Machine Intelligence and Pattern Recognition 7

PATTERN RECOGNITION AND ARTIFICIAL INTELLIGENCE

Edited by
E. S. Gelsema and L. N. Kanal

North-Holland

Pattern Recognition and Artificial Intelligence

Towards an Integration

Proceedings of an International Workshop
held in Amsterdam, May 18–20, 1988

Edited by

Edzard S. GELSEMA
Department of Medical Informatics
Erasmus University, Rotterdam

and

Laveen N. KANAL
Department of Computer Science
University of Maryland, College Park, MD

1988

NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD · TOKYO

ISBN: 0 444 87137 3

Publishers:

ELSEVIER SCIENCE PUBLISHERS B.V.
P.O. Box 103
1000 AC Amsterdam
The Netherlands

Sole distributors for the U.S.A. and Canada:

ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
52 Vanderbilt Avenue
New York, N.Y. 10017
U.S.A.

PRINTED IN THE NETHERLANDS

PREFACE

The excellent reception accorded to our 1980 and 1985 workshops on Pattern Recognition in Practice and to the resulting two books, encouraged us to hold the third Pattern Recognition in Practice workshop in May 1988 and to assemble the edited papers in this volume. The theme for this workshop was broadenend to include artificial intelligence methodologies, such as expert and knowledge based systems, which are increasingly being incorporated in pattern recognition systems. As in previous workshops, the focus was on bringing together researchers in methodology with those working on specific applications.
Thirty two papers were presented at the conference and have been organized into the following three parts:

1. Image analysis and processing software; systems and algorithms;
2. Pattern analysis and classifier design;
3. Model based- and expert systems, including uncertainty calculus methods in pattern analysis and object recognition.

The first two papers of part I are on image analysis and image processing software packages, respectively. Image analysis is distinguished from image processing in that the former is concerned with the extraction of information from an image, whereas the latter mostly involves manipulating one image to produce another image.
The image software package Acuity, presented by Young and Roos is written in C and makes use of the Apple Macintosh (tm) user interface. It is intended for use in automated cytology.
The image processing package Lily, described in the paper by Dewaele et al. is a hierarchically structured software toolkit from which the appropriate routines must be selected by the user. Two applications, viz. defect detection in textiles and in unexposed radiographic film are described.
The next paper, by Persoon, analyses the basic requirements for response time and types of algorithms in different real-time image analysis tasks, especially those encountered in industrial applications; it is argued that no existing system meets all the desired requirements and that it is necessary to re-examine known algorithms in terms of the stated requirements.

The next four papers in the first part discuss algorithms for finding edge, line, shape and symmetry features in images.
The paper by Danielsson and Ye proposes to use a combination of certain even order rotation invariant operators for detecting the magnitude and orientation of lines and shows an application to fingerprints.
The paper by Van Vliet and co-authors describes the performance of a non-linear Laplace operator for edge detection which has an adaptive orientation perpendicular to the local edge direction, decreased sensitivity to noise and ability to detect one pixel thick edges as well or better than competing detection techniques.
The following paper, by Bigün, presents a method to model symmetries of the neighbourhoods in grey value images based on the form of the iso-greyvalue curves. Through a convolution of a complex-valued image with a complex-valued filter, a complex valued number is obtained for each neighbourhood. The magnitude and argument of this complex valued number are then used to determine symmetry descriptions.

Experimental results on the application of the methodology to circular, linear, hyperbolic and parabolic symmetries are presented.
Accurate measurement of shape features in images is the subject of the paper by Smeulders and Worring. For a given contour, they find the set of predigitized figures, which could have generated it. By maximizing an accuracy criterion, arbitrary local shape features are then estimated. This feature definition is shown to be optimal. Practical algorithms, including one suited for parallel implementation are given.
Sometimes, objects of interest in pattern recognition problems are represented as polygons and their shape is measured in terms of visibility relations between edges. The paper by Toussaint surveys some recent results in computational geometry on efficient computation of visibility properties between edges of a simple polygon, and on computing other more general visibility relations for polygons.

The next two papers describe the use of image processing techniques for applications in robotics.
The paper by Nagy and Falsafi discusses a method to locate an object in three-dimensional space from an image, using a single camera. In general, the transformation describing the relationship between points in the image plane and in three dimensional space is described by a number of non-linear equations, and a closed-form solution cannot be formulated. The method presented in this paper makes use of the vanishing points (intersections of the projections of parallel lines onto the image plane). It is shown, that using such points, a closed-form solution is possible. This is of interest, e.g. for tracking rapidly moving objects, where an iterative solution of the transformation equations would be much too time-consuming.
The paper by Rijnierse and Groen is also concerned with stereo vision. A method is described to link image primitives (i.c. edges) in a relational graph. The graphs corresponding to the left and right images are then matched and depth is calculated from the disparity between the elements. Orientation and position of the object are then found by matching the 3-D graph with a 3-D object model. An error analysis of the method is presented in the paper.

The last three papers of Part I describe applications of image processing in medicine. The papers by Dumay and co-workers and by Van Cuyck and colleagues both deal with angiograms.
Dumay et al. describe the problem of three-dimensional reconstruction of the myocardial contrast distribution from two orthogonal cineangiograms. The bifurcations of the left coronary arterial tree are used for matching the two 2-D views. The contrast distribution is then reconstructed using cost coefficient matrices in a linear programming technique. The problem of properly choosing the cost coefficient matrix elements and of the lower- and upperbound capacities is still a matter of research.
The paper by Van Cuyck and colleagues discusses the performance of two algorithms, a "beam" algorithm and a minimal cost algorithm, for the automated tracing of centerlines in coronary angiograms. The beam algorithm performs well on all arterial segments, whereas the minimal cost algorithm does well on straight segments but rather poorly on curved segments and on those in which the contrast with respect to the background is low.
Finally, the paper by Stark and Peng discusses the efficient estimation of the shape of opacities in computer tomograms. In the paper it is shown that for this purpose only $O(N)$ operations are required, rather than $O(N^3)$ required for a complete reconstruction. A surprising result is that there exists an optimum in the number of views used for a given detector spacing: increasing the number of views beyond the optimum results in a deterioration of the estimation error.

In the first paper of Part II, Jain and Ramaswami discuss the design of classifiers using Parzen windows, giving special attention to the problem of the optimum window size. A number of window-width estimation techniques are compared. The Bootstrap technique is shown to be a reasonable strategy. The performance of the resulting classifiers is compared with other classifiers using several real data sets. The results suggest that in designing pattern recognition systems, Parzen window classifiers should be given serious consideration.

Duin and Backer discuss the problem of design and evaluation of classifiers based on learning sets that are not random samples of the universe. It is argued that in such cases the use of fuzzy labels may improve the quality of the classification.
In the next paper, Queiros and Gelsema review methods to estimate missing values. Furthermore, simulation experiments are described to determine an optimum strategy for the design and use of classifiers in the case of incomplete data sets. It is shown that a combination of the "delete" technique and the "estimate and replace" technique leads to a strategy that is better than using one of these techniques exclusively.
The paper by Morgera and Soleymani presents a rate-distortion view of pattern recognition. They describe the pattern recognition problem in terms of a communications channel model. Just as vector quantization, based on rate-distortion theory, has led to advances in speech recognition, the notion of vector recognition is advanced as it may well quantify the improvement of recognition performance for isolated versus strings of symbols in character and object recognition problems.
The paper by Siedlecki et al. presents a comprehensive overview of mapping techniques, which they systematize in eight groups. Such mappings may be used to enable user interaction in cluster finding processes or in classifier design. Two experiments to evaluate the usefulness of such user interaction, using various mapping techniques are described. It is shown that untrained humans aided by mappings may outperform automatic techniques. The difference of performance between non-experts and experts is also evaluated and it is shown that the ability to efficiently use the information carried by the mappings is enhanced by extended training.

The final two papers in Part II describe applications of pattern recognition methodology.
The paper by Timmers and Gelsema presents a model for the design of classifiers of object populations. Their application is automated microscopy of cytological specimens. The design of a cascade of classifiers on the basis of compound distributions is described. The model derived in this way is shown to unify the cell-class and cell-feature approaches. Existing models are extended by including between-specimen variability. It is also shown that pooling of the cells within diagnostic classes, as is done conventionally, leads to a suboptimal procedure.
In his paper Kurtz discusses the use of pattern recognition techniques in astronomy. While classification methods have already been used in astronomy for more than a century, the development of new generations of detectors create a new situation in which huge streams of information have to be handled, necessitating the development of feasible techniques of data reduction. The need for implementations on multiple parallel processors is clearly indicated.

Part III opens with an opinion paper by Musen and Van der Lei. They describe the phenomenon of brittleness and the knowledge acquisition bottleneck in the construction of expert systems as the major problems in the field of knowledge based systems. One of the major differences between models used in expert systems and models used in pattern recognition is that the former are usually qualitative as opposed to the explicit statistical theories implemented in pattern recognition models. Similarities among various expert-system models have been obscured by the fact that knowledge engineers have not viewed their work as the application of a "bag of tools". The focus on task knowledge has deemphasized the problem-solving strategies. It is argued that as the AI community begins to identify and apply generic problem solving models, knowledge acquisition and expert-system validation will be established on firmer theoretical foundations. A vivid discussion (in essence presented on pages 483-485) underlines the importance of the matter discussed and justifies the expectation expressed in the subtitle of this book.
In the next paper, by Bhatnagar and Kanal, a graph based formalism is presented that represents a link between observable variables and variables of interest in a given problem situation characterized by incomplete and/or uncertain knowledge. A number of different underlying causal structures may be possible and the objective is to provide a mechanical reasoner with the capability of constructing such problem models and to choose between them on the basis of their computed level of interest.
The paper by Van Ginneken and Smeulders analyses the propagation of uncertainty in five well known reasoning models. The combinatorics used in Bayes theorem, MYCIN,

Dempster-Shafer, Fuzzy logic and INTERNIST are compared and the different aspects of uncertainty expressed in the various systems are analyzed. These questions are addressed especially from the point of view of the use of such reasoning systems in pathology. Their conclusion is that the combinatorics of the Bayes' rule and of Dempster-Shafer are preferable, although the acquisition of the required information is not an easy task.
In the paper by Mandler and Schürmann the theory of Demster-Shafer is applied to the problem of combining the results of independent classifiers. In the use of different nearest neighbour classifiers for the same recognition task, the various distance measures are transformed into evidence functions, which are then combined. Results are presented on the application of the proposed technique to on-line script recognition. The resulting output vector of confidence values is shown to be compatible with a-posteriori probability vectors produced by statistically adapted classifiers.
In the next paper, Backer describes a knowledge based system for use in exploratory data analysis. Cluster analysis is presented as a complex and approximate reasoning process, which has been formalized in the Clusan/Delfi system. For any point configuration, the system produces an opinion about the configuration. Confidence values are given for the data being randomly distributed, for clustering tendency and for the number of clusters. The output is based upon the edge-length distribution of the minimum spanning tree and on k-means analyses with a varying number of candidate clusters. Finally, the system produces a measure of belief in its opinion. Three problem examples are discussed to show the system's capabilities.
In their paper, Cocca and Dubuisson present a decision method to detect the evolution of a system towards a possibly abnormal state. Their decision method is a two steps strategy. They define categories as groups of classes. They monitor a system's behaviour by combining classical pattern recognition decision rules for class assignment with production rules in an expert system for assignment to a category. Thus mixing quantitative with qualitative information, the evolution towards undesirable system states is detected before such states actually occur.
The last four papers in this book describe the use of combinations of pattern recognition and artificial intelligence techniques in bio-medical applications.
In the paper by Smets and co-authors, a system is described for the 3-D reconstruction of the cerebral blood vessels from a pair of angiograms. The stereoscopic matching process is guided by high level primitives, i.e. blood vessel segments, detected in both images separately. The processes of primitive detection and 3-D reconstruction are integrated by the use of a knowledge based system. This makes it possible to improve the detection of vessel segments in one image, using evidence obtained from the other image. This is demonstrated by showing the results on a set of clinical images.
The paper by Harauz offers an overview of the use of computers in molecular biology. A fundamental tenet of molecular biology is that the structure of a biological macromolecule or complex determines its function precisely. Therefore structure determination is very important. The paper describes structure determination in X-ray crystallography as well as in electron microscopy. Another section is devoted to a review of computational techniques in the analysis of sequences in nucleic acids and amino acids. Also, the planning of molecular genetics experiments is briefly described. In all these applications, knowledge representation remains a critical issue. Knowledge-based systems like MOLGEN and StrateGene are widely used in this area.
An application in chromosome analysis is discussed in the paper by Piper and co-workers. They describe a system which, although written in a traditional programming style, combines knowledge from multiple sources in the detection of dicentric chromosomes. They argue that a system for aberration scoring differs in many respects from a conventional karyotyping system. The system described here is intended to be a pre-screening system in which a low false positive rate is essential. In some instances, the authors deliberately reduce the system's sensitivity to achieve this goal. Results on a set of 457 cells, containing 20,093 chromosomes are presented. A false positive rate of about 0.10 chromosomes per cell is achieved.
Jordan and Perkins describe the use of expert systems in the analysis of biological images. The paper consists of two parts. The first part describes a project for the automated detection of malaria parasites in red blood cells. Following the course of a malaria infection over a period of weeks involves detection of parasites under widely varying conditions. New infections, progressive infections, the occurrence of anaemias

all have their own characteristic blood-smear composition. A knowledge base containig three levels of knowledge is intended to assist in the analysis of the slides. The second part of the paper describes the use of an expert system in the 3-D reconstruction of serial sections in electron microscopy. Features of nematodes are used to guide the reconstruction process in an interactive system. Also here, decisions at different levels for matching image features with a model are incorporated in a knowledge base.
The final paper, by Dawant and Jansen, outlines the software architecture of a system consisting of two interacting expert systems for signal understanding. In the application described, viz. the analysis of electroencephalographic data, knowledge about two completely different domains is required: the domain of signal analysis and the domain of EEG's. Two separate expert systems represent and apply these two kinds of knowledge, with interaction between them. The object-oriented approach results in a flexible system, allowing contextual information to be incorporated. Results of applying the system to 11 segments of 10 sec. each are given to indicate its potential value in practical use.

All papers in this volume indicate that pattern recognition and artificial intelligence are still expanding areas of research and of considerable importance in many, diverse applications. The advent of new detection techniques in many fields requires ever more powerful data reduction and -interpretation methods. It is our strong belief that the integration of techniques developed in pattern recognition and in artificial intelligence will result in systems which are more flexible and more powerful than those which the two methodologies in isolation are likely to produce. The papers in this book may give prospective system designers ideas for the construction of such systems and may give workers in application areas material to reflect on the usefulness of such systems in their field of expertise.

September 1988, Rotterdam, Edzard S. Gelsema
College Park, Laveen N. Kanal

ACKNOWLEDGEMENTS

This book is a compilation of the papers presented at the Conference "Pattern Recognition in Practice III", held in Amsterdam from May 18 to May 20, 1988.
We wish to thank the following Organizations, Societies and Companies for their financial support of the Conference:

The Royal Dutch Academy of Sciences

The Netherlands Organization for Scientific Research

TNO Division of Health Research

Océ-van der Grinten N.V.

IBM Nederland N.V.

Shell Nederland B.V.

We also want to thank Dr. T. Timmers and Professor J.H. van Bemmel for their active help and support during the preparatory phase of the conference.
Dr. Timmers carried out many organizational duties, also during the conference. His help is particularly acknowledged.

The Program Committee, besides the editors, consisted of Professor E. Backer and Dr. R.P.W. Duin. Their help in selecting the papers to be presented from a total number which was almost the double of what could be accepted in view of the time constraints, helped to establish the high-standard scientific contents of the conference.

During the conference, all discussions were recorded on tape and transcribed on the spot. The scientific secretaries Mr. J. van der Lei, Mr. G. Beckers, Mrs. N. Schipper and Mr. M. Worring coordinated the transcription process.

The secretariat of the conference was run by Mrs. Loes de Langen and Mrs. Jeanette Vos. Both of them worked hard to take a heavy burden off the shoulders of the conference organisers. Their diligence brought order to the hectic, behind the scene, activities involved in running a conference.

Finally, we owe a debt of gratitude to the participants whose contributions, after all determine the scientific quality of the conference. We thank the authors for their cooperation in providing their draft manuscripts and edited versions in a timely manner, leading to a relatively quick production of these Proceedings.

The editors.

CONTENTS

PART I

IMAGE PROCESSING

SECTION I

SYSTEMS

Pattern Recognition and Artificial Intelligence
E.S. Gelsema and L.N. Kanal (Editors)

ACUITY: IMAGE ANALYSIS FOR THE PERSONAL COMPUTER

Ian T. YOUNG and Roelof ROOS

Faculty of Applied Physics
Delft University of Technology
Delft, The Netherlands

The analysis of images has traditionally been confined to the domain of the minicomputer (and larger). With the advent of technology such as the MC-680x0 it is now possible to analyse complex images on personal computers in reasonable time. The **Acuity** software package provides such a possibility for the measurement of the properties of objects in an image.

The package, written entirely in C, consists of approximately 120 kB of code and provides image segmentation into the individual objects (automatic and/or interactive), measurement of object features, and measurement statistics. Automatic image segmentation is generally based upon histogram analysis with the possibility of pre-filtering. Interactive segmentation makes use of a mouse interface. After the image has been split into the individual objects, measurements are performed. The features fall into several categories: Position (2 measures), Size (2 measures), Shape (6 measures), Intensity (2 measures), and Texture (4 measures). The formulas used to compute the measures are based upon recent developments in digital–image measurement theory.
A number of utilities are available to define an experiment (e.g., to choose which features need to be measured), to print a summary of the measured data, to print a limited set of descriptive statistics, or to format the output data (feature vectors) in a manner that is compatible with commercially available data analysis software (e.g., StatWorks™, MacSpin™, Excel™).

The complete analysis of a 256^2 image that contains approximately 60 objects each with an approximate diameter of 25 pixels takes about 35 seconds. This is on a personal computer that uses a MC-68020 with a clock frequency of 16 MHz and a MC-68881 floating point co–processor.

1. Introduction

While it is not uncommon to find articles about specific hardware systems for image processing applications, it is somewhat more unusual to find articles about specific software systems (or packages) for similar goals. Nonetheless, an important part of image processing and analysis research revolves around software and this has been reflected in articles by Preston [1983], Krusemark and Haralick [1983], Tamura, et. al. [1983], and Duin [1983]. In this article we describe a software package for image *analysis*. We make a clear distinction, here, between image processing and image analysis; a distinction illustrated in Figure 1.