continuity can be modelled by a proper specification of q_i parameters.

The scene is corrupted by additive, independent, zero-mean Gaussian noise. In other words, the r.v.'s constituting the noise field are assumed to be iid and $N(0,\sigma^2)$. For the sake of the Computer Vision system used it is further assumed that the noise takes discrete values with probabilities proportional to that of $N(0,\sigma^2)$, upon proper normalization. The segmentation algorithm is applied to images with varying noise levels. The signal to noise ratio (SNR) of an image is defined as

$$\text{SNR} = \frac{\ell_2 - \ell_1}{\sigma} \tag{30}$$

where ℓ_1 and ℓ_2 are the levels of the binary scene. For the test images segmented the quantities ℓ_1, ℓ_2 and σ are all prespecified and hence known. For the real SAR images, they are estimated by assuming that the image is a mixture of two Gaussian distributions with means ℓ_1 and ℓ_2 and variance σ^2.

In all the segmentation examples reported here, the strips are taken as 3-row wide with the "center section" of the strip being the middle row. Algorithm can be implemented with greater strip widths but the computation times become prohibitive for strip width greater than 5. However, as will be illustrated, remarkably good segmentation results are attained even with 3-row wide strips. Both the test images and the SAR images segmented are taken to be 64 x 64. Larger images can be processed with computation times proportional to the total number of pixels in the image. Segmentation of 64 x 64 images with 3-row wide strips takes in the order of 30 seconds on a Computer Vision System supported by a VAX 11/780 computer.

Taking 3-row wide strips in Algorithm A the scene values are estimated along the center row using the (noisy) image values on the 3 row strip. In Algorithm B, similarly center row scene values are estimated using the 3 row strip and the estimate of the "row above". In Algorithm B', again using the image on the 3-row strip and the estimate of the "row above", scene values along the top row of the strip are estimated. In all cases, the estimate is taken as the mode of the corresponding a posteriori distribution.

Following the segmentation, the estimated scene is post processed by a 3 x 3 median filter. This post processing helps eliminate some of the specks, i.e. the isolated, presumed to be erroneous estimates. But while eliminating erroneous specks some new errors may be generated due to a resulting decrease in scene detail. A desired degree of scene detail together with possibly erroneous speckles can be attained by a proper choice of the q_i parameters. Different segmentation results based on different sets of q_i parameters are illustrated in the figures.

In Figures 7-22, segmentation of some test images with different SNR's using Algorithms A, B and B' are presented. The arrangement of the four sections of these images is described in Table I. The SNR of the noisy image, the type of the algorithm used and the q_i parameters used are all specified separately for each Figure. The three versions of the algorithm yield similar results for high to medium SNR's but show distinctively different characteristics for low to very low SNR's. It is seen from these Figures that segmentation results are excellent for high to medium SNR's (e.g., SNR = 7 or 1) and resonably good for low-to-very low SNR's (e.g. SNR - 0.7 or 0.5). Some directionality features are apparent in the segmentation for low SNR's. It is believed that this undesirable effect can be eliminated or reduced by increasing the strip width and by properly choosing the q_i parameters.

During experimentation with the segmentation algorithm, we observed that the results are somewhat sensitive to the parameters q_i's. The set of parameters which yields a "good" segmentation depends on the noise level and the level of detail (object sizes) in the image. In this study,

TABLE I

DESCRIPTION FOR FIGURES 7-22

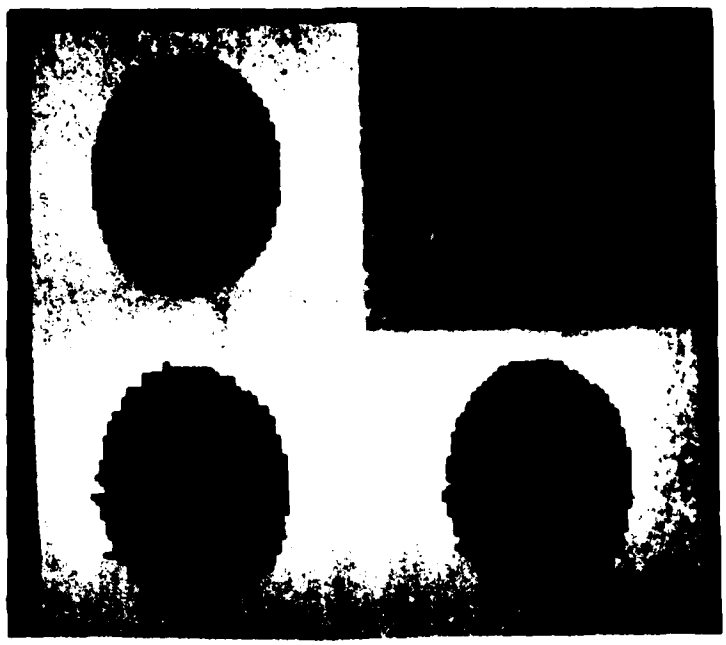
(a) Noiseless test image.

(b) Gaussian noise added to (a) with specified SNR.

(c) Segmentation of (b) by the specified algorithm.

(d) (c) Post-filtered by 3 x 3 median filter.

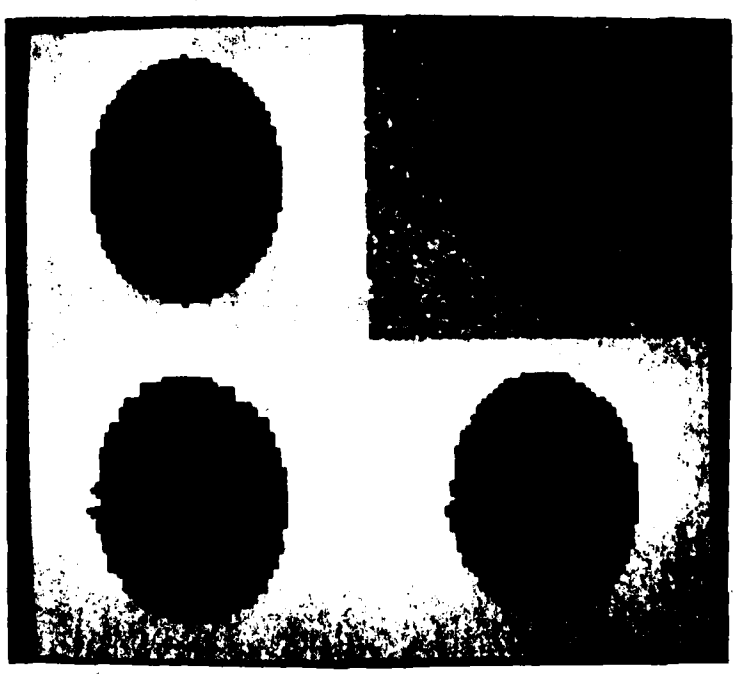
a b



c d

Figure 7. Segmentation with Algorithm A, SNR=2, q_1=0.44, q_2=q_3=q_5=0.005, a_4=ϵ and q_6=0.5.

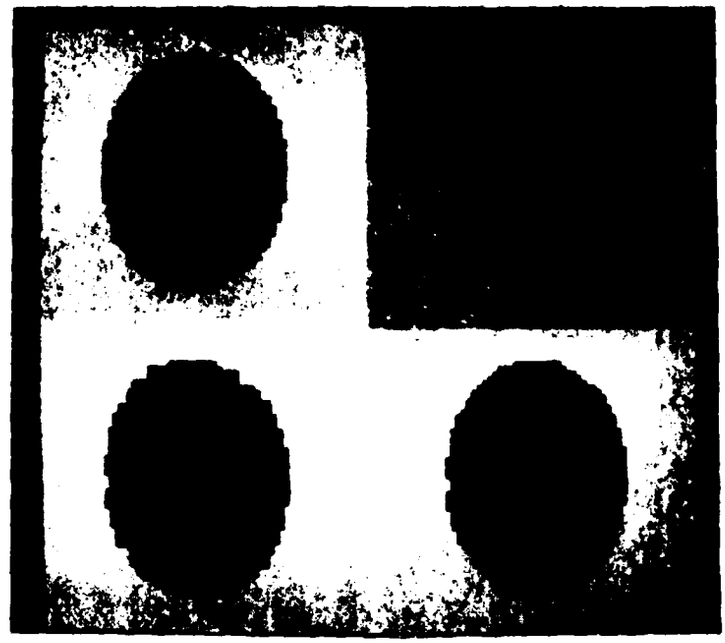
a b



c d

Figure 8. Segmentation with Algorithm B, SNR=2, q_1=0.42, q_2=q_5=0.005, q_3=0.01, q_4=ϵ, and q_6=0.5.

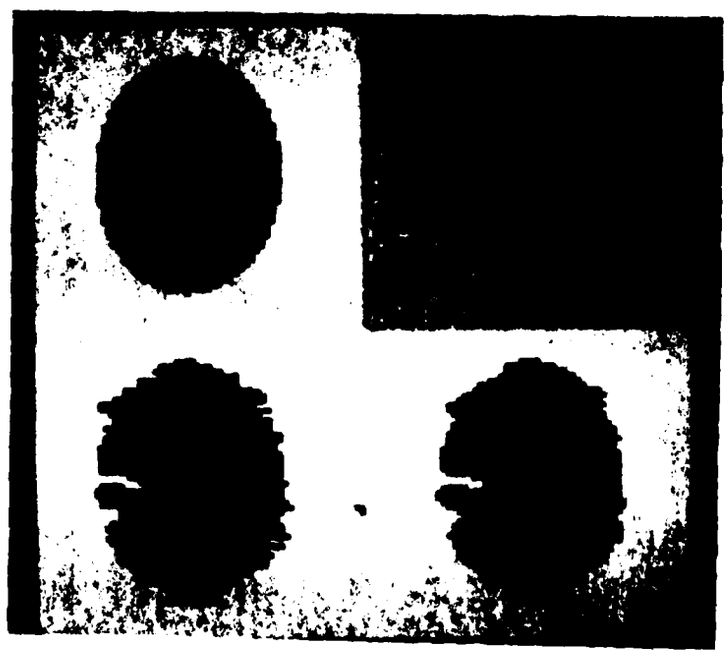
a b



c d

Figure 9. Segmentation with Algorithm B', SNR=2, q_1=0.42, q_2=q_5=0.005, q_3=0.01, q_4=ϵ, and q_6=0.5.

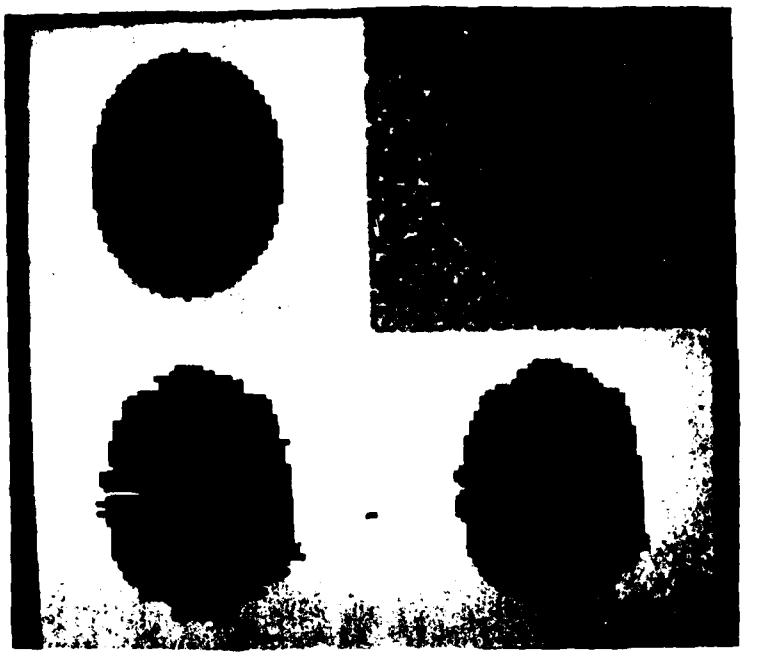
a b



c d

Figure 10. Segmentation with Algorithm A, SNR=1, q_1=0.44, q_2=q_3=q_5=0.005, q_4=ϵ, and q_6=0.50.

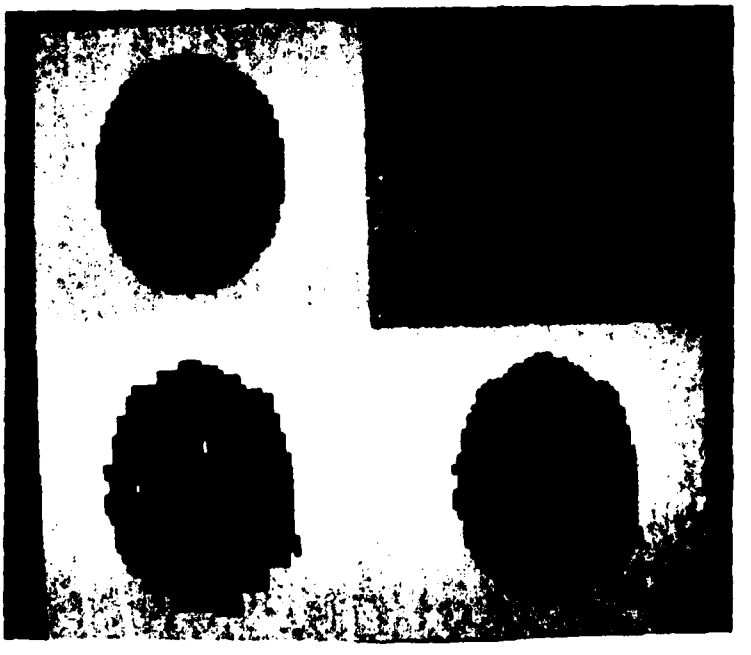
a b



c d

Figure 11. Segmentation with Algorithm B, SNR=1, q_1=0.42, q_2=q_5=0.005, q_3=0.01, q_4=ε, and q_6=0.50.

a b



c d

Figure 12. Segmentation with Algorithm B', SNR=1, q_1=0.38, q_2=q_5=0.01, q_3=0.015, q_4=ε, and q_6=0.48.

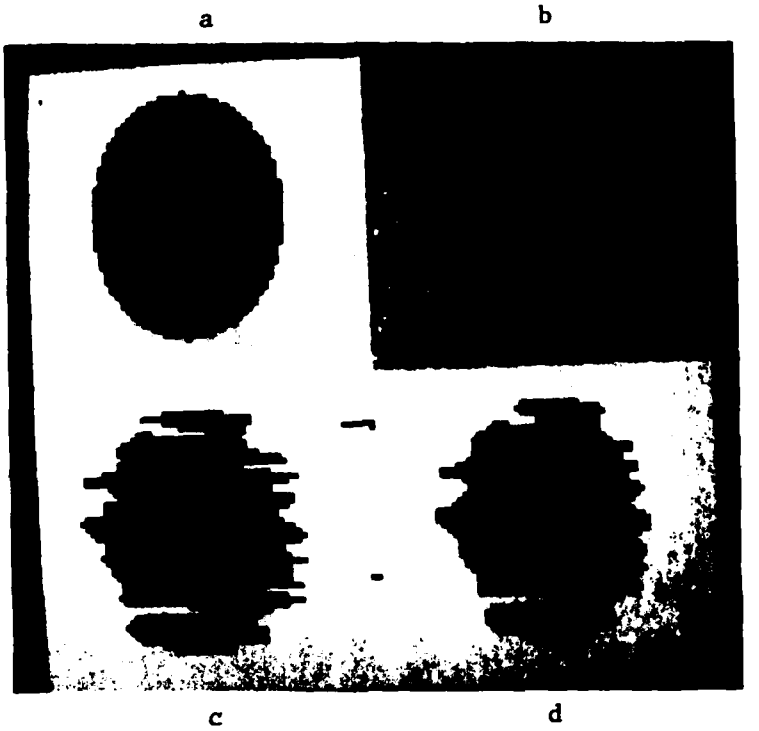


Figure 13. Segmentation with Algorithm A, SNR=0.7, q_1=0.3, q_2=q_5=0.005, q_3=0.01, q_4=ε, and q_6=0.62.

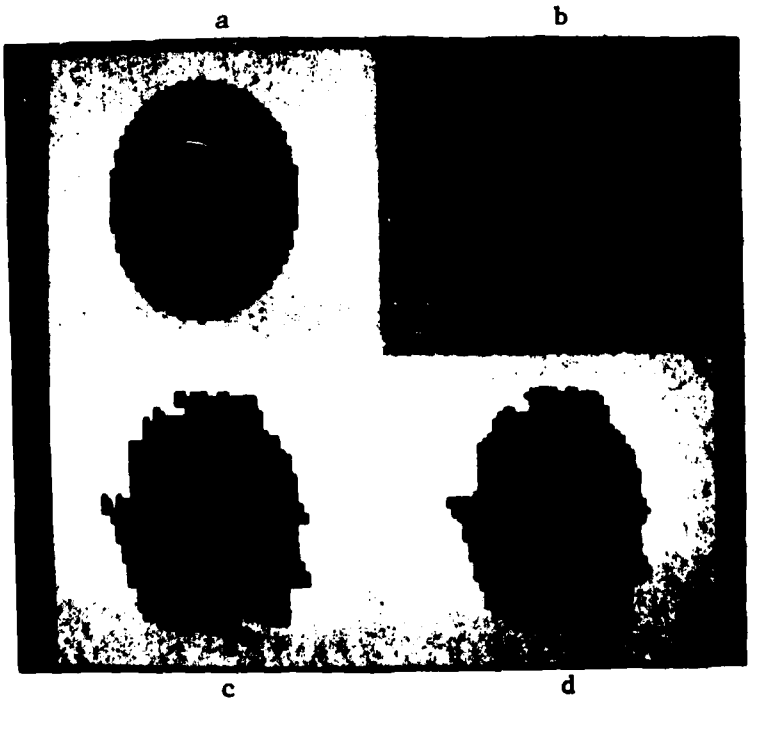
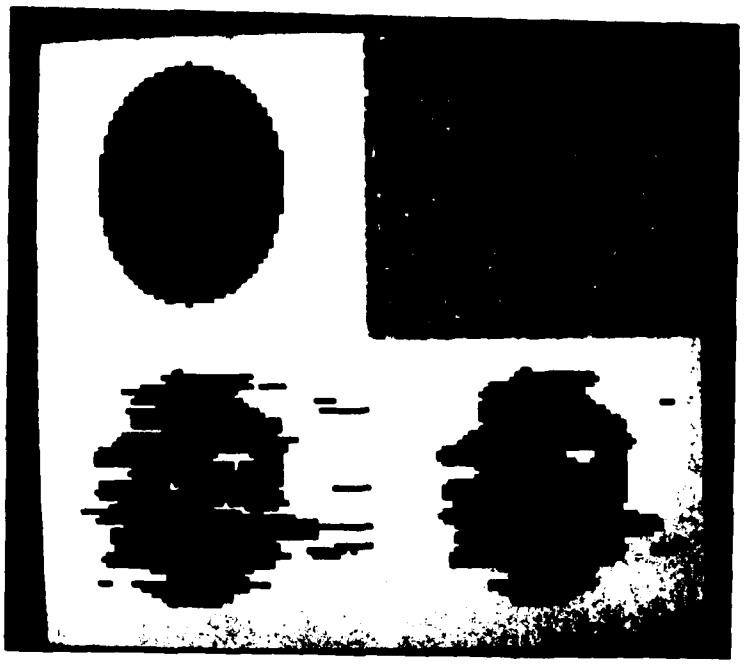


Figure 14. Segmentation with Algorithm B', SNR=0.7, q_1=0.3, q_2=q_5=0.02, q_3=0.025, q_4=ε, and q_6=0.44.

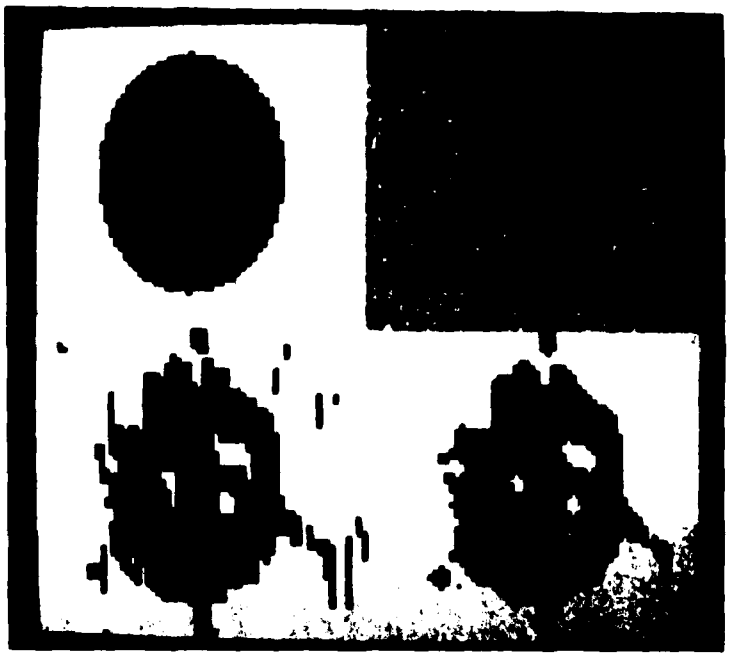
a b



c d

Figure 15. Segmentation with Algorithm A, SNR=0.5, q_1=0.22, q_2=q_5=0.005, q_3=0.01, q_4=ε, and q_6=0.70.

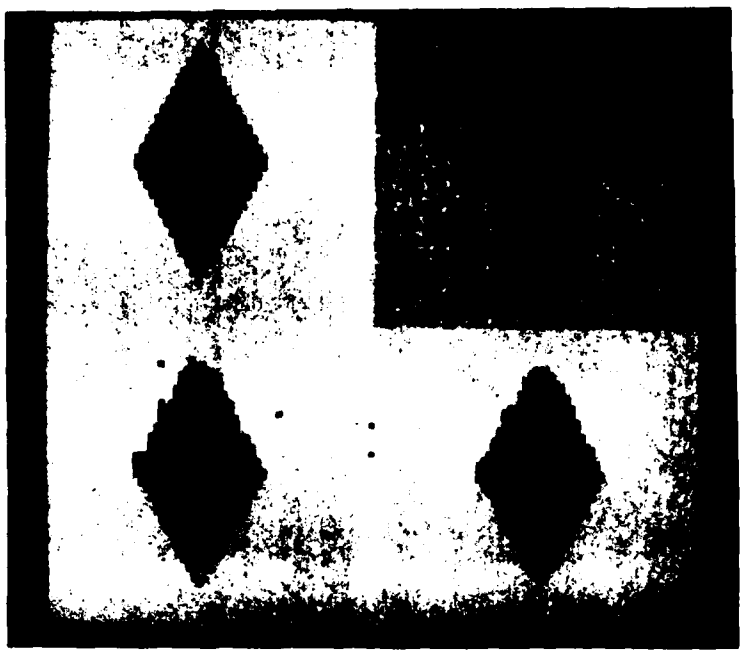
a b



c d

Figure 16. Segmentation with Algorithm B', SNR=0.5, q_1=0.28, q_2=q_5=0.03, q_3=0.04, q_4=ε, and q_6=0.32.

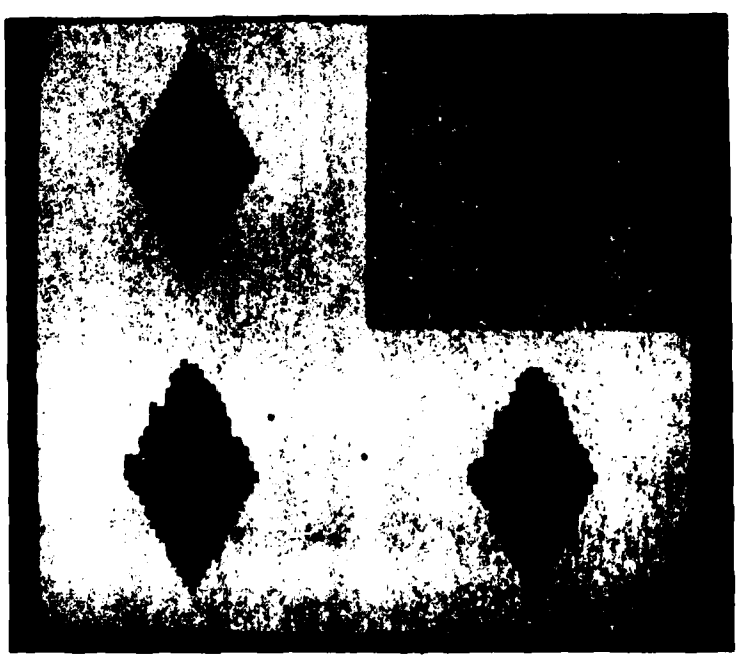
a b



c d

Figure 17. Segmentation with Algorithm A, SNR=2, q_1=0.38, q_2=q_3=q_5=0.015, q_4=ϵ, and q_6=0.44.

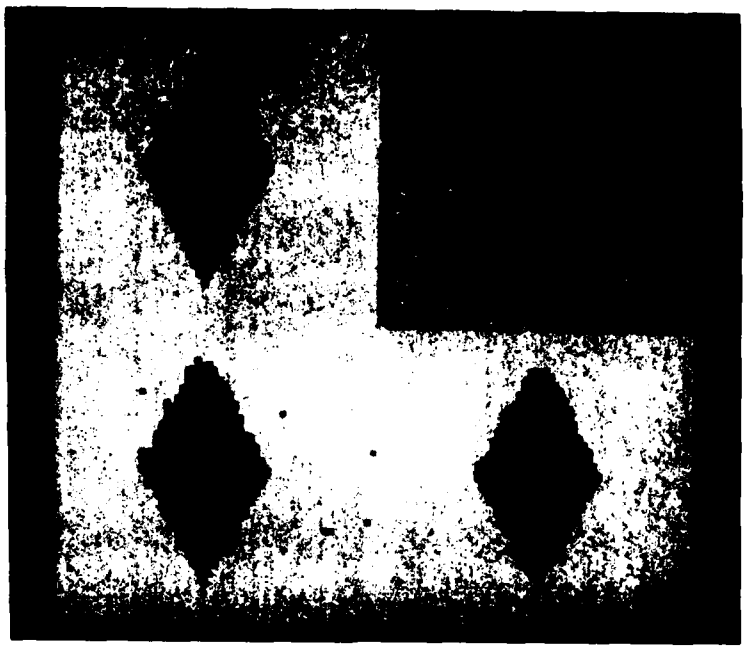
a b



c d

Figure 18. Segmentation with Algorithm B, SNR=2, q_1=0.38, q_2=q_3=q_5=0.015, q_4=ϵ, and q_6=0.44.

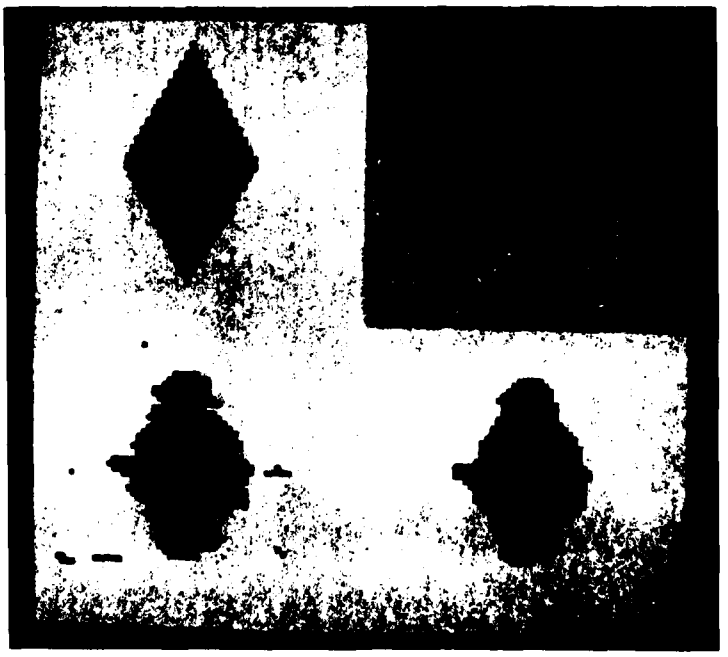
a b



c d

Figure 19. Segmentation with Algorithm B', SNR=2, $q_1=q_6=0.38$, $q_2=q_3=q_5=0.02$, and $q_4=\varepsilon$.

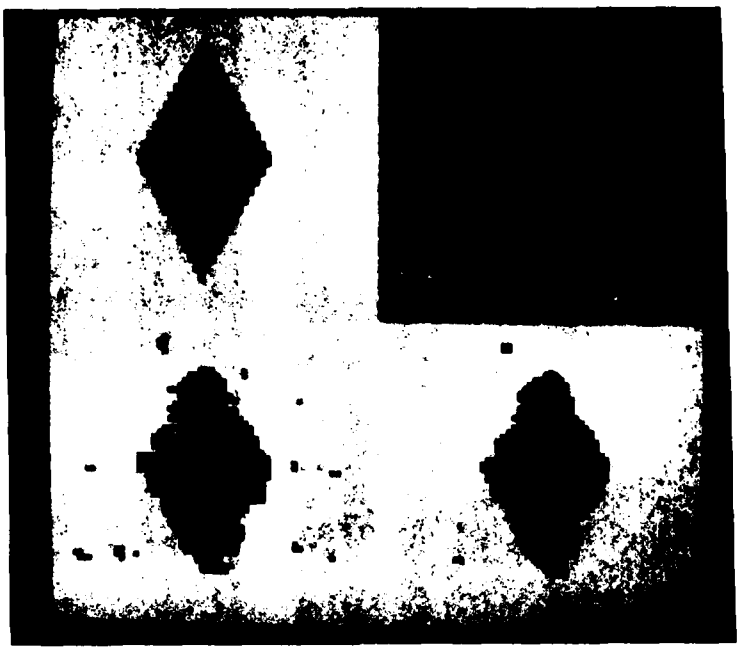
a b



c d

Figure 20. Segmentation with Algorithm A, SNR=1, $q_1=0.42$, $q_2=q_5=0.01$, $q_3=0.005$, $q_4=\varepsilon$, and $q_6=0.48$.

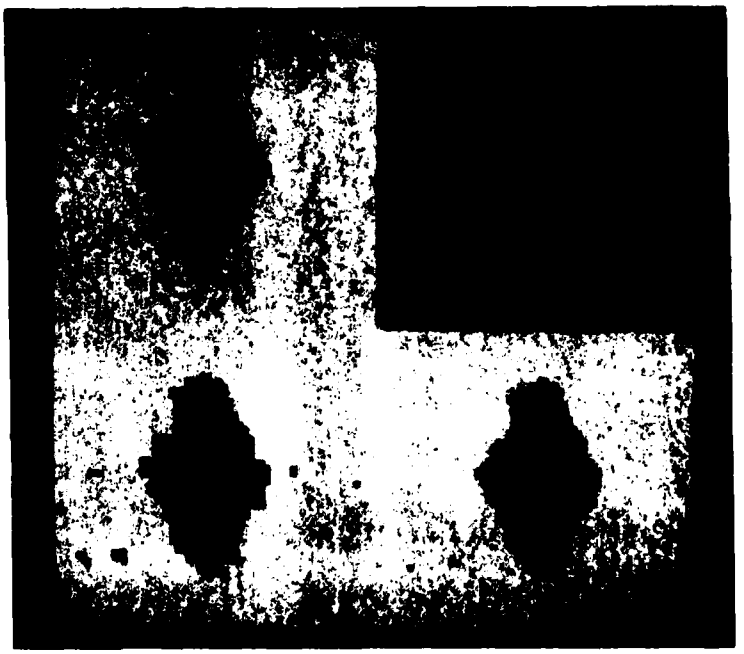
a b



c d

Figure 21. Segmentation with Algorithm B, SNR=1, q_1=0.36, q_2=q_5=0.02, q_3=0.015, q_4=ε, and q_6=0.42

a b



c d

Figure 22. Segmentation with Algorithm B', SNR=1, q_1=0.36, q_2=q_3=q_5=0.02, q_4=ε, and q_6=0.40.

we did not try to develop a scheme to estimate the parameters that will give the "best" segmentor. Instead, based on a few trials, we chose sets of parameter values that gave reasonably good segmentation results. Increasing q_1 and q_6 and equivalently decreasing the others has a stiffening effect which reduces the details and the specks presumed to be noise.

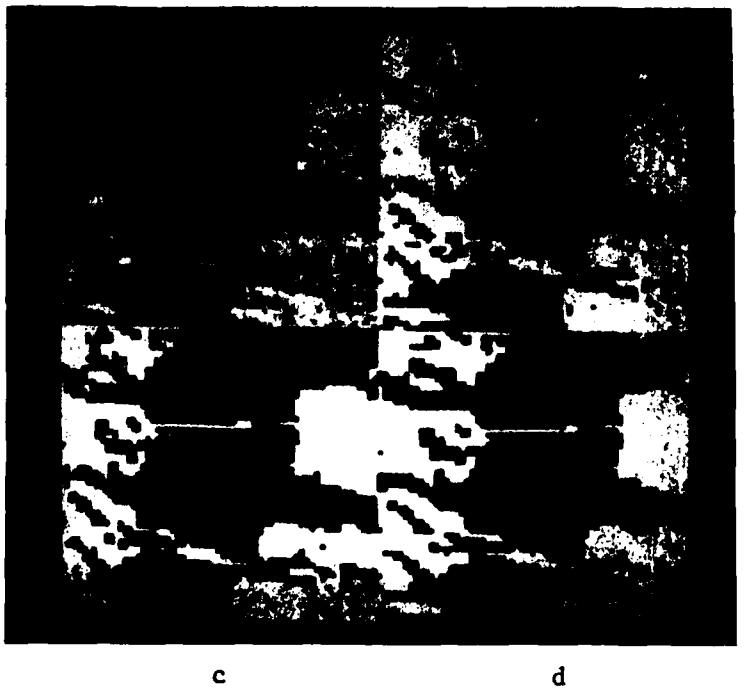
The segmentation algorithm is also applied to some remotely sensed data, namely to some synthetic aperture radar (SAR) data obtained by SEASAT. Such an SAR image provided by Office of Naval Research is shown in Figure 23. Segmentation Algorithm A is applied on three 64 x 64 portions of this image, shown in frames in Figure 23. The segmentation results of these three images by Algorithm A is presented in Figures 24-26. The sections of these Figures are arranged such that in section (a) is the actual SAR image, in sections (b), (c), and (d) are segmentations of (a) using q_i parameters getting stiffer in that order. We see that as the q_i parameters are made stiffer some of the image details as well as some specks presumed to be noise are eliminated. The q_i parameters can be chosen according to the level of detail desired in the segmented image. Again the actual image is assumed to be a mixture of two Gaussian distributions and the parameters of these Gaussians are estimated and used in the segmentation algorithm. The q_i parameters used are specified following each figure. In all cases $q_4 = 10^{-5}$ is taken and this amount is properly subtracted from other q_i's so that

$$q_1 + 4q_2 + 4q_3 + 2q_3 + 4q_5 + q_6 = 1$$

is insured. It is clearly seen in Figures 24-26 that the algorithm yields

Figure 23. SAR image of Chesapeake Bay area.

a b



c d

Figure 24. Segmentation of SAR image by Algorithm A. (a) image, (b) $q_1=q_6=0.2$ $q_2=q_3=q_5=0.05$, $q_4=\varepsilon$, (c) $q_1=q_6=0.32$, $q_2=q_3=q_5=0.03$, $q_4=\varepsilon$, (d) $q_1=q_6=0.44$, $q_2=q_3=q_5=0.01$, $q_4=\varepsilon$.

a b

c d

Figure 25. Segmentation of SAR image by Algorithm A. (a) image, (b) $q_1=q_6=0.20$ $q_2=q_3=q_5=0.05$, $q_4=\varepsilon$, (c) $q_1=q_6=0.32$, $q_2=q_3=q_5=0.03$, $q_4=\varepsilon$, (d) $q_1=q_6=0.44$, $q_2=q_3=q_5=0.01$, $q_4=\varepsilon$.

a b

c d

Figure 26. Segmentation of SAR image by Algorithm A. (a) image, (b)$q_1=q_6=0.20$ $q_2=q_3=q_5=0.05$, $q_4=\varepsilon$, (c) $q_1=q_6=0.32$, $q_2=q_3=q_5=0.03$, $q_4=\varepsilon$ (d) $q_1=q_6=0.44$, $q_2=q_3=q_5=0.01$, $q_4=\varepsilon$.

remarkably good segmentations of remotely sensed data as well. The algorithm can be incorporated with a parameter estimation scheme so that a completely automated implementation of the algorithm is feasible.

VI. CONCLUDING REMARKS

In this report we presented a new segmentation algorithm based on recursive Bayes smoothing of images modelled by Markov random fields corrupted with additive noise. Ideally, the algorithm yields the a posteriori distribution of the scene at each pixel, based on the whole noisy image. Computational concerns, however, necessitate certain simplifying assumptions on the model and some approximations during the implementation of the algorithm. In particular, the scene is modelled as a Markov mesh random field, a special class of Markov random fields, which allows for a causal conditional distribution characterization. Various properties of the Markov mesh random fields, their relationship to Markov random fields and to other classes of Markov random fields are investigated.

The algorithm is implemented using strip processing approach, where the estimate for each pixel value is determined based on all the data over a strip. This, near optimal version of the algorithm is applied to some test images of various noise levels and to some remotely sensed SAR data. The algorithm yields excellent segmentation results for high to medium SNR's and reasonably good results for low to very-low SRN's (up to 0.5).

Some areas for future work include extending the algorithm to segment multilevel images and to employ wider strips. As the strip width is increased the algorithm tends to the optimal Bayes estimate of each pixel based on the whole image. It is also of interest, to explore ways of determining the vector Markov chain transition probability distribution according to some

criterion. The transition distribution is to be determined using the noisy image and it should closely represent the scenes of interest.

REFERENCES

[1] M. Askar and H. Derin, "A recursive algorithm for the Bayes solution of the smoothing problem," IEEE Trans. Automat. Contr., Vol. AC-26, pp. 558-561, April 1981.

[2] N. E. Nahi and T. Assefi, "Bayesian recursive image estimation," IEEE Trans. Comput., Vol. C-21, pp. 734-738, July 1972.

[3] N. E. Nahi, "Role of recursive estimation in statistical image enhancement," Proc. IEEE, Vol. 60, pp. 872-877, July 1972.

[4] A. Habibi, "Two-dimensional Bayesian estimate of images," Proc. IEEE, Vol. 60, pp. 878-883, July 1972.

[5] N. E. Nahi and C. A. Franco, "Recursive image enhnacement—vector processing," IEEE Trans. Comm., Vol. COM-21, pp. 305-311, April 1973.

[6] J. W. Woods and C. H. Radewan, "Kalman filtering in two-dimensions," IEEE Trans. Inform. Theo., Vol. IT-23, pp. 473-482, July 1977.

[7] J. W. Woods, "Two-dimensional discrete Markovian fields," IEEE Trans. Inform. Theo., Vol. IT-18, pp. 232-240, March 1972.

[8] R. Chellappa and R. L. Kashyap, "Digital image restoration using spatial interaction models," IEEE Trans. Acous., Speech, Sig. Proc., Vol. ASSP-30, pp. 461-472, June 1982.

[9] R. L. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial-interaction models of images," IEEE Trans. Inform. Theo., Vol. IT-29, pp. 60-72, January 1983.

[10] A. K. Jain, "Advances in mathematical models for image processing," Proc. of IEEE, Vol. 69, pp. 502-528, May 1981.

[11] H. Kaufman, J. W. Woods, V. K. Ingle, R. Mediqvilla and A. Radpour, "Recursive estimation: a multiple model approach," Proc. 18th Conf. on Decision and Control, Fort Lauderdale, December 1979.

[12] C. W. Therrien, "Linear filtering models for texture classification and segmentation,"Proc. of 5th Int. Conf. on Pattern Recog., Miami, December 1980.

[13] C. W. Therrien, "Linear filtering models for terrain image segmentation," MIT Lincoln Laboratory, Tech. Rep. #552, February 1981.

[14] H. Elliott, D. B. Cooper, F. Cohen and P. Symosek, "Implementation, interpretation and analysis of a suboptimal boundary finding algorithm," IEEE Trans. PAMI, Vol. PAMI-4, pp. 167-182, March 1982.

[15] H. Elliott and L. Srinivasan, "An application of dynamic programming to sequential boundary estimation," Computer Graphics and Image Processing, Vol. 17, pp. 291-314, April 1981.

[16] D. B. Cooper, H. Elliott, F. Cohen, L. Reiss and P. Symosek, "Stochastic boundary estimation and object recognition," Computer Graphics and Image Processing, Vol. 10, pp. 326-355, April 1980.

[17] F. R. Hansen and H. Elliott, "Image segmentation using simple Markov random field models," Computer Graphics and Image Processing, Vol. 20, pp. 101-132, 1982.

[18] D. Geman and S. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," in preparation, August 1983.

[19] H. Elliott, H. Derin, R. Cristi, D. Geman and R. Soucy, "Experiments in the use of the Gibbs distribution for image segmentation," Tech. Rep. #UMASS-ECE-AU83-2, University of Massachusetts, Amherst, MA 01003, August 1983.

[20] Y. C. Ho and R.C.K. Lee, "A Bayesian approach to problems in stochastic estimation and control," IEEE Trans. Automat. Contr., Vol. AC-9, pp. 333-339, October 1964.

[21] L. L. Scharf and H. Elliott, "Aspects of dynamic programming in signal and image processing," IEEE Trans. Automat. Contr., Vol. AC-26. pp. 1018-1029, October 1981.

[22] E. Ising, Zeitschrift Physik, Vol. 31, p. 253, 1925.

[23] R. Kindermann and J. L. Snell, Markov Fields and Their Applications, Vol. 1, Amer. Math. Soc., Providence, RI, 1980.

[24] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," J. Royal Statist. Soc., Series B, Vol. 36, pp. 192-236, 1974.

[25] K. Abend, T. J. Harley, and L. N. Kanal, "Classification of binary random patterns," IEEE Trans. Inform. Theo., Vol. IT-11, pp. 538-544, October 1965.

[26] L. N. Kanal, "Markov mesh models," Image Modeling, Academic Press, 1980.

[27] D. K. Pickard, "A curious binary lattice process," J. Appl. Prob., Vol. 14, pp. 717-731, 1977.

END

FILMED

11-83

DTIC