

AD-A124 917

DEVELOPMENT OF PROCEDURES AND ALGORITHMS FOR PATTERN  
RECOGNITION AND IMAG. (U) ILLINOIS UNIV AT CHICAGO  
CIRCLE DEPT OF QUANTITATIVE METHODS S L SCLOVE

1/1

UNCLASSIFIED

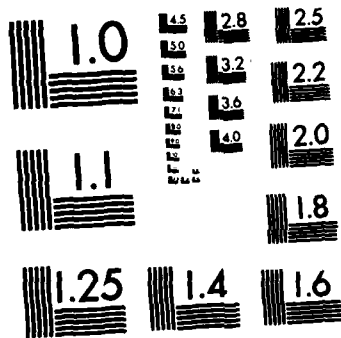
15 FEB 83 UIC/DQM/N82-8 N00014-80-C-0408

F/G 12/1

NL



END  
FILMED  
36



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

12

AD A124917

FINAL REPORT:  
CONTRACT N00014-80-C-0408, TASK NRO42-443

DEVELOPMENT OF PROCEDURES AND ALGORITHMS FOR  
PATTERN RECOGNITION AND IMAGE PROCESSING  
BASED ON TWO-DIMENSIONAL MARKOV MODELS

by

STANLEY L. SCLOVE

FINAL REPORT  
February 15, 1983

PREPARED FOR THE  
OFFICE OF NAVAL RESEARCH  
UNDER  
CONTRACT N00014-80-C-0408, TASK NRO42-443

DTIC  
SELECTED  
FEB 25 1983  
H

Principal Investigator: Stanley L. Sclove

Reproduction in whole or in part is permitted  
for any purpose of the United States Government.  
Approved for public release; distribution unlimited

QUANTITATIVE METHODS DEPARTMENT  
COLLEGE OF BUSINESS ADMINISTRATION  
UNIVERSITY OF ILLINOIS AT CHICAGO  
BOX 4348, CHICAGO, IL 60680

2/21/83

DTIC FILE COPY

008

FINAL REPORT:  
CONTRACT N00014-80-C-0408, TASK NRO42-443

DEVELOPMENT OF PROCEDURES AND ALGORITHMS FOR  
PATTERN RECOGNITION AND IMAGE PROCESSING  
BASED ON TWO-DIMENSIONAL MARKOV MODELS

STANLEY L. SCLOVE  
Department of Quantitative Methods, College of Business Administration  
University of Illinois at Chicago

CONTENTS

- Abstract
1. Introduction
  2. Major Research Results
    - 2.1. Multivariate analysis
    - 2.2. Cluster analysis
    - 2.3. Image analysis
      - 2.3.1. Background
      - 2.3.2. Segmentation
  3. Computer Programs Developed
  4. Plans for continuation of the work
  5. List of Technical Reports of the Project
  6. Abstracts of Technical Reports of the Project

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Avail and/or	
Dist	Special
A	



FINAL REPORT:  
CONTRACT N00014-80-C-0408, TASK NR042-443

DEVELOPMENT OF PROCEDURES AND ALGORITHMS FOR  
PATTERN RECOGNITION AND IMAGE PROCESSING  
BASED ON TWO-DIMENSIONAL MARKOV MODELS

STANLEY L. SCLOVE  
University of Illinois at Chicago

ABSTRACT

This report summarizes the major research work accomplished under the contract, including the topics of cluster analysis, time-series segmentation, image segmentation, and related software development. The period of the contract was April 1, 1980 to December 31, 1982. A list of technical reports written under the project is given. A list of programs in a software package developed under the project is also given.

1. Introduction

This is a final report for Contract N00014-80-C-0408, Task NR042-443, Development of Procedures and Algorithms for Pattern Recognition and Image Processing based on Two-Dimensional Markov Models. The period of this contract was 4/1/80 - 12/31/82. The report is organized according to the research topics pursued under the contract. A brief summary of each is presented. Brief descriptions of computer programs developed under the project are given. The program listings on magnetic tapes were sent in May, 1982, to Dr. Douglas J. DePriest for deposit in the Naval Research

Laboratories. A complete list of the project's technical reports (TRs) is provided in the References. Bibliographic information is given for those TRs which have been published or accepted for publication. Abstracts of the TRs are included.

## 2. Major Research Results

### 2.1. Multivariate analysis

TR 82-3 is a summary of earlier work on infinitely divisible distributions, together with some results not before presented. Infinitely divisible distributions may prove useful in modeling the signal-noise mixture in certain signal detection and image analysis contexts.

### 2.2. Cluster analysis

Earlier work on cluster analysis includes Sclove's 1977 paper "Population Mixture Models and Clustering Algorithms" (Communications in Statistics, Vol. A6, pp. 417-434). Extrapolating from that work, Sclove directed H. Bozdogan's doctoral dissertation, "Multi-Sample Cluster Analysis and Approaches to Validity Studies in Clustering Individuals" (Math. Dept., Univ. of Illinois at Chicago, 1981). TRs 82-1, 82-2 and 82-6 emanate from this Ph.D. thesis. These TRs relate to multi-sample clustering, the grouping of samples. In the problems of clustering and segmentation treated under this contract, statistical model-selection criteria were specialized for and applied to the various problems treated. Multi-sample cluster analysis is a case in

\*\*\*\*\*

point. Use of model-selection criteria in this context can provide an alternative to conventional multiple-comparisons procedures. The model-selection approach avoids the difficulty of choice of levels of significance in multiple-comparisons problems.

### 2.3. Image analysis

#### 2.3.1. Earlier work

Dr. Sclove entered the field of digital image processing through consultation on a project involving the probabilistic analysis of fingerprint characteristics. (J. W. Osterburg, T. Parthasarathy, T.E.S. Raghavan, and S. L. Sclove, "Development of a Mathematical Formula for the Calculation of Fingerprint Probabilities based on Individual Characteristics," Journal of the American Statistical Association, Vol. 72 (1977), pp. 772-778; Stanley L. Sclove, "The Occurrence of Fingerprint Characteristics as a Two-Dimensional Process," Journal of the American Statistical Association, Vol. 74 (1979), pp. 588-595; Stanley L. Sclove, "The Occurrence of Fingerprint Characteristics as a Two-Dimensional Poisson Process," Communications in Statistics, Vol. A9 (1980), pp. 675-695.) The approach used was to superimpose a grid of cells over the fingerprint, and to note the number of occurrences of the ten standard (Galton) fingerprint characteristics, cell by cell. The cells correspond to the pixels in digital image processing. TR 80-2 is a summary of this work.

#### 2.3.2 Segmentation

TR 80-1 (revised in TR 82-5) details the use in image segmentation

of the population mixture model for cluster analysis.

This project initiated the use of transition probabilities to model the labels of the classes of segment. This approach was used in both time-series segmentation and image segmentation.

Time-series segmentation (TRs 81-1, 82-4, 82-7) is interesting in its own right and also may be considered as a one-dimensional version of image segmentation.

Image segmentation is treated in TRs 80-1, 82-4, and 82-5. The approach, based on transition probabilities for the labels, utilizes both spatial and contextual information, in that the transition probabilities tell how frequently there is a change in label as one scans the image and also these probabilities of change are class-dependent.

### 3. Computer Programs Developed

PROG.IMSEG1 for segmenting uni-channel digital images using an algorithm based on a Markov model for the spatial dependence of the pixel labels

PROG.IMSEGP.COMMON.VRSN39 for segmenting multi-channel digital images using an algorithm based on a Markov model for the spatial dependence of the pixel labels

PROG.ISOLINE.TRANS.VRSN57 for segmenting time series using an algorithm based on a Markov chain model for the labels

PROG.ISOLINE.TRANS.SPARSE.VRSN24 for segmenting time series using an algorithm based on a Markov chain model for the labels, with sparse

\*\*\*\*\*

transition probability matrix

PROG.ISOLINE.MANUAL.VRSN21 for clustering univariate data by ISODATA procedure

PROG.ISOLINE.AUTO.VRSN21 for clustering univariate data by ISODATA procedure, automatically trying different numbers of clusters and scoring the results by Akaike's information criterion (AIC)

PROG.ISODATA.COMMON.VRSN48 for clustering multivariate data by a variant of the ISODATA procedure, using the within-groups covariance matrix

The model giving the algorithm used in PROG.ISOLINE.TRANS and PROG.ISOLINE.TRANS.SPARSE is discussed in TR 81-1, "On Segmentation of Time Series."

#### 4. Continuation of the work

So far the greatest attention has been paid to the case in which the observations are conditionally independent, given the labels. Algorithms for fitting the model in other situations will be developed.

The focus so far has been on Gaussian data. At the same time, however, a general approach to segmentation problems has evolved from the project. Using this general approach, algorithms will be developed which are specialized to particular situations of interest, including the case of binary data. Also to be considered are scanning problems, where one knows in advance that the segmentation is of A-B-A form.

Model-selection criteria other than Akaike's (AIC) will be applied and the results compared to those obtained with AIC.

\*\*\*\*\*

Consideration will also be given to use of the two-dimensional Markov model to study efficient compression for storage and archiving digital images.

Work on these and other problems is continuing under the support of Army Research Office Contract DAAG29-82-K-0155, Cluster Analysis and Image Segmentation, University of Illinois at Chicago, Stanley L. Sclove, Principal Investigator. It is hoped that further funding may eventually be obtained from ONR.

#### 5. List of Technical Reports of the Project

- No. 80-1. Stanley L. Sclove. "Application of the Conditional Population-Mixture Model to Image Segmentation." 8/15/80  
(See also No. 82-5.)
- No. 80-2. Stanley L. Sclove. "Modeling the Distribution of Fingerprint Characteristics." 9/19/80  
(Published in C. Taillie et al., eds., Statistical Distributions in Scientific Work (NATO ASI Series). D. Reidel, Dordrecht, Holland, 1981. Vol. 6, pp. 111-130.)
- No. 81-1. Stanley L. Sclove. "On Segmentation of Time Series." 11/30/81  
(Invited Paper, Special Session on Cluster Analysis, Meeting of the American Mathematical Society, Univ. of Mass., Amherst, MA, Oct. 16-18, 1981)
- No. 82-1. Hamparsum Bozdogan and Stanley L. Sclove. "Multi-Sample Cluster Analysis using Akaike's Information Criterion." 1/30/82
- No. 82-2. Hamparsum Bozdogan and Stanley L. Sclove. "Multi-Sample Cluster Analysis with Varying Parameters using Akaike's Information Criterion." 3/8/82
- No. 82-3. Stanley L. Sclove. "Some Aspects of Inference for Multivariate Infinitely Divisible Distributions." 6/15/82 (Accepted for

\*\*\*\*\*

publication in Statistics & Decisions, 1983.)

- No. 82-4. Stanley L. Sclove. "On Segmentation of Time Series and Images in the Signal Detection and Remote Sensing Contexts." 8/1/82
- No. 82-5. Stanley L. Sclove. "Application of the Conditional Population-Mixture Model to Image Segmentation." 8/15/82  
Revision of Technical Report No. 80-1.  
(Accepted for publication in IEEE Trans. Pattern Analysis & Machine Intelligence, 1983.)
- No. 82-6. Hamparsum Bozdogan and Stanley L. Sclove. "Multi-sample Cluster Analysis using Akaike's Information Criterion." 12/20/82  
(To appear in Annals of Institute of Statistical Mathematics, Series B.)
- No. 82-7. Stanley L. Sclove. "Time-Series Segmentation: a Model and a Method." 12/22/82  
(Presented at the Workshop on Applied Time Series Analysis, Munich, Germany, Oct. 22-23, 1982. Accepted for publication in Information Sciences.)

#### 6. Abstracts of Technical Reports of the Project

- No. 80-1. Stanley L. Sclove. "Application of the Conditional Population-Mixture Model to Image Segmentation." 8/15/80  
(Presented at the 5th International Conference on Pattern Recognition, Miami, Dec. 1-4, 1980. Revised as TR No. 82-5.)

Image segmentation is modeled as a statistical estimation problem by associating with each class of segment a probability distribution and with each pixel an identification (pixel-classification) parameter. The problem is then to estimate simultaneously the parameters of the distributions and the identification parameters. Segmentation algorithms are obtained by applying the method of maximum likelihood to the likelihood function resulting from the model. A relaxation method for accomplishing the maximization is presented. The similarities and differences between the resulting algorithms and some statistical clustering algorithms (isodata and k-means) are discussed. The relationship of the model to the conditional population-mixture model for statistical cluster analysis is explained. The conditional population-mixture model is contrasted with the standard population-mixture model, which would suggest modeling the pixel-classification parameters as random variables. Some discussion of alternative models is given. It is suggested that functional (spatial) dependence of the pixel-classification parameters be introduced. If they

\*\*\*\*\*

are then treated as random variables, they will then form a two-dimensional Markov-type process.

No. 80-2. Stanley L. Sclove. "Modeling the Distribution of Fingerprint Characteristics." 9/19/80 (Published in C. Taillie et al., eds., Statistical Distributions in Scientific Work (NATO ASI Series). D. Reidel, Dordrecht, Holland, 1981. Vol. 6, pp. 111-130.

Quantitative aspects of fingerprints are discussed. A study undertaken to develop methods for assigning probabilities to partial fingerprints is summarized, with emphasis on distributional aspects. Key Words & Phrases: fingerprints; two-way series; multinomial distribution; Markov process; Poisson process

No. 81-1. Stanley L. Sclove. "On Segmentation of Time Series." 11/30/81

The problem of partitioning a time-series into segments is considered. The segments fall into classes, which may correspond to phases of a cycle (recession, recovery, expansion in the business cycle) or to portions of a signal obtained by scanning (background/clutter, target, background/clutter again, another target, etc.; or normal tissue, tumor, normal tissue). Parametric families of distributions are considered, a set of parameter values being associated with each class. With each observation is associated an unobservable label, indicating from which class the observation arose. The label process is modeled as a Markov chain. Segmentation algorithms are obtained by applying a method of iterated maximum likelihood to the resulting likelihood function. In this paper special attention is given to the situation in which the observations are conditionally independent, given the labels. A numerical example is given. Choice of the number of classes, using Akaike's automatic (model) identification criterion (AIC), is illustrated. Prediction is considered.

Key Words & Phrases: forecasting; prediction; signal analysis; isodata procedure; Markov chains; maximum likelihood; Akaike's information criterion (AIC)

No. 82-1. Hamparsum Bozdogan and Stanley L. Sclove. "Multi-Sample Cluster Analysis using Akaike's Information Criterion." 1/30/82

Multi-sample cluster analysis, the problem of grouping samples, is studied from an information-theoretic viewpoint via Akaike's Information Criterion (AIC). This criterion combines the maximum value of the likelihood with the number of parameters used in achieving that

\*\*\*\*\*

value. The multi-sample cluster problem is defined, and AIC is developed for this problem.

The form of AIC is derived in both univariate and multivariate analysis of variance models. Numerical examples are presented and results are shown to demonstrate the utility of AIC in identifying the best clustering alternatives.

Key Words and Phrases: Multi-sample cluster analysis; Akaike's Information Criterion (AIC); ANOVA Model, MANOVA Model; maximum likelihood.

No. 82-2. Hamparsum Bozdogan and Stanley L. Sclove. "Multi-Sample Cluster Analysis with Varying Parameters using Akaike's Information Criterion." 3/8/82

Multi-sample cluster analysis, the problem of grouping samples, is studied from an information-theoretic viewpoint via Akaike's Information Criterion (AIC). This criterion combines the maximum value of the likelihood with the number of parameters used in achieving that value. The multi-sample cluster problem is defined, and AIC is developed for this problem. The form of AIC is derived in the univariate model with varying means and variances, and in the multivariate model with varying mean vectors and variance-covariance matrices. Numerical examples are presented and results are shown to demonstrate the utility of AIC in identifying the best clustering alternatives.

Key Words and Phrases: Multi-sample cluster analysis; Akaike's Information Criterion (AIC); Univariate model with varying means and variances, Multivariate model with varying mean vectors and variance-covariance matrices; maximum likelihood.

No. 82-3. Stanley L. Sclove. "Some Aspects of Inference for Multivariate Infinitely Divisible Distributions." 6/15/82 (Accepted for publication in *Statistics & Decisions*, 1983.)

Measurement of dependence in the infinitely divisible class of distributions, based on developments in probability theory for that class, is discussed. It has been shown that pairwise independence is equivalent to mutual independence in this class. When the infinitely divisible variables contain no normal component (in particular, when they are discrete), the cumulant of order (2,2) can be used as a measure of pairwise dependence; when a normal component is present, the appropriate measure also involves the covariance. Results for testing independence of infinitely divisible random variables are discussed. A method of testing normality against infinitely divisible alternatives is given.

- No. 82-4. Stanley L. Sclove. "On Segmentation of Time Series and Images in the Signal Detection and Remote Sensing Contexts." 8/1/82 (Presented at ONR Workshop on Signal Processing in the Ocean Environment, Annapolis, Md., May 11-14, 1982. To be published in a proceedings volume.)

The problem of partitioning a time-series into segments is considered. The segments fall into classes, which may correspond to phases of a cycle (recession, recovery, expansion in the business cycle) or to portions of a signal obtained by scanning (background/clutter, target, background/clutter again, another target, etc.), or normal tissue, tumor, normal tissue in medical applications. A probability distribution is associated with each class of segment. Parametric families of distributions are considered, a set of parameter values being associated with each class. With each observation is associated an unobservable label, indicating from which class the observation arose. The label process is modeled as a Markov chain. Segmentation algorithms are obtained by applying a method of iterated maximum likelihood to the resulting likelihood function. In this paper special attention is given to the situation in which the observations are conditionally independent, given the labels. A numerical example is given. Choice of the number of classes, using Akaike's information criterion (AIC) for model identification, is illustrated. Prediction is considered.

Similar ideas are applied in the image processing context, where a possible application is the construction of a topographical map by segmentation of the digital image according to spectral properties of the various topographical objects. Thus, in particular, the methods can be applied to LANDSAT (or SEASAT) data.

- No. 82-5. Stanley L. Sclove. "Application of the Conditional Population-Mixture Model to Image Segmentation." 8/15/82  
Revision of Technical Report No. 80-1.  
(Accepted for publication in IEEE Trans. Pattern Analysis & Machine Intelligence, 1983.)

The problem of image segmentation is considered in the context of a mixture of probability distributions. The segments fall into classes. A probability distribution is associated with each class of segment. Parametric families of distributions are considered, a set of parameter values being associated with each class. With each observation is associated an unobservable label, indicating from which class the observation arose. Segmentation algorithms are obtained by applying a method of iterated maximum likelihood to the resulting likelihood function. A numerical example is given. Choice of the number of classes, using Akaike's information criterion (AIC) for model identification, is illustrated.

Key Words and Phrases: Image processing, image segmentation, pixel classification; pattern recognition; mixtures of distributions; cluster analysis, isodata procedure, k-means procedure; Mahalanobis distance, multivariate statistical analysis; relaxation methods

No. 82-6. Hamparsum Bozdogan and Stanley L. Sclove. "Multi-sample Cluster Analysis using Akaike's Information Criterion." 12/20/82 (To appear in Annals of Institute of Statistical Mathematics, Series B.)

Multi-sample cluster analysis, the problem of grouping samples, is studied from an information-theoretic viewpoint via Akaike's Information Criterion (AIC). This criterion combines the maximum value of the likelihood with the number of parameters used in achieving that value. The multi-sample cluster problem is defined, and AIC is developed for this problem. The form of AIC is derived in both the multivariate analysis of variance (MANOVA) model and in the multivariate model with varying mean vectors and variance-covariance matrices. Numerical examples are presented for AIC and another criterion called w-square. The results demonstrate the utility of AIC in identifying the best clustering alternatives.

Key Words and Phrases: Multi-sample cluster analysis; w-square criterion; Akaike's Information Criterion (AIC); MANOVA model; multivariate model with varying mean vectors and variance-covariance matrices; maximum likelihood

No. 82-7. Stanley L. Sclove. "Time-Series Segmentation: a Model and a Method." 12/22/82 (Presented at the Workshop on Applied Time Series Analysis, Munich, Germany, Oct. 22-23, 1982. Accepted for publication in Information Sciences.)

The problem of partitioning time-series into segments is treated. The segments are considered as falling into classes. A different probability distribution is associated with each class of segment. Parametric families of distributions are considered, a set of parameter values being associated with each class. With each observation is associated an unobservable label, indicating from which class the observation arose. The label process is modeled as a Markov chain. Segmentation algorithms are obtained by applying a relaxation method to maximize the resulting likelihood function. In this paper special attention is given to the situation in which the observations are conditionally independent, given the labels. A numerical example, segmentation of U.S. Gross National Product, is given. Choice of the

number of classes, using statistical model-selection criteria, is illustrated.

Key Words and Phrases: Markov chains; maximum likelihood; maximum a posteriori estimation; Viterbi algorithm; relaxation methods; isodata procedure; model-selection criteria; Akaike's information criterion (AIC).

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER UIC/DQM/N82-8	2. GOVT ACCESSION NO. AD-A124 717	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Final Report: Contract N00014-80-C-0408, Task NRO42-443, Development of Procedures and Algorithms for Pattern Recognition and Image Processing based on Two-Dimensional Markov Models		5. TYPE OF REPORT & PERIOD COVERED Final 4/1/80 - 12/31/82
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stanley L. Sclove	8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0408 (NRO42-443)	
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Illinois at Chicago Box 4348, Chicago, IL 60680		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Statistics & Probability Branch Office of Naval Research/Dept. of the Navy Arlington, VA 22217		12. REPORT DATE February 15, 1983
		13. NUMBER OF PAGES i + 12
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) cluster analysis; pattern recognition; time-series segmentation; digital image processing; image segmentation; model-selection criteria		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report summarizes the major research work accomplished under the contract, including the topics of cluster analysis, time-series segmenta- tion, image segmentation, and related software development. The period of the contract was 4/1/80 - 12/31/82. A list of technical reports written under the project is given. A list of programs in a software package developed under the project is also given.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 68 IS OBSOLETE  
S. N 3102-LF-314-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END