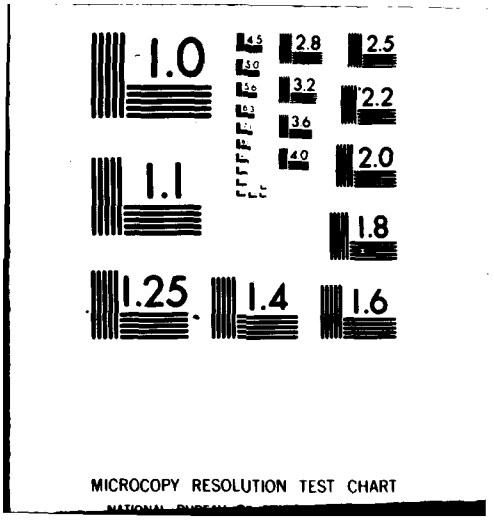
AD-A088 183 ILLINOIS UNIV AT CHICAGO CIRCLE DEPT OF MATHEMATICS F/G 12/1
APPLICATION OF THE CONDITIONAL POPULATION-MIXTURE MODEL TO IMAG--ETC(U)
AUG 80 S L SCLOVE N00014-80-C-0408
UNCLASSIFIED UICC-MATH-TR-80-1 NL

END
DATE FILMED
9-80
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART

(6) APPLICATION OF
THE CONDITIONAL POPULATION-MIXTURE MODEL
TO
IMAGE SEGMENTATION.

by

(10) Stanley L. Sclove

Accepted for presentation at the
5th International Conference on Pattern Recognition,
Miami Beach, Florida,
December 1-4, 1980

(9) TECHNICAL REPORT NO. 80-1

(11) 15 Aug 1980 (12) 6

PREPARED FOR THE
OFFICE OF NAVAL RESEARCH
(15) UNDER
CONTRACT N00014-80-C-0408
TASK NR042-443

Principal Investigator: Stanley L. Sclove

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS AT CHICAGO CIRCLE
CHICAGO, ILLINOIS 60680

APPLICATION OF THE CONDITIONAL POPULATION-MIXTURE MODEL TO IMAGE SEGMENTATION

Stanley L. Sclove

University of Illinois at Chicago Circle
Chicago, Illinois

Abstract

The problem of image segmentation is considered in the context of a mixture of probability distributions. A modification of the usual approach to mixtures of distributions is employed. Parametric families of distributions are considered, a set of parameter values being associated with each distribution. In addition, an identification parameter is associated with each observation, indicating from which distribution the observation arose. Thus, the segmentation problem is cast as a problem of statistical estimation. Segmentation algorithms are obtained by applying a method of iterated maximum likelihood to the resulting likelihood function.

1. Introduction

Consider a digital image, given as a set of p-dimensional vectors $\underline{x}_{ij} = (x_{1ij}, \ldots, x_{pij})$, $i = 1,2,\ldots,I$, $j = 1,2,\ldots,J$.

Examples. (i) $p = 3$, x_{1ij} = red level, x_{2ij} = green level, x_{3ij} = blue level of pixel (i,j).

(ii) $p = 1$ (monochromatic image), $x_{1ij} = x_{ij}$ = gray level of pixel (i,j).

The problem of image segmentation is, simply stated, the problem of putting the pixels (i,j) into groups (classes, clusters), i.e., the "segments."

Define a *segmenting* as a partition of the set of pixels, i.e., as a collection $\{C_1,\ldots,C_k\}$ of disjoint sets such that each pixel belongs to one and only one set C_g. Each set C_g is a *segment (cluster)*. Here we shall assume that the integer k is specified in advance. (A modification of the algorithm allows some of the segments to join or split, thereby permitting fewer or more than k clusters to form. See Sec. 6.2 below.)

In what follows we shall write x_i rather than x_{ij}, using a single subscript i rather than the double subscript ij for the pixels, even though they are a two-dimensional array.

It seems reasonable to consider the following model for segmentation problems:

Assumption (1). With the g-th segment $(g=1,\ldots,k)$ is associated a probability distribution with probability density function (p.d.f.) $h_g(\underline{x})$. The p.d.f.'s are generally unknown.

Assumption (2). With the i-th pixel $(i=1,\ldots,n=IJ)$ is associated a *group (segment) identification parameter)* γ_i which is equal to g if and only if pixel i belongs to segment g. Each pixel thus gives rise to a pair $(\underline{X},\gamma)$ where $\underline{X}$ is observable and γ is not.

Remarks. (i) In the context of this model "segmentation"is merely estimation of the parameters γ_i for the n pixels. (ii) In regard to Assumption (1), when we are working with some parametric family, indexed by a parameter, say $\underline{\beta}$, then h_g takes the form $h_g(\underline{x}) = h(\underline{x};\underline{\beta}_g)$. The parameters are generally unknown. (iii) This model is a *population-mixture model.*

It is convenient to reparametrize. Replace γ_i by a k-vector $\underline{\theta}_i$ which consists of k-1 zeros and a single 1, the position of the 1 indicating which segment pixel i belongs to; i.e., $\underline{\theta}_i$ has a 1 as its γ_i-th element and 0's elsewhere. The p.d.f. of $\underline{X}_i$, given $\underline{\theta}_i$, is

$$f(\underline{x}_i|\underline{\theta}_i) = \sum_{g=1}^{k} \theta_{gi} h_g(\underline{x}_i), \qquad (1.1)$$

where θ_{gi} is the g-th element of $\underline{\theta}_i$.

2. The Probability Model

The model of Sec. 1 should be contrasted with the usual population-mixture model, in which any observation $\underline{X}_i$ is chosen from Population g with probability π_g, so that in this standard population-mixture model the p.d.f. of $\underline{X}_i$ is

$$f(\underline{x}_i;\pi_1,\ldots,\pi_k) = \sum_{g=1}^{k} \pi_g h_g(\underline{x}_i), \qquad (2.1)$$

$i=1,\ldots,n$. This standard mixture model has been used for pixel classification; see, e.g., Eklundh, Yamamoto, and Rosenfeld.[5] The purpose of the present paper is to suggest the conditional mixture model as an alternative and to present some algorithms derived from it. Further discussion of the model, in the context of statistical cluster analysis, and further references are given by Sclove.[7]

A likelihood approach, whether based on the standard or the conditional mixture model, is illuminating in that it can show how *ad hoc* optimality criteria (objective functions) which have been proposed relate to likelihood functions in particular probability models.

Note that (1.1) can be written as a product

$$f(\underline{x}_i|\underline{\theta}_i) = \prod_{g=1}^{k} [h_g(\underline{x}_i)]^{\theta_{gi}}. \qquad (2.2)$$

The form (2.2) is often more convenient, and we shall use it in what follows.

3. The Segmentation Algorithm

Using the form (2.2), one sees that the joint p.d.f. of the $\underline{X}_i$'s, given the $\underline{\theta}_i$'s, is

$$\prod_{i=1}^{n} \prod_{g=1}^{k} [h(\underline{x}_i;\underline{\beta}_g)]^{\theta_{gi}}.$$

The likelihood is to be maximized over all assignments of pixels to segments and over all permissible parameter values. Many *ad hoc* schemes can be applied to this maximization problem. E.g., one way to maximize is to start with a given segmentation, take each observation successively and shift it to the first segment for which a shift results in an increase in likelihood, and loop through the data unitl no pixel changes segments.

The algorithm to be developed here is an iterative, back-and-forth procedure. We first maximize with respect to (w.r.t.) the θ's (holding the $\underline{\beta}$'s fixed at initial values), then w.r.t. the $\underline{\beta}$'s (holding the θ's fixed at the values obtained in the previous stage), then again w.r.t. the θ's (holding the $\underline{\beta}$'s fixed at the values obtained in the previous stage), etc. We stop when no θ changes, i.e., when no pixel changes segments, or when a specified amount of computer time is used.

An alternative of starting the procedure is to start with an initial segmentation rather than with initial guesses of the $\underline{\beta}$'s.

It is clear that, for fixed values of the $\underline{\beta}$'s, say $\hat{\underline{\beta}}$'s, the likelihood is maximized, for each i, by taking

$$\hat{\theta}_{gi} = \begin{cases} 1 \text{ if } h(\underline{x}_i;\hat{\underline{\beta}}_g) = \max\limits_{1\le\ell\le k} \{h(\underline{x}_i;\hat{\underline{\beta}}_\ell)\} \\ 0 \text{ otherwise.} \end{cases} \qquad (3.1)$$

(In case of ties an arbitrary choice is made.) In other words, segmentation proceeds by allocating pixel i to the group g for which the estimated probability density of the observation $\underline{x}_i$ is largest.

Note that, having tentatively estimated the θ's at any stage, i.e., having tentatively segmented the image, estimation of the $\underline{\beta}$'s is reduced simply to ordinary maximum likelihood estimation in the particular parametric family at hand. This is a particular advantage of this approach.

Let T denote the set of $\underline{\theta}_i$'s and B the set of $\underline{\beta}_g$'s. Let L(B,T) denote the likelihood. Let $B^{(s)}$ denote the value of B which maximizes L at the s-th stage of the iteration, and let $T^{(s)}$ denote the value of T which maximizes L at the s-th stage of the iteration. Then $T^{(s)}$ maximizes $L(B^{(s)}, T)$ w.r.t. T, and $B^{(s)}$ maximizes $L(B,T^{(s-1)})$ w.r.t. B. This back-and-forth maximization is an example of the *relaxation method (Southwell's method)*; see Ortega & Rheinboldt[7] (pp. 214ff.) and Southwell.[9,10] It is true that

$$L(B^{(s+1)},T^{(s)}) \ge L(B^{(s)},T^{(s)})$$

and

$$L(B^{(s)},T^{(s+1)}) \ge L(B^{(s)},T^{(s)}).$$

That is, at no stage of the procedure can the value of the likelihood decrease; however, there is no guarantee of convergence to the global maximum (neither do alternative clustering algorithms guarantee convergence to the global max of their objective functions). To see how the procedure can fail to converge to a global max, suppose it happens that $L(B^{(s)}, T^{(s)}) > L(B,T^{(s)})$ for all B, or $L(B^{(s)}, T^{(s-1)}) > L(B^{(s)},T)$ for all T. Then the procedure will terminate at the s-th stage, without having necessarily reached the global max. That is, if, having maximized w.r.t. one of the variables B and T, we happen to find ourselves at a (relative) max w.r.t. the other, we may not reach a global max.

4. Application to Particular Distributions

Now we consider application of this general method to particular families of distributions. First we consider normal distributions with common covariance matrix, for in this case it becomes clear how the model establishes a link with some existing clustering procedures.

4.1. Multivariate Normal Populations with Common Covariance Matrix

In the case of normal populations with means $\underline{\mu}_g$, g=1,...,k, and common covariance matrix $\underline{\Sigma}$, the likelihood takes the form

$$(2\pi)^{-np/2}|\underline{\Sigma}|^{-n/2}\exp[-\sum_{i=1}^{n}\sum_{g=1}^{k}\theta_{gi}q(\underline{x}_i;\underline{\mu}_g,\underline{\Sigma})/2],$$

where the quadratic form q is given by

$$q(\underline{x};\underline{\mu},\underline{\Sigma}) = (\underline{x}-\underline{\mu})'\underline{\Sigma}^{-1}(\underline{x}-\underline{\mu}),$$

the (Mahalanobis) distance between $\underline{x}$ and $\underline{\mu}$ in the metric of $\underline{\Sigma}$. Here (3.1) is equivalent to

$$\hat{\theta}_{gi} = \begin{cases} 1 \text{ if } q(\underline{x}_i;\hat{\underline{\mu}}_g,\hat{\underline{\Sigma}}) = \min\limits_{1\le\ell\le k} \{q(\underline{x}_i;\hat{\underline{\mu}}_\ell,\hat{\underline{\Sigma}})\} \\ 0 \text{ otherwise.} \end{cases} \qquad (4.1)$$

That is, pixel i is assigned to that group to whose tentatively estimated mean vector it is closest, where the distance is in the metric of the tentatively estimated covariance matrix. Having estimated the θ's, we have multivariate normal observations arranged into groups; maximization w.r.t. the $\underline{\mu}$'s and $\underline{\Sigma}$ is accomplished by taking the group mean vectors as estimates for the $\underline{\mu}$'s, and the within-groups sum-of-products matrix gives the estimate of $\underline{\Sigma}$. The procedure is iterated: using new estimates $\hat{\underline{\mu}}_g$, g = 1,...,k, and $\hat{\underline{\Sigma}}$, the rule (4.1) is applied again. Then new $\hat{\underline{\mu}}$'s and a new $\hat{\underline{\Sigma}}$ are calculated; etc. The Mahalanobis distances can be computed efficiently; see, e.g., Anderson[1], p. 107.

Relationship with the *isodata* procedure. This scheme is a Mahalanobis-distance version of Ball and Hall's *isodata* clustering procedure.[2] *Isodata* proceeds as follows. One starts with tentative estimates of cluster means and assigns each individual to the mean to which he is closest. (The *isodata* scheme uses Euclidean distance, or

modified Euclidean distance in which different weights are assigned to the p dimensions.) The cluster means are then re-estimated, and one loops through the data again, reassigning the individuals, etc. Note the similarity to our scheme: We start with tentative estimates of the $\underline{\mu}$'s and $\underline{\Sigma}$ and assign each individual to the mean to which he is closest, using Mahalanobis distance in the metric of the tentatively estimated covariance matrix. The $\underline{\mu}$'s and $\underline{\Sigma}$ are then re-estimated, the individuals (pixels) are reallocated to clusters (segments), etc.

An important difference is that our scheme employs Manalanobis distance rather than Euclidean or weighted-Euclidean distance. (It is worth emphasizing that it is the Mahalanobis distance based on the *within-groups* sum-of-products matrix that arises here; some data analysts use the *total* sum-of-products matrix, which is not appropriate; see, e.g., Chernoff.[3])

Some experiments with the algorithm, in the context of statistical cluster analysis, are reported in Sclove.[8]

Relationship with the *k-means* procedure. Arranging the computation a little differently, updating the estimates of the $\underline{\mu}$'s and $\underline{\Sigma}$ after each individual pixel is assigned rather than waiting until all have been assigned, produces a Mahalanobis-distance version of MacQueen's *k-means* procedure.[6]

4.2. Multivariate Normal Populations with Different Covariance Matrices

The algorithm generated for this case turns out not to be simply to use a different Manalanobis distance for each cluster. (The complication which occurs is analogous to that in classification--discriminant analysis--where one is led to quadratic discriminant functions if the covariance matrices differ.) The likelihood is

$$(2\pi)^{-np/2} \prod_{ig} |\underline{\Sigma}_g|^{-\theta_{gi}/2} \exp[-\sum\sum_{ig} \theta_{gi} q(\underline{x}_i;\underline{\mu}_g,\underline{\Sigma}_g)/2].$$

Equation (3.1) becomes

$$\hat{\theta}_{gi} = \begin{cases} |\hat{\underline{\Sigma}}_\ell|^{-1/2} \exp[-q(\underline{x}_i;\underline{\mu}_\ell,\underline{\Sigma}_\ell)/2] & \text{if setting } \ell=g \text{ maximizes} \\ 0 & \text{otherwise .} \end{cases} \qquad (4.2)$$

Maximizing the expression in (4.2) is equivalent to minimizing

$$\ell n|\hat{\underline{\Sigma}}_\ell| + q(\underline{x}_i;\hat{\mu}_\ell,\hat{\underline{\Sigma}}_\ell) .$$

It has been noted [see, e.g., Day[4]] that in the standard mixture model for this case the supremum of the likelihood is infinity. This is reflected in the fact that in our algorithm it would be possible that at some stage one of the clusters would consist of a single individual, so that the tentative estimate of the mean of that group would be the vector of observations for that individual, and the tentative estimate of the covariance matrix of that cluster would be undefined. It is also possible for the observations in a given cluster to be very close to lying on a lower-dimensional subspace, so that the tentative estimate of the covariance matrix could have an arbitrarily small determinant, and the maximized likelihood could be arbitrarily large, for the contribution of Group g to the maximized likelihood is inversely proportional to a positive power of the determinant.

5. Comparison with the Method Based on the Standard Mixture Model

Wolfe has considered clustering based on the standard mixture model.[11] Under that model the posterior probability that Individual i belong to Group g is

$$\pi_g h(\underline{x}_i;\underline{\beta}_g) / \sum_{\ell=1}^{k} \pi_\ell h(\underline{x}_i;\underline{\beta}_\ell) . \qquad (5.1)$$

If we can obtain estimates for $\underline{\beta}_g$, π_g, $g = 1,\ldots,k$, they can be substituted to provide an estimate of (5.1),

$$\hat{\pi}_g h(\underline{x}_i;\hat{\underline{\beta}}_g) / \sum_{\ell=1}^{k} \hat{\pi}_\ell h(\underline{x}_i;\hat{\underline{\beta}}_\ell) . \qquad (5.2)$$

Individual i is assigned to that group g for which the estimated *posterior probability* of group membership, (5.2), is largest. On the other hand, with the conditional mixture model Individual i is assigned to that group g for which the *estimated density* $h(\underline{x}_i;\hat{\underline{\beta}}_g)$ is largest.

Wolfe has provided computer programs for the case of normal distributions. As is well known, the maximum likelihood equations for mixture problems are messy. He solves them by a multivariate Newton-Raphson method of iterative solution. This involves the assignment of arbitrary initial values to start the iterative solution, as does the general method described here.

6. Some Remarks on Statistical Inference

The maximum likelihood estimate of (B,T) is the value $(\hat{B},\hat{T})$ for which the likelihood L is largest. The quantity $L(\hat{B},\hat{T})$ is the corresponding maximum value of the likelih od. To approximate $(\hat{B},\hat{T})$ one uses the algorithm. Let $\lambda(B,T) = L(B,T)/L(\hat{B},\hat{T})$. Let F denote the large sample c.d.f. of $-2\,\ell n\,\lambda$, i.e., $\lim_{n\to\infty} \Pr[-2\ell n\,\lambda(B,T) \le x] = F(x)$. Suppose that F is independent of the true values (B,T). E.g., it may be the c.d.f. of a chi-square distribution with an appropriate number of degrees of freedom; it is necessary to investigate the extent to which the large sample theory of the generalized likelihood ratio applies when there are incidental parameters.

6.1. Confidence Sets

Let x_α denote the upper α-th percentage point of F. Then $1-\alpha = F(x_\alpha) \doteq \Pr[-2\,\ell n\,\lambda(B,T) \le x_\alpha] = \Pr[-2\,\ell nL(B,T) \le x_\alpha + 2\,\ell nL(\hat{B},\hat{T})]$, so that $\{(B,T): -2\,\ell nL(B,T) \le x_\alpha + 2\,\ell n(\hat{B},\hat{T})\}$ is an approximate $100(1-\alpha)\%$ confidence set for (B,T). Denote by $(\tilde{B},\tilde{T})$ the estimates produced by the algorithm. Then $L(\tilde{B},\tilde{T}) \le L(\hat{B},\hat{T})$. Thus a conservative confidence set--one that contains more values of (B,T) than the true confidence set and has confidence coefficient at least $1-\alpha$ -- is $\{(B,T): -2\,\ell n\,L(B,T) \le x_\alpha + 2\,\ell n\,L(\tilde{B},\tilde{T})\}$.

6.2. Some Remarks on Choice of k

The algorithm can be run with different choices of k and the results compared. Note that the likelihood function is a different function for different values of k. Denote this dependence upon k by writing the likelihood as $L_k(B_k,T_k)$. Let $\hat{B}_k$, $\hat{T}_k$ denote the maximum likelihood estimates for fixed k. Following Wolfe's approach for the standard mixture model, one might make a sequence of hypothesis tests to decide on k, first comparing $L_2(\hat{B}_2,\hat{T}_2)$ with $L_3(\hat{B}_3,\hat{T}_3)$, then if necessary comparing $L_3(\hat{B}_3,\hat{T}_3)$ with $L_4(\hat{B}_4,\hat{T}_4)$, etc. Wolfe uses the asymptotic chi-square distribution of the generalized likelihood ratio here; even in the context of the standard mixture model this may not be the asymptotic distribution.

An alternative approach to choice of k is to follow MacQueen's suggestion of introducing refinement and coarsening parameters R and C such that two clusters join when their mean vectors are less than R units apart and a cluster splits when its diameter exceeds C.[6]

7. Conclusions

A modification of the usual mixture model has been employed to provide a probability framework for clustering/segmentation problems. A general method of producing algorithms which correspond to a method of iterated maximum likelihood has been given. The general method given here is plausible, is linked to a probability model, and is easy to program. In the case of multivariate normal distributions with common covariance matrix the general method produces schemes which can be viewed as improved versions of some existing schemes.

The focus here has been on the parametric case, but the methods discussed might be applied to the nonparametric case by estimating the p.d.f's $h_g(\underline{x})$ as the clustering proceeds, using standard methods of density estimation.

Algorithms based on a likelihood function are based on the raw data matrix, in contrast to many clustering procedures which are based on a matrix of pairwise similarities or distances. The latter procedures have the advantage of applicability to problems where a raw data matrix is not available. When the raw data are available, such algorithms have the theoretical disadvantage of not extracting all the information from the observations and the computational disadvantage of preliminary computation of all the pairwise distances (or similarities).

Alternative models for image segmentation. The focus here has been on a model in which the segment-identification (pixel-classification) parameters $\underline{\theta}_i$ are treated as functionally independent. In the standard mixture model they become random variables and are treated as statistically independent. To Assumptions (1) and (2) of Sec. 1 it seems reasonable to add

Assumption (3). Each segment consists of more than one pixel.

As a corollary to this assumption, it follows that the $\underline{\theta}_i$'s are functionally related, in as much as each $\underline{\theta}_i$ must be equal to one of its eight neighbors. It would be interesting to study the problem resulting from maximizing the likelihood function under this condition. Alternatively, if the $\underline{\theta}_i$'s are then treated as random, they would be a two-dimensional Markov process. It will be interesting to study the problem of estimating them in this model.

Acknowledgements

This report was prepared under Office of Naval Research Contract N00014-80-C-0408, Task NR042-443.

Bibliography

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis.* New York, Wiley, 1958.

[2] G. H. Ball & D. J. Hall, "A clustering technique for summarizing multivariate data, *Behavioral Sci.*, vol. 12, 153-155, 1967.

[3] H. Chernoff, "Metric considerations in cluster analysis," *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, vol. 1. Los Angeles and Berkeley: Univ. of Calif. Press, 621-629, 1970.

[4] N. E. Day, "Estimating the components of a mixture of normal distributions," *Biometrika*, vol. 56, 463-475, 1969.

[5] J. O. Eklundh, H. Yamamoto & A. Rosenfeld, "A relaxation method for multispectral pixel classification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-2, 72-75, 1980.

[6] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, vol. 1. Los Angeles and Berkeley: Univ. of Calif. Press, 281-297, 1966.

[7] J. Ortega & W. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* New York: Academic Press, 1970.

[8] S. L. Sclove, "Population mixture models and clustering algorithms," *Communications in Statistics (A)*, vol. A6, 417-434, 1977.

[9] R. Southwell. *Relaxation Methods in Engineering Science: a Treatise on Approximate Computation.* London: Oxford Univ. Press, 1940.

[10] R. Southwell. *Relaxation Methods in Theoretical Physics.* London and New York: Oxford Univ. Press (Clarendon), 1946.

[11] J. H. Wolfe, "Pattern clustering by multivariate mixture analysis," *Multivariate Behavioral Research*, vol. 5, 329-350, 1970.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: Technical Report 80-1 ✓	2. GOVT ACCESSION NO.: AD-A088183
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Application of the Conditional Population-Mixture Model to Image Segmentation	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s): Stanley L. Sclove	8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0408
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Mathematics ✓ University of Illinois at Chicago Circle Box 4348, Chicago, IL 60680	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE: August 15, 1980
	13. NUMBER OF PAGES: 4
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Office of Naval Research Statistics and Probability Branch Arlington, VA	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)

Unlimited distribution

18. SUPPLEMENTARY NOTES

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

Image processing, pattern recognition, pixel classification; mixtures of distributions, cluster analysis, isodata procedure, k-means procedure, Mahalanobis distance, multivariate analysis; relaxation methods.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The problem of image segmentation is considered in the context of a mixture of probability distributions. A modification of the usual approach to mixtures of distributions is employed. Parametric families of distributions are considered, a set of parameter values being associated with each distribution. In addition, an identification parameter is associated with each observation, indicating from which distribution the observation arose. Thus, the segmentation problem is cast as a problem of statistical estimation. Segmentation algorithms are obtained by applying a method of iterated maximum likelihood to the resulting likelihood function.

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

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)