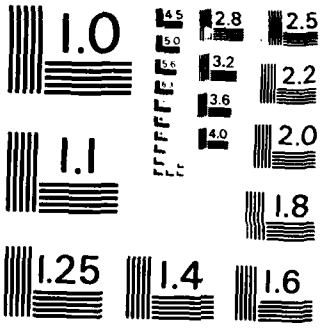
AD-A142 323 A MAXIMUM-LIKELIHOOD APPROACH TO IMAGE SEGMENTATION BY TEXTURE(U) GEORGIA INST OF TECH ATLANTA SCHOOL OF ELECTRICAL ENGINEERING J E BEVINGTON ET AL. 1984 1/1

UNCLASSIFIED AFOSR-TR-84-0337 DAAG29-81-K-0024 F/G 12/1 NL



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

AD-A142 323

SECURITY CLASSIFICATION

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 84-0337
6a. NAME OF PERFORMING ORGANIZATION: Georgia Institute of Technology; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code) School of Electrical Engineering Atlanta GA 30332	7b. ADDRESS (City, State and ZIP Code) Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR; 8b. OFFICE SYMBOL (If applicable): NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAAG29-81-K-0024
8c. ADDRESS (City, State and ZIP Code) Bolling AFB DC 20332	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO. A9; WORK UNIT NO.
11. TITLE (Include Security Classification) A MAXIMUM-LIKELIHOOD APPROACH TO IMAGE SEMENTATION BY TEXTURE	
12. PERSONAL AUTHOR(S) J.E. Bevington and R.M. Mersereau	
13a. TYPE OF REPORT: Technical; 13b. TIME COVERED: FROM ____ TO ____	14. DATE OF REPORT (Yr., Mo., Day): 1984; 15. PAGE COUNT: 4
16. SUPPLEMENTARY NOTATION PAGES 1-4 1984	
17. COSATI CODES: FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

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

This paper addresses the problem of segmenting an image by texture. Specifically, the investigators are concerned with estimating the trajectory of the boundary between the two regions characterized by different two-dimensional autocorrelation functions. An algorithm is developed which is an approximation to a maximum-likelihood estimator based on the assumption of Gaussian random fields with known mean and covariance. Preliminary experimental results are given.

DTIC FILE COPY

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Joseph Bram	22b. TELEPHONE NUMBER (Include Area Code): (202) 767-4939; 22c. OFFICE SYMBOL: NM

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE

84 06 18 161

UNCLASSIFIED
SECURITY CLASSIFICATION

To be presented at the 1984 International Conference on Acoustics, Speech, and Signal Processing.

AFCSR-TR- 84-0337

DTIC COPY INSPECTED 2

A MAXIMUM-LIKELIHOOD APPROACH TO IMAGE SEGMENTATION BY TEXTURE

J. E. Bevington and R. M. Mersereau

Georgia Institute of Technology
School of Electrical Engineering
Atlanta, Georgia 30332

ABSTRACT

This paper addresses the problem of segmenting an image by texture. Specifically, we are concerned with estimating the trajectory of the boundary between two regions characterized by different two-dimensional autocorrelation functions. An algorithm is developed which is an approximation to a maximum-likelihood estimator based on the assumption of Gaussian random fields with known mean and covariance. Preliminary experimental results are given.

INTRODUCTION

The problem of segmenting an image into more or less homogeneous regions arises in the areas of image coding and image understanding. Segmentation may be based on a number of local properties, one of which is texture. Although many definitions of texture are available, we will for this work assume that a textured region may be characterized by its statistical autocorrelation properties.

The specific problem addressed in this paper is that of estimating the boundary between two regions which have different textures but not necessarily different average gray levels. The algorithm which is presented assumes complete knowledge of the mean and autocorrelation functions which characterize the two textures. In practice, this information would be estimated from neighboring regions which are assumed homogeneous. We envision then that this algorithm might be used following an initial coarse segmentation procedure, which divides the image into frames and pre-classifies each frame as described in [1].

The boundary estimation algorithm is derived as a maximum- likelihood estimator on the assumption of jointly Gaussian statistics for the image data. The Gaussian assumption is made for mathematical tractability. Although actual image data may not fit the Gaussian model very well, the algorithm developed under this assumption has performed well in initial tests with real image data. The Gaussian assumption leads to an implementation which involves two-dimensional linear prediction. In this sense our work closely parallels the work of Quatieri [2], in which a statistical formulation of the problem of detecting an anomaly in an otherwise homogeneous textured region also led to a 2-D LPC implementation.

ESTIMATOR FORMULATION

The 2-D boundary estimation problem is that of finding a contour C which partitions a set of observed image samples $\{x(m,n)\}$ into two sets, where one set is presumed to be a realization of the discrete random field $\{x_1(m,n)\}$, the other of the field $\{x_2(m,n)\}$. We assume that $\{x_1(m,n)\}$ and $\{x_2(m,n)\}$ are independent homogeneous Gaussian random fields with known mean and covariance.

For this paper we impose the following constraint on the contour C and the resulting partition of the N X N set $\{x(m,n)\}$. We require that C be characterized completely by the N-dimensional vector $\underline{L} = (L(0),L(1),\ldots,L(N-1))$, such that

$$x(m,n) = \begin{cases} x_1(m,n), & 0 \le n \le L(m) \\ x_2(m,n), & L(m) < n \le N-1 \end{cases} \qquad (1)$$

where $0 \le m \le N-1$.

That is, we require that there be only one "boundary point" per line. (See, for example, Figure 1.) This requirement may at first seem rather severe, but bear in mind that we intend that the algorithm be applied only over small areas of an image, and that we have the freedom to rotate the frame.

To formulate the maximum likelihood estimator for $\underline{L}$, we need an expression for the joint probability density function (pdf) for the set $\{x(m,n)\}$ conditioned on $\underline{L}$. This function is most easily expressed in vector notation. Let $\underline{x}$ be a vector containing all of the observed samples $\{x(m,n)\}$. Now let us partition $\underline{x}$ by assigning to vector $\underline{x}_1$ those points determined by a particular $\underline{L}$ to be from $\{x_1(m,n)\}$, and to $\underline{x}_2$ those points from $\{x_2(m,n)\}$. The conditional pdf for $\underline{x}$, which we denote by $p(\underline{x}|\underline{L})$ is the joint density of $\underline{x}_1$ and $\underline{x}_2$ conditioned on $\underline{L}$. Since the random fields $\{x_1(m,n)\}$ and $\{x_2(m,n)\}$ are independent, this joint density is simply the product of the individual conditional densities. That is,

$$p(\underline{x}|\underline{L}) = p(\underline{x}_1|\underline{L})p(\underline{x}_2|\underline{L}) \qquad (2)$$

The Gaussian assumption then leads to ([3])

Approved for public release; distribution unlimited.

84 06 18 161

$$p(\underline{x}|\underline{L}) = \text{const} \cdot |K_1|^{-1} \cdot |K_2|^{-1} \cdot \exp\left\{-\frac{1}{2}\left[(\underline{x}_1-\underline{m}_1)^T K_1^{-1}(\underline{x}_1-\underline{m}_1) + (\underline{x}_2-\underline{m}_2)^T K_2^{-1}(\underline{x}_2-\underline{m}_2)\right]\right\} \quad (3)$$

where

$$\underline{m}_1 = E\{\underline{x}_1\} \qquad \underline{m}_2 = E\{\underline{x}_2\}$$

$$K_1 = E\{\underline{x}_1\underline{x}_1^T\} \qquad K_2 = E\{\underline{x}_2\underline{x}_2^T\}$$

Note that all of the vectors and matrices on the right-hand-side of (3) (even their dimensions) depend on $\underline{L}$. The explicit forms of these elements also depend on the order we choose in assigning the points of $\{x(m,n)\}$ to $\underline{x}_1$ and $\underline{x}_2$.

We now consider a whitening transformation on the observed data, which will lead to an expression for $p(\underline{x}|\underline{L})$ which is more easily evaluated for varying $\underline{L}$. Let

$$\underline{e}_1 = A_1(\underline{x}_1-\underline{m}_1) \quad (4)$$

where A_1 is such that

$$E\{\underline{e}_1\underline{e}_1^T\} = D_1 \quad (5)$$

for D_1 diagonal. Then from (4) and (5),

$$D_1 = A_1 K_1 A_1^T \quad (6)$$

Assuming A_1 and D_1 are invertible, this gives

$$K_1^{-1} = A_1^T D_1^{-1} A_1 \quad (7)$$

Using (4) and (7) we can rewrite the first half of the exponent of $p(\underline{x}\ \underline{L})$ in (3) as

$$(\underline{x}_1-\underline{m}_1)^T K_1^{-1}(\underline{x}_1-\underline{m}_1) = \sum_n e_1^2(n)/d_1(n) \quad (8)$$

where $e_1(n)$ is the nth element of $\underline{e}_1$, and $d_1(n)$ is the nth diagonal element of D_1, which is the mean squared value of $e_1(n)$. Using similar definitions for $\underline{e}_2$, A_2, D_2, $e_2(n)$ and $d_2(n)$ we have

$$(\underline{x}_2-\underline{m}_2)^T K_2^{-1}(\underline{x}_2-\underline{m}_2) = \sum e_2^2(n)/d_2(n) \quad (9)$$

Now let us address the construction of the matrices A_1 and A_2. All we have required to this point is that these operators whiten the vectors $(\underline{x}_1-\underline{m}_1)$ and $(\underline{x}_2-\underline{m}_2)$ and that they be invertible. One such construction is as follows. (See Therrien [4].) Let A_1 be lower triangular with 1's on the diagonal. In addition, let the non-zero off-diagonal part of each row j be the (j-1)st order linear predictor for the jth element of $(\underline{x}_1-\underline{m}_1)$ given all previous elements. In effect, we compute the jth element of $\underline{e}_1$ by subtracting from the jth element of $(\underline{x}_1-\underline{m}_1)$ the part predictable from all elements k for which k<j. Let A_2 be constructed similarly, using linear predictors for $(\underline{x}_2-\underline{m}_2)$. The linear prediction coefficients in the two matrices will depend on the statistics of the corresponding processes and on the ordering used to form the vectors $\underline{x}_1$ and $\underline{x}_2$.

Returning to $p(\underline{x}|\underline{L})$ in (3) we take the logarithm, multiply by -2, and drop the constant term to obtain the log-likelihood function $A(\underline{L})$:

$$A(\underline{L}) = \ln|K_1| + \ln|K_2| - (\underline{x}_1-\underline{m}_1)^T K_1^{-1}(\underline{x}_1-\underline{m}_1) - (\underline{x}_2-\underline{m}_2)^T K_2^{-1}(\underline{x}_2-\underline{m}_2) \quad (10)$$

Note that

$$|K_1| = |A_1^{-1}|\,|D_1|\,|A_1^{-T}| = \prod_n d_1(n) \quad (11)$$

(Similarly for $|K_2|$.)
From (8),(9),(10), and (11),

$$A(\underline{L}) = \sum_n \left[e_1^2(n)/d_1(n) + \ln(d_1(n))\right] + \sum_n \left[e_2^2(n)/d_2(n) + \ln(d_2(n))\right] \quad (12)$$

The maximum likelihood estimate for $\underline{L}$ is that value which minimizes $A(\underline{L})$.

COMPUTATION OF THE LIKELIHOOD FUNCTION

We now consider the computation of the linear prediction residuals $e_1(n)$ and $e_2(n)$ for a given $\underline{L}$. The first step is to establish an ordering for the elements of $\underline{x}_1$ and $\underline{x}_2$, which we show in Figure 1. For a given $\underline{L}$, the points assumed to be from $\{x_1(m,n)\}$ are scanned left to right, top to bottom, while those from $\{x_2(m,n)\}$ are scanned right to left, top to bottom. It will be shown that this ordering leads to a simple line-by-line procedure for estimating $\underline{L}$.

An exact evaluation of $A(\underline{L})$ in (12) would require that the residuals e_1 and e_2 be computed using spatially varying prediction masks for which the order increases with each new point as shown in Figure 2. Also, the shape of the masks for a given pixel location (m,n) would depend on L(k) for all k<m (the hypothesized edge-point locations for all previous rows). In order to arrive at an implementable algorithm we replace the ideal spatially varying masks with relatively small spatially invariant ones as shown in Figure 2. There are two sources of error introduced by this approximation. First, by restricting the mask size, we are not in general able to remove all of the correlation between the residual terms. Second, by not allowing the shape of the mask to vary with its location relative to the hypothesized boundary, we encounter situations like that shown in Figure 3, where part of the mask extends over the boundary, and we in essence attempt to

predict a sample from one process using samples assumed to be part of the other. There are obviously some tradeoffs to be made in selecting the mask shape and size.

The spatially invariant prediction mask approximation leads to some simplification of (12). Because the random fields $\{x_1(m,n)\}$ and $\{x_2(m,n)\}$ are homogeneous, the mean squared value of the prediction error is constant and we can replace $d_1(n)$ and $d_2(n)$ by σ_1^2 and σ_2^2, respectively. Also, each residual term e_1 or e_2 is now computed independently of $\underline{L}$, so the dependence of A on $\underline{L}$ is reduced to a form which may be expressed explicitly by the limits on the summations. Letting $A'(\underline{L})$ denote the approximate value of $A(\underline{L})$ we have

$$A'(\underline{L}) = \sum_{m=p}^{N-1-p} T(m,L(m)) \qquad (13)$$

where

$$T(m,k) = \sum_{n=p}^{k} e_1^2(m,n)/\sigma_1^2 + \ln(\sigma_1^2)$$

$$+ \sum_{n=k+1}^{N-1-p} e_2^2(m,n)/\sigma_2^2 + \ln(\sigma_2^2)$$

(The limits on the summations have been chosen such that we compute prediction residuals only to within a distance p of the frame borders, where p is the extent of the predictor mask. This measure, taken to prevent the mask from ever extending outside the frame, slightly decreases the area over which we can search for the contour.)
As implied in the discussion, the residuals $e_1(m,n)$ and $e_2(m,n)$ are given by

$$e_1(m,n) = x(m,n) - \sum_i \sum_j a_{ij} x(m-i,n-j)$$

$$e_2(m,n) = x(m,n) - \sum_i \sum_j b_{ij} x(m-i,n-j)$$

where $\{a_{ij}\}$ and $\{b_{ij}\}$ are the sets of 2-D LPC coefficients for the processes $\{x_1(m,n)\}$ and $\{x_2(m,n)\}$, respectively. The LPC coefficients are dependent on the prediction masks used [1].

EDGE SHAPE CONSTRAINTS

The likelihood function in (13) is minimized by choosing independently for each row m the value L(m) which minimizes the corresponding T(m,L(m)). It has been found in practice however that much better results are obtained when we introduce constraints on the shape of the boundary so that coupling is introduced into the row-by-row decisions. The need for shape constraints arises partially because of errors introduced by the small spatially-invariant prediction masks.

Constraints on contour shape may be imposed conveniently in the framework of the expression for $A'(\underline{L})$ in (13). In our current implementation, we first generate T(m,n) (the "likelihood terrain") for all (m,n) in the search area. The next step is to search for a path through this terrain which is in some sense optimal subject to a desired constraint. We require of course that this path touch exactly one point in each line. If, for example, the constraint is that the contour be a straight line, we might systematically generate all $\underline{L}$ which would correspond to a straight line, and evaluate A' for each such $\underline{L}$ by summing the heights of the terrain at points touched by the corresponding line. Another form the constraint may take is stochastic. For example we might model the sequence of line-by-line edge coordinates as a first order Markov process, as is done by Cooper and Sung [5]. In this case the optimal path is found using a dynamic programming procedure.

PRELIMINARY RESULTS

The results of some early testing of the algorithm appear in Figure 4. We show results for two different 32X32 textured images, one of which is repeated along the top of the figure, the other repeated along the bottom. The images were formed using regions extracted from a multiple-textured image in the USC data base. The true edge (the one used in constructing the images) is superimposed on the first frame in each row. The difficulty of finding the edge by naked eye (particularly for the second image) is illustrated in the second frame in each row, where we repeat the first frame without drawing the edge. In the third frame we superimpose the boundary found by our algorithm when we impose a straight-line constraint. The fourth frame shows the result of using a crude stochastic constraint, where the edge coordinate for line j is assumed to be the coordinate on line j-1 plus a bias plus a uniform random variable in the interval [-1,1]. The bias term is estimated by first finding the edge subject to a straight line consraint. The prediction masks used in all instances were 3X3 quarter plane.

We have found that, in general, the more we know about the edge beforehand, the more accurate is the resulting estimate and the more quickly it can be found. We are currently studying intelligent search procedures and ways of specifying and incorporating a priori information. We are also investigating the effects of prediction mask size and shape on the accuracy of the boundary estimate.

ACKNOWLEDGEMENT

This work was supported in part by the Joint Services Electronics Program under Contract #DAAG29-81-K-0024.

REFERENCES

[1] P. A. Maragos, R. M. Mersereau, and R. W. Schafer, "Two-Dimensional Linear Predictive Analysis of Arbitrarily-Shaped Regions", Proc. ICASSP 1983.

[2] T. F. Quatieri, "Object Detection by Two-Dimensional Linear Prediction", Proc. ICASSP 1983.

[3] H. L. Van Trees, "Detection, Estimation, and Modulation Theory, Part I", John Wiley and Sons, New York, 1968.

[4] C. W. Therrien, "On the Relation Between Triangular Matrix Decomposition and Linear Prediction", to be published in Proc. of IEEE, Dec., 1983.

[5] D. B. Cooper and F. P. Sung, "Multiple-Window Parallel Adaptive Boundary Finding in Computer Vision", IEEE Trans. PAMI, Vol. PAMI-5, No. 3, pp. 299-316, May 1983.

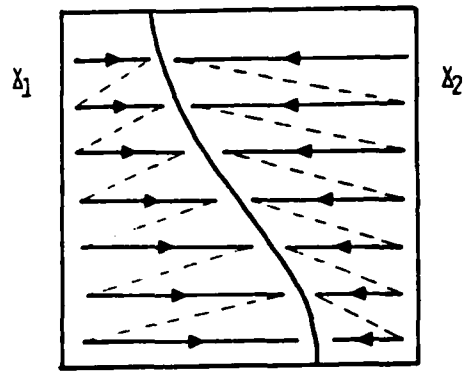


FIGURE 1 - PIXEL ORDERING

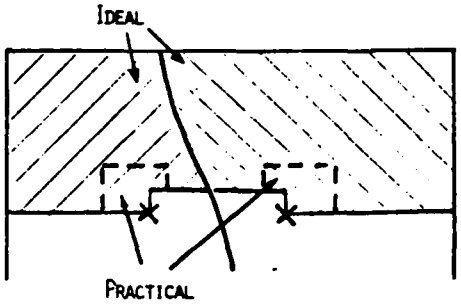


FIGURE 2 - PREDICTOR MASK SUPPORT

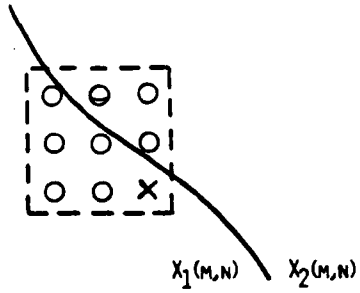


FIGURE 3 - BOUNDARY OVERLAP BY SPATIALLY-INVARIANT MASK

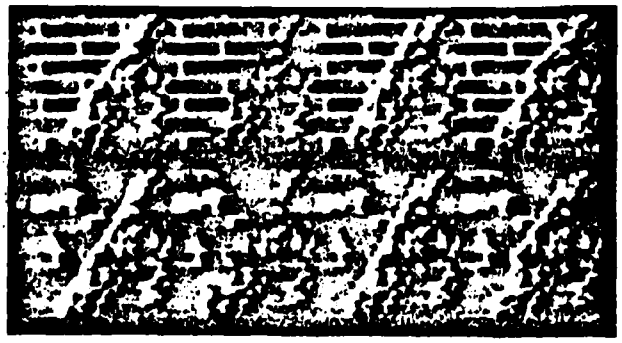


FIGURE 4 - PRELIMINARY RESULTS

END

FILMED

DTIC