A STUDY IN AUTOMATIC PHOTO-INTERPRETATION

I.D.G. Macleod, B.E. (Hons.)

Submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in The Australian National University.

March, 1970

Department of Engineering Physics,
THE AUSTRALIAN NATIONAL UNIVERSITY,
CANBERRA, A.C.T.
DECLARATION

I declare that this Thesis reports my own original work, that no part of it has previously been accepted or presented for the award of any degree or diploma by any University, and that to the best of my knowledge no material previously published or written by another person is included, except where due acknowledgement is given.

Iain D.G. Macleod
Canberra
March, 1970.
This Thesis is concerned primarily with problems in automatic photo-interpretation (API), with special emphasis on naturally-given images (using pollen micrographs and alpha-particle tracks in photographic emulsion), and draws on several fields, including Computer Science and Perceptual Psychology. The work presented commenced early in 1967 in the Department of Engineering Physics, Research School of Physical Sciences, The Australian National University (ANU), under the guidance of Dr. S. Kaneff.

Regarding the contents of the Thesis:- Chapter 1 outlines the field of picture processing by computer, and delimits the problems of interest herein; Chapter 2 identifies major areas of difficulty inherent in API; Chapter 3 evaluates current approaches (in the light of these difficulties), and underlines the requirement for a more adequate API model. Chapter 4 sketches a conceptual model for API (which forms a basis for discussion and for future research), and contributes numerous observations and suggestions relevant to this model's further development; Chapter 5 describes attempts at digital simulation of certain basic processes stipulated in the conceptual model, using pollen micrographs and alpha-particle tracks in photographic emulsion as the input images; Chapter 6 contains a discussion of the results of the present study, its relationship to other work, its implications and possible applications, and suggestions for future research; and Chapter 7 lists general conclusions drawn. The Appendices relate to the simulations described in Chapter 5: Appendix 1 describes a simple picture scanner; Appendix 2 outlines an improved technique for producing pictorial representations of data via a high-speed line printer; Appendix 3 explains the operation of a new algorithm for in-situ permutation of array elements; Appendix 4 discusses the implementation of a list-processing system (SLIP) on the ANU's IBM 360/50 digital computer; and Appendix 5 gives listings of a few of the more interesting programs developed in the course of the simulation experiments.
Originality is claimed in the following respects:-

1) To the author's knowledge no other similar study of the API problem as a whole has been reported.

2) Certain areas of difficulty inherent in the API problem have been identified and elucidated; insights into the nature and extent of this problem have been presented.

3) Various shortcomings of current approaches, with respect to complex PI tasks, have been identified.

4) The conceptual model presented as a basis for discussion and for further research, is novel both as a whole and in many of the details.

5) Original observations and suggestions are made with respect to further study.

6) Improved techniques for (i) detecting, encoding, and segmenting image contours, and (ii) identifying "objects" defined in terms of simple relationships between and within such contours, have been developed.

7) Implications and possible applications of the concepts and techniques developed, have been suggested.

The following publications have resulted from this work:-

a) Macleod, I. D. G.: "On finding structure in pictures";

b) Macleod, I. D. G.: "Comments on 'A high-speed algorithm for the computer generation of Fourier transforms' "; Trans. IEEE,


The content of (a) above has been revised and expanded in Chapters 2, 3, 4 and 5. Publications (b) to (f) arose from the simulations reported in Chapter 5. Further papers relating to the work reported in Chapters 2, 4, and 5, are in preparation.

Publications incidental to the present study are:-


ACKNOWLEDGEMENTS

The author wishes to acknowledge and express sincere thanks to his supervisor, Dr S. Kaneff, of the Department of Engineering Physics, The Australian National University, for valuable guidance, discussion, and constructive criticism throughout this project.

Appreciation is expressed to Professor Gordon Newstead, of the Department of Engineering Physics, The Australian National University, for his interest and for the use of departmental facilities.

Financial support provided by the Commonwealth Scholarship Board and The Australian National University, is gratefully acknowledged.
CONTENTS

1 INTRODUCTION

1.1 Picture Processing

1.1.1 Coding

1.1.2 Restoration and Enhancement

1.1.3 Computer Graphics

1.1.4 Automation of Photogrammetry

1.1.5 Analysis

1.2 The Present Investigation

1.2.1 Pollen-Interpretation

1.2.2 Discussion

2 SOME AREAS OF DIFFICULTY INHERENT IN AUTOMATIC PHOTO-INTERPRETATION

2.1 Three-Dimensional Fields

2.2 Segmentation

2.3 Image Complexity and Variability

2.3.1 Quantity of Detail

2.3.2 Organization of Detail

2.3.3 Image and Object Variability

2.3.4 Unfamiliar Objects

2.4 Ambiguity

2.5 Representation of Information
AN EVALUATION OF CURRENT APPROACHES TO AUTOMATIC PHOTO-INTERPRETATION

3.1 Template Matching
   3.1.1 Outline
   3.1.2 Examples
   3.1.3 Evaluation

3.2 Property List
   3.2.1 Outline
   3.2.2 Examples
   3.2.3 Evaluation

3.3 Articular Analysis
   3.3.1 Outline
   3.3.2 Examples
   3.3.3 Evaluation

3.4 Picture Parsing
   3.4.1 Outline
   3.4.2 Examples
   3.4.3 Evaluation

3.5 Discussion

TOWARDS THE FORMULATION OF AN ADEQUATE MODEL FOR AUTOMATIC PHOTO-INTERPRETATION

4.1 Preliminary Discussion
   4.1.1 The Nature of the API Task
   4.1.2 Approach Taken in Development of the Model
4.2 Outline of a Conceptual Model for API

4.2.1 Attention

4.2.2 Recovering Properties, Relationships, and Constraints
  4.2.2.1 Associations
  4.2.2.2 Discriminations
  4.2.2.3 Constraints

4.2.3 Selection from Alternative Organizations

4.2.4 Evaluation of Suggested Organizations

4.2.5 Resolution of Ambiguities

4.2.6 Abstraction and Economy

4.2.7 Factors in the Derivation of 3-D Organization
  4.2.7.1 Simplicity
  4.2.7.2 Experience
  4.2.7.3 Rotational Symmetry
  4.2.7.4 Other Assumptions

4.2.8 The Information Store

4.3 Pollen-Interpretation

5 IMPLEMENTATION EXPERIMENTS

5.1 Computer Simulations

5.2 Discussion

6 DISCUSSION

6.1 Results of the Present Study
SUMMARY

The results of a broad study of the problem of designing automatic photo-interpretation (API) systems capable of analyzing naturally-given images (using analysis of scanning electron micrographs of pollen and nuclear tracks in photographic emulsion as examples), are presented.

Areas of difficulty inherent in this problem which have been identified, include those relating to (i) the three-dimensional (3-D) nature of the fields portrayed, (ii) isolation of simple image components, (iii) the quantity, organization, and variability of image detail, (iv) ambiguities of association and segmentation, and (v) development of appropriate information representations. These areas are interdependent, and it is argued that a broad study of the API problem as a whole should precede detailed examination of particular aspects.

A conceptual model for API, which serves as a basis for discussion and for future research, is outlined. Although incompletely formulated, this model has potential for development so that it can cope with non-trivial PI tasks, and many observations and suggestions relevant to such development are presented.

Attempts to simulate certain basic processes (relating to the recovery of simple image organization) stipulated in the conceptual model are described. Problems related to the 3-D nature of the fields portrayed in pollen micrographs were found to be formidable, and in the simulations greater progress has been made with the nuclear track images (which are essentially 2-D in nature). Techniques for (i) detecting, encoding, and segmenting image contours, and (ii) identifying "objects" formed on the basis of relationships within and between contour segments, have been developed.

The difficulty and complexity of the API problem is such that this study has not been oriented towards the production of immediately-useful API systems, but rather towards an identification and elucidation of key factors and their inter-relationships. Applications of the concepts and techniques developed to problems such as character recognition and picture coding, have, however, been suggested.

A major problem for future research is elaboration, clarification, and refinement of the conceptual model, including (i) development of effective information
representations, (ii) further elucidation of the important types of image and object organization, and of image features which suggest underlying 3-D fields, (iii) integration of image and 3-D detail at various levels of resolution, (iv) clarification of the role of decision thresholds, and (v) development of simplicity measures for evaluating alternative descriptions.

It is concluded that the API problem is solvable, but that because of the nature and extent of the difficulties involved, substantial progress will be achieved only as a result of much further patient study.
Chapter 1
INTRODUCTION

The research reported herein concerns an investigation in the field of automatic photo-interpretation — part of the larger field of automatic picture processing. The subject and aims of this investigation are outlined in Section 1.2, but to put the present study in context, this outline is preceded by a short discussion of the field of picture processing and of the major areas of research within it. Because of the extensive literature, the references cited in this and following chapters are usually typical rather than comprehensive.

1.1 Picture Processing

Mechanical (as opposed to biological) processing of pictorial information will be referred to as "picture processing". Since the advent of the electronic digital computer as a general purpose information processing device, there has been increasing interest in its application to pictorial information (Butler and Butler, 1967; Rosenfeld, 1968a, 1969; Harmon and Knowlton, 1969). Interest has also been shown in the use of optical and other analogue and non-electronic computing techniques (Fischer et al., 1962; Pollock et al., 1963; Tippett et al., 1965; Vander Lugt, 1968) as well as in digital simulations of these (Andrews and Pratt, 1968).

The pictures considered for processing may be thought of as single-valued functions of two spatial co-ordinates, usually satisfying additional constraints, in that the "picture functions" are real-valued, static, bounded, and discrete (Rosenfeld, 1968a). Coloured pictures have only occasionally (Anonymous, 1969) been considered, and there is usually only one dependent variable of interest — gray level (also termed gray-scale value, albedo, density, tonal value, reflectance, etc.), on a gray scale extending from black to white. Pictures are distinguished from arbitrary mathematical functions meeting the constraints listed, by their representational nature, whereby the picture function portrays some possible scene in the real world (Rosenfeld, 1968a).

The extensive research into various aspects of picture processing has produced a voluminous literature, some of which is unfortunately not readily available. Several broad areas of picture processing research have emerged,
but there is a degree of overlap between these because of problems, concepts, and techniques which are common to more than one area. The major areas are summarized below.

1.1.1 Coding

When considered in the context of picture transmission or storage, three aspects of picture coding can be identified:

(i) encoding, i.e., analyzing the input picture and producing a coded representation of it;
(ii) transmission or storage of the encoding;
(iii) decoding, i.e., synthesizing an output picture on the basis of information contained in an encoding.

The coded representation is usually required to satisfy certain criteria such as accuracy of representation (reflected in the fidelity possible in decoded output pictures), simplicity of encoding and decoding, and economy (in terms of information cost). Coding schemes which attempt to minimize information cost while satisfying fidelity criteria, have been the subject of considerable interest, often with regard to reducing the bandwidth required for TV transmission (Cutler, 1967; Pratt, 1967; Rosenfeld, 1968a,b; Pratt et al., 1969). (Picture synthesis and picture analysis, per se, are discussed below in Sections 1.1.3 and 1.1.5 respectively).

1.1.2 Restoration and Enhancement

In the acquisition and transmission of pictures, various types of distortion may occur. A number of techniques for the restoration of variously degraded and distorted pictures has been developed. Among the degradations which can be at least partially compensated are:

(i) spatial frequency distortion resulting from lens aberrations, turbulence, diffraction, poor focus, large sampling window, and other effects (Tsuiuchi, 1963; Barkdoll and McGlamery, 1968);
(ii) geometrical distortion (Nathan, 1968);
(iii) non-linearity of gray-scale (Nathan, 1968);
(iv) noise (Nathan, 1968);
(v) blurring caused by image motion (Schroeder, 1969).

Apart from restoration of degraded pictures, techniques are also available for the enhancement or detection of certain features of a given picture. By means of the technique of matched spatial filtering (Aroyan, 1959), objects or features with known spatial characteristics (e.g., printed characters or contours) can be treated as signals in spatial noise, and so may be enhanced or detected (Hawkins and Munsey, 1963; Vander Lugt et al., 1965; Watrasiewicz, 1967). It is possible to enhance detail of small size or small variation in gray level, and this may be an aid to human interpretation of some photographs (Anonymous, 1967; Oppenheim et al., 1968; Selzer, 1968), as may the detection of changes in successive photographs of the same scene (Cathey and Doidge, 1966; Guignon et al., 1968; Hawkins, 1968).

1.1.3 Computer Graphics

Two objectives of computer graphics research are:-

(i) facilitation of graphical man-machine communication (typically via an on-line cathode-ray tube (CRT) display with some form of light pen), in applications such as computer-aided design (Jacks, 1964; Russo, 1968) and computer-assisted instruction (Suppes, 1966; Bitzer et al., 1967);

(ii) automation of graphical production, typical applications being computer generated movies (Knowlton, 1965), computer typography (Bozman, 1966; Mathews et al., 1967), and computer drafting (Fetter, 1965).

The images dealt with have typically been either symbolic (e.g., flow diagrams) or abstractions (e.g., line drawings of mechanical parts), as well as being line-like and thus binary (i.e., only two gray levels — black and white) rather than half-tone. Development of suitable representations for the class of graphical images being constructed and/or manipulated has been a major concern (Johnson, 1963; Sutherland, 1963; Gray, 1967), as has the development of suitable languages for specifying the range of constructions and manipulations required. The
representation for a particular image is usually constructed more or less directly, by user specification, rather than by analysis of an input picture.

Research on picture synthesis from non-pictorial encodings or representations is related to work in computer graphics. Eden (1968,p.415) suggests that an understanding of how a class of patterns may be generated, can shed light on the converse task of recognition or analysis of that class, and gives a generative model for handwriting. Workers at the National Bureau of Standards have produced a similar model for Chinese characters and for chemical structure diagrams (Anonymous, 1965; Anonymous, 1968), as have Narasimhan and Reddy (1967) for alphabetic characters. There have also been experiments on synthesizing pictures of simple solid objects under given conditions of illumination (Sutro and McCulloch, 1968; Warnock, 1968).

1.1.4 Automation of Photogrammetry

There have been a number of attempts to automate various aspects of photogrammetry (i.e., the science of obtaining reliable measurements by means of photographs). Notable success has been achieved in the derivation of contour data from stereo pairs of aerial photographs (Lawrence, 1967) and in the production of ortho-photos, i.e., "flattened" photographs in which geometric distortions resulting from relief, camera attitude etc., have been corrected (Bertram, 1967). Colwell (1965, p.212) has referred to photo-interpretation as "the act of examining photographic images for the purpose of identifying objects and judging their significance". Some of the attempts to automate the identification phase of various photo-interpretation problems are referred to in Section 1.1.5; only limited success has been achieved, however — see Chapter 3.

1.1.5 Analysis

An important aim of picture analysis research (i.e., the area of pattern recognition research (Nagy, 1968) concerned with pictorial images) is to develop techniques for describing a given input picture in terms of the objects represented within it. In so far as a representation of the input is required,
there are some parallels between this problem and that of encoding pictures for storage or transmission. However, the ultimate concern in picture analysis research is usually a description which emphasizes the significant aspects of the input picture, and not an output picture synthesized from this description. Thus, the complexity and types of descriptors employed in each case will differ. For some applications, such as the identification of printed characters, it may be sufficient to describe the picture (or its component sub-parts) simply as one of a chosen set of categories. This categorization process has been referred to as "pattern classification" (Rosen, 1967; Ho and Agrawala, 1968) and much of the research designated as pattern recognition falls within this framework.

There are two broad classes of input pictures which we will term "symbolic" and "natural" respectively. Symbols are considered here as purely pictorial objects which, by convention, represent some other (usually non-pictorial) object or concept. Typical examples are electronic circuit symbols such as "\(\text{IA}/\), and alphanumeric characters such as "A". Symbolic images are typically constructed more or less directly (by means of printing, drawing, light-pen, etc.) out of line-like segments which contrast strongly with the background such that the image is intended to be binary valued rather than half-tone. Most pattern recognition research on this type of image has been concerned with either printed or handwritten alphanumeric characters (Lindgren, 1965; Minneman, 1966; Ho and Agrawala, 1968; Holt, 1968; O'Callaghan, 1968), but work on other symbols, such as Chinese characters (Casey and Nagy, 1966), mathematical operators (Anderson, 1968), and geometric shapes (Evans, 1968a,b; Guzman, 1967) (if these are admitted as symbolic), has also been reported.

Apart from exceptions such as caricatures and fingerprints, pictures whose content is not symbolic have usually been generated indirectly, typically by the exposure of photographic film. Aerial photographs and photomicrographs are examples of this class of "naturally given" pictures, and the term "images" will (if not otherwise qualified) be understood hereafter to refer to such pictures, this term being preferred to "photographs" (i.e., photographic images) because the pictures in this class need not be formed by the exposure of photographic film.
Pattern recognition research concerning such images has been referred to as "pictorial pattern recognition" (Cheng et al., 1968) or "automatic photo-interpretation" (Holmes, 1966).

The objects in terms of which a given image is to be described are usually real-world (i.e., with an extra-pictorial existence). Thus, the variations in gray scale within the image are a (projective) representation of some object, but are not the object itself. In general, images tend to be more complex, and their successful description more dependent on full gray scale information, than is the case for symbolic pictures.

Among the pictorial pattern recognition problems which have previously been investigated are target detection in aerial photographs (Sebestyen, 1963; Smillie, 1966), analyses of biomedical pictures (Neurath et al., 1966; Ledley, 1969), analyses of spark-and bubble-chamber photographs (Gelernter, 1965; Alder et al., 1966; Brown, 1969; Shaw, 1968), interpretation of photographs and TV pictures of scenes containing simple three-dimensional (3-D) geometrical objects (Roberts, 1965; Forsen, 1968), cloud-photo interpretation (Rosenfeld et al., 1965; Joseph and Viglione, 1966; Darling and Joseph, 1968), terrain-type discrimination (Goldstein and Rosenfeld, 1964), recognition of human faces (Taylor, 1967; Sakai et al., 1969), fingerprint description and recognition (Yefsky, 1967, Ch. 5; Anonymous, 1968; Hankley and Tou, 1968), star-pattern recognition (Sowers, 1963), and crack detection in X-ray photographs of castings (Shirai, 1969).

1.2 The Present Investigation

The overall aim of this investigation is to make a broad contribution to the field of automatic photo-interpretation. The images studied in this field need not be photographic and the term "automatic image-interpretation" might better be employed, but in deference to accepted terminology, the term "automatic photo-interpretation" (API) will be used herein. A brief examination of this field is given in Section 1.1.5, in so far as API may be regarded as that area of pattern recognition or picture analysis research which concerns photographs and other pictorial images. Among the possible current and future motives for API research
There may be too many images and/or too much detail for economical human interpretation. "Huge masses of data are being continuously collected in science, industry, and government. Several hundred pictures of cloud patterns are transmitted daily to earth by weather satellites; thousands of frames of aerial-photography are collected every day by reconnaissance aircraft, hundreds of thousands of bubble chamber and spark chamber pictures are collected annually by high energy physicists; millions of photomicrographs of chromosomes, neurons, blood cells, "pap smears", etc., are taken by biologists every year; and millions of medical and dental x-rays are taken each year — just to name a few of the sources of this great mass of pictorial information" (Cheng et al., 1968, preface p. v).

Effective interpretation may depend upon accurate measurements of image features (such as track curvatures in bubble-chamber photographs or cross-sectional areas of chromosomes) which are more readily measured mechanically than manually. It is also possible that the subtlety or quantity of detail distinguishing certain objects could prevent human beings from effecting reliable discriminations (without mechanical assistance) and that, looking far into the future, an API machine with high measurement accuracy and sophisticated interpretation and classification algorithms, might be capable of better and more detailed interpretations than we are. The likelihood of errors resulting from the inattention or carelessness of a human observer may also favour API systems.

The source image may be formed in an environment which is hazardous or inconvenient for human beings, and there may be limitations on the data transmission rate and/or time which preclude remote human interpretation. An example of this situation occurs in unmanned exploration of extra-terrestrial environments (Asendorf, 1968).
4) Decisions made on the basis of visual data may be required very quickly, and there may be insufficient time for manual transcription or abstraction of the relevant data. (Consider a table-tennis machine (with a servo-controlled bat) playing a human opponent. On the basis of visual data available from a stereoscopic pair of dynamic images, the machine could measure the velocity, direction, and spin of its human opponent's return; by solving the relevant dynamical and fluid-dynamical equations, it could then predict the ball's subsequent path, and produce a counter stroke with unerring and frustrating accuracy!)

5) Depending to some extent on the approach taken, the development of a successful API machine could have implications with regard to biological visual perception, and might provide a useful tool for experimental perceptual psychologists. Conversely, in developing and evaluating theories of biological visual perception, the ultimate test of any theory is implementation as an API machine.

6) In the future, it is conceivable that a radically different visual environment, in which human perception is inadequate, might be encountered. A slight indication of this possibility is the stark contrast between objects' surfaces lit by sunlight, and those in shadow on the moon. With an environment in which the significant attributes and relationships differ from those on earth, the evolutionary development of the human visual system might be a hindrance and prevent the human observer from learning how to perceive effectively, but an API machine might more readily be modified to respond to the significant attributes and relationships in the new environment.

The problem of specifying general API systems\textsuperscript{1} has proved to be very

\textsuperscript{1}A general API system is thought of as one which approaches, in the range and complexity of image interpretation problems which can be handled, human perceptual ability. The "API problem" is regarded here as the problem of formulating such a system.
complex and difficult. The few at least partially successful API systems reported, have concerned restricted and relatively simple classes of images — see Chapter 3.

One approach to contributing towards eventual solution of the API problem, is to study tasks which are sufficiently restricted for a detailed practical solution to be achieved with a reasonable amount of time and effort, perhaps by studying artificial or synthesized images so that some of the difficulties incurred with real-world images are avoided or mitigated. With this approach, it is hoped that in addition to devising a solution for the restricted task, the concepts and techniques evolved will generalize to more complex tasks and thus contribute towards the general problem.

It is tempting to adopt the above approach because even if the contribution to the general problem is eventually shown to be minor, positive results will have been achieved with respect to the restricted problem. An alternative approach, which has not found much favour hitherto, is to make a broad study of the API problem and to attempt to contribute towards the formulation of a general model for API, in the hope that what contributions are made will at least be suggestive of future research. Despite the inherent difficulties and the small likelihood of achieving substantial short-term theoretical or practical results, this latter approach is the one followed in the present investigation, for reasons which are discussed further in Section 1.2.2. More specifically, a particular API task (that of pollen-interpretation — see Section 1.2.1) is used as an example, but this task is somewhat less restricted and more complex than most of those previously studied. The aim has not been to detail a solution to this particular task, but rather to use it to identify difficulties which will be confronted in attempts to solve the API problem, thereby enabling an evaluation of previous API research in the light of these difficulties, and facilitating some preliminary steps toward the specification of a general model for API.

1.2.1 Pollen-Interpretation

The photo-interpretation (PI) task used as an example in the present investigation, is the image interpretation phase of pollen analysis. A brief account of this task is given below, and the particular aspects of interest in the
present investigation are discussed in Section 1.2.2.

Pollen grains are the small cells which carry the male genetic material of plants — there are several hundred thousand varieties of pollen, with the individual grains ranging in size from a few microns up to a few hundred microns (Echlin, 1968). Fig. 1-1 shows four scanning electron micrographs of pollen grains; obvious features of the grains represented are their three-dimensionality and their clear but intricate surface structure.

Palynology (i.e., pollen analysis) is of interest in a number of fields. As examples, Flenley (1968) gives: allergy studies; plant pathology (fungal spores being counted); oil exploration (using analyses of fossilized pollen as a means of stratigraphic correlation); and biogeography (using analysis of fossilized pollen as an indication of former vegetation and climate (Faegri and Iversen, 1964)). The last two applications rely upon the remarkable durability of the outer wall of pollen grains; samples taken from the depths of ancient bogs have contained clearly recognizable pollen grains in spite of their having been buried for hundreds of thousands of years (Echlin, 1968). Because of the complexity and quantity of detail associated with individual grains, the large number of possible species, and the small differences between some species, the PI phase of palynology can be very difficult. In a study carried out by Prof. D. Walker, of the Australian National University's (ANU's) Dept. of Biogeography and Geomorphology, a microscope slide was prepared from reference stocks of four species of pollen grains found in the New Guinea highlands, there being a total of ca. 100 grains on the slide. Four observers (A, B, C, and D) with experience in palynology ranging from one year (A) to twenty years (D), viewed the slide with an optical microscope and attempted to ascertain the pollen species present, and the number of grains of each species. The results ranged from two out of four species correct in 1 hour 40 minutes for observer A, to all four species correct in 6 hours for observer D. The time taken by experienced human analysts is an indication of the difficulty of this task, the main difficulty being in selecting and testing possible species (checked by visual comparison with a slide of reference grains) rather than in perceiving the 3-D organization of the grains. Two problems

\footnote{Counting of grains once their species has been identified is relatively trivial.}
FIGURE 1-1 Scanning Electron Micrographs of Pollen Fields

Figs. 1-1c,d copyright Patrick Echlin and Cambridge Scientific Instruments; reproduced by kind permission.
which arise in connection with the use of optical microscopes in the analysis, are
that the pollen grains are partially transparent to light, and that the depth of field at
the high magnifications required may be only a small fraction of the pollen grain's
diameter (it being necessary to refocus several times to comprehend the detailed
organization of the grain, although some idea of the 3-D disposition of various
details may be gained in this manner). The optical micrograph of Morning Glory
pollen shown in Fig. 1-2a may be compared with the corresponding scanning
electron micrograph shown in Fig. 1-2b, to gauge the greatly increased depth
of field in the latter, and also the improved surface definition resulting from the
pollen grain's opaqueness to low energy electron beams. Echlin (1968) and
Hayes and Pease (1969) describe the operation of the scanning electron microscope
(SEM); in brief, one electron beam scans the specimen (coated with a thin layer
of evaporated gold to render it electrically conductive and to prevent charge
accumulation resulting from secondary electrons) which is mounted in an
evacuated chamber, and a second beam scans the face of a CRT in synchronism
with the first and with an intensity proportional to the number of secondary and
reflected primary electrons collected by a detector in the evacuated chamber, so
that a magnified image of the specimen is formed on the CRT face. Echlin (1968,
p. 81) says that "with the recent development of the SEM, we have been able to
examine pollen grains in unprecedented detail ... such an instrument is of immense
value in the study of specimens as complex as the surface of pollen grains".

Among the more obvious characteristics and restrictions of the pollen
grains, the 3-D fields of pollen, and the scanning electron micrographs of these
fields, are:-

1) The pollen grains are usually of a fairly simple overall 3-D shape
and are relatively robust (with no flimsy appendages), but their
surfaces may be very complex and finely structured, taking on a
textured appearance as the structure gets finer. In contrast to
objects such as houses, which are characterized partly by the mater-
ials of construction, pollen grains are of apparently homogeneous
construction (as far as can be determined from the image) and are
characterized principally by the shape of their surfaces.
FIGURE 1-2 Comparison of Optical and Scanning Electron Micrographs

Copyright Patrick Echlin and Cambridge Scientific Instruments; reproduced by kind permission.
2) The fields are not self-luminous and are evenly illuminated by the scanning electron beam. They consist typically of a distribution of several pollen grains and other vegetable or mineral material (particularly in the case of "fossil" rather than "new" pollen), resting on a support plane, which may or may not be visible. All surfaces are opaque to the electron beam and have reasonably constant electron reflection and emission properties (cf. pigmentation). Variations in the number of electrons collected are related primarily to the topography of the surface being scanned, with topographically high surfaces appearing lightest (Kelly et al., 1969). In more general 3-D fields, some objects may be self-luminous, the strength and direction of the illumination may change (such as in shadows), objects may be transparent to some degree, and surfaces may have differential pigmentation and microrelief (see Section 2.1).

3) The images are functions of one variable (i.e., the number of detected electrons) and so can be represented as gray scale rather than, in the more general case, as full colour.

For brevity, the image interpretation phase of pollen analysis will be referred to as "pollen-interpretation".

1.2.2 Discussion

Mechanization of pollen-interpretation is desirable because of the difficulty and laboriousness of this task, and the fact that skilled analysts are required. The same features which make mechanization desirable also indicate the improbability of achieving an adequate mechanization given the current state-of-the-art in API; none of the successful API applications yet reported even approaches the difficulty of pollen-interpretation, and complete mechanization of this task does not appear to be a reasonable short-term goal\(^1\). An outline of the objectives of the present

\(^1\)A more reasonable objective would be to provide mechanical assistance for the human analyst, via a suitably organized information retrieval system containing both pictorial and textual data; provision of such a system is currently being investigated at the ANU.
investigation, and of the aspects of pollen-interpretation which are of particular interest, is given below.

The pollen images are rather less restricted than the images studied in most previous API investigations. Pollen-interpretation can therefore be regarded as a reasonably "general" PI task, in that an API system capable of effective pollen interpretation would, no doubt, be capable of coping with a fairly wide range of other PI tasks. There are several ways in which study of general rather than restricted PI tasks could contribute to the field of API:

1) The problems involved in general PI tasks are not clearly isolated, and are not independent. The development of concepts and techniques for coping with these problems may well be interdependent, in which case consideration of reasonably general PI tasks is required.

2) API research has been characterized by an apparent divergence of approaches (see Chapter 3), depending to some extent on the characteristics of the (restricted) PI task studied. Consideration of a general PI task (of which the other restricted tasks studied form subsets) could possibly uncover relationships between superficially different current approaches, and also indicate (and give some unity to) future research.

3) Whatever progress is achieved will probably be relevant to an eventual solution of the API problem, whereas with a restricted PI task, the concepts and techniques evolved may not generalize to more complex tasks (Narasimhan, 1968).

Another aspect of interest in pollen images is that the objects represented within these images are clearly 3-D, and an essential part of the interpretation of these images concerns this underlying 3-D field. With few exceptions (e.g., Roberts, 1965), previous API studies have neglected this field and have treated the objects of interest as purely pictorial (i.e., 2-D). One of the objectives of this investigation is to account for the role that this 3-D field plays in the interpretation process, and to identify image features which suggest possible underlying 3-D organizations.

The pollen images exhibit a clear but delicate and rich organization
(usually manifested more clearly in the perceived 3-D field rather than pictorially). Human perceivers assign similar organizations to given pollen images, even if they are not familiar with the pollen grains represented and even if they do not know that the images are of pollen fields. It seems natural to regard the processes of assigning such organization as "interpretation", even if no objects are identified. There are some subtle but significant differences between the view of interpretation as the set of processes whereby the raw sensory data is organized to a greater or lesser degree, and the view that (photo) interpretation is the identification of known objects in photographs (Colwell, 1965; Holmes, 1966). The latter view emphasizes the results rather than the processes; a proposition entertained herein is that the way in which unfamiliar inputs are organized is in some sense fundamental to the later recognition of known objects, with identification or classification being of only secondary importance. Logicians speak of interpretation as being the process of using a syntactical system to denote or refer to objects outside the system (Carnap, 1958); in this view, the derivation of a perceived 3-D field from a 2-D image might also be regarded as interpretation. Kirsch (1964, p. 376) says that it is tempting to identify interpretation operations with the informal notion of "understanding".

The principal objective of this investigation is to examine the possibility of automatically interpreting a given image (i.e., perceiving the "simplest" 2-D and/or 3-D organization of the image in terms of the knowledge available), particularly concerning those processes which precede the utilization of specific knowledge regarding complete objects. Fulfilment of this objective depends upon an identification of the important forms of organization, and of the relationships and constraints underlying these. The author is not aware of any similar study concerning natural images having been performed, but some preliminary investigations with a restricted class of images have been reported (O'Callaghan and Maxwell, 1969).

Because of the preliminary nature of this investigation (which can be regarded as a prelude to further research) and the difficulty of the problem being studied, concepts rather than techniques, are the main concern. Development of implementation techniques must await further detailing and refinement of
Some restrictions inherent in the pollen-interpretation task have been mentioned in the previous section. Actually, with regard to the inherently ambiguous relationship between a given 2-D image and possible underlying 3-D fields (see Section 2.1), veridical perceptions (i.e., perceptions which are in accord with reality) are possible only to the extent that general assumptions made about the underlying fields hold true, or if they do not hold, lead to inconsistencies which are later satisfactorily resolved. Examples of plausible general assumptions are given in Section 4.2. If these assumptions (or restrictions) do not hold, an apparently satisfactory but nevertheless inaccurate perception may be derived, as in the case of the "Ames' room" (Munn, 1961). Hake (1957) has argued that the requirement for a satisfactory perception is that it should be consistent or coherent rather than faithful.

In view of the concern for organization of the underlying 3-D field, and of the successes achieved with regard to obtaining depth information from stereo pairs of images (see Section 1.1.4), it might seem desirable to consider pollen-interpretation in the context of a stereo pair, it being relatively easy to obtain such pairs from the SEM by tilting the specimen stage. If a pollen-interpretation machine were ever constructed, the use of stereo pairs (or possibly dynamic images) would no doubt be of great value, but the attitude taken in the present investigation is that binocular parallax is only one clue to depth information and, even though it may be a useful clue, it is in many cases not mandatory (witness the facility with which we can perceive depth in Fig. 1-1). The use of binocular parallax as a depth cue has been extensively studied with regard to stereo-compilation (Lawrence, 1967) and for this reason, and because it facilitates the recovery of depth information only for detailed surfaces which are visible in both images, its use is not considered in detail herein.

Pollen micrographs have been found to be a useful image class for exposing problems and aiding the formulation of a conceptual model for API (see Chapter 4). However, the problems to be accounted for in interpreting these images effectively, are such that a much more restricted class of images, viz., low-energy nuclear particle tracks in photographic emulsion, has been studied in attempts to simulate
some components of the conceptual API model. The restrictions of this second class are such that it is not as useful for evaluating and developing approaches for general-purpose API, as the pollen micrographs. Consideration of the nuclear track images is therefore postponed until discussion of the simulation experiments — see Chapter 5. In the following discussion, unless otherwise noted, references to an API machine imply a machine organization paralleling that of a conventional general-purpose digital computer.
Chapter 2

SOME AREAS OF DIFFICULTY INHERENT IN AUTOMATIC PHOTO-INTERPRETATION

Some of the principal difficulties inherent in the API problem (rather than in specific approaches to it), are discussed below. Many of these problems (particularly those related to the 3-D nature of the object-world) have not been identified or acknowledged in previous API investigations, and as such have not been directly addressed. Some of the problems identified herein have not been apparent in these earlier investigations, because of the restricted nature of the images and PI tasks studied. A further motivation for this identification is that it provides criteria for evaluating current approaches to API (see Chapter 3) and for developing a new approach (see Chapter 4).

The following identification of potential difficulties has been carried out primarily with regard to the pollen-interpretation task. Narasimhan (1966, p. 167) remarks that in a certain basic sense he was fortunate that his introduction to picture interpretation was via bubble-chamber pictures and not alphanumeric characters, because with the former, the concept of "categorization of images as belonging to one or another of a finite set of prototypes" became virtually meaningless. In a similar manner, the choice of pollen-interpretation as an example has helped clarify some of the shortcomings of currently espoused approaches.

In this chapter it is not intended to suggest approaches to solving the various problems identified, but in the case of Section 2.1 (3-D fields) the major problems inherent in the approach of trying to interpret images in purely pictorial terms (without explicitly accounting for an underlying 3-D field), are examined first, and then some of the difficulties arising with respect to approaches which attempt to account for this field are discussed.

In the following discussion, a degree of interdependence and overlap between the various sections, reflecting the interdependent nature of the problems discussed, will be apparent.

2.1 Three-Dimensional Fields

In the pollen-interpretation task, the scene represented in the input image consists typically of a number of pollen grains (possibly of different species),
together with mineral and other vegetable matter, distributed on a support plane which may be partially or totally occluded by the objects resting on it. The objects and arrangements represented in the input image are thus 3-D and not pictorial; they are represented pictorially by a projective transformation such that there is an area in the input image corresponding to each 3-D object. One common approach has been to assume that 3-D objects are effectively characterized by their projections, and to try to interpret images in purely pictorial terms. Some of the difficulties inherent in this approach are:

1) There is a very large number of different projections of a typical 3-D object onto a 2-D image plane; there will thus be many 2-D gray level distributions corresponding to the same object in different 3-D locations (changing 2-D location and size) and orientations (changing 2-D shape). These variations are lessened if the objects are symmetrical to some extent (e.g., a uniformly pigmented sphere looks the same from all aspects), usually rest on the background in a particular manner (e.g., the vertical orientation of aircraft on an aerodrome is nearly always the same), or are at fairly constant distances from the image plane so that size variations are minimized (e.g., objects on the ground relative to the image plane in aerial photography). Even for fairly simple objects such as the example from Guzman (1967) shown in Fig. 2-1, the topology as well as the shape of the 2-D projection can vary quite drastically as the object assumes different 3-D orientations. In the case of pollen grains, with their intricate structure, this problem is aggravated.

2) Any given object may be wholly or partially occluded by other objects, and will always be partially occluded by itself; an object's overall 3-D shape is, therefore, never available from its image alone. Similarly, the projected 2-D outline shape of an object will often be unavailable because its silhouette will be wholly or partially occluded. Another problem arising from self-occlusion is that of degeneracy, where less than the normal number of faces appear in
FIGURE 2-1
Change in Shape and Topology of a Simple 3-D Object's Projection, with Changing Viewpoint

FIGURE 2-2
Degenerate Projections of 3-D Objects

FIGURE 2-3  Mapping of a 3-D Surface onto a 2-D Image
objects' 2-D projections, as is illustrated in Fig. 2-2.

3) The intensity of the light projected onto the image plane is a function of the aspect (relative to the incident illumination and the image plane), of the surface from which it was reflected, of the magnitude and distribution of the incident illumination, and of the pigmentation and microrelief\(^1\) of this surface. 3-D shape, rather than the surface pigmentation or the distribution of the incident light, is usually what characterizes the objects of interest; as a result, their 2-D projections may not be good characterizers. With scanning electron micrographs, the flux of secondary and reflected electrons is largest when the surface being scanned is topographically high (Kelly et al., 1969), and this type of variation in an object's 2-D projection is not as extreme as in normally pigmented and illuminated 3-D fields.

It is argued in Chapter 4 that the above three areas of difficulty (which arise with attempts to interpret images purely pictorially, i.e., assuming that objects are characterized by their gray level distributions) constitute sufficient reason for attempting to perform interpretations in 3-D terms. This approach immediately raises the problem of deriving 3-D information from a single static input image.

The mean gray level of a resolution element (assumed to be a small square) in the input image depends on the light being reflected towards the focal point of the image-forming system from a square-prismatic volume in the 3-D field — see Fig. 2-3. The surface from which the light is reflected can be at any depth in this field, and the gray level of the image is a function of the pigmentation, microrelief and aspect of this surface, and of the distribution of the incident illumination.

\(^1\) Microrelief (or grain) is considered as object surface detail which is too fine to be resolved in the image, but which nevertheless causes observable effects in the way that it determines the relationship between the incident and reflected light distributions. The nap on velvet is an example of microrelief, whereas with a mirror, microrelief is effectively absent. Microrelief usually causes a Gaussian-type "spread" of the reflected illumination relative to the incident illumination.
If there are no restrictions on the nature of the 3-D field, extraction of veridical relative (let alone absolute) depth information is clearly not possible on the basis of the image information alone. Even with the restrictions inherent in the pollen interpretation task, mechanical extraction of relative depth information will be very complex and difficult, despite the apparent ease with which human beings derive consistent interpretations of depth for these images. Note the asymmetry, in that deriving a unique 2-D projection, given a specification of a 3-D field and illumination sources, is conceptually simple. However, because of secondary illumination, derivation of such an image may be computationally prohibitive (unless the image is formed via a lens and an actual 3-D field, of course).

Other difficulties likely to arise in attempts to extract 3-D information from a given 2-D image are:-

1) Because of the problem of occlusion mentioned above, only part of each object is directly represented in the image.

2) Because of deep shadows, glare, or chance compensating changes in any of the 3-D parameters determining the gray level at a point in the image, there is sometimes no pictorial evidence of actual boundaries in the 3-D field. Examples of this are the cases of a "black cat in a coal bin", a "white cat in a snowstorm", and the absence of pictorial indications of changes in depth in Fig. 2-4. (For a rather different reason, occlusion also leads to a lack of pictorial representation of 3-D boundaries).

2.2 **Segmentation**

The input image will usually have a number of structured 3-D objects represented within it. The problem arises of segmenting or isolating these objects from each other, and the component parts of individual objects from other parts of the same object. Segmentation is required in both the perceived 3-D field and the 2-D input image; there will clearly be interactions between these two domains.

The need for object isolation could be queried — it is conceivable that the whole image might be processed at once "in parallel". There are compelling
FIGURE 2-4 Non-Correspondence of Pictorial and 3-D Boundaries

Fig. 2-4c copyright Patrick Echlin and Cambridge Scientific Instruments; reproduced by kind permission.
arguments, based on combinatorial unmanageability, against this approach in favour of attempting to interpret the input image sequentially, by segmenting it into regions (corresponding to underlying 3-D objects) and attending to each of these in turn with the full power of the interpretation machinery (Minsky, 1961, pp. 16-17). Another objection to the parallel approach (with the class of images being considered) is that the 3-D objects of interest are characterized by the way in which their parts are assembled; in a basic sense these parts are therefore prior to the whole and have to be at least partially isolated first.

Neisser (1967, p. 88) considers that "visual objects are identified only after they have been segmented, one from the other", but in complex images simple "once-off" initial segmentation is only rarely possible. Some of the difficulties likely to be encountered are:

1) Boundaries of objects may not be directly represented in the image (because of confusing surroundings, shadow, glare, etc.). Two examples of this absence of a pictorial boundary, where there is nevertheless a clearly perceived change in depth in the 3-D field, are shown in Fig. 2-4.

2) There may be several types of pictorial boundary forming the outline of an object of interest. For example, in Fig. 2-5 some of the white bars are bounded partly by a sharp black-white contour, partly by a virtual or structural boundary (as shown in Fig. 2-6\(^1\)), and partly by junctions between areas of different gray level gradients (but with the same gray level at the junction).

3) The internal properties which define objects, and enable a segmentation of objects from their surroundings, can be quite complex and subtle. (and may not be known before processing commences, see Section 2.3.4). In Fig. 2-5a for example, the exterior boundary of the lacework region (or boundary at the edge of the furrow) is defined by a discontinuation of the structured region and not simply by the boundary of any of the white bars. In

\(^1\)In this and subsequent diagrams, a dotted circle is used to indicate the extent of the image area which is attended to in fine detail.
(a) Micrograph of Thrift. Copyright Patrick Echlin and Cambridge
Scientific Instruments; reproduced by kind permission.

(b) Boundary formed by
change in gray level.

(c) Boundary formed by change
in gray level gradient.

FIGURE 2-5  Gray Scale Boundaries in Thrift Micrograph

..."virtual" boundaries

FIGURE 2-6  Example of "Virtual" Boundaries Perceived on the Basis
of Constraints Between Gray Scale Contours
this example, the importance of considering the underlying 3-D field is indicated.

4) Objects may be differentially pigmented and/or illuminated, so that there could be boundaries within their 2-D projections which do not correspond to the object's 3-D shape.

It is clear that effective preliminary segmentation will be possible only in restricted API tasks. Nagy (1968, p. 850) observes that "it seems that the isolation problem can only be solved by including it in a loop with the recognition process by trying different partitions of the overall pattern until the individual components 'make sense'. The heuristic techniques involved here are now becoming known under the name of scene analysis".

Satisfactory object isolation or segmentation is sometimes achieved by human perceivers only as a result of identifying known objects in their context, rather than by any properties and relationships existing within the given image. Examples of this situation are connected handwriting and puzzle pictures (e.g., "find the hidden policeman by the side of the road" (Munn, 1961, p. 569)). In many cases, however, the fact that the objects represented are unfamiliar, does not preclude effective segmentation by a human perceiver. It is not anticipated that the reader will find much difficulty in discriminating the pollen grains represented in Fig. 2-7 from their surroundings, although he might not be familiar with these grains.

2.3 Image Complexity and Variability

A major difficulty in API is the possible complexity and variability of the underlying 3-D fields and their images. This difficulty is exemplified by the pollen-interpretation task. Fields of pollen and other accompanying material manifest several forms of complexity:— the quantity and organization of detail in any individual grain; the large number of pollen species; the variability of images of the same 3-D field and of pollen grains of a given species; the "open" nature of properties required for the appropriate description of given images; and the likelihood of unfamiliar pollen grains or other material being represented in the images. These and other aspects of image complexity are examined below.
(a) Micrograph of fossilized Lithocarpus and extraneous material

(b) Micrograph of fossilized Pasania and extraneous material

FIGURE 2-7 Discrimination of Unfamiliar Objects from Confusing Surroundings
2.3.1 Quantity of Detail

Problems related to this aspect of complexity arise with respect to the images being interpreted, and to the quantity of stored information necessary for their satisfactory interpretation.

To represent fine detail (i.e., detail of small size or gray scale excursion) adequately, the image must be of high resolution in both distance and gray scale. For a high resolution aerial photograph, it has been estimated (Harley et al., 1968, p. 18) that there are up to $10^8$ resolution elements. If these elements are assumed to be independent and 64 gray levels are resolvable, then the information content of such a photograph would approach $6 \times 10^8$ bits, a very large sum in the light of current digital processors. The SEM images of pollen fields are not coloured, but with fields in which colour is an important characteristic, difficulties arising from the quantity of image detail may be aggravated. There is usually much more detail in a typical image than is required for its satisfactory interpretation, but it is not known 'a priori' which detail will be important and which will not. In discriminating between two objects of similar classes, or in evaluating a hypothesized interpretation, fine detail often assumes an importance which is disproportionate to its obviousness or size.

With regard to the quantity of information which must be available to the API machine if it is to be able to make satisfactory interpretations, it is clear that a characterization of the object classes of interest is required. As already mentioned, there are several hundred thousand distinct (but sometimes very similar) pollen species, and there may be several tens of thousands in any chosen geographical region. As a result of this, a vast quantity of information is required for the characterization of likely species in even a single region, and a simple serial comparison of objects found in the image with the classes characterized is almost certain to be impracticable.

---

1 This is usually not a valid assumption; if it were, the image would resemble pure pictorial noise (Moles, 1966). There has been extensive research into techniques using dependencies between resolution elements as a basis for economical encoding of pictures — see Section 1.1.1.
2.3.2 Organization of Detail

In many image classes, particularly pollen images, the organization of detail is a very important characteristic. (For example, all the objects in "\[\triangle\]", "\[\downarrow\]", "\[\leftarrow\]", and "\[\rightarrow\]" are composed of the same parts, but their organizations differ.) The nature of this organization is often rather intricate and subtle, and the problem arises of recovering the organization of the 2-D images and also the perceived 3-D fields. The existence of constraints within an image decreases its objective (information-theoretical type) information content, but increases the subjective information content (meaningfulness), provided that the organization is not too redundant or too well defined (e.g., as it is in a checker-board (Moles, 1966)); compare the information that we derive from each of the two images shown in Fig. 2-8.

Relationships between and constraints within distinguishable components underlie all organizations, and there are several different types of organization corresponding to different types of relationships and components. A simple and basic type of organization is that resulting from associations or constraints between various parts of the boundary of some region, as shown in Fig. 2-9. A related form of constraint exists within, for example, a straight or gently curved contour (even though there are no distinguishable parts in this case), or within an undulating curve such as "\[\curvearrowleft\]".

A second important form of organization (which may be thought of as "structure") is the composition of an object as an arrangement of parts, which may themselves be composed of simpler parts and so on through several levels or depths of structuring down to the level of the simplest distinguishable parts (which cannot be seen as a composition of even simpler parts). Examples of this hierarchic organization are (i) the composition of strokes making letters, letters making words, words making lines, and so on; (ii) the configuration of individual bubbles making up tracks, and sets of several coterminous tracks (i.e., tracks with coincident ends) making up "events" in a bubble-chamber (Brown, 1969); (iii) the individual arms of a chromosome organized to give the characteristic chromosome shape; and (iv) hierarchies of textural elements such as in an aerial photograph of a forest, where leaves, branches, and trees all form elements of texture.
(a) Unstructured image (high objective information content)\n
(b) Structured micrograph of Mallow (lower objective information content)

FIGURE 2-8 Subjective Versus Objective Image Information

Fig. 2-8b copyright Patrick Echlin and Cambridge Scientific Instruments; reproduced by kind permission.

Note the tendency to group the middle boundaries and to perceive the shaded areas as holes.

FIGURE 2-9 Influence of Boundary Constraints on Perceived Organization
A third form of organization (which is in some respects similar to the second form but nevertheless distinct from it) is that where the total configuration can be perceived as a "superimposition" of two or more components of unequal status. Examples of this type of organization are: (i) a shadow on a surface; (ii) a bump in a straight line and other types of "figure on ground" such as the spines in Fig. 2-10 being seen as superimposed on a spherical base; (iii) the coexistence of different 2-D structures shown in Fig. 2-11; and (iv) the summation of components at different levels of resolution \(^1\) to yield an overall detailed organization (note the superimposition of both fine and coarse detail in the background of Fig. 2-4c). It is possible to view this third form of organization as "vertical" structuring compared to the "horizontal" structuring of the second form.

A major problem in recovering structural information is the vast number of interrelationships between each component and every other component. Just what relationships should be recovered? It is clear that only a small proportion of the possible set of relationships is of direct importance; specification of which relationships are likely to be important is not, however, any easy task.

2.3.3 Image and Object Variability

There are several forms of variability in both images and objects, which may have to be accounted for in interpreting given images. Considered below are: (i) variations between separate images of the same 3-D field; (ii) variations between 3-D fields containing objects of the same class; and (iii) the wide range of properties required for satisfactory image and object descriptions, as a result of these variations.

An obvious cause of variations in images of the same 3-D field is changes in the location of the image-forming device (i.e., viewpoint) relative to the 3-D field; the variations resulting from such changes have already been discussed in Section 2.1. Examples of other variations and imperfections which might

\(^1\)With perhaps relatively independent organizations within each resolution level, in so far as coarse-resolution boundaries may not be visible in a local fine-resolution area, and vice versa.
FIGURE 2-10  Superimposition of 3-D Components

FIGURE 2-11  Overlapping 2-D Components
arise during the image formation process, during encoding of the image, and possibly during transmission of this encoding are:

1) Overall changes in gray level and contrast, and non-linearities of gray scale can arise through, for example, differences in photographic film, exposure, or development.

2) There may be various imperfections in the image-forming system used; there may be a restricted depth of field, the grain size of a photographic emulsion may be apparent, image motion or diffraction effects may lead to blurring, and there may be several types of geometrical distortion (e.g., "pin-cushion") present.

3) In converting the image to the initial representation used in digital processing, or in transmitting such an encoding, several effects may be noticeable. A digital representation of the image is not continuous, but is quantized in both space and gray scale. To reduce the storage requirement of this representation, it is usual to employ quanta which cause barely noticeable effects, but in this case the question arises as to whether fine detail is a quantization effect or is actually present in the image. Fig. 2-12, taken from Nathan (1968, p. 259), illustrates the effect of errors in transmission of horizontal lines from a scanned image and also the effect of added fiducials.

The above imperfections and variations could be viewed as various types of "pictorial noise", i.e., noise which is not present in the 3-D field being imaged. A satisfactory interpretation of imperfect or variant images (or their encodings) may require that the imperfections or variations be recognized for what they are, rather than being interpreted as actually present in the underlying 3-D field. Thus, in Fig. 1-2a, the loss of focus at the grain’s periphery should not be seen as an attribute of the 3-D field. To recognize these effects as such, knowledge of how it is that they arose or with what process they are associated, is required. For example, knowledge of the way in which the images are digitized enables an association of the horizontal streaks in Fig. 2-12 with the scanning process. (Effective representation and utilization of such knowledge is a difficult problem; see Section 2.5).

Turning to the differences between 3-D fields containing objects of the same
FIGURE 2-12 Image Showing Effect of Errors in Transmission of Scan Lines

FIGURE 2-13 Interactions Between "Spots" and "Tracks" in Interpretation of Nuclear Particle Tracks

FIGURE 2-14 Obviously Non-Polliniferous Objects (enlargements from Fig. 2-4a)
class, an obvious possible variation is in the number and arrangement of these objects. Of more concern here are the variations within objects of the same class, and the presence of and variations in other material (which is not of direct interest), in the 3-D fields. These variations could be considered as "object-world noise" in contrast to the pictorial noise mentioned above.

With regard to inherent differences between objects of the same class, it is clear that for many objects found in nature, the 3-D cross-correlation between pairs of objects (as measured by positioning the objects optimally with respect to each other and comparing them on a volume for volume basis, i.e., 3-D "template matching") may be both small for objects which are accounted as being similar, and large for objects which are alike only in gross shape. 3-D correlation is usually not, therefore, a satisfactory measure of resemblance; in characterizing and comparing such objects, their organization may be far more important than their precise 3-D distribution. Some organizations, e.g., spiders' webs, or the distribution of trees in a forest, arise from constraints (which are often not apparent) present during their formation. The organizations of objects in the same class tend to be similar (in what may sometimes be rather obscure ways), but because of hereditary and environmental differences, their 3-D distributions may be markedly different, particularly if the objects are jointed, pliable, or flexible. (The situation with regard to cross-correlation of their 2-D images is even worse).

Pollen grains, in common with many other classes of objects, are characterized partly by their global shape, and partly by their internal organization. With such objects, identification (to a degree of confidence which is a function of the uniqueness of the internal organization and of the sub-parts) may be possible even if their projected boundaries are partially or totally occluded. Indeed, some objects such as forests are characterized primarily by internal organization rather than by their boundary shape

Even when organization is considered, instances of the same class may still vary quite drastically. For example, the number and disposition of limbs

---

1 Conversely, some abstract concepts, such as triangularity, are concerned with boundary shape rather than internal organization.
of a tree, or gaps in a bubble-chamber track, may be predetermined only within rough limits. Other common variations are that (i) an object may have extraneous material adhering to it or may have unexpected additional components; (ii) an object may have some expected parts missing, because of local damage or malformation; and (iii) an object may be globally or locally deformed, again as a result of damage or malformation. Variations such as these three usually qualify an object's description rather than disqualify it from class membership, but difficulties are likely to be experienced in assigning satisfactory interpretations to such objects, and in characterizing object classes to allow for these alterations.

Even more difficult is the case of objects whose class-membership is based on the way in which they could be employed, rather than their particular shape or organization. Bridges (used for crossing over obstacles) and chairs (used for sitting in) are examples of such objects; it is clear that interpretation on this basis by an API machine will be extremely difficult. The variations between repeated similar parts of a given object are in some sense comparable to the variations between complete isolated objects, but in the former case the variations may well be interdependent. For example, both pentagonal and hexagonal cells can be seen on the spherical surface of the pollen grain in Fig. 1-2b, but the number of sides in any particular cell is very much a function of the neighbouring cells, i.e., there is "context dependency".

As well as the material of direct interest in the 3-D fields, there may be material which is ultimately of no concern. It is often not possible to ignore this "uninteresting" material, however, because of interactions between the organizations of the two types of material, and consequent interactions in their interpretations. In Fig. 2-13, for example, the "tracks" are of direct interest, whilst the "spots" are ultimately of no concern, but when spots adjoin tracks or fall on top of them, a characterization of the spots is required to facilitate interpretation as adjacency or superimposition.

Another problem associated with the complexity and variability of objects, 3-D fields, and images, is the wide range of descriptors necessary for their appropriate or natural description. The behaviour of the various entities to be described seems to be continuous in most respects, so that it may be difficult
to derive consistent and natural descriptions using a reasonably small pre-chosen set of descriptors or properties. In attempting to describe contours in a given image, for example, it is clear that description of a contour as a sequence of straight line segments will usually be rather clumsy, in the sense that this description may indicate boundaries in the contour (at the end of each straight line segment) where none are pictorially evident, or may fail to indicate visible boundaries. If arcs of circles as well as straight line segments are allowed, a less clumsy description may be obtained, but there always seem to be situations which are describable only awkwardly with a small pre-chosen set of descriptors. This contention is perhaps better supported with respect to description of the gray scale variations of image areas; if gray level is the only property allowed (and therefore areas of constant gray level the only descriptor allowed), difficulties are encountered when two areas of differing gray levels merge smoothly into each other. Allowing gray level gradient as a property, areas could be described as having constant gradient in a given direction and a more appropriate description of the above situation might be possible, but difficulties will once again be encountered in trying to describe a white round spot falling away uniformly in a radial direction to a black background.

Much the same situation arises with respect to complex structural properties; it seems impossible to pre-specify all the possible organizations which might be encountered, particularly if unfamiliar material may be present — see Section 2.3.4.

The continuous nature of the behaviour of the objects, fields, and images introduces problems whenever a classification based on some property value is required; because the properties measured can take on a continuous range of values, two situations being classified may be assigned different classifications even if they are not perceptibly different. As an example, consider the relationship "near" defined for a pair of objects (A, B) such that only if the inter-object distance \( D_{AB} \) is less than some maximum \( M \) (which might be a function of a "frame of reference" such as the sizes of A and B), is A said to be near B; if \( D_{AB} = 0.99 \) \( M \) and a third object C is introduced such that \( D_{AC} = 1.01 \) \( M \), then A would be said to be near B but not near C, even though there would probably be no perceptible difference between \( D_{AB} \) and \( D_{AC} \).
2.3.4 **Unfamiliar Objects**

In the pollen-interpretation task, as in many other image interpretation problems, new species or classes of objects may be represented in a given input image and the action required in this case has to be considered. Even in the absence of such objects, the variability of objects, fields, and images is such that all images which have not been previously presented, or whose descriptions have not been retained for later interpretation, will be to some extent unfamiliar or novel.

Even if the range of possible object classes is bounded, there may be so many classes (e.g., the several hundred thousand species of pollen mentioned in Section 1.2.1), that it is not practicable to include detailed characterizations of all these classes within the API machine, and the range of inputs is therefore effectively unbounded.

It must be recognized that there are degrees of unfamiliarity, ranging from an object which is a variant of a particular known class, through an object which can be perceived as a member of a general family or as a broad type of object, to an object for which no resemblances are perceived. Examples from the pollen images range from:— (i) pollen grains which can be perceived as variants of a particular species; (ii) grains which cannot be seen as known species but bear a family resemblance to other species (e.g., prolate spheroids with three furrows); (iii) objects whose appearance characterizes them as pollen grains of some type; and (iv) other objects, as in Fig. 2-14, which are clearly not pollen grains. The task of an API machine can be viewed as describing input images in terms of the knowledge represented within it, and the descriptions in each of the above cases should utilize any relevant knowledge available, assigning, for example, a pollen grain to a family if not to a particular species, and using the interpretation "don't know" strictly as a last resort. Note that even in case (iv) above, it might be desirable to interpret the objects as non-polliniferous, partly to save an exhaustive search of possible pollen species and partly out of a desire to account for all image detail (for the purpose of ensuring consistency of interpretation — see Section 4.2.5).

All classes of objects are to some extent unfamiliar until they are characterized within the interpretation system; the question of formation of such character-
izations arises. One possibility is that the machine designer could attempt to specify these characterizations, either initially or as needed, on the basis of his impressions of the characteristic features. The veridicality of human perception indicates that the criteria employed in assigning objects to classes are effective, but there is some doubt as to the perceiver's ability to deploy these criteria satisfactorily (Clowes, 1968). Once an API machine has the ability to isolate objects and their sub-parts, and to recover organizational detail, then it is already capable of forming descriptions of individual objects, and if given the additional ability to abstract from object descriptions to class characterizations, it would be capable of "learning" new classes. In a complex and changing environment such a capability would be valuable, but there are problems with regard to formation of alternative taxonomies according to which features are regarded as significant. Through an unsatisfactory choice of such features, objects which in reality are different might be grouped, and objects which are otherwise known to be of the same class might be separated into different classes. As an indication of the importance of a correct choice of features, a Chinese acquaintance of the author's, on being told that Australians had difficulty in telling Chinese apart, remarked that there were (to him) obvious differences between Chinese but he had great difficulty telling Australians apart. Hebb (1949, p. 27) relates a similar experience with naive observers trying to tell chimpanzees apart, before they become aware of the significant discriminating characteristics.

2.4 Ambiguity

Ambiguity exists whenever there is more than one possible interpretation of either the complete input image or a local portion of it. Because the image itself can never provide more than partial evidence for the situation depicted, there are usually many possible interpretations, but the ambiguity may not be apparent to the human perceiver because, in the global context of what is known about images and objects, one of these interpretations is the simplest or least awkward.

There are many examples in the psychological literature of ambiguous images (the Necker cube, the Peter-Paul goblet, Boring's "my wife and my
mother-in-law", Schroeder's staircase, etc.) where two or more interpretations are tenable. Clowes (1968) gives the example shown in Fig. 2-15, which may be interpreted as either a bellying sail or as a stingray. As is pointed out in Section 2.1, there are many possible 3-D fields which could underlie any given 2-D image, so that the perception of this 3-D field is inherently ambiguous or uncertain. This is perhaps the most fundamental and difficult form of ambiguity in the class of images considered herein.

In processing a given image and trying to impose an organization on it, problems of ambiguity of association and segmentation may arise. As Evans (1968a) observes, there is a close relationship between association and segmentation; which way a given process is regarded is sometimes simply a matter of whether the region being articulated is initially regarded as an undifferentiated whole or as a conglomeration of unorganized parts. The class of images considered contains simple parts which can be seen as being organized into rich structures. In trying to group these parts into more complex parts and wholes, there will usually be several concurrent relationships between pairs of parts suggesting alternative associations or groupings. It is often the case that the best organization in a local context is not the best globally, because it leads to inconsistencies. In Fig. 2-16, for example, choosing the "best" local organizations leads to multiple global inconsistencies. (As is discussed in Section 4.2, detection of inconsistencies is an important tool in the resolution of ambiguities). The many possible articulations of a sinusoid illustrate the problems which may arise. In Fig. 2-17, the curve L can be interpreted as a contour separating two different regions, 1 and 2, or as a line on a background region 1 plus 2. Taking the first alternative, there are several possible organizations depending upon which points are chosen for segmenting L, which associations between and within segments are regarded as significant, and which associations are regarded as subordinate. If points of maximum curvature are chosen as "break-points" (Attneave, 1954), region 1 (or 2) can be seen as being on top of region 2 (or 1), and as having a straight boundary with "peninsulas" added to it or "bays" missing from it, according to which S-shaped segments are paired (as boundaries of common parts). If points of inflection are chosen (Freeman, 1967), the boundary between the two regions is seen as consisting
FIGURE 2-15  
Ambiguous Figure

FIGURE 2-16  
"Impossible" Figure  
(Penrose and Penrose, 1958)

region 1

region 2

curve L

• = possible break-points

FIGURE 2-17  Ambiguities in the Segmentation of a Sinusoid
of both "bays" and "peninsulas" at the one time.

It was mentioned above that ambiguities are often not apparent because, in the context of relevant knowledge, one particular interpretation is the "best". In the absence of such knowledge, no satisfactory global interpretation may be possible, because of the multiple local ambiguities which cannot be resolved and which lead to global inconsistencies. This situation is typified by the experience of a naive human observer viewing unfamiliar material, such as a microscope slide, but lacking relevant knowledge of what is represented; parts may be seen, but it may not be apparent which way the parts "go together", and a coherent overall organization may not be achieved. An experienced observer will probably not have the same difficulty; the application of relevant knowledge of objects is an important process in guiding interpretation and in resolving both local and global ambiguities (see Section 4.2).

A rather different form of ambiguity arises with respect to the question of which of several levels of resolution is significant. In this case, the organizations at each of several levels may be relatively independent, but are not mutually exclusive; in so far as only one organization may be attended to at any given time, there is a degree of ambiguity. Figs. 2-18a,b illustrate this form of ambiguity.

2.5 Representation of Information

It has been implied in previous sections of this chapter, that effective interpretation of images relies on a representation of various types of general and specific "knowledge of the world" being available within the API system. Some of the types of knowledge required are:

(i) characterizations of known objects and the possible arrangements of these objects in 3-D fields;

(ii) knowledge of important relationships upon which organizations of the 2-D image and the perceived 3-D field are based;

(iii) knowledge of how 3-D fields are related to their 2-D images, and of how an input image is related to its machine representation;

(iv) knowledge of unreconciled past experience (i.e., objects and configurations which could not be interpreted satisfactorily when
(a) "Study of differentiation and similarity in visual perception" by Parisian Op artist Lily Greenham (reproduced from Guilano, 1967, p. 41). Note the different organizations perceived when looked at closely (thus emphasizing fine-resolution contours), and when looked at from a distance or through half-closed eyes (thus emphasizing coarse-resolution contours).

(b) Line drawing exhibiting three different organizations at different levels of detail (squares, triangles, and an arc)

FIGURE 2-18 Separate Organizations at Different Resolution Levels
first encountered, but for which later experience might supply required information and allow satisfactory interpretation); and (v) information concerning the current input image, i.e., its description as a 2-D image of a 3-D field, in terms of the knowledge available.

Effective interpretation is therefore a function of the adequacy of an internalized "model of the world". The importance of suitable models has previously been emphasized: McCarthy and Hayes (1969) observe that "a computer program capable of acting intelligently in the world must have a general representation of the world in terms of which its inputs are interpreted"; Evans (1964) states that the critical factor is that the machine should have "a good internal representation of both its subject matter ('objects') and its methods ('transformations'), as well as an elaborate set of 'pattern-recognition' techniques for matching transformations to object pairs"; and Rosenfeld (1968a) says that "if we want to give our computers eyes, we must first give them an education in the facts of life". The adequacy of such models can be judged with respect to a number of criteria such as the extent and organization of the information represented, and the relative economy of the encodings used. The organization of the information in the data base or model is important if relevant information is to be readily accessible (or even indicated as being relevant).

The difficulties likely to arise in choosing a representation are greater than it might at first appear; effective information representation is one of the most difficult but important requirements in an API system. How, for instance, might the relationships between an input image and the machine encoding of it be represented, so that an isolated horizontal streak in this encoding could be perceived as resulting from noise in one of the associated horizontal scan lines of the digitization process? 3-D objects can, in general, be positioned anywhere in the 3-D field and can be randomly oriented with respect to the image forming device. The projected images of similar objects can therefore be of widely varying sizes, locations, and appearances (from different orientations). How should classes of such objects be characterized so that in comparing objects in the input image with known classes, variations such as these do not prevent detection of correspondences?
It seems clear that the characterizations should not be inextricably tied to fixed reference frames. In many cases the characterizations should not be completely free of reference frames either, because the objects being characterized may have expected sizes, preferred orientations, etc. For example, "A" may be described simply as an A, whereas "\forall" might be described as an inverted A. As a further example, it may be readily observed that it is much more difficult to recognize known faces in a group photograph if it is inverted, thus indicating that our internalized characterizations are not independent of reference frames.
Chapter 3

AN EVALUATION OF CURRENT APPROACHES TO AUTOMATIC PHOTO-INTERPRETATION

This chapter briefly surveys and evaluates current approaches to API. Work on character recognition and symbolic images has been considered when it appears to be of particular relevance, but the main emphasis is on approaches which have been implemented or advocated for PI. The evaluation is primarily with respect to the difficulties likely to be encountered in general PI tasks (as already outlined).

The principal aim of this survey and evaluation has been to facilitate the development of an approach to API which shows some promise of being able to cope with the problems identified, by exposing strengths and weaknesses of current approaches. Some initial steps toward this goal are described in Chapter 4. A further motive was to provide a rather more detailed background of previous work in API than is supplied by the brief outline of the field of picture processing by computer given in Chapter 1, so that the research reported in the following chapters can readily be related to and compared with other investigations.

Many API schemes have been reported in detail in the literature. The range of approaches is both wide and to some extent continuous or overlapping, in the sense that there is no clear unequivocal partitioning of the various schemes into separate areas; different "clusters" appear depending on which features are considered significant. Nevertheless, for the purpose of simplifying discussion, the range of approaches is divided herein into four areas: template matching, property list, articulation analysis, and picture parsing. Some approaches (e.g., the "perceptron" approach (Rosenblatt, 1960)) seem to span more than one of these areas, and there may be some question as to the suitability of the choice and delimitation of areas and the assignment of individual schemes to these. As Uhr (1963, p. 45) notes, there are "confusing similarities that exist between superficially different methods".

Discussion of each of the four areas commences with a summary of the main distinguishing features of the general approach in each area. Next, brief
descriptions of hopefully representative samples of specific schemes and applications in each area are made. In an attempt to get some "depth" as well as "breadth", these are followed by more detailed descriptions of two or three schemes selected according as they (i) exemplify the general approaches, (ii) are applied to API tasks which are in some respects similar to the pollen-interpretation task, or (iii) exhibit features of particular interest. In discussing various schemes, generality and capability are of more interest than implementation, but there is of course concern that they should be implementable. The discussion of each area concludes with an evaluation.

The chapter finishes with a short discussion of the four general approaches examined, individually and as a whole, in the context of complex PI tasks — see Section 3.5.

3.1 Template Matching (Minsky, 1961, sect.IIC; Uhr, 1963)

3.1.1 Outline

In this approach it is implicitly assumed that objects are characterized by their 2-D images, and that if a set of characteristic or prototype images (i.e., "templates") is available, a given image can be described in terms of the sub-images within it which match templates according to some acceptance criterion. The prototypes usually include some "background" or "context" for the projected images of the interesting objects.

A common technique for computing the match between a template and a sub-image is that of 2-D template matching or cross-correlation. In computing the match, varying weights may be assigned to different parts of the template, or two binary fixed-weight templates (one positive and the other negative) may be employed (Andrew, 1969). In the latter case, the sub-image must correlate well with the positive template and poorly with the negative template. The templates are sometimes formed by a process of averaging over a large set of sample images of objects in the desired classes, perhaps by means of a training or "learning" procedure, rather than by predetermined initial specification.

The input images are often "preprocessed" for the purpose of reducing the
"nuisance variables of brightness and contrast so that consistent, detailed, binary pictures of the original image are obtained" (Abend et al., 1965, p.538), or alternatively, for standardizing the variations in image parameters such as contrast and gray level (e.g., by automatic gain control (Sebestyen, 1963)). Enhancement of contours (Kanal and Randall, 1964) is a common operation, and with this and other types of preprocessing, the templates reflect the appearance of the images of interest after such processes have been performed. Segmentation of the image into "objects" (Holmes, 1966) might also be considered as preprocessing.

Cross-correlation of a template with an input is similar to the process of spatial filtering, where the sub-images of interest are regarded as spatial "signals" embedded in background "noise", and a matched filter is employed to maximize the response to signals relative to noise (Vander Lugt, 1964). The filters employed here are closely related to those employed for "enhancing" the appearance of interesting image detail — see Section 1.1.2.

Many experiments with "perceptrons" (Rosenblatt, 1960; Murray, 1961) and modifications of these (Joseph et al., 1963; Widrow, 1964; Nilsson, 1965) have been reported. In essence, these devices compute matches with a large number of "sparse" templates and classify according to more or less complex functions of these matches, which can be considered as simple "properties" of the input image. Such devices, therefore, could be considered for inclusion in Section 3.2, but they are considered here because of the restricted nature of the properties, and because the locations of the various templates are fixed with respect to each other and, to this extent at least, the properties are not independent.

In general, templates are by their very nature, tied to reference frames for gray scale, size, location, and orientation. To facilitate generalization over these parameters, gray scale excursions may be standardized by preprocessing (typically to a binary image, i.e., black and white), and attempts may be made to extract "objects" from the image and normalize these with respect to size, orientation, and location, before attempting to match the isolated object against prototypes (Holmes, 1966). Alternatively, templates may be correlated with all possible positions in the input image (this being readily achieved with optical processing (Vander Lugt, 1964)), and at a number of orientations (Harley et al., 1968),
although templates of different sizes are only occasionally (Sakai et al., 1969) employed.

The objects of interest (or at least their projections) are treated as "wholes", i.e., they are not considered as consisting of a number of organized or unorganized sub-parts. Even in perceptron-type devices, with their multitude of separate templates, there is no effective detection of separate sub-parts, because the templates are fixed relative to each other. Note that the input image itself is sometimes implicitly segmented into sub-parts when "objects" (i.e., sub-images) are extracted either by preprocessing or through matches with templates.

3.1.2 Examples

Much template matching work has concerned character recognition (Bledsoe and Browning, 1959; Kamentsky, 1961; Dye, 1969), but several API investigations have also been based on this general approach. Kabrisky (1966, 1967) proposes 2-D cross-correlation as a basic model for biological information processing, and tries to reconcile known physiological facts of the visual cortex with a possible neural structure for computing such correlations. Sakai et al. (1969) have developed a rather sophisticated template matching scheme which uses contour extraction and extension preprocessing followed by template matching with variable size templates, and have applied this scheme to detection of human faces (in a vertical orientation and frontal aspect, but of no fixed size) in photographs. Vander Lugt (1964) has applied linear matched-filtering concepts to character recognition and the detection of a plane-like shape embedded in a noisy background, but Trabka and Roetling (1967) discuss some of the conditions required for the applicability of linear matched-filtering theory and, arguing that these conditions are usually not satisfied in API tasks, give an example of non-linear filtering applied to the detection of trucks in an aerial photograph. Kanal and Randall (1964) use

---

1 An apparent error in Sakai et al.'s computation of gray level gradient (by subtracting gray levels of adjoining image matrix cells), is that no allowance is made for the fact that the distance between diagonally adjacent cells is $\sqrt{2}$ times the distance between directly adjacent cells.
Laplacian contour-enhancement preprocessing followed by a perceptron-type discriminator, for detection of tanks in aerial reconnaissance photography. Joseph and Viglione (1966) have investigated the application of perceptrons to the discrimination of vortex from non-vortex cloud patterns in TIROS satellite cloud photography, but with only limited success. Darling and Joseph (1968) have used a number of different decision functions within a perceptron-type structure, in discrimination experiments concerning images of lunar topographic features and satellite cloud patterns, with somewhat better results.

Kovalevsky (1968) has investigated the problem of "describing a noisy picture in terms of certain elementary parts of which the picture is composed", using the examples of describing individual handprinted characters and lines of typed text. The letters in the text were not clearly separated, and Kovalevsky presents a formulation of the question of finding the best segmentation of the text into individual characters, as a problem of finding the longest path through an oriented graph. The length of each branch in a sequence forming a path corresponds to the correlation value of a particular character prototype with the image at a corresponding location in the input text sequence. For example, the prototype for the character "F" might correlate well with the vertical arms of a given character "N", but this segmentation leaves the diagonal stroke unaccounted for and so, in the global context, is rejected in favour of classifying the given character as an "N". While based on template matching of individual characters, Kovalevsky's technique is fairly sophisticated and the method of finding the best global interpretation is rather interesting.

Joseph et al. (1962, 1963) report an "advanced image filter" which is supposedly capable of "automatically recognizing and classifying three-dimensional objects regardless of their size and orientation with respect to the viewing mechanism" (1962, p. 1). A lens images the object to be recognized (which is light coloured and mounted on a black background at a fixed distance from the lens) onto a $20 \times 20$ matrix of photo cells, whose outputs are applied to a perceptron-type structure. The system is said to be remarkably accurate and reliable, regardless (within broad limits) of the size, orientation, or position of the objects being recognized, but there are quite severe restrictions on these parameters in the experimental set-up and the system in its present form.
the experimental set-up used, because the objects have to sit on a small table, with the result that their projected sizes and locations will not vary greatly.

Holmes (1966) considers the problem of locating and classifying "targets" (i.e., objects of timely military interest) in aerial reconnaissance photography. The projected images of targets are embedded in background "noise" and may assume any orientation and a range of sizes, although Holmes suggests that size variations could readily be reduced because the taking height and camera characteristics are known. Target location is viewed as a pattern recognition problem and (Holmes assumes) can therefore be represented by the simple block diagram shown in Fig. 3-1; whether or not this is an adequate representation of pattern recognition processes is a moot point. Holmes' investigations seem to have been primarily concerned with techniques for "object isolation", where it appears that he defines objects as sub-images which have reasonably consistent internal gray level statistics relative to their surroundings. Object isolation is considered to be a non-adaptive preprocessing operation, and two techniques for isolation, a "Kolmogorov-Smirnov filter" and a "picture-frame filter", are described. These filters produce binary silhouettes, thus discarding gray level information. The silhouettes tend to be "shredded" if shadows fall across the targets, and a gap-filling routine is invoked to fuse the shredded portions. The resulting silhouettes are passed, one at a time, to a program which standardizes their orientation via rotation of the major axis. Similarly, it is also possible to standardize the location and size of individual silhouettes prior to further processing. Holmes reports some apparently successful experiments with synthesized silhouettes and a perceptron-type classifier.

3.1.3 Evaluation

The template matching approach is conceptually simple and easy to implement, but is clearly inadequate in at least two very important respects (viz., objects are treated as 2-D, and there is no consideration of image or object organization) and can hardly be considered seriously as a basis for a general API system. There seems, therefore, little point in detailing all the shortcomings of this approach, but other major weaknesses in addition to
FIGURE 3-1  Block Diagram for Pattern Recognition (after Holmes, 1966)
those pointed out above are:

1) With regard to segmentation, the objects of interest often do not have constant internal gray scale statistics relative to their surroundings (e.g., because of shading, shadows, and pigmentation), and cannot be satisfactorily isolated by techniques based on an assumption that these statistics will be relatively constant.

2) Fine detail is submerged in averages, or deleted in drastic preprocessing.

3) The representations employed (i.e., prototype sub-images) are incapable of representing such simple classes as the class of figures with an even number of disconnected parts (Minsky, 1961, p. 12; Minsky and Papert, 1969), and usually do not capture the essential features of the objects of interest. Several template matching investigations have concerned target detection in aerial photography; an unco-operative enemy could play havoc with such systems simply by decorating each potential target with different abstract motifs! As representations, templates are deceptive because the human observer articulates them and sees more in them than does the machine.

The main strength of template matching is its performance with regard to fixed spatial signals in noise. As a component in a more sophisticated scheme, template matching could well play a useful role in detecting simple unstructured image features such as contours and spots.

3.2 Property List (Minsky, 1961; Sebestyen, 1962; Abramson et al., 1963; Uhr, 1963; Nagy, 1968; Levine, 1969)

3.2.1 Outline

In this general approach (which is sometimes referred to as the "parametric" approach (Sebestyen, 1963)), a number of property values are computed for the input image (or some selected portion of it), and a classification into one of a chosen set of categories is made on the basis of these values, the idea being that
similar objects will have similar property values if the selected properties are appropriate. This is basically the receptor-categorizer model outlined by Marill and Green (1960). As with template-matching, most property list work has been concerned with character recognition, usually with isolated characters, and when applied to API, it is implicitly assumed that 3-D objects are effectively characterized by their projected images. With images containing projections of a number of objects, it will usually not be practicable to classify all the objects present at the one time (Minsky, 1961), and some form of figure extraction is often attempted (Muerle and Allen, 1968; Rosenfeld, 1968a, Ch. 8), although by measuring suitable properties at all points in the image and classifying individual points (e.g., as "urban" versus "non-urban") preliminary segmentation may sometimes be avoided (Hawkins et al., 1966; Rosenfeld, 1962). The property list approach has been predominant in character recognition and API work for some years and several surveys of this work have been published (e.g., references given at heading).

Three processing phases have been identified in this general approach (Davis, 1967a,b):— (i) preprocessing, (ii) property measurement or feature extraction, and (iii) classification on the basis of property values. The first two phases are viewed as being rather more application-dependent than the third. The preprocessing phase is not always present, but when employed the intention is the same as that in the template-matching approach, i.e., to reduce the "trivial" variations and the information content of the input image, while retaining the "significant" information, a typical operation being binary quantization of the input (Butler et al., 1967). Alternatively, the image may be transformed into another domain (e.g., the spatial frequency plane) in which unwanted variations in the input image are less evident. Another important preprocessing operation is image segmentation prior to measurement of property values for extracted individual "objects" (Butler et al., 1967).

A correct choice of properties is very important; the success of the classification phase is dependent upon the chosen properties being both appropriate and adequate. Many types of properties have been considered, ranging from cross-correlations with a number of independent templates, (either pre-chosen (Hawkins
et al., 1966), or generated randomly and retained if they are shown to be useful (Uhr and Vossler, 1961), through computation of central moments of various orders for silhouette-type figures (Butler et al., 1967), to gray scale statistics (Rosenfeld, 1962; Prewitt and Mendelsohn, 1966). If, however, a given approach employs relational or articular properties, then that approach is considered herein as articular analysis (examined in Section 3.3 following) rather than as property list. The properties selected should ideally be "rugged" in the sense that they are relatively insensitive to "trivial" variations of the input images (e.g., size or orientation of the figures of interest). Plausible properties are discussed in Nagy (1968), Rosenfeld (1968a, Chs. 7 and 9), and Levine (1969).

With regard to the classification phase, there has been extensive research into methods of classifying objects on the basis of their measured property values, with the aim of minimizing misclassifications (Ho and Agrawala, 1968). A basic principle of the property list approach is that it is often possible to get reliable decisions from a combination of individually unreliable measurements; Teitleman (1964) says that "it has been observed in many pattern recognition projects that a reliable decision may be possible even where individual tests are poor, provided each test contributes some different fractional bit of information". From the standpoint of simplicity, it is tempting to derive the final classification by means of a decision tree, with intermediate decisions based on individual property measurements, but Sebestyen (1963, p. 31) warns that the accuracy of final classification cannot be superior to, and will usually be inferior to, a final decision which is reached in a single step taking all the measurements into account simultaneously. A useful way of visualizing the complete set of N measurements, is as a point in N-dimensional hyperspace. If the individual measurements are good characterizers of the figures of interest, then by a suitable transformation of the property space, similar figures will be represented by points which are close together (in the Euclidean sense) in this space (Sebestyen, 1962). The classification problem can then be viewed as one of partitioning this space into regions, by means of decision surfaces which separate "clusters" of figures in each class from clusters of figures in other classes. For convenience of analysis and design, it is often assumed that the individual properties are independent and that the property values
of figures in a given class form a Gaussian distribution; these assumptions are usually not applicable in practical PI tasks.

3.2.2 Examples

Levine (1969) reports experiments by Smith and Wright in which ship silhouettes were classified on the basis of central moments as properties. Butler et al. (1967) have also employed central moments for the recognition and pairing of chromosome silhouettes previously isolated from a micrograph; quite useful results seem to be obtained provided that a satisfactory initial segmentation has been achieved. Uhr (1963) describes a successful character recognition scheme (patented by Rabinow) which employs templates for straight lines, curves, angles, etc., forming sub-parts of characters, and treats matches with these templates as properties. Rosenfeld et al. (1965) discuss properties such as brightness, brokeness, texture, connectivity, and shape, which are apparently useful in cloud pattern recognition. A technique for discriminating broken from unbroken cloud is described, but techniques for measuring some of the other properties had yet to be developed. Hawkins et al. (1966) correlate local areas of aerial photographs with small templates designed to respond to spots, contours, lines, and arcs at various orientations, using above-threshold responses to these templates to classify each local area as an orchard, forest, lake, railroad yard, etc. The authors claim 85% recognition over all classes in 1124 classification decisions.

Prewitt and Mendelsohn (1966) are concerned with methods for extracting meaningful information from optical data such as cell images, and they report experiments in discriminating four different types of leukocytes in micrographs of human blood. They see the analysis of digitized images as involving five principal stages:- (i) delineation of figures (i.e., objects of interest) from the background; (ii) description of the figures by numeric, non-numeric, and relational properties; (iii) determination of the range of variation and discriminatory power of these properties; (iv) development of appropriate decision functions and taxonomies for classification; and (v) identification of unknown specimens. A flying-spot scanner digitizes the $50 \mu \times 50 \mu$ input image at $0.25 \mu$ intervals into an 8-bit gray scale. The leukocyte images contain several sub-areas of constant but differing
gray levels, which are all darker than the background. Prewitt and Mendelsohn report an interesting and successful technique of setting thresholds (for discriminating between the several sub-areas) equal to the gray levels of "valleys" in a histogram of the image gray levels. A large number of statistical properties can be computed from the gray levels of the several sub-areas in each extracted leukocyte image, and a selection of 35 possibilities is listed. Excellent discrimination between the four types of leukocytes was obtained using only a few of these 35 properties.

Rosenfeld (1962) considers the problems of characterization and segmentation of highly detailed images such as aerial photographs. Many of the properties considered for simpler images (e.g., typewritten characters) are not applicable in this case because of the lack of a "natural distinction between figures (= shapes or patterns) and background". He suggests characterizing the various regions of a given image in terms of statistical moments of the local gray level distribution, and using sudden changes in these properties as a basis for image segmentation. These moments characterize visual texture to some extent, and much of the detail in a given image is seen as being of little significance by itself except in so far as it contributes to this characterization. The image might then be described in terms of variously textured regions within it. Rosenfeld considers mean gray level and contrast frequency as the two most important textural characteristics, and contends that "the variance of the density will usually be closely related to the mean contrast frequency, since a high density variance means a high degree of fluctuation of density, which is essentially equivalent to a high frequency of contrasts".\(^1\)

Experiments were performed with 1-D statistical properties computed along two scan lines in an aerial photograph; all "obvious" boundaries were identified and it was assumed that the few spurious boundary indications could be eliminated by comparison with adjacent scan lines.

Rosenfeld's technique represents an interesting departure from the commonly adopted definition of figures as areas of constant gray level relative to their

---

\(^1\) This contention needs further qualification however; for example, two checkerboards of different cell size may have equal variances but rather different contrast frequencies.
surroundings, but Selfridge (1962), in a critical review of Rosenfeld’s paper, questions (i) the generality of the technique; (ii) the assumption that the information discarded is irrelevant; and (iii) whether the statistical properties employed capture the fine differences in texture which humans can respond to. The work reported is apparently of a preliminary nature, and the following comments may be relevant to further studies:—

(i) where the "textural elements" have directional properties, characterization by 1-D statistics in a single direction is inadequate;

(ii) textural organization, in terms of an underlying 3-D field and relationships between elements, may be important (see Section 6.2.2); and

(iii) a difficulty with finding individual boundary points is that even though such points may group into boundary segments, these segments will often fail to form closed figures.

3.2.3 Evaluation

Whilst the property list approach is more general and more capable than template matching, it still cannot be seriously considered as a basis for a pollen interpretation system, because of major inadequacies; this evaluation is therefore not detailed. Among the inadequacies are:—

1) Objects are treated as 2-D, with the attendant problems outlined in Section 2.1. In simple cases it might be possible to conceive of 2-D properties which are characteristic of 3-D objects, but this is unlikely with more complex objects.

2) Independent image components may be considered, but there is no attempt to capture their organization.

3) As with template matching, satisfactory initial segmentation on the basis of gray scale statistics will usually not be possible. The alternative technique employed by Hawkins et al. (1966), applies only to objects characterized by their internal properties rather than by outline shapes.

4) The representation of instances and classes as points and volumes in hyperspace respectively, does not seem an effective characterization
for the patterns of interest. The properties employed are usually chosen for their ability to contribute to correct classification, rather than for their ability to describe effectively; these abilities are not necessarily equivalent (Uhr, 1968, p.175). Properties such as normalized central moments, which are useful for classification because they are insensitive to changes of pattern size, orientation, or location, convey little information to us about the shape of the pattern being classified.

Given suitable properties, property list techniques enable very successful classifications, and these techniques could well be useful in a comprehensive API system, particularly if the objects of interest are characterized by metrical properties (such as the ratios of lengths of sub-parts to their widths).

3.3 **Articular Analysis** (Levine, 1959, sect.III.D; Uhr, 1963; Lipkin et al., 1966).

3.3.1 **Outline**

This general approach is distinguished from the two previously examined by attempts to deal with image organization. Articulation can be defined as the process of recovering or imposing organization, and an articular image description is thus one which specifies image organization. At one extreme, articular analyses might resemble property list approaches augmented with relational properties. Thus, structured figures such as alphanumeric characters could be represented in terms of detected sub-parts, together with sufficient structural information to "tie down" the arrangement of these sub-parts to some extent. At the other extreme, the "picture parsing" processes examined in the next section could also be considered as articular analyses.

Lipkin et al. (1966) maintain that "one can be fully as precise in explicating pattern and structure as in performing measurements which yield numerical quantities". They distinguish "metrical" properties, which are concerned with numerical values, from "articular" properties, which are concerned with structural information.

Articular analysis shows much greater latent promise of being able to cope
with complex images (in which image and object organizations are important) than either the template matching or property list approaches. The difficulty of specifying and recovering important relational information has, however, been such that few attempts have been made to apply this approach to other than relatively simple images.

The concept of pattern syntax (i.e., rules for the construction of patterns out of simpler components) is fundamental to articular analysis, but the extent, method of representation, and part played by these rules vary considerably. The syntax rules (no matter what form they take), together with other information regarding the components out of which the patterns are constructed, form a more or less detailed "model" of the patterns of interest. The adequacy of such models may be evaluated with respect to their ability to generate patterns which are similar to the patterns of interest (Lipkin et al., 1966). The patterns specified may be specific classes of objects (as in Roberts, 1965) or more general classes (as in Guzman, 1968). The "picture parsing" or "linguistic" approach to picture interpretation examined in the following section, can be considered as a subset of the articular analysis approach. A perhaps somewhat arbitrary distinction will be made herein on the basis that in picture parsing, the syntax specifications are explicit and directly control the interpretation, with the depth of structuring of patterns in terms of parts, sub-parts, etc., possibly extending to many levels, whereas in articular analysis, the syntax rules aid rather than direct the analysis, and there are typically only one or two levels of structuring. This distinction is not always made (e.g., Levine (1969) considers picture parsing as a form of articular analysis), but because of the wide range of approaches encompassed otherwise, a division seems profitable. The articular analysis approach (as delimited here) has alternatively been referred to as the "syntax-aided" approach (Narasimhan, 1968) or the "strong analytic" approach (Uhr, 1963).

The work in this area has been predominantly concerned with line-like elements as the basic parts, although these elements are not always of negligible thickness and Narasimhan (1962) refers to them as "roads". The input images have typically been line-like as presented, or have been such that they can be satisfactorily represented as line-drawings. There may be extensive preprocessing
involved in automatically obtaining line-like representations of the inputs (Roberts, 1965), or the analysis may commence with a list of line segments and vertex locations as input, with the assumption that suitable preprocessing could have produced these (Evans, 1964; Guzman, 1968). Because of the preoccupation with idealized line-like elements, most of the relationships considered have been those which obtain between such elements at the points of coincidence (e.g., "joined to", "crossing", and "coterminous"), although more general relationships such as "inside" and "above" have sometimes (Evans, 1964) been employed.

In this general approach, classification of objects may be secondary to their description. In some applications, classification is not evident (Roberts, 1965) or not even meaningful (Evans, 1964; Guzman, 1968), and flexible data structures may be employed to facilitate hierarchical machine representations which reflect the organization of the objects described.

An important advance over the previously examined approaches is the explicit acknowledgement in the work of Roberts (1965) and, to a lesser extent, Guzman (1968), of a 3-D reality underlying the input images. Unfortunately, this acknowledgement is not general, and other articular analysis schemes (Olson, 1968) still treat the projected images as the objects of ultimate concern.

3.3.2 Examples

The first clear example of an articular analysis scheme appears in the work of Grimsdale et al. (1959). Whilst they have investigated isolated hand-drawn letters and other line-like figures, they claim that their technique is applicable to any form of spatial pattern. The input is binary quantized and segmented into line segments, of relatively constant width, which are then assembled into longer elements such as straight lines and curves, making allowance for ragged edges, gaps, and other imperfections. A description of the figure in terms of these longer elements, their connectivities, and their relative lengths, is formed and compared with similarly formed descriptions of known figures (or stored for comparison with later inputs). An interesting aspect of the comparison technique is that known descriptions are grouped according to certain basic features, and the presence or absence of these features in a given input figure helps determine the order of
comparison. Other work concerning articular analysis of characters is reported in a recent special issue of the Marconi Review (Byatt, 1969; Hosking, 1969).

Hankley and Tou (1968) describe a system for automatic description and identification of fingerprints, based on the recovery of topological features from fingerprint images which have been subjected to fairly sophisticated and extensive preprocessing. Topological relationships have also been employed by Sherman (1960) in character recognition experiments.

Evans (1964), in his program for solving geometric-analogy IQ test problems, assumes that suitable preprocessing has produced a list of basic elements (lines, spots, and circular arcs) for all the pictures in the given test question, and derives descriptions for each picture in terms of its connected subfigures, their compositions from basic elements, and their relationships one to another. In addition to relationships like "inside" and "above", the important relationship of similarity between subfigures and groups (i.e., identity under simple transformations such as rotation and translation), is also recovered.

Lipkin et al. (1966) suggest that computers might well be valuable allies in analyzing structural as well as metrical information, and introduce the concept of "pictorial augmentation" in the analysis of deficient images, whereby parts "which from biological considerations are present, though pictorially manifestly absent" (p. 996), are supplied. Such inference on the basis of prior knowledge is also referred to by Rosenfeld (1968a, Ch. 10, pp. 2-3).

Olson (1968) has studied the problem of detecting man-made artefacts in aerial photographs and (assuming that such objects are characterized by straight lines) describes a system for detecting straight line features and their connectivities, using a special contour-line detector followed by a "path processor" which attempts to construct long straight lines or gentle curves out of detected line-segments. Classifications as freeways, streets, automobiles, etc., are to be performed on the basis of the intersections, junctions, and lengths of lines detected, but Olson does not give practical results.

Further examples of articular analyses are found in the work of the groups at Stanford University (McCarthy et al., 1968), Stanford Research Institute (Forsen, 1968; Raphael, 1968), and MIT (Davis, 1969), on robots with visual input; some
of this work is based on the earlier investigations of Roberts (1965) and Guzman
(1967).

Roberts (1965) describes experiments in "machine perception of three-dimensional solids". An input photograph is taken of a 3-D field containing white solids resting on a plane black background (so that problems with image contours resulting from shadows cast by objects on the backgrounds are minimized). The solids must be composed of plane-surfaced models such as cubes, wedges, and hexagonal prisms. The photograph is digitized, scaled psychophysically (to increase the correspondence between perceived and measured gray level differences) and differentiated so that contours are emphasized. Line templates at each of four orientations are correlated with small square areas of the differentiated image, and if the ratio of best to worst correlation is greater than a preset threshold, then a line segment is "detected" and is later joined to adjacent segments. A sophisticated line extension; and "noise" cleaning process is followed by a routine which fits straight lines to the groups of line segments and eventually produces a line-drawing abstraction of the input photograph. Information regarding contained interior polygons and their topology is obtained from this abstraction. Using the idea of "approved" polygons, which are those formed by projections of the simple basic models (and which for cubes, wedges, and hexagonal prisms must have either 3, 4, or 6 sides), a reference point is chosen and models are tested against the input by listing topologically equivalent point pairs and seeing if there is a projective (and scaling) transformation which can take the model vertices into the picture vertices. If the model "fits", its orientation in the 3-D field is specified and its depth is found later by use of the camera parameters and the knowledge that the objects rest on a support plane. If the fitted model does not account for all the lines in the abstraction, it is accepted as being part of a more complex object and an internal "joint line" is constructed. The model fitting process is then repeated until all information in the abstraction is accounted for. An internal description of the objects perceived in the input photograph is encoded in a ring-type data structure (Sutherland, 1963). This description is a full 3-D representation, including surfaces not directly portrayed in the input photograph, and the interpreted scene can be manipulated and output on a graphic display, to verify
that the program perceives the same underlying situation as a human observer does. A number of significant contributions have been made in this work:— (i) the 3-D nature of the field portrayed in the input photograph is convincingly accounted for via the use of homogeneous co-ordinates and projective transformations; (ii) a "model" of the world, as contained in the representations of cubes, wedges, and hexagonal prisms, together with a scaling and projective transformation matrix equation, is used to guide the analysis; and (iii) complex objects are articulated into simpler components. Roberts' work is not referred to in a recent survey of linguistic methods in picture processing by Miller and Shaw (1968), but his models are reasonably explicit and have a strong influence on the analysis, and his scheme could possibly be regarded as picture parsing.

Turning to the work of Guzman (1968) on "decomposition of a visual scene into three-dimensional bodies", there are similarities in the problems investigated by Roberts and Guzman, but their solutions are rather different. Guzman considers fields of 3-D plane-faced solids but, rather than starting with a photograph, he commences with a line drawing representation coded as a list of vertices, their co-ordinates, and the connections between vertices, together with a list of regions (i.e., projected surfaces limited by a simple closed curve) adjoining each vertex. After some preliminary processing, vertices are classified into one of eight types according to the number of lines meeting and their angles relative to each other (and in one case, "PEAK", relative to a base line).

The objective of the program is to link or associate pairs of regions and to form groups of associated regions into "objects", corresponding to the projections of 3-D objects which are perceived by human beings. A number of heuristics is employed in this process, some of which form "strong" links, and others "weak" links. Most links are formed at vertices (with the particular regions grouped depending on the type of vertex), with one exception being pairs of "T" joints; the eight vertex types, and the strong links set up, are shown in Fig. 3-2. A nucleus is defined as a region or a set of nuclei which are formed as follows:— if two nuclei are connected by two or more strong links, they are merged into a larger nucleus by 2-D concatenation. Nuclei are allowed to grow in this manner until no new nuclei are formed, thus generating a "maximal" set. Any nuclei
'L'; vertex where two lines meet.

'ARROW'; three lines meeting at a point, with one angle $\leq 180^\circ$.

'K'; four lines, one pair of which is collinear, with the remaining lines falling on the same side of the collinear pair.

'PEAK'; four or more lines whose uppermost ends are concurrent.

'FORK'; three lines forming angles $\leq 180^\circ$.

'T'; three concurrent lines, two of which are collinear (two separate T's shown).

'X'; four lines, one pair of which is collinear, with the remaining lines falling on opposite sides of the collinear pair.

'MULTI'; four or more concurrent lines not forming one of the other seven vertex types.

Note: Strong links between surfaces are shown by dotted lines.

FIGURE 3-2  Guzman's Vertex Types (after Guzman, 1968)
linked by one weak and one strong link are then merged, but this may introduce additional strong links. The original criterion is readopted and new merges are attempted, alternating between the weak and strong criteria until no further merging occurs. Guzman's program is reasonably successful, even with quite complex input scenes, when compared to the interpretations assumed by human beings. There are several assumptions made about the nature of the 3-D object-world depicted in the input image, which are reflected in the heuristics employed, but Guzman does not list these assumptions. One assumption exhibited in a heuristic concerning parallel lines, seems to be that the input image is an orthogonal (rather than perspective) projection of a 3-D field. A further assumption, indicated in the definition of the vertex type "PEAK", is that pyramidal objects will usually have their vertices uppermost.

3.3.3 Evaluation

As outlined herein, articular analysis is more a collection of related schemes, rather than a cohesive, clearly defined and delimited approach. However, of the four examined, this approach shows the most promise of being eventually able to cope with reasonably complex API tasks in which consideration of a structured 3-D object-world is required. Identification and recovery of the important relationships, and development of suitable representations, have to date been demonstrated only for relatively simple image classes. Considerable generalization and augmentation of the currently formulated articular analysis approach is required before PI tasks of the difficulty of pollen-interpretation can be considered, but because of its ultimate promise, a detailed evaluation of the current strengths and weaknesses of this approach is given below.

3-D fields

Roberts (1965) has made a significant contribution with regard to accounting for the 3-D fields portrayed in the input images. Within the restricted context of his fields, he demonstrates that interpretation in 3-D terms avoids the problems of changing 2-D shape, size, orientation, location, occlusion, etc. Guzman (1968) implicitly accounts for 3-D fields, but no explicit 3-D description is formed. Unfortunately the techniques employed by Roberts and Guzman do not generalize
to less restricted fields in which the surfaces may be curved, differentially pigmented, shadowed, etc. When articular analysis schemes are applied to these more general fields, the objects may still be treated as 2-D (Olson, 1968).

Object isolation

In articular analysis schemes the early stages of object isolation or image segmentation are usually similar to those employed in template matching or property list approaches, and the same problems arise with boundaries which fade out or which are based on differences in complex gray scale properties. The type of structural segmentation employed by Grimsdale et al. (1959), wherein strokes of constant width are separated from other strokes to which they are joined, is applicable to articulation of images such as that shown in Fig. 2-6, and might help overcome errors in the initial image segmentation by "breaking off" extraneous pieces. Image segmentation on the basis of internal structure of sub-areas has not yet been considered, and involves more sophisticated structure abstraction and comparison techniques than those currently implemented. The use of underlying "models of the world", to enable object isolation despite confusing surroundings and occlusion, is demonstrated by Roberts (1965) and Guzman (1968); it is interesting to note that their models are general enough to enable segmentation of objects which are to some extent unfamiliar, and that in Roberts' case, further segmentation of the actual object into familiar components takes place.

Complexity and variability

Image complexity tends to be a serious problem in articular analyses, and there has been a marked tendency to investigate API tasks in which the images are more or less adequately represented by line segments, and in which such segments are relatively easy to derive by preprocessing. Fine detail is usually suppressed by spatial differentiation and thresholding or by binary quantization of the input image, so that the image components used in later processing are simple and not too numerous or detailed. With regard to the number of possible class characterizations required in pollen-interpretation, some parallels with fingerprint characterizations (Hankley and Tou, 1968) can be seen concerning the number of classes and the fine differences. Selection of the order of comparison of known classes with an input object, on the basis of structural and other properties of the input,
is suggested by Grimsdale et al. (1959) and Hankley and Tou (1968): Evans (1964) and Roberts (1965) search for topological equivalences before attempting further comparisons, but no use is made of hierarchic class structure in reducing the information cost of representing large numbers of classes.

There is an evident concern in this approach both for component parts of objects and for the organization of these, but (as noted earlier) the components of interest have usually been restricted to line-like elements, and the relationships recovered have primarily been those obtaining between connected line segments at their points of coincidence, rather than between the segments as wholes. Partly as a result of this limited range of relationships, the depth of structuring is very limited, and organizations of constructs (rather than of simple elements) are usually not considered, so that the levels of structure exhibited in the micrograph of Thrift (Fig. 2-5a) could not be captured. Recovery of the organization of superimposed components has been considered by Evans (1964) for the case of overlapping geometrical figures; with regard to levels of resolution, however, some of the troubles with "gapped" contour lines experienced by Roberts (1965) and Forsen (1968), seem to be related to their processing the input images at only one resolution level. Superimposition of components such as Zinnia's spines on a spherical base (Fig. 2-10), does not seem to have been considered; the various model components comprising Roberts' (1965) objects have equal status, rather than that of base component plus detail.

The potential difficulties regarding the large number of possible relationships between image and object components have not been evident in the investigations discussed. This is primarily because the relationships considered have typically been between only connected components, and secondarily because the number of components involved has been relatively small as a result of inherent image simplicity or preprocessing.

The concern for connectivities of components rather than more basic 2-D relationships, makes some of the techniques employed rather sensitive to any pictorial or object-world "noise" which alters connectivities. Extensive noise cleaning and gap-filling preprocessing may be invoked in order to reduce this sensitivity, but the effort in devising such ad hoc processes seems misdirected
consideration of various levels of resolution and of relationships more general than connectivities, is surely more important.

Object variations which preserve structural similarities are, to the extent that significant structure is recovered, dealt with satisfactorily. With regard to added or deleted components, Roberts (1965) ignores any isolated line segments which do not form polygons, but his and other schemes fail to recover structural similarities and differences when examining objects which have extraneous or missing components. Simple transformational relationships (such as scaling, rotation, and perspective projection) between regularly deformed or transformed objects are recovered by Evans (1964) and Roberts (1965), but further research is needed before more complex or less regular deformations and transformations can be accounted for.

Problems with respect to the open nature of image behaviour have not been directly confronted because of either the restricted images considered, or the concern for simple features (e.g., contours or lines) in more complex inputs.

With regard to processing of unfamiliar objects and formation of initial characterizations, both Grimsdale et al. (1959) and Roberts (1965) demonstrate that effective descriptions of unfamiliar objects can be formed, provided that these objects are composed of known components whose significant relationships to each other are those recovered within the analysis procedure. This approach to characterizing new objects is intuitively more satisfying than "learning" via averages, as employed in template matching and property list schemes.

Ambiguity

The fundamental problem of 2-D to 3-D ambiguity is avoided either by ignoring the 3-D field, or by choosing 3-D fields which are so restricted that any contours in their 2-D images correspond to the junctions of plane surfaces. Ambiguity of association (or segmentation, depending on the point of view) is exhibited with respect to the many possible groupings of lines and surfaces in the images considered. The importance of implicit or explicit "models of the world" in resolving these ambiguities has been convincingly demonstrated. Partly because of the limited depth of structuring considered, inconsistencies (wherein incompatible interpretations of different areas of the input image are derived)
either do not arise, or are detected before further structuring takes place, so that no re-interpretation is required. With more complex images, inconsistencies are usually not so readily detected and resolved. Only one resolution level has been considered to date, and no question arises of this form of ambiguity.

**Representation**

Effective representations of objects, inputs, and transformations have been developed for rather limited problem domains. Roberts' (1965) basic models of cubes, wedges, and hexagonal prisms are fixed with respect to a 3-D reference frame. By representing transformations for 3-D scaling, rotation, and translation, together with the perspective transformation, any instance of these models in an input image is viewed as the basic model plus these transformations, thereby generalizing over the changes (despite the fixed nature of the basic models) and indicating how expected orientations, sizes, etc. of the basic models might be accounted for in descriptions of instances.

The importance of effective models of the world in enabling successful interpretations is acknowledged in this approach. With other than relatively simple domains, however, there are difficulties in developing such models; for example, Guzman (1967) feels that "there are major conceptual problems to be faced if we are to find really good models for intricately curved surfaces".

To the extent that the models employed are appropriate to the input images, satisfactory descriptions of the inputs (or at least line-drawing abstractions of them) may be derived. In Roberts' (1965) case, the input image is described as a perspective transformation of an interpreted 3-D field; he demonstrates the generative adequacy of his descriptions on a graphic display.

The relationships between 2-D images and the 3-D fields portrayed have been represented as perspective transformation, but this is clearly not an adequate representation for fields which are characterized by surface shape, pigmentation, and illumination as well as by the locations of junction lines between surfaces.

Representations of unknown objects as articular descriptions (in terms of the machine's current knowledge of basic components and relationships) are employed by Grimsdale et al. (1959), but structural similarities with known objects are not employed in these descriptions.
3.4 Picture Parsing (Feder, 1966a,b, 1968; Miller and Shaw, 1968; Rosenfeld, 1968a,Ch. 10; Narasimhan, 1968,1969).

3.4.1 Outline

This general approach, which is also known as the "linguistic" (Miller and Shaw, 1968) or "syntax-directed" (Narasimhan, 1968) approach, is based on the concept of "parsing" (i.e., delivering a generative specification of) the given picture with respect to a (generative) model of the class of well-formed pictures, in much the same way that sentences in natural or artificial linear languages may be parsed with respect to a suitable grammar (i.e., a set of syntax rules). Just as ordering and relationships between words are important in linear languages, and form an integral part of the parse, so too are pictorial relationships seen as important in picture parsing.

Analogously with linear languages, formulation of an explicit model for the desired class of pictures is usually in terms of a "picture language" specified by a "picture grammar"; any valid "pictorial expression" (Clowes, 1968) in this language is thus a well-formed picture. The syntax or composition rules specify the ways in which the basic pictorial elements or "primitives" (cf. phonemes in a spoken language) may be put together to form constructs, and the ways in which these constructs may be put together to form more complex constructs, and so on up to the level of the figures of interest (and to the picture itself as a composition of these figures). The aim of the parsing procedure is to find a valid sequence of composition rules which, if followed, produces the given picture; such a sequence forms a hierarchical "structural description" of the input picture, in which the more complex constructs may be labelled as instances of the figure classes of interest. An adequate picture sub-grammar for each class of such figures generates all and only the acceptable figures in that class and assigns them an appropriate structural description. By merging all such sub-grammars, a grammar for the given picture class may be formed. Early exponents of this approach were Minsky (1961, sect. II. H), Kirsch (1964), and Narasimhan (1962). Work in computer graphics (Sutherland, 1963; Stanton, 1969a,b) has influenced the formulation of the picture parsing approach, with the representations of pictures
as hierarchical compositions of basic elements being common to both areas.

Paralleling the idea of a syntax-directed programming language analyzer, which accepts a syntactic specification (in an appropriate meta-language or formalism) of a given programming language as data, and then proceeds to parse a string of text according to that syntax (Cheatham and Sattley, 1964), there is the concept of a syntax-directed picture analyzer, which accepts a syntactic specification of a given picture class, and parses input pictures with respect to this specification (Evans, 1968a, b). Much of the research in picture parsing has concerned the development of appropriate formalisms (or meta-languages) which facilitate the writing of grammars for fairly general classes of pictures, and in the implementation of analyzers which accept the grammar rules written in these formalisms (Evans, 1968a,b; Shaw, 1968; Stanton, 1970). To the extent that these formalisms allow natural specification of fairly general classes of pictures, they reflect something of the designers' intuitions regarding pictorial organization. The formalisms were initially based on phrase structure grammars (PSG's) (Narasimhan, 1966; Ledley, et al. 1965; Shaw, 1968), but more powerful notations have been developed (Clowes, 1968; Evans, 1968a,b; Stanton, 1970). As in programming language analyzers, both "top-down" (Shaw, 1968) and "bottom-up" (Narasimhan, 1962) picture analyzers have been reported. It will be observed that in this approach, the analysis is completely and directly under the control of the grammar rules, whereas in articular analysis, these rules are used to augment rather than to direct the analysis.

The primitive pictorial elements are usually conceived of as simple shapes whose internal structure is of no direct interest (Kirsch, 1964) and, as such, are seen (Shaw, 1968) as being suitable candidates for simple recognition techniques such as template matching or property list. In most investigations, the primitives have been line-like elements (Evans, 1968a) or boundary curves (Ledley, et al. 1965), although Shaw's (1968) line elements can represent more general shapes. Narasimhan and Fornango (1964) report some preliminary experiments in labelling classes of pictures "in which shaded regions play a defining role". Lipkin et al. (1966) introduce the idea of an "iconic microgrammar" whose basic primitives are simple line-like elements and more general shapes.
which may assume arbitrary gray level and texture. Clowes (1969a) takes a rather
different view of primitives and regards "distinct positions in a two-dimensional
array" as the primitive elements, and locations of such positions (relative to a
reference frame) as the primitive relationships. The relationships considered
to date have usually been those obtaining between line segments, and even then
have been primarily strictly local relationships concerning connectivities of line
segments, rather than general 2-D juxtaposition as suggested by Kirsch (1964).
Because of this reliance on connectivity, difficulties may be experienced with
pictorial or object-world "noise", and extensive "noise-cleaning" preprocessing
may be invoked (Narasimhan, 1964).

The concept of the "semantics" or the meaning of pictures is introduced
by Kirsch (1964) and elaborated by Clowes (1968, 1969a) and Stanton (1969a, b).
They see the semantic component of pictures as concerning the "mapping" relation-
ships between the picture itself and the object-world situation portrayed in it.
Alternatively, Shaw (1968) sees the semantic component of pictures as concerning
the metrical properties of the various pictorial elements and constructs. The
former view of semantics is intuitively more satisfying, and provides a framework
in which it is possible to view the relationships between a 3-D field and its 2-D image.

3.4.2 Examples

Narasimhan (1962, 1964) pioneered the application of linguistic techniques
to picture interpretation problems, and has considered the analysis of bubble-chamber
photographs and hand-drawn characters. His techniques are directly applicable
only to figures which can be characterized by connections between "road"-like
elements and in which figures with unconnected sub-parts do not appear.

Kirsch (1964) considered the relationships between two or more descriptions
of the same physical object or event (e.g., syntactic descriptions of a circuit
diagram and of some English text describing the diagram). He argues that
pictures have a syntactic structure analogous to that of English text and gives
an example of a context sensitive picture grammar which generates all and only
45° right-angled triangles, but this grammar fails to provide a satisfactory
structural description of the generated triangles (Clowes, 1968).
The work of Ledley et al. (1965) in syntax-directed chromosome analysis, clearly illustrates the basic concepts of the picture parsing approach. Individual chromosomes on a microscope slide are isolated and their boundaries traced. Boundary segments (whose lengths are fixed to a preset number of traversed matrix cells) are classified into one of five basic curve types, and the string of segments forming the outline of a single chromosome is parsed by a syntax-directed analyzer, according to a grammar for two different types of chromosomes — submedian and telocentric. Objects which are well-formed according to this grammar have their areas and arm lengths measured to enable later pairing-off. This scheme cannot cope with touching or overlapping chromosomes; these problems are investigated in later work (Ledley et al., 1968; Hilditch, 1968).

It seems that the same chromosome shape could easily be assigned different descriptions using this technique, because the positioning of the boundary segments relative to the outline will vary as the chromosome is rotated (the uppermost part of the outline being chosen as the start of the first boundary segment), and the actual lengths of the segments could vary by a factor of \( \sqrt{2} \) depending on whether the adjacent boundary cells in the image were diagonally or directly adjacent. Clowes (1968, p. 19) points out that Ledley's chromosome grammar is not generatively adequate because on the basis of this grammar alone, one can generate shapes which are not chromosome-like.

Clowes (1968, 1969a) has investigated formalisms for picture grammars, and the manner in which these reflect human perception of pictorial relationships and organization. The type of formalism currently being developed by Clowes and his co-workers seems among the most powerful and natural (in the sense that a wide range of pictorial organizations can readily be expressed) yet devised, and an analyzer has been constructed (Stanton, 1970). As well as involving the capacity to name or classify, recognition is seen as implying a capacity to form and compare descriptions of (perhaps previously unknown) objects, and to identify the aspects in which objects differ and in which they are the same. Descriptive adequacy is judged not according to accuracy of classification, but rather insofar as a description manifests such similarities and dissimilarities, and is generatively adequate. Clowes identifies the notion of articulation of pictures into
parts as the most basic feature of human apprehension of shape, and regards the identification of pictorial relationships on which such articulations are predicated as a central problem. He suggests using ambiguity, anomaly, and paraphrase to "fleece out" these pictorial relationships (as advocated by Chomsky (1965) for the identification of linguistic relationships).

Clowes rejects PSG's as suitable formalisms, on the grounds that the only relationships explicitly exhibited in PSG's are "parts of" and "followed by", and that even these relationships are represented indirectly by the same relationships existing between the symbols of the meta-language. Instead, he proposes a notation in the spirit of, and isomorphic with, Chomsky's (1965) transformational grammar notation, in which (i) composition rules can be conditional on the satisfaction of explicitly specified relationships between the picture parts involved in the composition, (ii) information regarding any attributes of the constructs can be appended to their representation in the structural description (such information possibly being used in determining the applicability of further compositions), and (iii) complex relationships may be specified in terms of simpler relationships. Clowes (1968) and Stanton (1970) give examples of other picture grammars (such as Shaw's (1968) and Anderson's (1968)) translated into their formalism.

Evans (1968a, b, c) has developed a formalism for the description of complex structured patterns, and a picture analyzer which accepts grammar rules written in this formalism, together with an input list of primitives, their attributes, and relationships to other primitives present in the picture to be analyzed. This list is assumed to have been derived as a result of suitable preprocessing. The analyzer produces structural descriptions of any patterns defined by the grammar rules and found in the input list. Evans' analyzer is written in LISP and his formalism mirrors LISP notation. A grammar for each pattern of interest consists of a set of rules, each of which specifies (i) the name of the construct formed, (ii) the constituent sub-patterns and any other contextually required patterns, (iii) a list of the various relationships between subpatterns which must exist for the rule to be applicable, and (iv) any attributes to be appended to a successful composition (such as attributes required for determining the applicability of some higher-order rule). Evans' notation is similar to that of Clowes (1968),
but is not as powerful in the sense that the relationships specified in (iii) above are built into the analyzer and cannot be specified by grammar rules in terms of simpler relationships. Evans (1968a, p. 3) claims that the approach of "emphasizing the generation and manipulation of appropriately-chosen descriptions is a highly promising avenue of attack on the machine processing of such complex patterns as aerial photographs and physics photographs".

Shaw (1968, 1969) has developed a picture description language (PDL), applicable to both picture analysis and generation, in which pictures are described as hierarchic compositions of a single type of primitive which can be thought of as an arrow \( \rightarrow \) with only two distinguished positions — a head and a tail. Any primitive will usually represent one of a set of basic picture patterns, which are considered to be the simplest pictorial units that it is convenient to recognize (or generate) independently. For picture analysis, Shaw thinks of these patterns as those recognizable by, for instance, Marill and Green's (1960) receptor-categorizer model (RCM). PDL is based on a restricted form of context-free PSG, but Shaw maintains that a reasonably large and interesting class of picture analysis and generation problems can be handled with this notation.

Each primitive has associated with it a "terminal semantic description" indicating the pattern class represented (e.g., an arc or a straight line) and other information such as the tail and head locations, arc radius, etc. Primitives are assembled into more complex constructs in one of only four ways:

\[
\begin{align*}
& (S_1 + S_2) \\
& (S_1 \times S_2) \\
& (S_1 - S_2) \\
& (S_1 * S_2)
\end{align*}
\]

with the new tails and heads as indicated; more complex entities are formed by applying these four concatenation operators in the same way to constructs. To broaden the descriptive capability of PDL, "null", "blank", and "don't care" primitives are allowed. Two more operators, negation ("\(\sim\)"), which interchanges tail and head, and superposition ("\(\cap\)"), which allows the same primitive or construct to be used in more than one composition, are introduced. This latter operator is required whenever more than two primitives or constructs join at a point, because the concatenation operators are only binary. The concepts of a "standard form" and "valid" PDL expressions are defined, and Shaw derives a number of properties of the concatenation operators, and theorems about PDL expressions.
Corresponding to the hierarchical structural description (expressible in tree form) of a picture in PDL, Shaw describes a hierarchic semantic interpretation (also expressible in tree form) with the terminal semantic interpretation as the basic nodes, but only the class names of the various constructs (i.e., no attribute values apart from head and tail locations) are appended to higher-level nodes in the semantic hierarchy.

Shaw has developed a top-down (i.e., goal-directed) analyzer which accepts (i) "primitive recognition" routines (e.g., RCM) for detecting instances of the basic pattern classes in the input image, and (ii) a PDL grammar describing allowable compositions of primitives and constructs, and the names to be appended to these compositions. The analyzer directs the recognizer routines over the picture trying to find instances of the prescribed objects (i.e., "goals"), and if successful, delivers a structural and semantic interpretation of the picture. This analyzer and the associated recognition routines have been successfully applied to the interpretation of bubble-and spark-chamber photographs.

A number of specific comments concerning Shaw's system are relevant here. Despite the completeness and mathematical "nicety" of his theorems and corollaries, PDL has a number of intrinsic weaknesses, and certain classes of pictures and pictorial relationships can be described only awkwardly. Among the major shortcomings are:

1) In the semantic interpretations, the only attributes which constructs may assume are a name, and tail and head locations; in this respect his system is not as general as Evans' (1968a,b) or Clowes' (1968), and composition rules can be predicated only on component names and head and tail locations. As noted previously, Shaw's notion of picture semantics is at variance with other current views.

2) The only pictorial relationships exhibited explicitly with PDL are "parts of", "connected to", and possibly "relative position". Other relationships and transformations such as "above", "inside", "rotated", etc., can be expressed only with difficulty, if at all.

3) The binary nature of the concatenation rules can lead to decidedly awkward articulations of even simple figures; see Fig. 3-3.
Description in PDL of a 4-node directed graph

given the primitives

\[ "a" = \rightarrow, \quad "b" = \downarrow, \quad "c" = \bigcirc, \quad "d" = \rightarrow \]

\[ (((b^i + a) \times (((/b^i) + d) + (/b^j))) \times ((a + b^j) \times c)) \]

Notes: (i) The multiple instances of primitives "a" and "b" are distinguished by their different semantic interpretations.

(ii) Superscripts are employed to allow cross referencing.

FIGURE 3-3 Example of an Awkward PDL Expression (after Shaw, 1968)
3.4.3 Evaluation

There is currently a great deal of interest being shown in the picture parsing approach; E. M. Braverman feels that "we are now entering the era of the linguistic approach to pattern recognition" (Nappelbaum, 1967), and Rosenfeld (1969, p. 163) expects the syntax-directed approach to API "to become a major theme in picture processing research during the coming years". Although the picture parsing approach is appropriate for richly structured well-formed pictures, major problems arise with regard to the highly variable nature of naturally given images. This class of pictures seems to be inherently "ungrammatical", and problems analogous to those of accounting for "degrees of grammaticalness" in grammars for natural languages are encountered. Minsky (1968) warns that while parsing with respect to a set of rules is a good approximation for simple cases, it is "a very dangerous idea to fix on so firmly that the problem is not thought about in any other way". It is possible that the main contribution of picture parsing research will eventually be with respect to analysis of restricted or synthesized pictures such as those employed in man-machine graphical communication, but even in this application, some serious questions have been raised as to whether the production of structural descriptions (via analysis with respect to picture grammars) should be the main objective (Stanton, 1970). In so far as this approach emphasizes pictorial relationships and organization, it will no doubt be a major influence in API research, but some of the problems encountered with real-world images result from the picture parsing framework itself, as well as from the PI tasks.

Considerable generalization of the picture parsing approach is required if it is to be successfully applied to complex PI tasks (Macleod, 1969a); an evaluation of this approach with respect to such tasks follows — because of the current interest in picture parsing, this evaluation is reasonably detailed.

3-D fields

No method of accounting for such fields within the picture parsing frame-

---

1 From the context of this remark, it appears that Rosenfeld is referring to articulatory analysis in general and not just picture parsing.
work has been demonstrated; the concept of "mapping" espoused by Clowes (1969b) might be applicable here, but the mapping rules required must inevitably be ambiguous, complex, and context sensitive, as a result of these attributes of the 2-D \(\rightarrow\) 3-D relationships. The difficulties encountered with picture grammars, relative to string grammars, do not augur well for grammars of structured 3-D objects. McCarthy et al. (1968, pp. 332-333) point to the problems of sensitivity to image imperfections and to the difficulty of describing occluded and degenerate projections, and suggest that considerable generalization of the linguistic approach is required before it can be effectively applied to 3-D scene analysis.

**Figure extraction**

As with articualar analysis, either initial segmentation is assumed, or the images processed have been relatively easy to segment. Segmentation of structured objects from complex surroundings, by trying various compositions until a known object is formed, has been demonstrated by Evans (1968a,b), but most other problems of missing boundaries, segmentation on the basis of complex and structural properties, etc. have not been addressed.

**Complexity and variability**

The complexity of natural images presents almost insuperable problems with straightforward parsing techniques, in which the number of operations required increases exponentially with the number of individual basic components in the data being parsed. This problem might, however, be mitigated to some extent by detecting simple unambiguous constructs formed from basic components, in an initial processing phase similar in principle to the "recognizer" phase in some syntax-directed compilers. Another approach is to reduce the image complexity by drastic preprocessing, but this is not acceptable if fine detail can be important. With the very large number of possible classes, the goal-directed (top-down) parsing approach requires hierarchical organization of the most specific classes into more and more general super-classes, if much fruitless backtracking is to be avoided; similar remarks apply to "bottom-up" parsing, but the organization required here is different. The hierarchical organization of classes is reflected in the use of common rules of composition in more than one class. This multiple use of rules permits the addition of characterizations of new and complex classes without much
additional specification.

Consideration for relationships and organization of image and object components is central to picture parsing, and complex hierarchies readily fit this framework. Unfortunately, the range of relationships considered has been very limited, and the comments made with regard to this shortcoming of the articular analysis approach apply here. Through restriction to connectivity relationships and simple images, the potential problem regarding the large number of possible interrelationships between components has not arisen, but an important aspect of picture grammars is that they do specify which relationships are relevant.

Perhaps the major weakness of currently formulated linguistic models is their sensitivity to image and object-world imperfections and variability. The inclusion of recursive syntax specifications allows finite picture grammars to generate an infinite variety of pictures, but there are problems of context sensitivity — a string of text can readily be "opened up" for the insertion of additional material, but such material cannot always be inserted in a picture without disturbing the spatial relationships of components already present. As in articular analysis, imperfections which alter connectivities of line segments may preclude effective interpretation. Kovalevsky (1968, p. II46) considers that linguistic models are "not applicable to the problem of describing pictures distorted by random interferences", but does not elaborate. In picture parsing work, attempts to improve the input pictures (e.g., by "thinning" and "gap-filling" (Narasimhan, 1964) before parsing, are common, although Shaw (1968) directs "primitive recognizers", which are presumably noise tolerant, over the input during parsing. Solomonoff (1966, p. 1695) feels that more powerful grammars should eventually be able "to generate the noise as well as the signal", but that tentative removal of noise, prior to analysis with respect to an "approximate" noiseless grammar, might be a more practical initial approach. Clowes (1969b) presents an interesting attempt to try to formulate the relationship between the idealized patterns described in picture grammars and the imperfect representations of them in typical images, as a "mapping" problem; the implications of this formulation, with regard to coping with image variability, have yet to be explored.

One of the main sources of difficulty is that current parsing algorithms are essentially decision trees, and one wrong decision prevents satisfactory
analysis. Evans (1968a) suggests the use of decision theory to reduce the chance of such errors. There is a clear analogy here with error-recovery problems in syntax-directed compilers — in trying to provide syntax rules for erroneous constructions, it is impossible to anticipate all the quirks of programmers or input images.

Current schemes can readily cope with object variations or deformations which preserve structural similarities, provided that connectivities do not change; added components tend to be ignored, but in so far as missing components change the organization, these usually prevent effective analysis. Freiberger and Grenader (1968) suggest possible deformation models (based on ideas from statistical linguistics) for use in the analysis of deformed images.

A conceptual difficulty arises with regard to attempts to apply the picture parsing approach to naturally given images. It was argued in Section 2.3.3 that such images form an "open" set not easily describable within a "closed" formalism. The concept of pre-chosen primitive elements (even matrix cells, as advocated by Clowes, 1969a) does not seem very useful here, and other difficulties arise with regard to description of gray scale behavior. Rosenfeld (1968a, Ch. 10 pp. 20-21) refers to the difficulty of formulating a grammar for visual texture. The problem of deciding on the truth or falsity of relationships between image components, because of the "continuous" nature of image properties, was alluded to in Section 2.3.3. This difficulty seems so basic that the question arises as to whether there is an alternative approach, but current picture grammar rules are solidly based on relationships which are simply true or false.

Unfamiliar objects are not accounted for in picture parsing schemes, except to the extent that variants of known objects (e.g., sub-median chromosomes of different arm lengths and ratios) are allowed, and previously unencountered productions of the grammar rules are successfully parsed. Class characterizations are provided by the designer rather than by "learning" new grammar rules.

Ambiguity

The question of 2-D to 3-D ambiguity has not been considered. The picture grammars may be ambiguous, in that a given input can be successfully parsed in more than one way. Indeed, images are themselves often ambiguous.
and adequate grammars should be capable of producing alternative parsings, by assigning different structural descriptions to the same "surface" pattern. Evans (1968a, b) and Shaw (1968) have demonstrated the use of models (as represented explicitly in picture grammars) to resolve ambiguities of association. Because the analysis process is directed by the grammar rules, and is essentially a sequential rather than parallel process, inconsistencies of interpretation in adjacent areas do not arise; "backtracking" usually takes place because no further progress can be made following the current path, rather than because of global inconsistencies.

**Representation**

Classes of objects are characterized by explicit models embodied in grammar rules, and there has been much concern with the development of suitable formalisms, but only for a limited class of pictures (typically binary and line-like). Some picture languages are intended as languages of discourse about pictures (for man–machine communication) as well as drivers for analysis or generation procedures. Narasimhan (1968, p. 46) observes that "there is a danger that this intense preoccupation with a rather narrowly delimited class of pictures might result in the development of tools, notations, and procedures which are technique-oriented rather than concept-oriented". Relationships between components are largely independent of external frames of reference and the main problems here are encountered in the detection of basic elements by RCM or other techniques, thus suggesting that pictorial relationships are still important at even this low level.

The importance of object-world knowledge in the interpretation of pictures has been emphasized by several workers, but much further work in the use of such knowledge during analysis is required. Transformations between domains such as the 3-D field, its 2-D image, and the digitized image, might be representable as mapping rules, but no work along these lines has yet been reported.

There is a concern in picture parsing for explicit and reasonably complete descriptions of input images. These descriptions are representable as successful paths through a grammar, together with any additional metrical or structural properties of the primitive elements or constructs. The completeness of descriptions, as exhibited by their generative adequacy, has been demonstrated by Shaw
(1968) in his application of PDL descriptions to picture generation.

There is a degree of organization of knowledge about object classes as represented in grammar rules, but accessing relevant information tends to be on a trial and error basis.

3.5 Discussion

The status (both collective and individual) of the four general approaches examined above (in Sections 3.1 to 3.4), is summarized below with respect to the difficulties outlined in Chapter 2 and the problem of designing a reasonably general API system. With many of the difficulties identified, little progress has been made in these approaches. This situation is partly a result of the interdependent nature of the problem areas, such that it is not easy to develop solutions for one area independently of the others, and partly a result of the extent and intractability of the overall API problem.

With regard to the 3-D nature of the fields portrayed in input images, attempts to interpret in purely pictorial terms have been criticized already. Despite the difficulties involved in obtaining 3-D information from a 2-D image, the objects of ultimate interest are 3-D and should be characterized in this domain; effective interpretation is unlikely to be achieved without consideration of the 3-D fields. Roberts (1965) has demonstrated interpretation in 3-D terms, but his techniques are applicable only to very restricted images and objects.

Difficulties of segmentation and object isolation are related to problems in other areas (e.g., recovery of structural information). Attempts have been made to compensate for missing or imperfectly defined boundaries via gap-filling operations which are based on purely local image detail, but these operations are information-destructive, and are usually unsatisfactory in that they group (by filling gaps) image elements which (in the global context) should not be joined, and fail to group elements which should be. Grimsdale et al. (1959) have demonstrated image segmentation on the basis of simple articular properties, but serious problems arise with regard to segmentations that depend on metrical and/or articular properties which have not been "built into" the interpretation system. The range of articular properties, in particular, is effectively unbounded —
consider, for example, the properties characterizing the surface structure of Thrift (Fig. 2-5) and Morning Glory (Fig. 1-2).

Roberts (1965) and others have demonstrated the utility of suitable object models in effecting accurate segmentations (of both the image and the perceived 3-D field) despite confusing surroundings. Where less-well-defined objects can be present, rather more comprehensive models (of the type implicit in Guzman's (1968) system) are required.

Problems arising from the mass of raw image information have been attacked by means of drastic information-reducing preprocessing operations. As is observed earlier, it is usually not possible to say at the start of processing which detail will be important and which will not, so that any reduction of detail at this stage may involve deletion of information which will be required in subsequent processing.

The template matching and property list approaches do not account for image and object organization. Concern for structural information is, however, evident in the articular analysis and picture parsing approaches, but the range of relationships considered has been very limited, and recovery of complex unfamiliar (i.e., novel) organizations has not been attempted.

With regard to the inherent variability of naturally-given images and objects, one approach has been to try to reduce variations by standardizing the input images in a preprocessing stage. With complex images, in which the objects of interest are not readily isolated, the improvement which may be effected in this manner is very limited, and the dangers of standardizing variations by simply deleting the variable information have been pointed out above. A second approach is to try to allow for variations in objects' class characterizations. In this case, problems arise analogous to those encountered in trying to provide a syntactic specification of erroneous constructions for the purpose of error control in syntax-directed compilers — it is not possible to anticipate all possible variations, and the class characterizations employed may be either too strict (in that they do not allow instances which are clearly class members), or too loose (in that they admit instances which should be rejected), or possibly both at once.

Resolution of ambiguities by the use of implicit or explicit "models
of the world", has been demonstrated by Roberts (1965), Evans (1968a,b), and others. In these cases, however, either the models employed allow only one coherent interpretation of the images (this being possible only if there are severe restriction on the objects depicted), or alternative interpretations are simply listed. There are inherent ambiguities in the class of images considered herein, but the "best" interpretation is required rather than a listing of alternatives (with reasonably general 3-D fields, an infinite set of coherent interpretations is possible).

Suitable representations of the various types of knowledge are essential, but such representations have been demonstrated only for simple images (e.g., Evans, 1964). Generalization over external reference frames has been demonstrated via Euclidean transformation and a concern for inter-component relationships.

Turning to individual consideration of the four general approaches, in the context of complex PI tasks such as pollen-interpretation, the template matching and property list approaches suffer from serious shortcomings and cannot seriously be considered as a basis for a general API system.

The articular analysis approach has not as yet been clearly formulated and delimited, and it is therefore difficult to make objective statements about its limitations. Some of the concepts exhibited (such as consideration of 3-D fields and of image and object organization) are very important, but few more-specific guidelines for the design of general API systems are given. This approach seems to be capable of being developed to the point where it can be applied to complex PI tasks, but design of a successful system relies upon such development.

The picture parsing approach is more clearly formulated and delimited than articular analysis, and some interesting work has been reported (e.g., Shaw's (1968) synthesis of picture parsing and property list techniques). Although an overall design procedure is indicated, there are severe problems in developing suitable grammars for other than very simple objects, and it is difficult to envisage how this approach might be applied to complex images.

In view of the problems which remain to be addressed and/or solved, none of the four approaches examined can (as currently formulated) be regarded as suitable for PI tasks and images as complex as those encountered in pollen analysis.
If such tasks are to be considered for API, there is a clear requirement for the formulation of a more adequate model.
Chapter 4

TOWARDS THE FORMULATION OF AN ADEQUATE MODEL FOR AUTOMATIC PHOTO-INTERPRETATION

The need for API models more suitable for complex PI tasks than those currently specified, has been established in the previous chapter. Contributions towards the formulation of an adequate model, and an account of some initial attempts to come to grips with the problems involved, are given below.

After preliminary discussion (Section 4.1), a conceptual model for API is sketched, and certain of the components are examined in detail (Section 4.2). An outline of the envisaged operation of the model, with respect to the initial processing stages in pollen-interpretation, is also included (Section 4.3).

4.1 Preliminary Discussion

The nature of the API task, and the approach followed in the model's development, are discussed below. Further background information is provided by the discussion in Section 3.5 of the remaining problem areas.

4.1.1 The Nature of the API Task

In Section 1.2.2 it was remarked that "interpretation" can be regarded as the process of coming to "understand" the sensory input. The task of an API system is thus viewed herein as that of explaining (and thereby imposing an organization on) an internal encoding of a given image, in terms of the knowledge represented within the system. This explanation is in terms of an underlying 3-D field, a projected image, and a digitization process. Perception, when viewed as the process of referring sensations to their external causes or of organizing sensory impressions (Sartain et al., 1962, preface p.vii), is equivalent to the above view of interpretation. There are, of course, limitations on the range of acceptable explanations or organizations. The most important requirement derives from the principle of inductive inference — "let the machine make..."
the simplest hypothesis from which it can derive all the accumulated information at hand" (Andrews, 1962, p. 11). Simplicity is not an easy attribute to define, and has many manifestations, but there is a close relationship between simplicity and economy of description.

A second important requirement is that the organization assigned be both internally and externally consistent. Interpretation takes place on the basis of reduced cues and is always fallible — Gregory (1966, p. 12) observes that "a perceived object is a hypothesis suggested and tested by sensory data". Formulation of the API task in terms of consistency of organization rather than of fidelity (so that, for example, Ames' room should be perceived as rectilinear) was suggested in Section 1.2.2, but as external information from other senses or from separate images becomes available, then an externally consistent interpretation becomes more likely to agree with reality, and the credibility of the hypothesized organization improves. As discussed in Section 2.5, effective interpretations can be derived only if extensive knowledge of the world is represented within the API system, and if relevant information can be selected and used appropriately.

4.1.2 Approach Taken in Development of the Model

With symbolic input images, the ultimate meaning of any symbols is that assigned by the human perceiver, as there is no direct link with extra-pictorial situations or concepts — an effective interpretation system must, therefore, be based on consideration of human perceptual processing. On the other hand, with naturally-given images, it is possible to conceive of a machine which derives satisfactory interpretations of input images without there being any consideration of human perception (the interpretations being checked against actual underlying 3-D fields rather than against human perception of the given images). The attitude adopted herein, however, is that the problem of designing an adequate API system is so formidable that any assistance which can be gleaned is worthwhile, and that because biological (particularly human) visual perception is so remarkably successful, anything which can be learned about the perceptual principles and processes employed, is likely to be relevant.
Studies of biological processing of visual information have included both neurophysiological and psychological; neurophysiological studies (Hubel and Wiesel, 1962) have been concerned primarily with the early stages of visual information processing in animals such as cats and monkeys, whereas psychological studies (Gibson, 1950; Kolers, 1968) have concerned the overt response of complete animal or human perceptual systems. Unfortunately, there is an apparently broad unknown area of visual information processing between these two types of studies; it is hard to extrapolate from perceived outputs to assumed input processes, and vice versa. Research in both areas of study is far from complete; Hunt (1962, p. 105) remarks that "how humans will structure any particular stimulus is an unsolved problem of perception". Consideration of reported neurophysiological and psychological studies has therefore been supplemented by a study of human perception of various drawings and image areas, and of the similarities and differences between original images and versions reconstructed from memory. The difficulty experienced in understanding the results of such experiments has been commented on by Wolters (1933), who suggests that by studying instances in which perception is uncertain or grossly at fault, we can judge how correct perception is brought about.

The development of the model has, of course, been influenced by previous research in API, in which many useful observations and suggestions have been made, and some of the problems to be faced have been identified; of the four general approaches examined in Chapter 3, that of articular analysis has been the most influential here.

Concerning the historical development of the model, the importance of interpretation in 3-D terms and of image and object organization, was underlined by an initial examination of many pollen images, as a result of which a very rough model was formulated. The difficulties involved in deriving 3-D information from a 2-D image were clear at this stage, and in attempts to simulate some of the processes envisaged, this aspect was neglected, the recovery of image organization being the main concern — see Chapter 5. There has been an interaction between the model's conceptual development and the simulation attempts, in that the failures and successes of the simulations have led to revisions of the model,
particularly with regard to the processes proposed for dealing with the mass of raw image data, with the variability of image detail, and with image organization covering a range of resolutions.

4.2 Outline of a Conceptual Model for API

This section sketches the conceptual model and discusses the envisaged components and processes. As a result of the magnitude and difficulty of the task of developing an adequate API model, the model presented here is far from complete, and much further clarification, refinement, and elaboration is required before a successful API system can be implemented. This model serves primarily as a basis for further research, but also as a framework in which to present observations to be accounted for in the formulation of more complete models, and in which to suggest possible strategies for coping with problems involved. Because of the preliminary nature of the current study, there is little consideration for eventual implementation techniques (but there is of course concern that the proposed components and processes should be implementable).

The model's development has been heavily influenced by consideration of human visual perception, and the pictorial examples given are in these terms. As a result, it is anticipated that the model may have implications with regard to human perception (although it was not developed with this specific intention).

One principle of the model is that the underlying 3-D fields should be explicitly considered. Thus, instead of shadows "shredding" silhouettes and hindering processing (as in Holmes (1966) — see also Fig. 4-1), they should assist interpretation through helping to indicate the 3-D shape of the shaded objects.

A second fundamental concern is for image and object organization, particularly with regard to the processes required for characterizing novel organizations and effecting the type of discrimination indicated in Fig. 4-2 (without the particular organizations involved having been pre-specified). Recovery of novel organizations is seen as being based on the application of general rules of organization (defined in terms of relationships between the components being organized) of the type studied by Gestalt psychologists.
Note that the left- and right-hand sides of the two regions shown differ in the manner in which their constituent parts are organized.
By searching for simple relationships and constraints in letter sequences, Simon and Kotovsky (1963) have been able to recover possible organizations of novel sequences, and thus predict continuations. If two sequences with different internal structures were concatenated (cf. adjacent regions in Fig. 4-2), Simon and Kotovsky's program could readily be extended to effect structural segmentations. Within the model, formation of appropriate metrical and articular properties (thereby allowing 2-D and 3-D segmentations on the basis of these properties), depends on the detection of similar constraints and relationships in the given image and the perceived 3-D field.

The basic premise here is that given sufficient external information about the way in which the discretized representations are related to continuous images, the way in which these images are related to underlying 3-D fields, and the nature of the possible 3-D fields, then it is possible to suggest:

(i) a continuous 2-D image organization underlying a given discrete representation; and

(ii) a continuous 3-D field underlying this 2-D image, with the inherent 2-D $\rightarrow$ 3-D ambiguity being resolved by choosing the simplest coherent interpretation in terms of the knowledge represented within the API system.

Given a satisfactory initial image representation, step (ii) above is much more complex and of more interest than step (i). It will usually be assumed below that the image itself is available for direct examination, except when some comment relevant to its machine representation is called for.

The overall approach in the model is that of:

(i) iteratively processing local areas of the input image (over a range of resolutions);

(ii) imposing the locally simplest organization on these areas, in both 2-D and 3-D terms;

(iii) reconciling the organizations imposed at various resolutions and in adjacent areas; and

(iv) gradually building up a coherent interpretation of the image (as a
projective transformation of a 3-D field) in a number of small steps (changing local organization as required to resolve inconsistencies), so that a single "best" global interpretation is eventually achieved.

A block diagram summarizing the model's operation is given in Fig. 4-3; it must be emphasized that the proposed components are rather interdependent and cannot readily be represented as clearly delimited boxes with few inter-connections.

Processing is seen as being based on detail observable in a concentric set of arrays, covering a range of resolutions, which can be indexed over the input image; the coarsest array covers the whole image, while the finest array encompasses only a small sub-image. The image is regarded at the start of processing as an undifferentiated and unorganized whole (despite its probable representation as a matrix of discrete gray levels). The first stage of processing is to take a global view of the image with the coarsest arrays, looking for detail (such as spots and contours) which calls for closer attention (i.e., a local view) with the finer arrays. Restriction of the area of attention necessitates iterative processing, but helps reduce problems associated with the mass of raw image information — see Section 4.2.1.

Having attended to a small local area, various features of this area which suggest segmentation of components (e.g., image contours), groupings of previously segmented components (e.g., proximity and/or similarity), or interpretations of the image in terms of an underlying 3-D field (e.g., suggests one object passing behind the other — see Section 4.2.7), are searched for. Thus, in attending to an image area such as , boundary constraints suggest segmentation into three bars, which are later seen as a group because they are joined. At later stages of processing, when a perceived 3-D field is gradually being built up, similar features of this 3-D field are also searched for. The features searched for at this stage of processing are those which suggest fairly general organizations (i.e., not according to specific experience), and are related to the laws of sensory organization studied by Gestalt psychologists — see Section 4.2.2. The organization-suggestive features detected sometimes reinforce, in that they suggest the same image organization as each other, but they frequently conflict,
G
INFORMATION STORE ("MODEL OF THE WORLD")

Knowledge of the general nature of the three domains* and of the relationships between these

Knowledge of specific classes of objects

* i.e., the 3-D fields, 2-D images, and digitized images.

H
Is any progress being made in the current local area?

Yes

No

A
Select a new area for attention (a low resolution global area initially, then high resolution local areas, and (as abstraction proceeds) finally back to global areas)

B
Search for organization-suggestive relationships within the raw or partially structured 2-D image or the perceived 3-D field

C
Select from the alternative organizations suggested, and elaborate and evaluate the selection

D
Abstract from detailed local organization and update the image representation

E
Has overall progress come to a standstill?

No

Yes

F
Has a satisfactory global interpretation been derived?

No

Yes

STOP

FIGURE 4-3 Block Diagram of the Main Processes in the Model
in that they indicate alternative organizations.

The next stage of processing involves selection of a possible organization from any alternatives suggested — see Section 4.2.3. An important information selection technique is the preference for simple or for known configurations of 2-D or 3-D components, and one function of this stage is to access information in the model of the world regarding known configurations which are relevant to the area attended to. Accession of relevant information is facilitated by organizing information about known objects in such a way that component parts, and the ways in which they are related, are linked to the more complex known configurations in which they participate — see Section 4.2.8. Note that, in so far as any known specific configurations will have been derived from processing similar to that being performed with the current image, one of the possible organizations of the current image which have been suggested by the general rules, will probably correspond to the organization of any relevant known configuration found. Retention and subsequent modification of specific configurations, and detection of correspondences between retained and currently derived configurations, are parts of the processes of "learning" and "recognition", respectively.

Although prior knowledge of objects, or of inherently simple configurations, will usually not be exactly applicable to given instances because of object and image variability, such organizations can still be used effectively as "base components" in the description of instances, with any variations and necessary additional specifications being appended to this base (i.e., describing as "a schema + correction" (Oldfield, 1954)). Selection of the correct base component is important; for example, "B", "R", and "ρ" could each be described as a "P" + variations, but better base components (i.e., the concepts of a "B" and an "R") are evident for the first two characters.

Elaboration of the selected base component (in terms of additions, deletions, and transformations) so that it accounts for all the image detail observed in the current area, and satisfactorily explains any unexpected differences or associations, is attempted — see Section 4.2.4. It is at this stage that boundaries which are not directly represented in the image (e.g., because of occlusion or poor lighting) but are contextually suggested, are imposed, but only if the imposition leads to a
simpler interpretation and does not introduce inconsistencies — see Sections 4.2.4 and 4.2.7.

The detailed organization is then evaluated in terms of its relative simplicity, its internal consistency, and its agreement with organizations previously imposed on the surrounding context; these requirements of simplicity and of internal and external consistency, are major tools in the resolution of ambiguities — see Section 4.2.5. If the suggested organization is unsatisfactory, a closer look is taken at the current area, and a new choice of possible organizations is made, taking into account any additional information gained during elaboration and evaluation of earlier suggestions, and in re-examining the current area. It is primarily to reveal inconsistencies that detailed local descriptions are attempted, but because of the mass of fine image detail, retention of fully-detailed descriptions makes extreme demands on storage and processing capacity. To reduce these demands, and to enable larger and larger image areas to be attended to at successive stages of processing, an information abstraction phase is included, in which insignificant detail not required for subsequent processing is deleted — see Section 4.2.6.

The sequence of searching for organization-suggestive features and then selecting, elaborating, evaluating, accepting, and abstracting higher and higher levels of organization, continues in the current area until no further progress is being made — with information gained from subsequent processing, further progress with this area may later be possible. New areas are selected for attention, and the processing sequence described above repeats. These new areas may be either as yet unexamined, or already partially organized. Higher and higher levels of organization of the image data (including the perceived 3-D field) are attempted until a successful global interpretation is obtained, or overall progress comes to a standstill (e.g., with unfamiliar input material, processing may not be satisfactorily completed because of multiple ambiguities and inconsistencies which cannot be resolved). Thus, the initial global (low resolution) look at the image is followed by attention to a small local area, but as processing continues and image detail is organized and abstracted, so larger and larger image areas can be attended to (via the organized image representation) until finally the whole image can be comprehended at once.
More detailed discussion of the model's principal components and processes, and their interactions, is given below in separate sections. In addition to this discussion of the overall framework, a number of possible strategies for coping with many of the problems involved is suggested. Further research is required to determine how some of these suggestions can be incorporated in the overall framework.

4.2.1 Attention (Block A in Fig. 4-3)

Restriction of the "span of attention" of the API system to a limited number of previously perceived components (which may within themselves be relatively large and complex), and to a limited fine resolution ("foveal") area of the input image, is one possible method for coping with the mass of image detail. This type of "attention" can be considered as a form of information selection, as can abstraction (see Section 4.2.6). If most of the significant relationships in the class of images considered are between components which are proximate relative to their size, restriction on the span of attention may lead to a linear rather than a square-law growth in processing time with image complexity. Significant relationships between components outside the current span may thus be recoverable only through "memory" rather than directly, and proximity therefore becomes an important factor in grouping components.

There is usually so much detail present in the input image that it is worthwhile devoting a substantial proportion of the total processing to selection of interesting areas and detail. There are several factors determining which detail is interesting and which is not; for example:

(i) a previously examined area is usually uninteresting unless there is some new information relevant to this area, or some question about it or an adjacent area arises;

(ii) structured, unique, distinct, and/or large detail is preferred\(^2\); and

---

\(^1\) For N components, there are N(N-1) possible related pairs, if relationships are not restricted to only proximate components.

\(^2\) Detection of structure or uniqueness of detail results from processing iterations (around the major loop) at coarse resolution, while a local area is being selected for attention.
(iii) frontal objects in a 3-D field are usually attended to first, so that detail at the bottom of an image, or occluding rather than occluded areas, may be preferred.

In this way, detail which is barely discernible but is adjacent to more obvious detail, or which supports or refutes a suggested organization, may be attended to, whilst other detail which is by itself more obvious, may go unnoticed. Note that because of the iterative nature of the processing, the order of selection of areas is not of prime importance.

The question of attention arises with respect to both the raw and the organized image data. Following Holmes (1966), the image is used as its own store of raw data because it will probably be impracticable to store all fine image detail. The image is processed via a set of concentric arrays, covering a range of cell sizes, such that the coarsest array spans the complete image whilst the finest array spans only a small area. Registration problems are circumvented by associating previously derived image organization with the corresponding detail in the finest array spanning the relevant image area.

The image details of initial interest are simple features such as contours, lines, and spots, and are detected by templates for these features (at differing resolutions, locations, and angles) — note the parallel here with neurophysiological findings (e.g., Hubel and Wiesel, 1962). "Zooming-in" on a small image area with the finest resolution array, is guided largely by the presence of coarser resolution detail, with a degree of organization of this coarse detail being achieved during the process of selecting the centre of attention. Apart from the initial coarse resolution view, the span of attention is smallest at the start of processing, and grows as image organization and abstraction proceeds (note that previously-viewed objects displaced from the centre of attention do not appear blurred in human vision, although the area of sharp focus subtends only 1° or so).

Selection of new areas for attention, on the basis of partially organized image information, depends very much on the immediately prior processing, and on any unresolved questions or difficulties arising therefrom. Areas adjacent to that last examined would normally be attended to next, but associations between the organizations assigned to spatially remote image areas might cause a sudden
change in the centre of attention, as for example when information found in one part of the image is relevant to the interpretation of an area which has previously been unsuccessfully examined.

4.2.2 Recovering Properties, Relationships, and Constraints (Block B in Fig. 4-3)

Measurement of property values, recovery of relationships between components, and detection of constraints within components, all play an important part in facilitating description of the image, by providing details for filling in the description, by suggesting organizations which could be imposed on the image (particularly organization of the image as a projection of a 3-D field, see Section 4.2.7), and by enabling the development of new properties (or descriptors) which fit the region being described. Recovery of many relationships and constraints is based on comparison of property values measured for distinguished components, or for different regions of the same component. There is a close interaction between the processes described in this section, and the processes of selecting and evaluating possible organizations, in that measurement of a property value is sensible only if that property is an appropriate component in the description of the region being measured.

The properties, relationships, and constraints can apply to the discrete values in the initial machine representation of the image, to boundary curves and areas of the 2-D image, and to surfaces and volumes in the perceived 3-D field. The properties may be both metrical and articular, and may cover a range of resolutions in any of these three domains. Only clearly defined features (of the type considered above) are detected initially, with less obvious features being searched for if a satisfactory local organization cannot be achieved. Finding associations, discriminations, and constraints, and the role that these play in suggesting organizations, will now be discussed.

4.2.2.1 Associations

Relationships between distinguished components which suggest that the components are associated in some manner, and should be grouped in any higher-level organizations of the image or 3-D field, form an important class. It is
not required that the components involved in these "associations" be completely bounded, as the associations may be between sub-regions. Indeed, it will often be the case that a component being examined is too large to be attended to at the present stage of processing or resolution level, and associations considered as being between sub-regions of separate components may later be interpreted as being between separate sub-regions of the same component, see Fig. 4-4; there are implications here with regard to approaches which try to completely isolate components before searching for relationships between them.

There are many ways in which associations are manifested, with several different concurrent associations between one component and others being usual. The concurrent associations may reinforce each other or may conflict in that they suggest alternative groupings, in which case ambiguity arises — see Section 4.2.5. When an element is associated with a large number of adjacent elements, it may appear as a member of a group (e.g., an element of a textured surface) rather than as being ambiguously associated with individual neighbouring elements. For a component or its sub-parts to be involved in significant associations, it is necessary that these be distinguishable in some sense from their surroundings, perhaps partially by the boundary of the current area of attention. Thus, points 1, 2, 3, and 4 in Fig. 4-5a are not regarded as significantly associated even though they lie on a straight line, but points a, b, c, and d in Fig. 4-5b are regarded in this way, although it is only through their associations to each other that they are distinguishable.

In the proposed model, the most basic associations are those studied by Gestalt psychologists (Katz, 1950), viz. — similarity, proximity, continuity, experience, and other associations derived from these four. In essence, all these associations are derivatives of similarity (e.g., proximity is similarity of position), and the strength of association is a function of the similarity of the entities being compared, in the context of the dissimilarity of their surroundings.

Associations may be based on any properties which are applicable to the components being compared. For example, in associating two short dark bars on a lighter background, attributes such as location, orientation, width, length,
(a) "A" seen as being distinct from, but similar to, "C"

(b) "A" and "C" seen as two sub-regions of the same component

**FIGURE 4-4** Association Between Distinguished Sub-Regions of a Single Component

(a) Points which are not significantly associated even though they lie on a straight line

(b) Associated points which are distinguishable only through their associations with each other

**FIGURE 4-5** Requirement that Entities Associated be Distinguishable
sharpness of boundary, and gray level of the bars and background, all contribute to the overall association between the two bars. Note that some properties are more meaningful in comparisons between objects, rather than as properties of the objects themselves. For example, the relative orientation of the two figures in Fig. 4-6 is clear, but the property of orientation (relative to a reference frame) does not seem applicable to the individual figures. Examples of typical relationships which suggest groupings of components are shown in Fig. 4-7.

Association according to experience, suggests that components should be perceived as being organized in a manner which is similar to previously encountered and remembered organizations. This type of association is mediated by specific knowledge of objects in the world rather than by the more general knowledge underlying the first three associations. The main problems in associating according to experience are those of accessing the relevant experience in memory and of applying it effectively; these problems are discussed in Sections 4.2.3 and 4.2.4.

4.2.2.2 Discriminations

Discriminating between components (which are already separate but are associated in some manner), on the basis of differences in corresponding property values, is part of the process examined above of getting an overall association between such components. Of more interest here is the process of taking components which are regarded as undifferentiated wholes at the current stage of processing, and seeing if there are differences in their (effectively continuous) interior property values which suggest that various sub-regions of these components should be discriminated from each other, and possibly segmented. Associations between any sub-regions so discriminated might still be present, and have to be accounted for in suggesting possible organizations—see Section 4.2.3.

The differences searched for may be in either articular properties (e.g., as in Fig. 4-8a where a boundary constraint ceases), or metrical properties (e.g., as in Fig. 4-8b, where there is a distinct difference in gray level), in any of the three domains*, and over a range of resolutions. The accuracy of location of a detected difference is, of course, a function of the corresponding resolution level. Low resolution and complex properties require examination of large areas

* i.e., the 3-D fields, 2-D images, and digitized images
FIGURE 4-6 Figures for which "Absolute Orientation" Does Not Seem to be an Applicable Property

(a) Associating small dots into a gentle curve on the basis of similarity (and contiguity)

(b) Association of vertically adjacent dots, rather than horizontal, on the basis of proximity

(c) Associating the horizontal bars on the basis of continuity

(d) Grouping of strokes according as they form previously experienced (and remembered) configurations

FIGURE 4-7 Associations which Suggest Groupings of Image Components
(a) Cessation of parallel boundaries
curve perceived as straight even though orientation relative to "ω" is increasing

(b) Gray level contour

(c) Differences in coarse-resolution behaviour

(d) Difference in average orientation of bars

(e) Cessation of boundary constraint
(distance of boundary from a centre of curvature = constant)

(f) Sudden change in boundary orientation

FIGURE 4-8 Property Changes which Suggest Image Segmentation
to ascertain if there are genuine differences. For example, a change in angle between two straight line segments, can be detected with less context than that required for detecting a change in curvature between two smoothly-joined curves. To be noticeable, a change in a continuous property has to be of sufficient magnitude to be perceptible, and also has to be sudden. If a change is gradual rather than sudden, then it suggests that an inappropriate property is being measured — a more complex property (e.g., spatial derivative of the current property) or a coarser resolution may be required. As a consequence of the requirement for a property change to be both sudden and perceptible, "anomalous" differences (leading to isolated boundary segments) may occur where a boundary fades out.

Points of absolute maximum or minimum property values tend to be significant, even when the properties involved are changing smoothly at such points, because there will be a repetition of (at least the magnitude of) the property values on either side of the extremum point, leading to a form of symmetry and to discrimination of the regions on either side of such points.

Examples of the types of discriminations and segmentations envisaged are shown in Fig. 4-8.

4.2.2.3 Constraints

Finding constraints and finding differences within regions and components are almost converse processes; the motivations for these two processes are, however, somewhat different. The primary aim in searching for constraints is to enable the development of properties which are appropriate to the region being examined, rather than having to select properties from a pre-chosen set. The difficulties inherent in trying to describe a naturally given image with a fixed set of metrical and articular properties have already been stressed, particularly with regard to the latter. The proposal here is that by defining a suitable basic set of attributes (for both the 2-D image and the 3-D field), it might well be possible to construct appropriate attributes out of this basic set, on the basis of detected constraints. Typical basic attributes for the 2-D image are gray level, distance, and orientation, with a more extensive set being required for the 3-D field.

Comparison of appropriate properties is required for the detection of
significant associations and discriminations. As a simple example of the way in which metrical properties can be developed, consider the processing envisaged for the boundary curve AB shown in Fig. 4-9. The simplest constraint is that of orientation being constant, i.e., "straightness". In attempting to describe AB in this manner, it is readily detected that the curve's orientation does not remain constant. The curve could still consist of two or more straight line segments, but no sudden and distinct change in orientation can be detected (except perhaps at a coarse resolution), and it is therefore evident that description in terms of only straight line segments is not appropriate. The next simplest property is based on the orientation of the boundary curve changing steadily with distance along the curve, i.e., constant curvature. (For small radii of curvature, this constraint might alternatively appear as the curve being a constant distance from a point). The curvature does change perceptibly over the curve AB, and points X and Y are identified as possible break points between segments of constant curvature. Each of the segments AX, XY, and YB is then examined to see if a property simpler than constant curvature can satisfactorily describe it. The complex curve AB is thus eventually described as a curve AX which joins a curve XY (with a sudden change in curvature but no sudden change in angle at the joint) which itself smoothly joins a straight line segment YB — each segment thereby being described in terms of the simplest property which does not change perceptibly within that segment. This description need not be the final description of the curve, of course; relationships between the segments other than at their junctions have to be recovered, and might suggest alternative descriptions or possibly even alternative segmentations (e.g., because of symmetry).

It might be argued that the properties employed in the above example could well have been pre-chosen. This example has been selected because of its simplicity and the consequent ease of explanation; in more complex cases, particularly those concerning articular properties (see below), image areas, and 3-D volumes, the need for development of properties is more obvious. Consider a black spot falling away uniformly to a white background; by detecting that the gray level is constrained to decrease uniformly with distance from a point, an appropriate description can be formed — with a pre-chosen set of descriptors, the possibility of an image area
FIGURE 4-9 Boundary Curve Segmentation and Description

FIGURE 4-10 Evenly-Structured Regions
such as this might have been overlooked. It is important to search for constraints in the right domain. In Fig. 4-8c, for example, it is tempting to think of the constraint in the bottom curve as being based on low resolution orientation of the curve versus distance along the curve (i.e., a 1-D constraint). The nature of the upper and middle curves indicates, however, that this type of constraint is better thought of in terms of 2-D areas occupied by the curve.

The approach indicated above relies on fairly sophisticated processing, but has more promise of coping with naturally given images than the use of a closed formalism. A similar process is envisaged for derivation of articular properties (i.e., properties for describing evenly structured regions), but in this case the development is based on the perception of constraints between elements (i.e., associations) rather than within them. In this way, appropriate descriptions of regions such as that shown in Fig. 4-10 could be formed, without the likelihood of the particular structures portrayed having been considered in the design of the interpretation system.

4.2.3 Selection from Alternative Organizations (Block C in Fig. 4-3)

There are so many possible alternative organizations which can be applied to typical image areas, that there is a clear requirement for selection of those which are potentially appropriate (particularly with regard to the perceived 3-D field — see Section 4.2.7). Such potential organizations are selected on the basis of previously recovered features of the image or the perceived 3-D field, acting in concert with the external information represented in the information store. Following Gregory (1966), the image organization finally chosen is regarded as being suggested by and tested against the image data. Evaluation and elaboration of selected organizations is considered in the following section; there is an interaction between selection and evaluation in that if one suggestion is not tenable, alternative suggestions are required (taking into account any information gained during evaluation of the initial suggestion).

The influence in human perception of guided selection of possible image organizations, is illustrated by the difficulty experienced in locating the six triangles and nine quadrilaterals in Fig. 4-11. Evans' (1968a) picture analyzer
Find the six triangles and nine quadrilaterals present.

**FIGURE 4-11** Example of Guided Selection in Human Interpretation

Three prolate spheroids

Object raised out of a plane surface

**SIDE ELEVATION**

Note: Edges such as "a", which are curved in the horizontal and vertical planes, lie in a plane passing through the viewpoint "α".

Three prolate spheroids

**PLAN**

(a) Perspective projection, from viewpoint "α", of an "impossible" figure

(b) Plan and side elevation of a possible, but perceptually unstable, 3-D interpretation

**FIGURE 4-12** A Coherent Interpretation of a So-Called "Impossible" Figure
experiences no such difficulty, because its analysis is not guided in this manner; this apparent strength of his analyzer would be revealed as a serious shortcoming, if the same techniques were applied to complex images. Further evidence for guided selection in human interpretation is provided by the apparent anomaly experienced in trying to interpret "impossible" figures such as Fig. 2-16 and Fig. 4-12a. Coherent 3-D interpretations of these figures are possible, but the image features which suggest inappropriate depth interpretations, are so compelling that we "zero-in" on an inconsistent interpretation, and it is rather difficult to perceive other organizations (even if they are coherent). In Fig. 2-16, for example, a coherent interpretation is possible if the straight projected lines are seen as boundaries of curved rather than plane surfaces; Fig. 4-12b gives an impression of a possible 3-D field underlying Fig. 4-12a.

The main aim at each stage of the iterative processing, is to group components (in both the 2-D image and perceived 3-D field) on the basis of relationships between them, into more and more complex configurations, which then serve as components for the next stage of processing. In this way, larger and larger units are built up until the level of objects (i.e., units which appear to have a separate existence) is reached. Relationships between units are useful for positioning these units with respect to each other, even if a more complex unit is not perceived; for example, individual objects in a 3-D field may not be grouped into more complex objects, but the relationships between objects are important in describing this field. Figural units arise in one of two basic ways:

1) If by grouping various components, a region which is internally consistent (relative to inconsistent surroundings) is formed, then that region appears as a figural unit — see Fig. 4-13a. The properties which define such units may be rather complex, and it is common for there to be several different types of regions adjacent to the figural unit, in which case isolation of the unit must proceed by determining the extent of the consistent region, rather than by detecting its boundaries directly.

2) A small region which is embedded in and interrupts a consistent larger region, tends to appear as a figural unit even if it is not
(a) Figure on inconsistent ground  
(b) Interruption to consistent ground  
(c) Combined effect

**FIGURE 4-13** Formation of Figural Units

---

(a) Undulations assumed perpendicular to a curved base  
(b) Tendency to group parallel rather than proximate lines

**FIGURE 4-14** Influence of Simplicity on Perceived Organization
internally consistent, although if it is, this tendency is reinforced — see Figs. 4-13b,c.

The features (i.e., associations, discriminations, and constraints) detected in the previous stage of processing are, of themselves, suggestive of certain organizations, but there will often be so many alternatives suggested, that selection of a single possible organization for evaluation is difficult. An important technique for resolving this ambiguity is to select organizations which resemble known configurations (see Fig. 4-7d, for example), i.e., to apply previous experience to interpretation of the current image area; this technique is also important in so far as it enables a simplification of description. The main problem is that of bringing relevant information to bear on the situation being interpreted.

It is proposed that by suitably organizing the information in the data base, so that simple components, constraints, and relationships are linked to any more complex known organizations in which they are involved, a process of "guided interpretation" (in which the final complex interpretation is arrived at via a large number of simple steps) can be achieved. For example, in processing an image of a capital "A", slanted strokes (corresponding to the two sides) are linked to the representation of an "A", so that having found such strokes in the input, the possibility of interpreting the image as an "A" is suggested. See Section 4.2.8 for further discussion of the role of data organization in guiding the interpretation process.

Having selected a possible interpretation of the given image area in terms of previous experience, the features found in earlier processing are passed across (with information about the suggested interpretation) to the evaluation phase, so that unexpected associations and differences might be accounted for, rather than ignored. In "A" for example, the short bars might be interpreted as unexpected additions to an "A", rather than as unrelated detail.

With unfamiliar input material, and during the early stages of processing material which is later found to be interpretable in terms of prior experience, relevant specific experience may not be found, and in this case organizations are suggested purely on the basis of similarity, proximity, contiguity, etc. Simplicity is an important factor in selecting possible organizations, and simple hypotheses are preferred; even if such hypotheses do not fit exactly, they might still be satisfactory in so far as any discrepancies might be readily explainable during
evaluation. Examples of such organizations are spheres, circles, straight, parallel, perpendicular, and right-angled. Thus, in Fig. 4-14a the axes of the undulations suggest that they are superimposed on a base component which intersects these axes at right angles, and in Fig. 4-14b, lines A and B tend to be grouped in preference to lines B and C, because A and B are parallel.

Other factors which introduce a predisposition to suggest certain organizations are (i) the tendency to interpret according to organizations already accepted in other parts of the given image (particularly if they are adjacent to the current area and continue into it), and (ii) knowledge of the external context — for example, if the image is known to be textual, interpretations in terms of alphabetic characters are suggested first.

An important operation in this phase of the processing, is that of suggesting organizations which are contextually indicated but not directly represented pictorially, where this leads to a simplification of the overall description. For example, in Hellebore (Fig. 2-4c), a boundary is perceived in an area which is locally homogeneous, and in Zinnia (Fig. 2-10) the shape of the occluded surfaces of the spines is inferred.

4.2.4 Evaluation of Suggested Organizations (Block C in Fig. 4-3)

In the previous section, selection (on the basis of incomplete evidence) of a hypothesized local image organization, was discussed. Before being accepted as a satisfactory description of the relevant locality, it will usually be necessary to elaborate the hypothesized organization by filling in details, accounting for any unexpected differences or associations, and judging its appropriateness (e.g., by gathering further evidence and trying to fit this organization in with the surrounding context). If a hypothesis is found untenable, it is necessary to return for alternative suggestions.

It will often be the case that there are several locally tenable organizations, but unless these are internally inconsistent, they cannot be regarded as either right or wrong — they must be regarded as simply more or less appropriate (e.g., a rectangle with all sides equal is more appropriately described as a square). The objective of the selection and evaluation phases of processing, is to arrive at a description which is the most satisfactory in the local context (it does not follow
that this will automatically be the best in a more global context — see Section 4.2.5). Even if a suggested local organization is inconsistent with its surroundings, it may be the surroundings which are at fault, and final evaluation may have to await much further processing.

The organizations suggested as base components for the description of the locality attended to, will usually have been selected on the basis of their being simple or within prior experience (possibly gained in processing another area of the current image). These organizations will usually not be specified in detail sufficient to fully describe the locality. In particular, metrical properties may have to be determined; some of these may be inherited in the suggested organization (e.g., the characterizations of A's, 6's and 9's include expected orientations), but others may not be inherited (e.g., the size and location of an A, and the width of its constituent strokes). Note that inherited properties may also have to be qualified, as in the case of describing "A" as a "rotated A". Although it may later be abstracted from (see Section 4.2.6), a reasonably complete local description (including a description of the background and of any detail which is not of direct interest) is attempted at each stage of processing, so that local and contextual inconsistencies might be revealed. Information regarding various modifications to the suggested organization (to make it fit the image data) will therefore usually be required. The main objective is to use the suggested (simple or known) organization effectively in describing the given locality. This organization can be regarded as a base component in the description, onto which are appended additional specifications and modifications (i.e., "schema + correction").

In comparing a known or simple organization with the image, it is important that corresponding parts, relationships, and constraints be detected, so that meaningful comparisons of articualr and metrical properties can be made, and so that the presence and extent of any differences can be noted in the description. In choosing among possible organizations, an important attribute is simplicity, but this is difficult to define objectively — the information cost of alternative descriptions (i.e., the number of required additions to, deletions from, and transformations of known or simple organizations) is one possible measure. When metrical properties are important, the classification techniques described in Section 3.2 are seen as
It is important that significant underlying relationships and constraints (rather than some of their surface effects) should be recovered, or generalization over minor variations may be prevented. For example, in Fig. 4-15 the relationship of strokes A and B goes beyond "A joined to the centre of B at right angles" — witness the perceived similarity of Figs. 4-15a and 4-15b. A known organization may still be a useful component in the description of an image area, even when the parts are different, if the relationships between these parts are similar — see Fig. 4-16.

The suggested organization will usually not fit exactly, and a principal requirement in evaluating suggestions is that any differences observed should be satisfactorily explained in terms of unexpected additions, deletions, deformations, transformations, etc. Thus "\wedge" might be described as a "\wedge" with a missing stroke or as a "\vee" with an added stroke, "\Lambda" might be described as a large "\Lambda", and "\bigtriangleup" might be described as a triangle with smoothed corners. Explaining deformations of simple arrangements, such as parallel and perpendicular, in terms of a projective transformation, may assist 3-D interpretation (see Section 4.2.7). Accounting for the similarities of configurations being compared, requires that the relationships between various properties or descriptors should be available. Thus, in describing "\Lambda" as an "A" with curved sides, the transformation "curved" has been invoked to explain the difference (i.e., arc of circle = curved (line segment)).

Just as unexpected differences have to be accounted for, so do unexpected associations and similarities. It has already been noted that in any set of components there will often be many associations between components, some of which reinforce and some of which conflict, in suggesting groupings. In accepting one organization, the associations which suggest alternatives have to be accounted for as indirect, subordinate, or coincidental — see Section 4.2.6. Another form of unexpected similarity occurs when a boundary which is indicated by the context (especially 3-D) as being present, is not present in the image, so that the regions adjoining the indicated boundary are unexpectedly similar. An example of this case occurs in Fig. 4-17; the different perceived depths of components which are not discrimi-
(a) Coincident strokes

(b) Separate strokes

FIGURE 4-15 Importance of Spatial Relationships Other Than Those Based on Coincidence of Strokes

(a) Known configuration "L" (b) Configurations whose description can be based on the concept of an "L"

FIGURE 4-16 Use of One Configuration in Describing Others Having Similar Organizations but Differing Components
(a) Scanning electron micrograph of an ant's head

Note the absence of a contour where spines cross.

(b) Enlargement from top right-hand corner of Fig. 4-17a

FIGURE 4-17 Image in Which Perceived Depths, and Other Contextual Information, Suggest Presence of a Boundary Which is Not Pictorially Visible
ated pictorially, suggests that there is nevertheless a 3-D boundary present. In situations such as this, the indicated boundary is imposed on the image if this imposition leads to a simplification of the description and the pictorial absence can be explained satisfactorily (e.g., as a chance equality of reflected light intensity from both sides of an actual 3-D boundary).

In comparing image properties with external knowledge, the problem arises of making choices between alternative descriptions; the difficulties which may ensue when a threshold value is employed in effecting such choices, have already been referred to. In the proposed model, it is suggested that these difficulties can be mitigated by using thresholds only as limits of perceptibility (rather than as decision boundaries), and by comparison with (and description in terms of) other similar situations in the current image, before referring to previous experience. The thresholds of perceptibility vary with the degree of attention, so that in examining (and therefore attending to) discontinuities arising with a high threshold of perceptibility, these "discontinuities" may be seen as continuations which are only barely discernible. The use of thresholds is discussed further in Section 6.1.

Having elaborated a suggested organization until it forms a detailed and accurate description of the given area, an evaluation of this description's appropriateness must be made. The appropriateness of the base component employed in the description is a function of the amount of image information accounted for by this component, of the number and extent of modifications required to "fit" the base to the image detail, and of the existence of suitable explanations for any unexpected differences and similarities. Also to be taken into account is the internal consistency of the detailed description, and the ease with which this description fits into the surrounding context.

After finding a suitable detailed description of a local image area, this

---

1 It is of course possible that in the process of elaboration and evaluation, a closer examination of the original area (searching for barely discernible detail which supports or argues against the suggested organization) may have been made, or adjacent areas may also have been examined, as is required in determining whether the arrangement perceived in "\( \vee \)" is part of a "\( V \)" or of a "\( W \)".
description is abstracted (see Section 4.2.6) and fitted into the current description of the total input image, before attempting higher-level organizations of the local area (by searching for organization-suggestive features in the current description of this area, selecting and evaluating hypotheses, etc.). This processing cycle repeats with the current area until no further progress is made, whereupon a new area is selected for attention, and processing continues.

4.2.5 Resolution of Ambiguities (Block C in Fig. 4-3)

Unless there are strict limitations on the nature of the underlying 3-D fields (as in the case of Roberts, 1965), the information available from an input image constitutes only incomplete evidence for the field portrayed in it, and the interpretation of this field is inherently uncertain or ambiguous. For the image class considered herein, the limitations on possible 3-D fields are insufficient to ensure unique interpretations, and there will always be many tenable (even if awkward) interpretations.

As organization of the image data proceeds, image areas are described in terms of more and more complex properties, and there are usually several properties (possibly at several levels of resolution) which can be concurrently applied to these areas. As a result, two areas can often be either grouped on the basis of similarities between some properties, or differentiated on the basis of differences in others, thus introducing an ambiguity that depends on which properties are considered to be important. Examples of possible ambiguities are given in Fig. 4-18. There are two main requirements of interpretations, which aid in resolving such uncertainties — simplicity and coherency. These are essentially assumptions about the nature of the 3-D world, which therefore should be reflected in the interpretations derived, if these are to be in reasonable agreement with the actual 3-D fields portrayed.

Simplicity of interpretation, as effected by the use of prior knowledge or of basically simple configurations as base components in descriptions, has already been discussed in Section 4.2.3. An example of the way in which specific experience can resolve ambiguities of association between various components is shown in Fig. 4-18a; note that the right hand vertical stroke of the M is not grouped with
(a) Ambiguous associations between strokes

(b) Is it A on B, B on A, or A adjacent to B?

FIGURE 4-18 Examples of Local Ambiguity

FIGURE 4-19 Disconcerting Figure
the nearby parallel stroke of the B, although on a strictly local basis there is a
strong tendency to group these two strokes. Situations such as this, where the
simplest local interpretation is not the simplest in the global context, are
common.

Detection of inconsistencies (either local or global) is another important
technique in revealing either current or earlier poor choices of local organization.
When such inconsistencies are found, it may be necessary to "backtrack" and
choose alternative organizations for one or both of the inconsistent regions. If no
suitable organization can be found, a form of anomaly results — note the discomfort
which results when an attempt is made to perceive the organization of Fig. 4-19
or of "impossible" figures.

When, in the global context, several possible organizations are equally
tenable (as in the case of the Necker cube), we seem to adapt to (and grow tired of)
the currently held organization, so that this becomes less satisfying and yields to
other organizations. This process could be understood as a form of precaution
against locking onto one organization, when other globally more appropriate alter-
natives are possible. Provision of a similar process in an API system seems to
be worthwhile, so that in ambiguous cases, a sequence of alternative organizations
could be produced, one of which might (through trying various combinations of
local ambiguities), be "locked onto" in so far as it appears the most appropriate.

4.2.6 Abstraction and Economy (Blocks D and G in Fig. 4-3)

Because of the vast quantity of pictorial information present in complex
images, and the extent of information (about the nature of the fields portrayed in
such images) which is required for effective interpretation, the question of economy
of representation arises. Economy can be achieved by selecting the most significant
information from that available (i.e., "abstracting"), and by employing descriptive
strategies appropriate to the situations represented. The whole interpretation
process can, of course, be viewed as one of abstraction.

Economy is necessary in describing the current image, and also in repre-
senting previous experience (possibly gained during earlier processing of this image),
particularly with regard to abstracting from instances to class characterizations.
Simplicity (and therefore economy) of representation is achieved when appropriate descriptors are employed; for example, representation of a circular arc as a sequence of short line segments (Freeman, 1961), is clearly more complex and less economical than representation as an arc. For the class of images considered, the most appropriate representation of the image is usually in terms of a projective transformation of an underlying field, and known objects also are most effectively characterized in 3-D terms. The image may be represented as either a discrete initial digitization, as a continuous 2-D gray scale pattern, or as a projection of a continuous 3-D field — once a satisfactory 3-D interpretation is obtained, the other two representations may be deleted.

In abstracting from a local or from a global description, inferred information may be added (in the sense of augmenting deficient images and imposing boundaries which are not pictorially represented), as well as information being deleted. There is a clear tendency in human perception to neglect details of minor imperfections and remember an idealized version, possibly noting however that there are imperfections (e.g., encoding a gapped bubble-chamber track as a continuous track and noting that it is gapped, but omitting the locations and lengths of the individual gaps). In the terminology of Section 4.2.4, there is a tendency to retain only the base components in the local descriptions, and to delete (or characterize only statistically) any modifications to this base. Significant detail in any local description can be regarded as that which participates in, or determines the applicability of, higher-level organizations, or as that which is required for descriptive purposes.

Ultimate causes are usually of more concern than extraneous effects, such as quantization errors and blurring, which do not usually form an important part of the final interpretation (although they may have to be considered in its derivation). Thus, in abstracting from a 3-D description, the shape of the objects represented is of more importance than the variations in surface illumination (and possibly pigmentation), so that information concerning shading, shadows, and surface pigmentation in the perceived 3-D field, might be deleted. Without consideration of the perceived 3-D field, the significance or otherwise of the corresponding image detail would not be apparent.

In encoding relational information for complex images, there is a serious
problem with regard to the multiplicity of possible relationships between pairs (and triples, and so on) of image components. One approach to reducing the severity of this problem is to assume that components can be organized in a hierarchy of more and more complex groups, with significant relationships obtaining only between components within a group, and between groups as wholes, rather than between individual components in separate groups. The component relationships on which a group is based are regarded as being direct or significant, whereas any relationships noted between components in separate groups are regarded as being accidental or concomitant with other relationships between these groups as wholes. Thus in "ME" the relationships between the strokes composing the "M" and the "E" are regarded as direct, whereas the parallel relationship existing between the two central vertical strokes, is regarded as concomitant with the relationships between the "M" and the "E" as wholes. The above remarks suggest an explanation of the human penchant for perceiving hierarchical organization in images.

In abstracting relational information, indirect rather than direct relationships may be deleted, except as the former appear to capture significant relationships in the situation being described; for example, the relationship between the horizontal strokes in "A-T" might be viewed as significant, in that it specifies the relative vertical positions of the two characters.

It is envisaged that the representation of internally uniform regions (i.e., contours or areas in 2-D, and surfaces and volumes in 3-D) will be in terms of the metrical and articular properties which remain relatively constant throughout each region, together with a specification of the shapes and types of boundary, and of any significant relationships between regions. The internal properties on which such regions are based may be quite complex, and in the case of textured regions, might already be abstractions from the actual image detail in terms of microstructure and statistics — see Section 6.2.2. Note that several superimposed regions might be involved in forming the description at any point (e.g., a shadow on a textured surface, or a summation of resolution levels).

As an example of the way in which superimposition of resolution levels can effect economies, consider the derivation of a description for the curve shown in Fig. 4-20a; if each of the individual segments (such as a, b, and c) is represented as a
(a) Curve to be described

(b) Description as a superimposition of components at different resolutions

FIGURE 4-20 Economy of Representation Through the use of Several Levels of Resolution

(a) Curve to be described

(b) Least-square fits of parameterized curves

(c) Modification to fit boundary constraints

FIGURE 4-21 Economy of Representation Through the use of Inter-Element Constraints
straight line segment of a measured length and orientation, and the curve AB is described as a sequence of these segments starting at a given point, then it is clear that (in so far as individual errors accumulate) each segment's length and orientation has to be specified very accurately, if precise representation of AB is required (accurate specification of such property values implying a high information cost). If, however, curve AB is described in terms of a summation of components of several resolution levels, and if the overall behaviour of this curve is of more interest than the exact lengths and orientations of the individual segments, then the required resolution of the individual components can be relaxed, because errors will not accumulate in the same manner as with a single level of resolution  

— see Fig. 4-20b. Thus, a reasonably accurate representation may be achieved even if the resolution of individual measurements is poor — some such coding strategy is implied in human perception, because judgement of magnitude is limited to an accuracy of 1 part in 7 to 10 or so (Miller, 1956). With regard to the limited resolution of human judgements, it is interesting to observe that given sufficient information capacity, an API system might be capable of much finer judgements than 1 in 10, and might therefore be able to effect discriminations which we could not. This possibility is indicated in the superior accuracy of a stereo plotter (Lawrence, 1967) as compared to a human being, in deriving depth information from stereo image pairs.

With regard to the use of constraints and relationships in effecting economies, consider curve AB shown in Fig. 4-21a, which is to be described in terms of the three segments a, b, and c. The descriptors chosen for these segments (e.g., curvature changing at a constant rate) do not fit exactly, and the residual (when a least mean-square error fit is attempted) might suggest more complex descriptors or further segmentation. By regarding these descriptors as pliable, in the sense that they are not assumed to be exact and may have to be modified to agree with any constraints perceived, unnecessary complication and segmentation can be avoided. Fig. 4-21b shows the assumed least-squares fits of segments of constant curvature; these segments do not exhibit the relationships apparent in the actual

Note the parallel between this form of representation and the representation of numbers to an arbitrary resolution in a digital computer, using low resolution (e.g., binary) components.
curve (i.e., the ends of the segments join smoothly). If the fitted segments are modified to satisfy the relationships inherent in the specification of "break-points" 1 and 2, the description of curve AB as three smoothly joined segments of constant rates of change of curvature, is a reasonable "fit" — see Fig. 4-21c.

4.2.7 Factors in the Derivation of 3-D Organization (Blocks B, C, D, and G in Fig. 4-3)

It has been argued in Section 4.1.1 that because the objects of ultimate interest are 3-D, interpretation in 3-D terms is essential. Consistent with the philosophy of the proposed model, if a 2-D image area is more simply described as a projection of some 3-D field, than as variations of image gray scale, then that area should be interpreted in 3-D terms. Witness the difficulty experienced in trying to perceive the Zinnia micrograph (Fig. 2-10) as variations in the pigmentation of a plane surface (which the micrograph actually is), rather than as a projection of a 3-D field.

This section discusses 3-D interpretations within the model (given a single static 2-D image), and contributes a number of observations and suggestions, elaborating on remarks made in earlier discussion of the model, as these bear on the problem of deriving 3-D interpretations. That such derivations are possible is demonstrated by the ease with which we can interpret the Zinnia micrograph, and the objects portrayed in Fig. 4-22, in 3-D terms. It has already been noted that the interpreted descriptions should be coherent, and as simple as is consistent with fitting the facts of the 2-D image. Because the objects of interest are characterized by their organization, the descriptions should also be articular. The basic attributes within the 3-D descriptions (for the class of solid opaque objects considered herein) are simply solid and void, together with 3-D boundary surface and illumination properties.

Another requirement, which may not at first be apparent, is that a plausible (if only approximate) description of the shape of occluded surfaces should be inferred. Human beings do in fact make such inferences, as is evidenced by the phenomenon of "disappointed perceptual expectation"; we should be very surprised if Zinnia's spines in Fig. 2-10 turned out to have a half-moon shaped (instead of round) cross
FIGURE 4-22  Images Which are Most Appropriately Perceived in 3-D Terms

FIGURE 4-23  Value of Inferring Occluded Surfaces, in Revealing Inconsistent Interpretations

FIGURE 4-24  Simplification of Description by Omitting Specification of Boundaries of Visibility for Occluded Surfaces
-section, when viewed from another direction. There are at least two reasons, derived from the requirements of coherency and simplicity of descriptions, for inferring the shape of occluded surfaces:

1) To ensure coherency of the overall interpretation and to detect any inconsistencies, a fairly complete description of the 3-D field (including occluded portions) is desirable. Thus, in Fig. 4-23, the two pyramids might initially be interpreted as being of the same size and at the same depth in the 3-D field, but this is inconsistent with the inferred continuation of the square bar.

2) The 3-D field will usually not change unexpectedly at the boundaries of the directly visible portion, and in a description of an image as a projection of an interpreted total 3-D field, knowledge of the viewpoint and of the projective transformation, enables approximate location of these boundaries without explicit specification, if this is required. These boundaries seem to be more a characteristic of the 2-D image than of the 3-D field, and do not therefore seem of direct importance in the final interpretation. The interpreted 3-D field, and not the image which enabled its derivation, is the eventual concern, and in many cases specification of only the directly visible portions may be more complex than an inferred description of the total field. For example, in Fig. 4-24, A and B can be described as parts of a single solid cylindrical bar continuing unchanged behind C — a somewhat simpler description than one that explicitly specifies which parts of the bars are visible from the given viewpoint.

Some of the major problems which arise in trying to derive 3-D information from a single static 2-D image have been summarized in Section 2.1. Briefly, these problems are (i) the inherent 2-D → 3-D ambiguity or uncertainty, (ii) occlusion, and (iii) the occasional absence of a pictorial representation of 3-D boundaries which are not occluded. In human perceptual experience, ambiguity with respect to the 3-D interpretation of 2-D images is usually not apparent, except in contrived situations such as the Necker cube. It is interesting to observe that some perceptual psychologists have taken such ambiguities "to mean that there is no hope of
discovering a lawful correspondence between stimulation and perception" (Gibson, 1966, pp. 246-247). The achievement of stable perceptions in human experience usually results from a process of "zeroing-in" on a satisfactory interpretation, such that it is hard to imagine alternatives. In attempting to formulate an approach to mechanical derivation of depth information, the problem of ambiguity is all too obvious: what particular combination of surface lighting, aspect, pigmentation, microrelief and depth, produced the gray level of each resolution element in the digitized image?

Occlusion presents problems in the derivation of reasonably complete 3-D descriptions, because of the lack of direct (even if ambiguous) information about the nature of the occluded portions. Unless there is evidence to the contrary, occluded surfaces and objects can be assumed to continue in a manner consistent with the characteristics attributed where they are visible, so that in the overlapping bars of Fig. 4-24, there is no uncertainty in inferring the shape of the occluded portion. In general, the inferred 3-D shape of occluded surfaces is bounded by what can be seen of the surrounding volume, and the inferred continuations of other surfaces and objects such as the background.

Absence of a pictorial representation of 3-D boundaries presents formidable problems, and requires the imposition of "invisible" boundaries — see Section 4.2.4. There is of course, a close relationship between inferring the continuation of an occluded surface or boundary, and inferring the continuation of a boundary which is not occluded but which, nevertheless, "fades out" or is not represented pictorially.

Previous work in mechanical derivation of depth information has been concerned mainly with detection of relative disparity of corresponding areas in stereo image pairs. These techniques are now well-developed, but there are obvious limitations — depth information can be derived only for detailed surfaces visible in both images, and the depth information is not structured. The use of such techniques would no doubt be valuable in an eventual API system, but of more interest here are the processes required for derivation of depth information from a single image. The work of Roberts (1965) and Guzman (1968) has been discussed in Section 3.3; whilst they demonstrated the importance of effective models of
the world in effecting 3-D interpretations, their techniques do not generalize to complex images and objects.

No detailed generally accepted theory of human space perception has evolved as yet, but the work of Gibson (1950, 1961, 1966) has been predominant. He argues (1966) for the wealth of information present in the ambient light available to a human perceiver, which is specific to certain 3-D arrangements, and proposes a theory of information-invariant based perception, rather than the traditional theory of energy- or sensation-based perception. Gibson (1966, p. 238) remarks "the problem of depth perception considered as the conversion of two-dimensional experience into three-dimensional experience seems to me quite insoluble.

In this book, and implicitly in my earlier book (Gibson, 1950), the problem disappears. If sensations are not entailed in perception at all, why speculate about how they might be changed into perceptions". These remarks revolve around Gibson's contention that the senses (considered as perceptual systems) respond directly to the higher-level information-variants specific to certain 3-D arrangements (e.g., gradient of visual texture specifying a receding surface). This formulation avoids the problem (of vital concern here) of specifying how it is that invariants are responded to, and also fails to account for the fact that these so-called invariants are merely fallible clues. Nevertheless, much of Gibson's and other work on human space perception is relevant here.

In the proposed model, derivation of 3-D organization is seen as being an iterative process intimately involved with other image processing, rather than as a separate step. Local 3-D organizations are suggested on the basis of image configurations indicative of certain 3-D arrangements, and also by the (3-D) organizations assigned to adjacent areas of the image. These suggested organizations are gradually merged into a coherent global interpretation (perhaps rejecting many suggested local organizations in the process).

With regard to the problem of ambiguity of 3-D information, it is clear that there is insufficient information in the image alone to deduce a unique underlying 3-D field; the derivation of 3-D information must proceed on the basis of "reduced cues". By itself, the image information allows an unlimited number of solutions in terms of distribution, illumination, pigmentation, etc., of the 3-D field.
To reduce the solution space, additional constraints must be imposed on acceptable solutions, principally through the use of extra-pictorial information such as assumptions about the 3-D fields portrayed (e.g., their general nature, the classes of objects likely to occur, and the relationships that these can enjoy).

Some apparently worthwhile assumptions about the nature of the 3-D world, and the way in which these assumptions are manifested in the suggestion of possible 3-D organizations on the basis of image features, are discussed in the following sections. This discussion is primarily addressed to the problem of ambiguity (i.e., what contributions to the image's gray scale behaviour are made by aspect, pigmentation, and illumination structuring of the 3-D field portrayed?), but comments relevant to the inference of occluded surfaces are also included.

4.2.7.1 Simplicity

An important assumption (derived from the principle of inductive inference cited in Section 4.1.1) is that the 3-D field underlying an image is as simple as is consistent with the image detail; in fact, all the assumptions discussed in the following sections are based on simplicity. One aspect of this assumption is that the objects in the 3-D field are presumed to be opaque and compact (i.e., separate parts of the same object are connected rather than dispersed), and the surfaces of simple component parts of objects are presumed to be evenly illuminated and pigmented, so that any variations in gray level within their projections are initially assumed to result from aspect changes.

Simple 3-D arrangements, such as parts being parallel or perpendicular to each other, or lines lying in a plane, usually do not appear in the projected image as parallel, perpendicular, or straight respectively. A strong tendency in human perception is to interpret nearly parallel or perpendicular projected parts as being perpendicular or parallel in the 3-D field, but located and oriented relative to the viewpoint to give the observed projection. Thus, in Morning Glory (Fig. 1-2), the spines are seen as being perpendicular to the body of the grain, and their perceived orientation suggests the shape of the body. It is difficult to achieve a stable perception of Fig. 4-25a as a solid object, because, in contrast to Fig. 4-25b, it is not possible to derive a coherent interpretation in terms of
(a) Perspective of non-rectilinear object
(b) Perspective of cube

FIGURE 4-25 Preference for Parallel and Perpendicular 3-D Plane Surfaces

FIGURE 4-26 Car and Head Silhouettes

FIGURE 4-27 Perception of a Changing Pattern as Similar Components from Different Aspects
perpendicular and parallel plane surfaces. Similarly, lines which are curved in the 2-D image may be perceived as lying in a plane cutting a curved surface; for example, Hellebore (Fig. 2-4c) is perceived as having straight furrows on a spherical surface. The simplicity assumption is also manifested in the previously discussed tendency to perceive surfaces as being spherical, conical, cylindrical, etc., or as simple transformations of these, such as oblate or prolate spheroids, cylinders and cones with curved axes, etc. — possibly as simple basic surfaces with additions or deletions (such as the spines on a spherical base in Zinnia (Fig. 2-10) or the furrows in Hellebore (Fig. 2-4c)). Note that perception in terms of simple configurations such as cubes and spheres is, in a rather important sense, "prior" to specific experience of them.

4.2.7.2 Experience

Another important assumption is that nature is repetitive (i.e., there are classes of similar objects), and that it is useful to attempt 3-D interpretations in terms of specific objects and configurations which have already been encountered (in either the current or previous images), allowing for transformations, additions, deletions, etc. In interpreting some area of the image as a projection of a known 3-D object, the 3-D nature of the occluded portion of the interpreted object is thereby inferred.

A particular example of some interest is the perception of depth in silhouette-type figures, in which case the usual depth cues of the type that Gibson (1966) seeks, are not present. Interpretation of depth for a cube silhouette (Fig. 4-22) might be explainable in terms of perceiving 2-D boundaries as corresponding to edges of surfaces which are parallel and perpendicular in 3-D, but successful interpretation of the car and head silhouettes shown in Fig. 4-26, implies that our characterizations of cars, heads, and other 3-D objects include some 2-D information (perhaps in the form of cross-sections through major axes) specific to these objects' images, as well as 3-D information. Thus, represented in a comprehensive API system we would expect to find features which are suggestive of particular 3-D organizations (such as a head), as well as features which suggest more general organizations (such as one object partly obscuring another).
Interpretation of new areas of a given image in terms of 3-D objects and components perceived in other previously examined areas of the same image, is illustrated by:-

1) Surface curvature may be perceived via changes in projected size and shape of repeated similar components. In Fig. 4-27, for example, a series of regular hexagons on a curved surface is seen (the perceived aspect of each hexagon corresponding to its projected shape). Another example of this tendency is the perception of surface orientation and curvature via changes of projected element size and inter-element spacing, where a large number of similar elements (e.g., textural elements) are present on the surface. On the assumption that the average 3-D element size and spacing are reasonably constant, a uniform decrease in size and spacing suggests a receding plane surface (Gibson, 1950), while a decrease which is not uniform with direction, suggests a curved surface.

2) There is an interaction between perceptions of repeated components and objects in any given image. In Mallow (Fig. 2-8b) and Morning Glory (Fig. 1-2), the interpretation of the spines which are seen end-on is refined by an oblique view of other spines; in Thrift (Fig. 2-5a) the spines on the straight bars appear to be only small undulations when viewed end-on, but a side view refines this interpretation.

4.2.7.3 Rotational Symmetry

A further apparently useful assumption is that, in the absence of contrary evidence, both the basic parts and the global shape of natural objects are often rotationally symmetric about a possibly curved axis of symmetry. Taking the silhouette shown in Fig. 4-28a for example, the symmetry of the 2-D silhouette about a curved axis suggests 3-D rotational symmetry, thus inferring the 3-D shape of both the visible and the occluded parts. The suggestion and subsequent evaluation of a plausible organization in terms of rotational symmetry, may require rather sophisticated processing. There are, for instance, two axes of symmetry in the 2-D
(a) Object perceived as "bean"-shaped
(b) Object perceived as rotationally symmetric despite unmatched detail
(c) Inhibition by deep unmatched nick
(d) Inhibition by enclosed straight line

FIGURE 4-28 Perception of Rotational Symmetry

FIGURE 4-29 Enlargement from Morning Glory
object portrayed in Fig. 4-28a, a vertical straight axis and a horizontal curved axis. Regarding this 2-D object as a silhouette, choice of the horizontal axis rather than the vertical, is based on the fact that the latter axis would lead to a mushroom-like 3-D object, whose silhouette could not be concave at the bottom. It is not necessary that there be perfect symmetry between corresponding parts of the 2-D outline; with symmetrical low-resolution "base" components of the outline, "lumps" which do not match up may be interpreted as local additions to a rotationally symmetric base — see Fig. 4-28b. Deep unmatched "nicks" in the boundary tend to inhibit the suggestion of rotational symmetry, because unless they arise from an elongated hole in the object, they will not be visible — see Fig. 4-28c. The presence of image detail within the 2-D outline shape can play an important part in elaborating, and confirming or refuting, the suggestion of rotational symmetry. For example, in the enlargement of a Morning Glory spine shown in Fig. 4-29, the faint curved contour within the spine's projection confirms the suggestion of rotational symmetry, and indicates that the axis of symmetry points out of the image plane at $45^\circ$ or so. The bottom contour also indicates this, on the assumption that the spine is perpendicular to a relatively flat surface. The presence of straight lines within a symmetrical outline shape tends to inhibit the suggestion of rotational symmetry (see Fig. 4-28d), although straight boundaries by themselves usually do not have this effect (e.g., the parallel-sided objects in Fig. 4-24 are seen as cylindrical).

4.2.7.4 Other Assumptions

A very common 3-D relationship or organization, is that of one object or part passing behind another relative to the viewpoint, as suggested by the interruption of one image area, boundary, or part, by a second. This suggestion is confirmed by a later continuation of the interrupted part, as demonstrated in Fig. 4-24.

As previously mentioned, component parts of objects are assumed to be relatively evenly pigmented and illuminated, so that gray scale variations within their projections will usually be interpreted as resulting from changes of aspect, unless this leads to awkward or inconsistent descriptions. Thus, the gray scale variations in Fig. 4-30a are seen as resulting from changes of aspect, but the variations in Fig. 4-30b cannot readily be interpreted in this way. Such variations
(a) Gray scale variations which are readily interpretable as 3-D undulations

(b) Gray scale variations which are more readily interpretable as changes in pigmentation than as 3-D undulations

FIGURE 4-30 Perception of Gray Scale Variations as 3-D Undulations
may be present at several resolutions, corresponding to levels of detail in the 3-D field — see Fig. 4-30a. Shading is an important clue to relative surface aspect (Hess, 1961) and when other clues are not abundant, is frequently relied upon in our perceptions of 3-D organization, as is demonstrated by the pseudoscopic (i.e., reversed) perception of depth when images such as Fig. 4-31 are inverted. This reversal results from our default assumption that the 3-D fields portrayed are illuminated from the direction of the upper half of the image.

As noted in Section 1.2.1, in the SEM the relationship between surface aspect and image gray level is rather different from the usual optical situation, with topographically high surfaces appearing lightest. It is of interest to note that this difference is not obvious, and has only a minor effect on our perception of 3-D organization in scanning electron micrographs.

Shadows themselves can, as "attached" shadows, indicate the projected shape of the object casting them, and the direction of illumination (this information possibly being of use in refining the perception of aspect changes via image shading). Discrimination of shadows from changes in pigmentation is often possible, on the basis that (i) there may be a continuation of structure, texture, or image contours across the boundary of the shadow (Gibson, 1966), and (ii) the light source is usually diffuse, so that shadows may have penumbras.

Other possible image clues to underlying 3-D organization are:

1) In naturally given images, sharp contours usually arise from sudden changes of depth or pigmentation, rather than aspect. This may not be the case, however, if the objects portrayed can have sharp joints between surfaces (e.g., Roberts' (1965) plane-faced 3-D objects).

2) The presence and extent of poor focus in some image area, suggested by the absence of sharp image contours, implies that the surface imaged is some distance in front of or behind the region of clear focus.

1 The nearest surfaces in 3-D fields do however tend to be brighter than remote surfaces, and it has been shown (Hagan et al., 1968) that modulating the intensity of lines in a graphic display, according to their depth in the 3-D field portrayed, is a useful aid to human depth perception.
Note reversed perception of depth when this image is inverted.

FIGURE 4-31 Pseudoscopic Image

Note the facilitation of 3-D interpretation when this silhouette is inverted.

FIGURE 4-32 Silhouette of Cube Viewed from the Underside
3) Objects are usually supported in some manner and often rest on a relatively flat background, so that objects which contact the background towards the bottom of the image are perceived as being closest; this effect is clear in the micrograph of Drimys (Fig. 2-4a). The difficulty in arriving at a 3-D interpretation for the silhouette in Fig. 4-32, suggests that the support plane assumption is important in human perception.

4) Straight lines suggest plane surfaces, particularly if a surface's projection has a polygonal boundary or straight lines within it.

4.2.8 The Information Store (Block G in Fig. 4-3)

The central component in the model is an information store or data base, in which is represented the "knowledge of the world" required for effective image interpretation. This knowledge of the world is seen as consisting of an initial component concerning the general nature of the 3-D world, and an acquired component concerning specific objects which have been encountered (cf. evolutionary versus individual experience in biological perception).

Of concern with regard to the information store are:-
(i) the types and extent of information represented;
(ii) the organization of this information;
(iii) the method of representation; and
(iv) the nature of possible implementations.

The problem of how to represent information effectively, is one to which the ultimate performance of an API system is tied, particularly with regard to the use of prior knowledge in describing a given image. Effective representations have been demonstrated so far only for very restricted object-worlds (e.g., Roberts, 1965). Development of suitable representations for more complex situations is likely to be a major stumbling block in the formulation of general API systems. Identification of the types of information to be represented, and the way in which this information is to be used, must precede such development. Some observations relevant to the development of representations have been made in previous sections, but this problem will not be discussed in detail here.
An important type of information concerns the classes of objects known to exist in the 3-D world. These classes are characterized in terms of total 3-D hierarchical construction, not just that visible from a single aspect. The organization of objects plays a dominant role in their representations. The shape of a house, for example, is represented in terms of plane surfaces, relationships between surfaces as wholes (e.g., perpendicular or parallel), and descriptions of the way in which adjacent surfaces join. Also included in the characterizations of 3-D objects are associations between objects (e.g., a cup and a saucer), and 2-D features characteristic of objects' projections (see Section 4.2.7.2). The simplest component parts are described in terms of constraints existing between their boundaries, rather than as "primitive" objects. Generalization over changes in size, location, orientation, etc., is thereby facilitated. If properties such as size and orientation relative to some reference frame vary little between instances of some class, then the characteristic values of these properties can usefully be specified in the class characterization (in so far as a simplification of the description of typical instances can be achieved). In any instance being described, variations of the values of these properties are accounted for in terms of an appropriate transformation (e.g., rotation or change of scale).

To the extent that class members are often highly variable (at least as far as location and illumination is concerned), class characterizations are abstractions from instances, such that they form useful base components in the description of any given instance. With this approach, the class characterization includes the organization and property values which are usually (but not necessarily always) present in class members, particularly those features which distinguish similar classes. A given object is therefore regarded as a member of that class whose characterization forms the most suitable base component for the object's description.

Another type of information in the data base concerns relationships and transformations between and within the three domains (viz., the digitized image, the image itself, and the 3-D field). This category includes relationships between components (such as proximity and continuity), transformations within a domain (such as bent, curved, and rotated), and transformations between domains (such as digitization and perspective projection). Certain features in one domain which are
suggestive of particular organizations in an underlying domain, are also represented — see Section 4.2.7 for a discussion of image features suggestive of 3-D organizations. Also included are knowledge of the basic properties of the three domains, and the relationships between basic properties and more complex properties derived from them.

Some of this information might be represented as procedures for carrying out the transformations or recovering the relationships, rather than as data, but the distinction is obscure. Roberts (1965), for example, has represented the projective transformation as a procedure for mapping points in the 3-D field into points in its projected image. It is not at all clear how much information about the world should be embodied in the processing stages, and how much should be represented in the information store (so that it can be employed by more general processing stages). Probably, the relatively fixed information should be embodied in procedures, and more transient information should be represented in the information store.

Another type of information in the data base is that concerning the current image, as expressed in its basic encoding, and in the 2-D image and 3-D field derived from this (depending on the stage of processing). To the extent that a meaningful interpretation of the input image has been achieved, its description will be largely in terms of information represented elsewhere in the information store.

The organization of information is crucial, from the points of view of both economical representation and facilitation of interpretation (an exhaustive search of the store for relevant information at each stage of processing being unacceptable). Information concerning both instances and classes could be richly structured in a hierarchical manner, with very simple components being configured into more and more complex arrangements up to the level of objects, with further associations between sub-classes, classes, super-classes, and so on. Economy of representation results from the repeated use of the same "chunks" of information in encoding many classes and instances. When characterizing a new class of objects, a specification of the super-class or family allows many characteristics to be "inherited", and only the distinguishing features of the new class need be added.

The associations between items in the memory are bidirectional whenever
there is a symmetric implication, with both excitatory (positive) and inhibitory (negative) associations being encoded. In this way, having interpreted a local image arrangement as a known part, the known configurations in which this part participates can be indicated, thus facilitating the selection of potentially applicable organizations at each stage of processing.

With regard to implementation of the information store, there has been considerable research into devising flexible data structures for representing organized information in a conventional memory vector (Gray, 1967; Rutovitz, 1968). These data structures free the data representation from the linear memory structure (at the expense of additional memory used for indicating linkages between data items), although the memory structure can still be used where it is appropriate (e.g., for a simple tabulation of information (Stanton, 1969b)). The manner in which information is represented and employed in the model, indicates a rather different implementation to the conventional "passive" memory structure. An "active" memory, with a degree of distributed processing, could aid selection of potentially relevant information. Consider a complex arrangement of known component parts, related in certain ways, and assume that these parts and relationships are linked (bidirectionally) to more complex configurations in which they are involved. If there is a known configuration resembling the current image arrangement, and the associative links are "active", there will be a summation of activity in the representation of the known configuration, and it will thereby be "brought to mind", and suggested as a possible organization. Prefacilitation of certain configurations and objects by "set" and contextual information, could also be accounted for within such a memory organization.

4.3 Pollen-Interpretation

To clarify the operation of the model, and to illustrate the type of descriptions that arise, the envisaged operation of the model with regard to images of pollen grains is summarized below. The following discussion assumes the order of processing indicated in the block diagram of Fig. 4-3, but the main interest here is in the processes which do not rely on the application of prior knowledge of specific objects; the part that such knowledge plays in the interpretation is discussed
Three pollen grain images are used as examples, those of Thrift, Morning Glory, and Zinnia — see Fig. 4-33. These images are of interest because of the obvious 3-D nature of the fields portrayed, and also with regard to the extent of image organization which human beings can readily and confidently impose without prior knowledge of the objects represented. The discussion concentrates on the processes performed during the interpretation of the Thrift micrograph, with processing of the other two images being considered in less detail. The terms "Thrift", "Morning Glory", and "Zinnia" will be understood here to refer to the relevant micrographs in Fig. 4-33.

The first stage of processing is that of selecting a small area of the image for attention, as discussed in Section 4.2.1. At the coarsest resolution, the whole image is examined for features such as low resolution contours which are then attended to at a slightly finer resolution, and any interesting features, organizations, or unanswered questions (such as what has caused an interruption in a contour found at a given resolution), call for closer attention, and so on. During selection of an area for close attention, the processing cycle of blocks A, B, C, D, E, and F of Fig. 4-3, might be repeated several times with coarse resolution detail. The area eventually attended to in fine detail may thus be selected as a result of a considerable amount of processing. Because of these processes of selective attention, certain fine detail in the image might never be attended to unless accompanying coarser resolution detail calls attention to it. With regard to the three examples, if Thrift is examined at a coarse resolution, a series of connected white "bars" on a black background (rather than black hexagonal "blobs" on a white background) emerges because of the obvious constraints existing between the gray scale contours, and because of the repetition of the adjoining black regions on both sides of the bars. One of these bars will probably be selected for closer examination. In Morning Glory, the only clear structure at a coarse resolution is an ordered arrangement of white "blobs" on a black background, and one of these blobs is

---

1 Coarse resolution detail can be exposed by viewing the given images through several thicknesses of tracing paper (placed either in contact with the images or slightly above them). Restriction of the area of attention can be simulated by viewing the images through a circular hole in an opaque sheet which masks the surrounding image area.
FIGURE 4-33 Pollen Micrographs used as Processing Examples

Figs. 4-33a,b copyright Patrick Echlin and Cambridge Scientific Instruments; reproduced by kind permission.
attended to initially. With Zinnia, the spine at the centre bottom is reasonably obvious at low resolution and will probably get the initial attention, but one of the light areas projecting into the black area at the top of the image might be chosen instead.

The next stage of processing is an attempt to impose an organization on the local area attended to, by taking abstracted features and evaluating and choosing from any suggested organizations (attempting to resolve any questions which arise), with this sequence repeating in the local area until no further local progress is being made. It is not anticipated that (except for simple configurations such as printed characters) specific prior experience of complete objects will be suggested (or will even be relevant) in these early stages of processing. As noted above, a measure of organization will have already been imposed at coarser resolutions, and the task now is also one of fitting in fine resolution detail with this coarser organization. The processes are, in this sense, parallel rather than sequential. Thus, assuming that one of the white bars (apparent at coarse resolution) in the centre of Thrift has been attended to, it will have already been suggested at a coarser resolution that this is a white bar in front of (or on) a black background, and the problem arises of fitting in the fine detail with this scheme. An enlargement of one of the bars (corresponding to an assumed local area of high resolution visibility) is shown in Fig. 4-34a. The variations in gray level within this bar can be interpreted as either changes in pigmentation or as 3-D undulations, but interpretation as undulations would be preferred. The 3-D cross-section of this bar is not clearly indicated, but an assumption of rotational symmetry (with the undulations added to a cylindrical bar) is possibly the simplest in the local context; little further interpretation is possible without consideration of other areas of the image.

For Morning Glory, one of the low-resolution white areas (seen as light

---

1 If selected for attention, the confusing surroundings at the top of this spine would soon cause a shift in attention because this area is difficult to interpret, without having first interpreted one of the other white areas, as a conical spine perpendicular to a spherical surface.

2 In the case of the scanning electron micrographs of pollen fields, it is known that the local variations in gray level will almost certainly result from 3-D undulations.
Note: This image has been inverted deliberately to enhance the ambiguity.

Are there light figures on a dark background or vice-versa?
blobs on a dark background) will have been attended to, say the area depicted in
Fig. 4-34b. The left hand boundary of this blob is clear, but parts of the bottom and right hand boundaries are vague at fine resolution (although they are reasonably clear at a coarser resolution). Nevertheless, symmetry is observable and suggests a 3-D shape for the blob, as discussed in Section 4.2.7.3. As with Thrift, further local processing is difficult at this stage.

For Zinnia, there is initially a certain amount of ambiguity at the top of the image, with the possibility of dark pointed areas on a light background or vice-versa, as is illustrated in Fig. 4-35. At fine resolution, the black areas may be seen as figures with "halos" of light adjoining their boundaries, but when a configuration such as that shown in Fig. 4-34c is encountered, organization as dark areas superimposed on light becomes awkward, and is rejected in favour of the converse organization; other factors, such as the lack of detail in the dark areas (or perhaps even the fact that they are dark), may also indicate that these areas are background. Accepting the light areas as figures, the question of 3-D shape arises. Consider Fig. 4-34c; the association of boundaries (on the basis of orientation, position, sharpness, and other attributes) is very strong, and in the absence of further information, rotational symmetry might be inferred, this inference being supported by low resolution gradients inside the boundaries, suggesting a changing 3-D surface aspect. In this case, however, the image boundaries are nearly parallel, and the question arises as to whether in the 3-D field they actually are parallel, so that the bounded object is cylindrical but receding, with the axis of symmetry pointing away from the image plane. Such an organization is consistent with the local detail and, although to the human perceiver it is clearly awkward in the global context, rejection in favour of an alternative interpretation has to await further processing.

At this stage, the image description is already partially in terms of projections of a 3-D field, which is itself described in terms of constrained surfaces of solid volumes, and possibly (as in the case of Thrift) 3-D additions to these surfaces. At this stage the descriptions have many "loose ends" and unspecified properties and relationships (such as how far one object is in front of another), to be "tied up" during later processing. Note that associations between objects interpreted as
being similar at coarser resolutions, such as the white blobs in Morning Glory, suggest that the organization perceived for the blob examined closely, also applies to the other blobs. As noted in Section 4.2.6, once a satisfactory local interpretation of the image as a projection of a 3-D field has been derived, the variations of gray scale in this area may well be unimportant, and might not be encoded.

When processing of a local area has proceeded as far as is currently practicable, there is the problem of selecting the next area to be attended to. In many cases, some question related to how the current area fits into its surroundings will have arisen, and the next area examined will probably be adjacent to the current area. In other cases, such as Morning Glory, the surroundings are of less immediate interest than another area previously associated at coarse resolution with the area just examined, but which has not yet been attended to in detail, and such an area may well be selected for subsequent processing in preference to an adjacent one.

With Thrift, it is anticipated that the next stages of processing involve examination of several more bars, seeing how they fit together into a larger 3-D structure. During these stages, several points might be noted:

1) The shape of the hexagonal and pentagonal cells becomes elongated towards the left hand side, suggesting that if the individual cell shapes are the same, then the surface aspect is changing.

2) This suggestion of changing aspect is reinforced by the changing 2-D shape of the undulations on the bars. The side view of these undulations suggests that they are conical in shape, thus refining the perception of the undulations on the bar first attended to (on the assumption that all the instances of bars and undulations are similar in form), and suggesting that boundary undulations such as "" in Fig. 4-33a, result from silhouettes of undulations which are not visibly distinguished from the bar.

3) There is evidence (at the sides of the image) which suggests that the "bars" are actually parallel-sided plates, and that the initial interpretation of the bars as rotationally symmetric, is not globally the simplest — the bars are better interpreted as parallel-sided plates with spines added to their outer edges.
4) The repeating structure of parallel-sided plates is found to cease just to the right hand side of the image's centre, and also to repeat on the other side of an elongated hole or furrow, which can thereby be seen as a type of figure embedded in this structure.

5) The overall boundary shape of the structured region suggests that the grain approximates either an oblate or a prolate spheroid, but the latter suggestion is reinforced by the curvature of the furrow, by the ways in which the projected shapes of the individual cells vary, and by an assumption that the parallel-sided plates are vertical to the surface. Perception of the overall shape as a prolate spheroid seems so natural to us, that it is interesting to identify the image features which suggest this interpretation.

Even in the absence of relevant prior experience, the final description derived for Thrift goes beyond what is visible in the image, and is a fairly drastic abstraction, using the prolate spheroidal surface shape as a base component in the description. Also included is a specification of the parallel-sided plates and the way in which they are conjoined and constrained in three dimensions (e.g., the tops of the plates being approximately level with, and perpendicular to, the prolate spheroidal basic surface), a rough description of the conical undulations (together with their dispositions and sizes relative to the plates), and a specification of the furrows interrupting the surface structure. The lengths and widths of the various components are specified in relation to the more complex components of which they form part. The heights and base widths of the conical undulations are, for example, specified in terms of the widths and lengths of the plates on whose outer edges they are superimposed, as is their orientation.

If a class of grains whose overall organization is similar to that perceived in the input image is represented in the information store, then the relevance of this class characterization will be suggested partly by the inferred global shape, partly by the parallel-sided plates plus spines, and partly by the way in which these plates are structured. In the case of relevant prior experience being indicated, any remaining ambiguities will probably be resolved, and much of the organization of the perceived input object can be implied by employing the class characterization.
as a base component in this object's description.

Assuming that the next areas of Morning Glory examined in detail are other white blobs close to the top of the image, there is an interaction between the interpretations of the various blobs — as the silhouettes of spines become visible at the top of the image, the central white areas are interpreted as projections of similar spines viewed end-on. The interposition of some spines in front of others, and the assumption that they are perpendicular to their base, leads to the suggestion that the underlying surface is spherical. This suggestion is reinforced when the changing shape of the 5- and 6-sided cells on the surface is noted. With Zinnia, the preferred organization of the spines as being perpendicular to their base, and inconsistencies of relative depths, leads to rejection of the organization in which the objects joined to the spherical surface are seen as being cylindrical but receding in depth. As an example of a depth inconsistency, the top of the bottom-most spine has to be much further rearward than the base of the next lowest spine (which it is obviously in front of) if the spines are interpreted as cylindrical but receding.

Thus, by means of processes similar to those outlined, it is anticipated that it will eventually be possible to mechanically interpret images as complex as pollen micrographs. Interpretation in terms of specific knowledge of classes of pollen grains has not been considered in detail above, but revolves around the use of prior knowledge in suggesting groupings and in simplifying descriptions. It should be observed that in the process of interpreting a given image, there is a very close parallel between using specific knowledge obtained from previous images, and in using characterizations of detail in other parts of the current image, to resolve questions such as those of the 3-D shape of Thrift's "bars", or of Zinnia's "spines".

A conceptual model for API, which serves as a basis for discussion and for further research, has been outlined in the previous sections; as noted earlier, much of the discussion accompanying this outline is directed to further development of the model, rather than to explanation of its current specification.

The roles of simplicity and coherency in resolving the ambiguities inherent
in interpretation of the chosen class of images, have been stressed. Simplicity (and economy of representation) is achieved to the extent that the descriptive organization parallels that of the data being described. For naturally-given images, representation in terms of projected 3-D fields, superimposition of detail and resolution levels, direct versus subordinate relationships, boundary constraints, etc., is therefore indicated.

Formulation of adequate API systems has proved to be a very complex and difficult task, and the model presented is incomplete. Nevertheless, this model, together with the above discussion of numerous suggestions and observations relevant to its further development, is suggested to be a step towards the eventual specification of adequate systems.
Chapter 5

IMPLEMENTATION EXPERIMENTS

The previous chapter outlined proposals for a general API model. Attempts at simulating some very basic processes required in this model are described below. Referring to the block diagram (Fig. 4-3), the processes studied in detail (bearing interrelationships with other processes in mind) have been:-

(i) searching for features such as associations, differences, and constraints in the raw or partially processed image data (i.e., part of block B); and

(ii) suggesting possible elementary image organizations on the basis of any features noted (i.e., a smaller part of block C).

In the context of the overall API model, the scope of these experiments is thus rather limited, but interesting results have been obtained.

As the successes and failures have played an important part in the development of the proposed model, these experiments are therefore also of interest historically. Because of this developmental role, in which practical results have led to revisions of the underlying model, the experiments described do not adequately reflect the current specification of this model.

All the simulations have been performed on the ANU's IBM 360/50 computer, which has 256K bytes of core storage (where \(1K = 2^{10} = 1024\), and 1 byte = 8 bits). The simulations were initially programmed in PL/I, but it was found that the code produced by early versions of the PL/I compiler was rather inefficient, and later simulations were programmed in FORTRAN (despite its limitations as a programming language when compared to PL/I) with occasional ASSEMBLER subroutines (e.g., for addressing individual bits in arrays of 1-bit elements). Listings of a few of the more interesting subroutines developed are included in Appendix 5.

One decision made at the start of these experiments was to work with real-world images in all their complexity, rather than synthetic or artificially simplified images. Evans (1968a,b) and others have experimented with synthesized inputs, so that they could concentrate on certain aspects of the processing.
There are several dangers in this simplified approach:

(i) the problem areas in API are not independent – the concepts and techniques derived from study of one area independently of the others, may not generalize (Narasimhan, 1968);
(ii) the techniques developed with simplified images may not be able to cope with the flood of raw information in typical real images; and
(iii) the processes postulated for deriving simplified images from the real-world inputs, may be impossible to implement as a separate stage independent of later processing (refer to (i) above).

One class of real-world images studied has been scanning electron micrographs of pollen fields (see Section 1.2). Whilst pollen micrographs have been very useful for conceptual development of the proposed model, processes for deriving satisfactory 3-D interpretations for this class of image seem to be very difficult to implement; for this reason it was found expedient to consider a second, more restricted class in the experiments, viz., micrographs of tracks made in photographic emulsion by magnetically-deflected low-energy alpha particles (resulting from nuclear interactions) — see Fig. 5-1. By counting the number of tracks per unit surface area of the emulsion (discarding extraneous tracks which are characterized by incorrect orientation and/or length), an energy spectrum for the alpha particles can be determined. If the thickness of the emulsion is neglected, the objects portrayed in the input images can be regarded as 2-D — a major simplification over the pollen images. The interest here with regard to the nuclear-particle tracks, is not so much in counting genuine tracks, but rather in identifying (and developing recovery techniques for) relationships between and within image entities, which cause us to perceive these images in terms of track segments, tracks of several segments, intersections of tracks, "blobs", and so on.

The implementation experiments are described in detail below. A brief discussion of these experiments is given in Section 5.2 (more detailed discussion being postponed until the following chapter).

5.1 Computer Simulations

Having decided to experiment with real rather than synthesized or artificial
Notes: (i) Tracks of interest are approximately vertical.
(ii) Magnification \(\approx 1000 \times\).

**FIGURE 5-1** Tracks made by Low-Energy Alpha Particles in Photographic Emulsion

<table>
<thead>
<tr>
<th>+1</th>
<th>+1</th>
<th>+1</th>
<th>-1</th>
<th>-1</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

Detector for vertical contour

**FIGURE 5-2** Equal-Weight Edge-Detector
images, the problem of digitizing input photographs arose. It was considered that at least 32 gray levels, and a matrix of order 128 x 128 points \(^1\) were required for satisfactory representation of the inputs. Although picture scanners are available commercially (Butler and Butler, 1967; Pearcey, 1968), the expense of obtaining and interfacing a standard picture scanning system could not be justified, because of the small number of pictures to be scanned in the developmental stages of this project. The practicability of constructing an inexpensive (but possibly slow) picture scanner, was therefore examined. It was decided to develop a scanner using a modified analogue X-Y recorder (whose pen assembly was replaced with a scanning head) connected to the IBM 360/50 computer via an analogue to digital converter (ADC) and digital to analogue converters (DAC's) which had previously been installed for seismic signal processing. The scanning head of the instrument consists of a light source and a lens which images the photograph onto one of two matched photo field-effect transistors. The complete system, which is described in more detail in Appendix 1 (published in Macleod, 1970a), has a usable maximum speed of 200 points per second (digitized on a 0.010" grid), and is estimated to be capable of resolving at least 64 gray levels. This rate of digitization is adequate for the purposes of the present investigation (but is very slow when compared to certain other systems, e.g., Ledley et al.'s (1965) FIDAC can digitize 35mm. transparencies at approximately \(10^6\) points per second).

To complement the picture input system, a picture output system was required for examination of processed images. A CRT was connected to the 360/50 computer via the DAC's, and the beam was modulated by the element values of the digitized image to be displayed, while sweeping the screen in a TV-type raster. Good quality images were obtained (and could be photographed for a permanent record), but in the ANU's batch-processing computing environment, this system

---

\(^1\)The dimension 128 (= \(2^7\)) was chosen for compatibility with fast Fourier transform processing (Cooley and Tukey, 1965). Elements in the array are referred to below by I and J co-ordinates for rows and columns respectively, with \([I = 1, J = 1]\) at the top left-hand corner, and \([I = 128, J = 128]\) at the bottom right-hand corner.
was found to be rather inconvenient, and an alternative output system, using the computer's line-printer, was developed. The quality obtainable from this system is inferior to that of the CRT system, but the convenience outweighs any loss of quality, and the line-printer system has normally been used. By the use of up to eight overprinted characters (to give a reasonable black-white contrast), linear interpolation (to overcome the distortion arising from the printer's 10 characters per inch versus 8 lines per inch), and statistical rounding of the desired gray level to one of the two nearest corresponding overprinted character combinations (to reduce the obviousness of quantization errors), acceptable hard-copy images are obtained — see Appendix 2 (published in Macleod, 1970b) and the program listings of subroutines PICOUT and SQPIC in Appendix 5.

The input image itself is regarded as an undifferentiated whole, with the segmentation into digitized picture points being viewed as simply a coding convenience. The most basic property of the input image at any chosen point is gray level, and the simplest feature to be searched for within the initially undifferentiated image is a sudden and perceptible change in gray level with distance. Such changes have usually been referred to as "edges", although "contours" might be a better term. Several different forms of "edge-detector" were tested by correlating them with digitized images and comparing the magnitude of correlation with perceived contour magnitude. Hawkins et al. (1966), Rosenfeld et al. (1969), and others have employed equally-weighted edge-detectors of the form shown in Fig. 5-2, but in the present study, it was found that these detectors were either too sensitive to small local variations (if they were made to cover only a few matrix cells), and/or too sensitive to variations near their periphery rather than close to their axis (if they were made larger). These shortcomings were found to be mainly due to the use of equal magnitude weights; although the use of such weights simplifies computation, a better correspondence with perceived edges has been achieved by using variable weights (with small weights being given to cells near the periphery and along the axis). Kulikowski and Parks (1967) report experiments with simple

---

1 Hochberg (1964) points out the difference between an edge and a contour — a contour can exist by itself, but an edge is a boundary of some object, and is not an edge without that object.
variable-weight detectors, but the edge-detector finally developed in the present study is more complex than theirs, and consists of superimposed displaced exponentials such that the correlation mask, centred on \( [x = 0, y = 0] \), is

\[
F(x, y, \theta) = e^{-\left[ \frac{x^2+y^2}{d_r^2} \right]} \left[ e^{-\left[ \frac{x \sin \theta + y \cos \theta + d_p}{d_p} \right]} - e^{-\left[ \frac{x \sin \theta + y \cos \theta - d_p}{d_p} \right]} \right] 
\]

Eqn. (1).

where \( \theta \) is the orientation of the edge, and \( d_p \) and \( d_r \) determine the rate of decay perpendicular to the axis of the edge, and radially, respectively. A typical mask is illustrated in Fig. 5-3. The advantages of this type of detector relative to fixed-weight masks, are that:

(i) the magnitude of correlation is a function of detail close to the point on which the detector is centred rather than of that close to its periphery (as a result of the \( e^{-\frac{x^2+y^2}{d_r^2}} \) decay); and

(ii) because the maximum weights are not immediately adjacent to the detector's axis, a detector designed for coarse resolution contours gives only a small correlation with strictly local sharp edges, even if these do lie along the axis.

The interpretation of a given image should not change with overall variation in the illumination of the 3-D field portrayed. The ratios of projected light energy at different parts of the image remain relatively constant as this illumination changes; ratios of gray level (on a gray scale which is linear with the energy of the light incident on the image) are therefore more important than absolute gray level or arithmetical differences in gray level. To improve the response to contours projected from poorly illuminated parts of the 3-D field, a non-linear scaling of the digitized image (whose original element values are linear with the energy of light reflected from the image) is carried out prior to edge-detection, because the cross-correlation of an edge-detector mask with the digitized image is a function of arithmetical differences of the digitized values. A pure logarithmic scaling should ideally be employed, but quantization and noise errors in dark regions of the image...
(a) Element values for $11 \times 11$ detector, with $\theta = \pi/4$, $d_p = 2.0$, and $d_r = 3.0$ in Eqn. (1)

Note: Gray = 0.0, black = +1.0, and white = -1.0.

(b) Line-printer representation of a $128 \times 128$ detector, with $\theta = \pi/4$, $d_p = 15.0$, and $d_r = 22.5$ in Eqn. (1)

FIGURE 5-3 Examples of Variable-Weight Edge-Detectors
are abnormally accentuated in this case. By the addition of a constant to all element values prior to logarithmic scaling, this accentuation is reduced, while the desired effect (of making edge-responses a function of gray level ratios, rather than of arithmetical differences) is largely achieved.

It is well known that human perception of sensory differences is, over a wide range of magnitudes, proportional to ratios rather than differences of sensory energy (Munn, 1961, p. 585), and the logarithmic scaling process helps establish a closer correspondence between perceived and computed contours. This process can therefore be regarded as "psycho-physical" scaling. Roberts (1965) employs a square-root scaling for this purpose, but in the present experiments, the modified logarithmic scaling gave better results.

The form of edge-detector employed is inherently directional, and gives a maximum correlation with an edge in the image at the same orientation as the detector. Non-directional edge-detectors such as a "Laplacian Filter" (Kanal and Randall, 1964) were tested, but were found to be overly sensitive to "spots" relative to "edges", and did not provide any information regarding the direction of detected edges (such information being required for later processing). As a result of the inherent directionality and scale of the chosen detector, several detectors (corresponding to edges at different orientations and at different resolutions) should be used. By suitable choice of $d_p$ relative to $d_r$, the response of the edge-detector for edges aligned at up to $22\frac{1}{2}^\circ$ from the axis, can be kept to 80% or so of that for "on-axis" edges. Four edge-detectors (aligned at $\theta = 0$, $\pi/4$, $\pi/2$, and $3\pi/4$ counterclockwise relative to the I axis) were sufficient to give a good response to edges at any orientation (negative responses corresponding to $\pi$, $5\pi/4$, $3\pi/2$, and $7\pi/4$ respectively). The results of correlating a typical mask with Thrift are illustrated in Fig. 5-4; in contrast to subsequent processing examples (which are in terms of nuclear-track images), the Thrift image has been used here to illustrate the dependence of correlation magnitude on contour orientation. "Spot-detectors", formed by the addition of radially symmetric exponentials having different radii, have also been developed, but these, together with edge-detectors covering a range of resolutions, have not yet been incorporated in the processing.

Typical correlation masks used up to 15 x 15 cells, and it was found that the time to correlate these with the 128 x 128 cell digitized image, using the conventional technique (i.e., superimposing the mask matrix on the image matrix, multiplying
FIGURE 5-4 Correlation of Edge-Detectors with Digitized Thrift Image: (a) Line-Printer Representation of Digitized Image (portion of Fig. 2-5a)

(in this and subsequent printouts, the aspect distortion resulting from the line-printer's 10 chars./in. vs. 8 lines/in., has not been reduced by interpolation)
FIGURE 5-4. Correlation of Edge-Detectors with Digitized Thrift Images: (b) Correlation with Edge-Detector at $\theta = \pi/4$, $d_p = 1.5$, and $d_r = 2.25$

(white = correlation value of 0.0, black = correlation value of 0.70, in the array represented pictorially above)
FIGURE 5-4: Correlation of Edge-Detectors with Digitized Thrift Image: (c) Correlation with Edge-Detector at $\theta = \pi/4$, $d_p = 1.5$, and $d_r = 2.25$

(white = correlation value of 0.0, black = correlation value of -0.65, in the array represented pictorially above)
FIGURE 5-4 Correlation of Edge-Detectors with Digitized Thrift Image: (d) Union of Four Correlation Arrays at $\theta = 0$, $\pi/4$, $\pi/2$, and $3\pi/4$

(this array has been formed by setting each element to the maximum absolute value of the corresponding elements in the four correlation arrays; white = correlation value of 0.0, black = correlation value of 0.73)
corresponding cells, summing the products, and repeating this process for each possible superimposition) was impractically long — approximately 8 minutes per correlation on the 360/50. The potential of Cooley and Tukey's (1965) fast Fourier transform (FFT) for reducing the number of arithmetical operations in correlation of time series (and data of higher dimensionality) has been pointed out by Stockham (1966). By using FFT, substantial time savings have been realised in computing the correlation of edge-detector masks with the digitized images — the correlation of a 128 x 128 image taking less than 60 seconds on the 360/50.

HARM, the IBM-supplied radix-4 FFT routine, requires auxiliary storage (for tables of sines and bit-reversed numbers) equal to one quarter of the storage required for the vector of values to be transformed. With the chosen size of 128 x 128 for the digitized image, and in the normal configuration of the ANU's 360/50 (i.e., two fixed size core partitions, of 106K bytes and 108K bytes respectively, with independent tasks in each), this auxiliary storage is not readily available. For this reason, a 1-D radix-8 FFT routine, which has a number of advantages over radix-2 and radix-4 algorithms (Macleod, 1969b), was developed — see listing of subroutine COOL in Appendix 5. The main advantage for the picture correlation application is that the number of references to sine and cosine subroutines, and the number of bit-reversed numbers required, are reduced. The overhead in indexing tabulated sinusoid values for one quadrant (0 to π/4) when computing sines and cosines for all quadrants, is such that little if any penalty is incurred in the radix-8 algorithm by making direct reference to sine and cosine subroutines for one sine-cosine pair, and computing the other six pairs required for each iteration by sums and products rules of the form sin 2A = 2 sin A cos A; similar remarks apply with respect to the bit-reversed numbers. Other advantages are that the number of arithmetic operations in the body of the algorithm is reduced,

Stockham's claim (1966, p. 232) that, for data of dimensionality > 1, the "time savings depend on the total number of data points contained within the entire data space in question, and that they depend on this number in a manner similar to that characterizing the one-dimensional case", is not correct — in so far as higher-dimensional discrete Fourier transforms are equivalent to multiple summations, and are effected by repeated transforms of 1-D vectors in the higher-dimensional space, the savings are more closely related to the maximum dimension of the data space.
and that in computing the 2-D Fourier transform by first transforming the columns of the data array and then transforming the rows, the computation can be performed in little more storage than that required for the original (real) data. COOL has been developed to compute the FFT of a vector of adjacent data elements; in trying to compute the FFT of the rows (or the columns, depending on the array mapping function employed) the problem arises that the row elements are not adjacent in storage. One solution to this problem is to transpose the matrix, recompute the FFT of each of its columns, and transpose it again. Because of storage limitations, in-situ transposition (without an auxiliary array) is desirable. A published in-situ transposition algorithm (Boothroyd, 1967) was modified (Macleod, 1969c) and used initially for this process. Although in the current application the matrices were square, the inefficiencies of this modified algorithm, when applied to certain non-square matrices, motivated the development of a new algorithm for in-situ transposition (and other permutations) which is very much faster for some non-square matrices, and slightly faster for square matrices — see Appendix 3 (published in Macleod, 1970c).

Using the approach outlined above, the digitized input picture is psycho-physically scaled, transformed, and written onto disc backing store. For each of four orientations, an edge mask is generated in the original data array, and is then transformed, multiplied (a column at a time) by the complex conjugate of the transformed input, and inverse transformed, to yield the correlation array for that orientation (which is then written to disc for later processing); see Fig. 5-4 for examples of typical correlations.

The next stage of processing invokes a routine RATND (see listing in Appendix 5) which searches through the correlation matrices, one at a time, calling attention to certain "edge-response" points in these arrays by flagging the corresponding locations in one of eight 128 x 128 arrays of 1-bit elements, depending on the orientation at which the edge-response is detected, and on the correlation polarity (e.g., negative response at \( \frac{\pi}{4} \) = positive response at \( \frac{5\pi}{4} \)).

Points are flagged only as they satisfy (for any of the four correlation matrices) all of the following criteria:–

1) The magnitude of correlation, \( M \), must be greater than a threshold
value, \( A \), i.e., the contour must be perceptible.

2) The magnitude of the maximum second derivative of correlation must be greater than a threshold value \( B \) multiplied by \( M \), and its polarity must be opposite to that of the correlation, i.e., there must be a sharp peak in the amplitude of the correlation in this direction — the contour must be sudden.

3) The correlation magnitude at the point being examined must be no smaller than that of adjacent points on a cross-section of the correlation array taken in the direction of the maximum second derivative, i.e., this point should be a local maximum across the direction of the contour.

4) The magnitude of the derivative of the correlation along the contour (perpendicular to the maximum second derivative) must be less than a threshold value \( C \) multiplied by \( B \), i.e., the correlation must not be decreasing too quickly along the contour, because this suggests that the correlation is due to some extraneous effect. For example, in Fig. 5-5, it is possible for a point in the correlation matrix at point 1 to satisfy the first three criteria, so that without this additional criterion, point 1 could be flagged as an edge-response.

5) The magnitude of the first derivative across the edge must be less than a threshold value \( D \) multiplied by \( M \), i.e., the edge-response should correspond to a contour between two areas of distinct but relatively constant gray level, rather than to the boundary between an area of constant gray level and an area of constant gray level gradient, which joins the first one without a sudden step in gray level.

6) The magnitude of the second derivative along the contour must be less than a threshold \( E \) multiplied by \( M \), or the polarity of the second derivative must be opposite to that of the correlation at the point, i.e., this point must not lie on a distinct "saddle" such as might result in the situation portrayed in Fig. 5-6, where criterion (4) might still be satisfied.

The above criteria are not independent, and interact to some degree, thus making the choice of suitable thresholds difficult, but when suitable thresholds have been
Note: Edge-detector oriented at $\theta = \pi/4$, i.e.,

**FIGURE 5-5** Requirement for Limit on the Rate of Change of Correlation Along an Edge

Note: The correlation array is not shown, but is considered to be superimposed on the image detail.

**FIGURE 5-6** Saddle-Point in Correlation Array
chosen, the edge-responses flagged correspond reasonably well to perceived contours in the input images. Fig. 5-7 shows the responses for one correlation matrix, and the union of responses in all matrices, for a typical edge-detector and threshold settings, using a nuclear-track image as input (refer to p. 156 para. 2).

The next stage of processing involves the concatenation of edge-responses into boundary strings, according as adjacent edge-responses are flagged as being of similar directions. Note that the edge-responses themselves are not pictorially identifiable entities, so that the boundary strings formed cannot be regarded as organizations (i.e., groups) of simpler units. It is at this stage that the question of data representation becomes important, because the boundary strings are of unknown lengths and will eventually become associated with other boundary strings, by virtue of certain relationships (which have to be represented) obtaining between them. The flexibility required indicates some form of data structure, and Weizenbaum's (1963) symmetric list processor "SLIP" (a set of FORTRAN-callable FORTRAN and ASSEMBLER subroutines) was chosen for its flexibility, efficiency, and ease of implementation (rather than for its symmetry, which is intra-list but not inter-list). SLIP is intended primarily for tree-type data structures, but these were found to be inadequate for the present application, and additional cross linkages between branches have been employed. In the process of implementing SLIP on the ANU's IBM 360/50 computer, a number of corrections and improvements to the original was made — see Appendix 4.

Edge-responses at each of the eight orientations are flagged in a 128 x 128 x 8 array (POSBIT) of 1-bit elements. The routine TRACE (see listing in Appendix 5) is the major component of the contour tracing program and forms SLIP lists of cells constituting a boundary segment, and places a pointer to these individual lists on a master list (of boundary segments) as each segment is recovered. TRACE sweeps through POSBIT in a TV-type scan until a flagged point is found which has not already been accounted for in tracing out an earlier boundary segment (as indicated by the corresponding bit being flagged in another bit array (PICBIT)). A new list is formed and this point is placed on it, with the 32-bit data field of the SLIP cell containing an 8-bit direction field, an 8-bit correlation magnitude field, and two 8-bit I and J location fields. The magnitude and direction data are
FIGURE 5-7  Edge-Responses in Nuclear-Track Image: (a) Line-Printer Representation of Digitized Image (portion of Fig. 5-1)
FIGURE 5-7. Edge-Responses in Nuclear-Track Image: (b) Responses at $\theta = \frac{3\pi}{2}$

(threshold values used are: $A = 0.15$, $B = 0.10$, $C = 0.10$, $D = 0.50$, $E = 0.375$)
FIGURE 5-7  Edge-Responses in Nuclear-Track Image: (c) Union of Responses at Eight Orientations ($\theta = 0$ to $\pi/4$, in steps of $\pi/4$)
obtained by correlating a series of four edge masks with the input image at the selected point, deducing the orientation of the actual contour by finding the phase of the fundamental component of the correlation magnitude from the four measured magnitudes, and by finding the correlation of a specially generated edge mask at this orientation. To facilitate later processing, code is included to assign a direction of greater than 2\(\pi\) or less than 0, if this is necessary to prevent a sudden jump in direction magnitude (e.g., 6.1, 6.2, 0.02, 0.12, etc. as the angle exceeds 2\(\pi\), is coded as 6.1, 6.2, 6.3, 6.4). Having added the first cell to the new list, a search is made for continuations, tracing initially in the direction of black on the left. The criteria for a cell to be considered as a continuation of a boundary string are that:

1) It must not already be accounted for on another boundary string.

2) It must be one of the three cells adjoining (perhaps diagonally) the current cell, on the side to which the current edge-response points.

3) It must be flagged in one of the same bit arrays that the current cell is flagged in, or in another bit array corresponding to a direction of \(\pm \pi/4\) to one of these currently flagged arrays.

All the three possible candidates are examined as continuations, and any satisfactory cells are added to the top of the current list, with the search for new continuations being re-initiated from the current cell with the largest correlation, thus avoiding problems of bifurcation. When no satisfactory continuation is detected, one end of the current boundary string has been found. The initial entry cell is returned to, and a trace in the other direction (with black on the right) commences, this time adding new cells to the bottom of the current boundary list. This reversed search is necessary because the entry cell may be in the middle of an arched boundary segment. When no continuation can be found in the reversed search, the TV-type scan for an entry cell on a new boundary segment recommences, and tracing of boundaries continues until all bits flagged in POSBIT have been accounted for in boundary strings. A printout of the various boundary segments traced in processing an input image is shown in Fig. 5-8; each boundary segment is indicated by a continuous string of identical characters, starting with an asterisk unless there is only one cell in the segment. This method of plotting the cells on the boundary lists is very useful for displaying the
FIGURE 5-8 Boundary Strings Formed by Associating Edge-Responses
(threshold values employed in deriving these edge-responses differ from those employed in Fig. 5-7)
results of processing, but some interesting problems arose in ensuring that two adjacent boundary curves were not represented by the same printed character — see listing of subroutine LSTPLT in Appendix 5. A 128 x 128 array of characters is initialized with blanks, and LSTPLT is called to plot each list of interest, before printing the array out.

Because of the method of searching for continuations in TRACE, a boundary curve drawn along a boundary string by joining cells which are adjacent on the SLIP lists, may be kinked, as illustrated in Fig. 5-9a. A routine STRTN (see listing in Appendix 5) examines the boundary strings one at a time to see whether it is possible, by rearranging the order of the cells, to straighten out the boundary curve. First of all, a check is made for "gaps" in the given curve, these being defined as points in the boundary string where the centres of cells adjacent on the SLIP list are greater than $\sqrt{2}$ apart in the image matrix. Unless there has been an error in tracing out the boundary string, it will always be possible to eliminate these gaps by reordering the cells, as is shown in Fig. 5-9a. Even after the gaps are eliminated, it may still be possible to reduce the "kinkiness" by further re-ordering, and a second pass through the boundary string tests for this possibility and effects any re-ordering required, see Fig. 5-9b.

The boundary strings formed by TRACE are only a start; some of these should be joined to other strings without differentiating the junction point, and others should be segmented because of sudden and obvious changes in their internal properties. With regard to segmentation, there are parallels between the boundary strings at this stage of processing and the initial input images; these parallels should be apparent in the following description.

The most basic property of the boundary strings is direction at some point relative to an axis or to the direction of another point. At this stage, this property is not available from the boundary strings, apart from the computed orientations of the contour angles of individual cells (but these are very variable). A routine RECODE (see listing in Appendix 5) is invoked to form strings of unit vectors scribed on the boundary curves. (The starting point and directions of the individual vectors form an alternative characterization of the curve, which is useful for the next stage of processing). RECODE operates by first scribing variable length
(a) Elimination of "gaps"          (b) Reduction of "kinkiness"

**FIGURE 5-9** "Kinked" Boundary String

(a) Boundary string 2 cells thick          (b) Vectors scribed by RECODE

**FIGURE 5-10** "Thick" Boundary Strings
vectors on the straightened boundary curve, joining either the centres of cells or the mean positions of pairs of adjacent cells (weighted by the individual correlations of the adjacent cells), according to the geometric disposition of the cells being joined. These initial vectors are usually not of unit length, and a second stage of RECODE scribes unit vectors on the initial vector string, and returns a SLIP list of the unit vector angles.

Because the boundary strings are formed using a union of the eight edge-response bit matrices, a boundary string such as that portrayed in Fig. 5-10a (with more than one cell across the width of the contour) is possible, even though in any of the four correlation matrices the local maximum across the contour is the only cell flagged. An interesting result in such cases is that RECODE, in scribing vectors between the weighted means of adjacent cells, places the boundary curve to an accuracy which may be better than ±1 cell spacing, see Fig. 5-10b.

A 1-D "angle-detector", composed of superimposed exponentials in the same fashion as the edge-detector described earlier, but with only one parameter (equivalent to resolution level), is correlated with the unit vector representations of the boundary strings, and a routine BREAK1 (see listing in Appendix 5) examines the 1-D correlation vector for "angle-responses", i.e., looks for sudden changes of angle between relatively straight boundary segments (cf. RATND). Angle-responses have to satisfy certain criteria, the three most important being:–

1) The change of angle must be large enough to be perceptible.
2) The change must be sudden and not part of a continuous curve.
3) The magnitude of correlation with the angle-detector must be a local maximum.

The results of applying BREAK1 to the boundary curves in Fig.5-8 are shown in Fig. 5-11; observe the segmentation of the curves labelled "A", "B", and "C" in Fig. 5-8.

The boundary curves are broken into separate segments (by a routine BRKLST) at the detected angle responses, and linkages between adjoining segments are set up by placing pointers on the boundary segments' attribute value lists. It is assumed that the individual segments can be adequately characterized as curves whose rate of change of curvature with distance is perceptibly constant, this character-
FIGURE 5-11 Results of Processing Boundary Curves with BREAK1 and OVLAP

Notes: (i) "Objects" defined by boundary constraints, are indicated by pairs of thick lines | |; only a few of the objects found are so indicated. (ii) Ambiguous boundary associations are indicated by j + + or + i.
erization including straight lines and arcs of circles as special cases. Individual segments are fitted, via a linear least-squares routine, for initial angle, initial curvature, and rate of change of curvature. The fitted curve is positioned relative to the actual curve for minimum mean-square error, and the co-ordinates of the start of the fitted curve, together with the three characterizing parameters, are added to the actual boundary curve's attribute value list. The characterizations of the boundary segments are completed by computing the ranges of I and J co-ordinates that the segments traverse (i.e., a rectangular "occupancy box"), and adding a SLIP cell containing $I_{\text{max}}, I_{\text{min}}, J_{\text{max}},$ and $J_{\text{min}}$ (as 8-bit fields in the 32-bit datum) to the top of each boundary segment list.

Having characterized the segments, the next stage of processing involves a search for various constraints between and within boundary segments. One constraint searched for is a tendency towards "closure" of single segments, or of several adjacent segments. This tendency is detected by computing distance "along" the curve versus Euclidean distance (distance "across") between pairs of points lying on the curve. The computation applies only to pairs of points separated by a specified minimum distance along the curve, to prevent small kinks being regarded as closures. A large ratio between the distance along and the distance across is an indication of closure — see the listing of subroutine NEAR in Appendix 5.

Another important constraint searched for is referred to herein as "correspondence" (for want of a better name), and is illustrated by the relationships between boundary segments depicted in Fig. 5-12a. The technique adopted for finding corresponding segments which in some sense "go together", is to examine pairs of nearby boundary segments\(^1\), identifying corresponding point pairs according as a line through the pair of points subtends nearly equal angles with the "insides" of the two boundary curves, as illustrated in Fig. 5-12b. The mid-points of the lines joining corresponding points can be thought of as forming an "axis of correspondence" (which need not be straight), whose characteristics might be used for predicting continuations of the "object" formed by the existence of the correspondence

\(^1\)Nearby segments are identified on the basis of the minimum distance between their occupancy boxes, relative to the segment lengths.
axis of correspondence

if $\alpha = \beta$, points 1 and 2 are said to be in "correspondence"

(a) Examples

(b) Identification of corresponding points

FIGURE 5-12 "Correspondence" Between Boundary Curves
constraint between nearby contours (which then become edges of this object). The tendency to interpret such constrained contours as edges of an object, is a function of the length of the corresponding parts of the contours, relative to the distance between them and also of the presence of other "corresponding" contours. A routine OVLAP (see listing in Appendix 5) searches through the boundary segments, identifies corresponding contours, and measures the distance between them and the length of correspondence (not necessarily the same for each contour in a corresponding pair, because of curvature).

"Objects" are defined as the areas bounded by pairs of corresponding contours for which the ratio of the length of correspondence to the distance apart is greater than a specified value. A few of the objects identified in this manner are indicated in Fig. 5-11 by pairs of heavy lines; ambiguous associations are shown by rows of dashes or crosses. It will be observed that this method of image articulation is not affected by changes of the size, orientation, or location of the "objects" detected.

The final output at present is a list of objects identified by their boundary constraints. Further image organization is required (e.g., associating objects (i.e., track segments) into longer tracks), but the necessary processes have not yet been included.

5.2 Discussion

The simulations described were developed over a period of more than 18 months, and involved a substantial programming effort (> 100 subroutines, not including the SLIP implementation). As implemented on the ANU's 360/50 computer, the final system was both slow (taking about 15 minutes processing to achieve the limited degree of organization indicated) and very demanding of core storage (extensive overlaying of programs being required). These experiments indicate, therefore, that rather more adequate simulations will be achieved only with the expenditure of considerable effort, and also that, given current information processing and storage technology, economical API systems are unlikely to be achieved except with restricted PI tasks or simple images.

The experiments have not been pursued further because of the obvious need
for revision, and the limited time available; there is, however, a clear need for the simulations to be extended and modified so that they reflect more faithfully the extent and nature of the proposed model — possible extensions are discussed in Section 6.2.4.

The simulations demonstrated that, at least for the class of images considered, the objects of interest are usually not outlined by a distinct image contour, despite the fact that much API research has, however, been based on an assumption that the objects of interest will have either consistent gray scale statistics relative to their surroundings, or will have sharply-defined closed boundaries.

The processes employed here do not rely upon objects having closed boundaries of the one type. Parts of objects' boundaries can be "virtual", in the sense that they are not present as image contours, but are defined structurally by the cessation of boundary constraints.

With regard to the presence of isolated boundary strings and gaps in objects' boundaries, it may be observed by comparing Fig. 5-8 with Fig. 5-7a, that these gaps do not always correspond to perceptible changes in the nature of the local image contour. This situation arises from the use of decision thresholds, whereby imperceptible differences in property values can lead to a change of classification; further comments regarding thresholds are given in Section 6.1.

Additional discussion regarding (a) the experimental results, (b) the relationship of the current simulations to the conceptual model, and (c) the nuclear-track interpretation task, is included in the following chapter.
Chapter 6

DISCUSSION

The research reported herein represents an attempt to identify and to come to grips with the many difficulties inherent in the overall API problem. Discussed below are:

(i) the results of the present study, its relationship to other work, and the contributions relative to the pollen-interpretation task (Section 6.1);

(ii) implications arising from these results and possible applications of the concepts and techniques developed (Section 6.2); and

(iii) the position of the present research in the light of the overall API problem, and recommendations for future research (Section 6.3).

6.1 Results of the Present Study

This has been a broad study of the problem as a whole, rather than an intensive investigation of a clearly delimited particular aspect; because of interdependencies between the various aspects, it has been considered that a global study should precede detailed investigation of selected sub-problems out of adequate context. The main contributions of the present study are an increased understanding of the nature and extent of the API problem, together with an elucidation of relevant factors for future research, rather than the production of systems with immediate practical applications.

One important result is an identification and clarification of the difficulties which must be faced in attempts to solve the overall problem — see Chapter 2. The main areas of difficulty include:

1) 3-D Fields:

If the objects of interest are characterized in terms of their 2-D projected images, problems arise of variations in their projections with changes in viewpoint, illumination, pigmentation, etc. Alternatively, if interpretation in 3-D terms is attempted, the question arises of inherent 2-D → 3-D uncertainties (which are
resolvable only through restrictive assumptions about the nature of the 3-D fields portrayed).

2) **Object Isolation:**

The properties on which segmentations of the image or of the perceived 3-D field are based, may be 3-D, articulal, contextual, and/or at various levels of resolution.

3) **Quantity of Detail:**

There are problems related to the amount of raw image detail, and to the extent of knowledge necessary for effective interpretation. At the start of processing, it is usually not possible to state which detail will eventually be unimportant, and abstraction of significant detail in an independent preprocessing phase may not be possible.

4) **Image and Object Organization:**

The organizations exhibited cover a wide range, and may be both complex and subtle. Identification of the inter-component relationships on which such organizations are based, is required, as is the development of techniques for recovering these relationships.

5) **Variability of Detail:**

The objects of interest form an "open" set, describable only awkwardly within a "closed" formalism. Metrical properties may assume values in an effectively continuous range, and anomalies can arise in basing classifications on such values — two situations which cannot be discriminated in any important respect, may be assigned different classifications.

6) **Ambiguity:**

Apart from the inherent 2-D → 3-D uncertainty, it is often not apparent in which way several components should be grouped, or at what points individual components should be segmented.

7) **Representation:**

Formation of information representations which enable relevant information to be accessed and used effectively, is an important but difficult task.
An examination of currently formulated approaches to API has revealed that little real progress has been made towards coping with the difficulties identified, and that all current approaches suffer basic inadequacies with respect to the API problem — see Chapter 3. It is argued here that template matching approaches are effective only for very limited classes of images. The property list framework is not as restricted, but fails with respect to the need to interpret in 3-D terms and to respond to image organization. Rather more capability is evident in the articulation analysis and picture parsing approaches, but both require much further development (particularly with regard to the range of organizations considered), and the latter is (as currently formulated) far too sensitive to image imperfections.

Another result of the present study is the development of a new and comprehensive model for API, which takes into consideration the difficulties identified, and serves as a framework in which a number of observations and suggestions relevant to further research are presented. Although it is as yet incompletely specified, this model is expected to be valid (at least in general terms), in so far as its development has been based on an evaluation of many possible approaches to coping with the problems involved, on a consideration of numerous examples taken from real-world images, and on the results of computer simulations. The processes proposed may, at first, appear to be overly complex, but in view of the intended generality and the problems addressed, this complexity seems necessary; it might even be shown by future research that further complexity is required.

Among the principles exhibited in the model are:-

1) Effective interpretations should be in terms of 3-D objects and arrangements, if the fields portrayed in the 2-D image are 3-D.

2) Interpretation should proceed iteratively over restricted (but growing) areas of the input image, so that by organizing image detail into more and more complex configurations (which function as units in subsequent processing), only a limited number of (possibly within themselves quite complex) components need be attended to at any one time. A final coherent interpretation is seen as being derived via many small steps, with alternative image organizations being attempted when inconsistencies arise — many small
pieces of local evidence are gathered and fitted into a global organization.

3) Generally applicable rules of organization are employed in the initial stages of processing each image area, and also at later stages if no relevant specific prior experience is found. In this way, objects which are unfamiliar (except to the extent that general rules of organization can be applied) can nevertheless be assigned an articular characterization, but the extent of processing possible may be limited by unresolved ambiguities and inconsistencies (which might, however, readily be resolved, given the necessary specific experience).

4) By suitable organization of the information relating to known objects, information relevant to an image area being processed may more readily be accessed and used to guide the interpretation process.

5) Either simple or known configurations should, where possible, be employed as "base" components in the description of a given image area, such components being "fitted" to this area by any required additional specifications or modifications.

6) Ambiguities may usually be resolved by appeal to simplicity (in terms of the knowledge represented within the API system) and consistency of description; to reveal inconsistencies, reasonably complete descriptions (which might later be abstracted from) should be derived at each stage.

7) If a particular image organization is suggested contextually rather than more directly (e.g., the suggested continuation of a boundary which has "faded out", or the inference of occluded surfaces of objects), then the suggested organization should be tentatively accepted, provided that it leads to a simplification of description, and that the absence of a more direct indication can be satisfactorily explained (e.g., as occlusion).

8) It is possible, by perceiving constraints within regions and relationships between distinguished regions (i.e., image or object components),
to develop descriptors or properties which are appropriate to the situation being described. To the extent that the descriptive system parallels the image organization (e.g., with superimposition of detail at various resolution levels, and specification of constraints existing at the boundaries between adjacent components), economy of representation is achieved.

9) The concept of pre-chosen primitive objects is rejected — the element values of cells in a digitized image are employed in the processing sequence envisaged (these values can thus be regarded as "processing primitives"), but in so far as these cells are usually not pictorially identifiable, they should not form part of the final description. Even in this final description, the simplest component parts are defined relationally (e.g., in terms of constraints between their boundaries, which are themselves defined in terms of the relationship "different from" applying to attribute values on either side of these boundaries). Attributes and relationships are therefore seen as being more basic than objects.

Experiments with computer simulations of some of the proposed initial stages of processing have aided in the development of the conceptual model, and in exposing some of the difficulties to be encountered. In the course of these experiments a number of contributions has been made. A simple and inexpensive picture scanner, based on an analogue X-Y plotter, has been developed and this has proved to be very satisfactory for the present purposes. To complement the picture input system, an output system which generates pictorial representations of data via the computing system's line-printer has also been developed. The quality of the output pictures is superior to that of previously reported line-printer techniques.

Experiments with "edge-detectors" resulted in the development of a detector suitable for a range of resolutions, and more responsive to "edges" (relative to other extraneous detail) than the fixed-weight detectors employed hitherto. In correlating edge-detectors with the input pictures, considerable savings of computer time were effected via the use of FFT. As a result of the correlation experiments, the following were devised:— (i) an improved FFT algorithm (COOL), based on 8-point
iterations, (ii) improvements to an existing in-situ transposition algorithm, and (iii) an improved algorithm for carrying out permutations of elements (including in-situ transposition).

Techniques for (a) detecting edge-responses in the arrays formed by correlating edge-detectors at several orientations with the input picture, (b) grouping individual responses into longer boundary curves, and (c) segmenting these curves into simple components, have also been developed. These simple components are described as curves with a constant rate of change of curvature, and relationships between components which suggest image organizations are searched for. In particular, a technique was developed for recovering the "correspondence" relationship between nearby curves.

The techniques described above have been applied to the recovery of simple organization in images of nuclear particle tracks in photographic emulsion, in the following manner:

(i) the input image is digitized and "psycho-physically" scaled (to compensate for the small numerical differences of gray level adjacent to contours in dimly illuminated parts of the image);
(ii) variable-weight edge-detectors at each of four orientations (0, \(\pi/4\), \(\pi/2\), and \(3\pi/4\)) are correlated with the input image by using FFT;
(iii) a search is made for individual positive or negative "edge-responses" in each of the four correlation arrays (negative responses corresponding to edges at angles \(\pi\), \(5\pi/4\), \(3\pi/2\), and \(7\pi/4\)), and these responses are flagged in arrays of 1-bit elements;
(iv) individual edge-response cells are associated into boundary strings, which are "straightened" if possible by rearranging the order of cells;
(v) unit vectors are scribed on the straightened boundary strings, and "break-points" are detected by correlating the sequence of unit vectors with an angle-detector and searching for sudden peaks in the 1-D correlation array;
(vi) the boundary curves are segmented at the break-points, and the segments are fitted (via linear least-squares routines) for five parameters — I
and J locations of the segment starting point, initial angle, initial curvature, and rate of change of curvature;

(vii) pairs of nearby parameterized boundary curves are compared to find if they have mutually "corresponding" parts, and "objects", defined by the ratio of the length of any corresponding parts to their distance apart being greater than a specified value, are indicated.

Further details of these processes have been given in Chapter 5.

Partly because of the developmental role that these experiments played in the conceptual model's development, and partly because of the difficulty of simulating some of the processes required, the simulations do not adequately reflect the model's nature and extent. A basic difference between the model and the simulations is that no processes for deriving 3-D interpretations have yet been included in the latter. Other important differences are that in the experiments described:

1) Only a single resolution level is employed in searching for edge-responses and angle-responses.
2) No attempt is made to develop properties appropriate to the regions being described, there being effectively only two main properties considered — gray level, and rate of change of curvature.
3) All boundary segments are treated as being of the one basic type (i.e., segments whose rate of change of curvature versus distance along the segment does not change perceptibly). It is now clear that these more complex segments should be distinguished from straight lines and arcs, although there are (transformational) relationships between these three types.
4) Only a small number of relationships, associations, and constraints is searched for, and the depth of organization achieved is very limited (e.g., no grouping of basic image components, such as track segments, into larger constructs is attempted). In particular, the important relationship of continuation after an interruption of an image area or boundary curve, is not searched for.
5) In the routine which searches for relationships between and within boundary curves, connectivity plays a much more important part
than it should. These curves are, after all, still 2-D entities, and connectivity is an additional relationship to proximity, rather than the only one.

6) In deciding whether or not a point qualifies as an edge- or angle-responsese, pre-chosen decision thresholds are employed, with the result that of two apparently identical local image regions, one might be classified as a response and the other might not.

With regard to 6) above, the use of thresholds (in one form or another) seems unavoidable, and further research into avoiding anomalies arising from their incorrect use is required (refer to suggestions in Section 4.2.4). Two types of threshold which can be distinguished are:

(i) a "limen" or threshold of perceptibility, e.g., the degree of curvature required for a line to be seen as curved rather than as straight;

(ii) a decision boundary which discriminates between two classes on the basis of the magnitude of some attribute, e.g., the magnitude of the allowable deviation from a straight line, if a line is to be classified as straight (even though it may be visibly curved).

The concept of a limen seems intuitively plausible, and much psychophysical research has concerned the identification of thresholds for absolute magnitude and just-noticeable-difference of various attributes. Nevertheless, there are problems involved with this concept (Corsco, 1963); Swets (1961) has introduced a decision-theoretic approach based on the idea of a response threshold as well as a sensory threshold, allowing for "noise" of various forms (including neural noise and external effects such as "set"). The problems involved with the second type of threshold are, for the present application, more serious. Gaps in a boundary curve which are not pictorially identifiable, are typical of the effects which result from the use of pre-chosen decision boundaries. Such boundaries may force an unqualified yes-no choice before all the information required for a satisfactory decision is available (if indeed the decision is meaningful).

There is a clear need for extension and revision of the simulations, taking account of the shortcomings noted above, in so far as further practical results will lead to clarification (and possibly revision) of the conceptual model — see Section 6.3.
6.1.1 Comparison with Other Work

There have been, to the author's knowledge, no previous studies similar in scope and motivation to the present investigation, so that direct comparison (as a whole) with other work is difficult. Previous API work (discussed in Chapter 3) has usually concerned relatively simple PI tasks or images. Several API models, specified in sufficiently precise terms to allow computer implementation and testing, have been developed, but important problems have either not been acknowledged or have not been satisfactorily accounted for. The present study has been addressed primarily to these problems, but the model developed is too complex (and as yet insufficiently detailed) to allow ready implementation and testing of a complete system.

Two problems of central concern in this study have been (i) identification of image features which suggest underlying 3-D configurations, and (ii) clarification of the various forms of image organization and of the relationships between (and constraints within) image and object components, which underlie the perception of the different types of organization. Problem (i) above has been studied by perceptual psychologists (notably Gibson, 1950, 1961, 1966), but whilst their findings are valuable, they have frequently been presented in a form which makes it difficult to relate them to the formulation of an API system (e.g., Gibson's (1966) treatment of the senses as responding directly to information-invariants which are specific to certain arrangements). Our tendency to perceive depth via rotational symmetry of 2-D outlines (see Section 4.2.7.3), does not appear to have been previously discussed in detail. Guzman (1968) has identified features of a restricted class of images, which suggest underlying 3-D arrangements; his techniques are rather more sophisticated in this regard than Roberts' (1965), who employs a brute-force matrix inversion to see if there is a possible projective transformation taking known 3-D model vertices into the 2-D vertices located in the input image. Problem (ii) has been studied by other workers (e.g., Evans (1964), O'Callaghan and Maxwell (1969), and Clowes (1968, 1969a)) in pattern recognition, computer graphics, and API, but only with respect to fairly restricted image classes; as a consequence of this restriction, some of the relationships and forms of organization identified in the present study (such as superimposition of levels of resolution, and constraints within and between boundary curves or surfaces) have not previously been considered to any extent.
The problems of image segmentation considered herein have also been of interest to other workers. Segmentation on the basis of metrical properties (typically gray level statistics) is considered by, for example, Holmes (1966), Rosenfeld (1962), and Muerle and Allen (1968). Only a few studies, however, have concerned segmentation on the basis of simple articulation properties (e.g., Grimsdale et al., 1959). Simon and Kotovsky (1963) demonstrate with letter sequences, the type of organization abstraction proposed here for images. This process is much more difficult for images than it is for letter sequences, but the concepts employed are similar. Simon and Kotovsky recover relationships between adjacent characters and look for repetitions of relationships; these processes are related to the Gestalt principles of association which are employed within the proposed API model for effecting image organization.

Neisser (1967) has argued for processes of selective attention similar to those recommended herein; he maintains (p. 86) that "there must be a way to concentrate the processes of analysis on a selected portion of the field". In the model presented here, the notion of a base component being modified with additions, deletions, and transformations, derives from the idea of a "schema plus correction". The tendency for human beings to describe variations of known or simple configurations in such terms is well known (e.g., Oldfield, 1954), but few attempts have been made to employ this idea in mechanical interpretation systems; Evans (1964) has made a preliminary step in this direction by developing techniques for detecting if a simple transformation carries one figure into another.

Finally, as regards contributions to the problem of pollen-interpretation (discussed in Section 1.2), the present model's application to pollen images was briefly discussed in Section 4.3; the contributions of this study tend to be more towards the API problem in general, rather than to the pollen-interpretation task in particular. Within the context of this model, it is at least possible to think about an API system which will be capable of managing this and other complex PI tasks, and to see directions in which research and development can proceed. With respect to the general approaches to API examined in Chapter 3, a template matching pollen-interpretation system is relatively easy to design but poor results are assured. Alternatively, specification of a pollen-interpretation system within a more powerful
approach (which has greater ultimate promise of satisfactory results), such as picture parsing, may be "too difficult to even think about" (M. B. Clowes — private discussion).

6.2 Implications and Applications

While the present study has not been directed towards immediate practical applications, it is anticipated that the concepts evolved and the techniques developed in the computer simulation experiments (reported in Chapter 5) will be found useful in other investigations, despite the limited scope of these techniques with regard to the conceptual model. The techniques for edge-detection, and for construction, segmentation, and characterization of boundary curves, might well be adequate for restricted image classes.

Because the conceptual model has, in part, been developed by consideration of human perceptual processing of a large number of selected image areas, and because its operation is explained and justified by comparison with human perception of pictorial examples, it is expected that the model will have implications with regard to the study of human visual perception. One aspect of human perception which has no parallel in the proposed model, is perceptual learning of the type demonstrated by Kohler (1962) in his experiments with goggles. In these experiments, human subjects wore goggles (which displaced, inverted, or distorted the visual field) continuously over a period of up to several weeks. The subjects' visual space at first appeared abnormal, but after varying periods of time (depending on the magnitude and form of the distortion), their visual space gradually began to appear normal again until, when the goggles were removed, a complementary distortion was perceived. In an API system, such perceptual learning or adaptation would be equivalent to learning how the 3-D fields map into the system's initial representation. Such learning is possible only through proprioception (Kohler, 1962; Gyr et al., 1966), and so could not take place in an API system whose only input was single static images. In such systems the relevant knowledge must be represented initially, rather than "learnt". The problem of finding out how it is that we learn to perceive, seems even more difficult than the problem of finding out how we perceive, having already learnt.
The idea of elaborating a simple or known base component with additions, deletions, and transformations (i.e., "schema + corrections" (Oldfield, 1954)), so that it forms an appropriate description of some image area, seems a useful concept in applications beyond API. For example, knowledge of the shapes of interesting signals in seismic traces and electrocardiograms, together with a representation of expected transformations (such as changes in time of occurrence, amplitude, and polarity) and a characterization of possible additions to the base component (e.g., "noise"), could be used in interpreting these records. Similarly, this concept could also be applied to the detection and correction of errors in compiler source text, by interpreting erroneous constructions in terms of known variable names and statement forms, and errors of substitution, insertion, or deletion of characters (e.g., WRITE (3,100)A might be interpreted as WRITE (3,100)A with a missing character). It is clear, however, that practical applications of this concept require much further research. Another principle of the conceptual model which has applications beyond API is that of searching for relationships and constraints within data. Further development of this principle might lead to techniques for discovering, within data, structure of the type not exposed by conventional statistical analyses.

A number of implications and possible applications of the present research are now examined in more detail.

6.2.1 Encoding and Decoding of Pictorial Information

The processes described in the conceptual model could well be applied to the problem of economical encoding of pictures, but, depending on the degree of abstraction from the original information, there are likely to be difficulties in synthesizing pictures from the coded representations.

Automatic development of descriptors appropriate to the image areas being described (together with effective segmentation into suitable areas), detection of constraints between adjacent areas, and the description of the image in terms of superimposed levels of resolution (cf. Graham's (1967) use of coarse and fine detail in image encoding), would all help effect worthwhile economies in the cost of storing or transmitting given pictures, without discarding much of the original detail.
Such economies could be effected only at the expense of considerable computation and information storage in both encoding and decoding the given pictures; this cost must be balanced against the cost of transmitting or storing less economical representations.

Further economies could be realized by interpretation of an abstraction from the given picture, but only at the expense of considerable additional processing and storage for the information relating to the objects portrayed in the input pictures (so that given pictures can be interpreted with respect to this information). A picture regenerated from such an abstraction would be similar to the original in the sense that they both lead to the same interpretation, rather than in the sense that there is a small RMS error between the gray levels of corresponding points in the original and regenerated pictures.

There are rather severe difficulties which will be encountered in trying to regenerate a picture from a highly abstracted representation. These difficulties should be taken into account when applying the criterion of generative adequacy (as espoused in the picture parsing approach) for evaluating picture descriptions. With complex images, an essential part of the interpretation process is that of abstraction of significant information from detailed local descriptions. These abstractions could be sufficient to allow satisfactory interpretation, in the sense that the important differences and similarities between two image areas could be exhibited by comparing their abstracted descriptions, but regeneration of the original areas, or areas similar in all important respects, would be difficult and somewhat of a trial and error process. Consider a lace-like surface such as that of Drimys (Fig. 1-1b), described in terms of 3-D structural elements (e.g., their shapes, and statistics of their relative lengths and thicknesses), together with structural information regarding the way in which the elements are configured, and constraints perceived for the surface as a whole (e.g., roughly spherical). In trying to reconstruct a 3-D object which is structured in the same manner as the original, and which has similar

\[1\] Witness the ease with which most of us can recognize a friend's face, relative to the difficulty that we experience in drawing such faces, typically via an iterative trial and error process.
statistical properties for the structural elements, there are obvious interactions between these elements, and a process of iterative refinement of the regenerated object (and thereby of its image), until the various constraints and statistics are satisfied, is indicated.

6.2.2 On the Characterization of Texture

Texture is understood to be a partially ordered arrangement of a large number of similar elements, in which there is a degree of randomness. Thus, the dimples in orange peel and the individual strands in the pile of a carpet are elements of texture. Other examples of textured surfaces are shown in Figs. 2-4c and 4-30a.

The characterization or description of textured images and surfaces is of interest for several reasons; within the proposed model, such characterizations are required for discriminating between areas having different textural properties, for noting changes (such as texture gradient) which are suggestive of underlying 3-D structure, and for effecting substantial economies in the description of textured surfaces (by avoiding descriptions of the individual elements and the relationships they enjoy with other individual elements). Another possible reason for interest in the characterization of texture concerns studies in visual texture synthesis (perhaps for perceptual research (Pickett, 1967)) and manipulation, including the interesting problem of filling in a "hole" in a textured area so that the synthesized portion merges smoothly with the surroundings and cannot readily be discriminated from them (Butt et al., 1968, 1969). Another problem in which the analysis of textural information is important, is that of examining mixtures of two or more components, for the purpose of quantifying their "mixedness", describing both the subdivision and spatial distribution of the mixture components (Hall, 1968).

Previous work with regard to the characterization of textural properties has been primarily concerned with statistical properties such as spatial power spectra (Julesz, 1962; Rosenfeld, 1962; Pickett, 1967). Levine (1969) includes a partial survey of this work, and suggests that in biomedical pattern recognition problems, consideration of texture might well be profitable. Rosenfeld (1962, p. 115) recommends background texture characterization as a technique useful for detailed
images such as aerial photographs, and states that the visual texture of any given portion of a photograph can be objectively described in terms of statistical parameters such as mean gray level and mean contrast frequency (i.e., inter-element spacing or granularity). Julesz (1962) gives some examples of synthesized textures, and makes the interesting observation that whilst differences in first order joint probability distributions are easily noted by humans, differences in the second order joint distributions are more difficult to perceive, and differences in the third order joint distributions are very hard to perceive (but could readily be detected by computations using the gray levels in the synthesized texture).

Julesz's examples expose one of the major weaknesses in the assumption that texture can be effectively characterized by simple gray scale statistics. Textural properties include element shape, inter-element organization, and statistics. Textural organization (i.e., "microstructure") is an important property, but there are many problems to be solved regarding isolation of textural elements and recovery of their organizations. Significant statistics usually concern properties such as inter-element spacing or element orientation, rather than mean gray level or mean contrast frequency. Rosenfeld (1967) acknowledges the importance of properties other than simple gray scale statistics, and observes that a complete model for visual texture perception and discrimination does not seem to be realizable at present. Julesz (1962) emphasizes the importance in texture discrimination of clusters or lines formed by proximate points of uniform brightness, and notes that other properties of clusters such as size, length and width, orientation, gray level, etc., may also serve as discrimination cues.

With regard to the textures arising in the class of images considered herein, it is apparent that only infrequently can the texture be regarded as purely 2-D, the more usual case being that the gray scale variations in the image result either partially or totally from 3-D undulations (differential pigmentation also being of importance in some textures such as pebbles on a beach, or faces in a crowd). In interpreting and characterizing such textures, the concern is for the underlying 3-D undulations or elements, rather than for clusters of similar points in the image. The corresponding elements in the image may well consist of both light and dark areas, as in the textures shown in Fig. 4-30a and Fig. 4-31. (That such textures
are usually perceived in 3-D is evidenced by the reversed perception of depth which may arise when these images are inverted.) The first problem in characterizing texture is the identification of the basic units or elements of texture. The elements identified may (in hierarchies of texture such as leaves, trees, and forests) be at several different levels of complexity and resolution. The isolation and characterization of elements will in many cases require a structural segmentation, because elements frequently merge into one another, and attempts to isolate by grouping adjacent points of similar gray level may lead to complex patchy configurations. The detection of various constraints within and between contours and image areas is required, so that configurations such as those shown in Fig. 6-1 may be segmented into the elements readily apparent to us (considering a 3-D organization if applicable).

The next stage of characterization is an attempt to interpret isolated elements (there being perhaps several distinct types of elements involved) in terms of each other, thus abstracting typical elements and the various statistical and articulart properties pertaining to groups of these elements (e.g., projected size, orientation, inter-element spacing, and periodicities). Formation of means and variances for the various statistical properties is required, if gradual or sudden changes in these properties between various areas of the image are to be detected (and used in suggesting changes of aspect in a textured 3-D surface, for example).

The processes envisaged above for characterizing textured areas imply a vast amount of computation if each textural element is to be closely examined. It is anticipated that it will usually not be necessary to examine the whole textured area in detail, and that having attended to a portion of it and formed a characterization, then other surrounding detail is attended to only as contours at a coarser resolution are noted. With this approach, changes in textural properties which do not lead to (or are not associated with) changes in the low resolution behaviour, might not be detected.

6.2.3 Character Recognition

Application of the proposed model to the interpretation of images in terms of alphanumeric characters (refer, for example, Holt (1968), Lindgren (1965) and Minneman (1966)) is examined below — examples of the class of image considered
The image contains text and diagrams. The text is as follows:

FIGURE 6-1  Structural Segmentation and Characterization of Textural Elements

The image is a black and white diagram showing various textural elements. The text is not fully legible due to the nature of the content.
here are shown in Fig. 6-2. This application illustrates the part that prior specific knowledge of objects plays in the model, the extent and organization of this knowledge (as it concern characters), and the model's generality.

The images considered here differ from pollen micrographs in the following ways:— (i) there is usually no concept of an underlying 3-D field (one exception being the "shadow" lettering shown in Fig. 6-2a); (ii) the number of possible objects is very much smaller, although there may be many stylistic variations of any of the characters, some of which form distinct subsets (e.g., "a", "a", and "A"); (iii) the images are less complex and detailed, and are usually intended to be binary rather than full gray scale (effective interpretation may still require full gray scale information, as in the case of interpreting a local image area as a smudge); and (iv) the pictorial units of eventual interest are dots and lines, although as is illustrated in Fig. 6-2b, the suggestion of line-like elements may result from repetitions, continuations, or axes of grouped elements which are not necessarily line-like themselves. Nevertheless, automatic recognition of "imperfect" and overlapping characters has proved to be a difficult task, and the problems of object isolation and resolution of ambiguities may still be formidable. To simplify the following discussion, the character "A" will be used where suitable as an example. Possible variations within characters and relationships between characters, which an effective general character recognition machine will have to account for, include:—

(i) imperfections in individual characters, such as smudges, spots, gaps, deformed characters, and ill-formed, misplaced, or extraneous strokes;

(ii) stylistic variations of idealized characters — for example, the characters may be skewed or embellished with serifs, or the individual strokes may be suggested by groups of separate elements;

(iii) inter-relationships between characters, except when single-character images are presented — characters may be overlapping, fused, or deliberately joined (as in running handwriting), thus leading to problems of segmentation and ambiguity. Contextual relationships, both internal to and external to the image, play an important part in resolving ambiguities, and should be accounted for. The internal context of
(a) "Shadow" lettering

(b) Strokes formed by grouping elements

(a) is this value

(b) a maximum or

(c) a minimum?

(c) Influence of extra-pictorial knowledge on interpretation

(d) Inter-stroke relationships which are additional to coincidence relationships

(e) Various representations of characters

FIGURE 6-2 Character Images
adjacent characters determines a reference frame for size, orientation, position, boldness, and style for the character being examined, and mediates description of a character as being, for example, unexpectedly long, high, and/or rotated, or as the start of a new word. The external context of known words, subject matter, grammatical context, etc., underlies the interpretations assigned to Fig. 6-2c.

If the above effects are to be accounted for in interpreting a given input image, then it is clear that a wide range of knowledge must be represented within the character recognition machine, and organized such that relevant knowledge can be accessed without an exhaustive trial and error search. Some forms of extra-pictorial knowledge, such as knowledge of the subject matter being described (in the text to be recognized) and knowledge of the grammatical context, will be very difficult to represent and use effectively. Although the importance of this type of knowledge is clear, its role is not central to the present discussion, and this role is not considered further herein.

A prerequisite to recognition of Fig. 6-2b as an "A", is a capability for detecting associations between pictorial entities, as is discussed in Section 4.2.2.1. Another requirement is that some representation of what "being an A" entails, can be brought to bear on the partially organized input. From one viewpoint, "A-ness" seems to be characterized by the presence of three line segments (whether actual or suggested), standing in certain relationships to each other and to a reference frame.

This object-specific knowledge is seen as being composed out of elements of more general knowledge, all information ultimately being specified in terms of simple attributes and relationships, such as distance, orientation, gray level, "same", and "different". These simple descriptors might be organized in a 2-way associative hierarchy, as follows:-

(i) the concept of strokes, either actual (in terms of constraints existing between the boundaries of some region distinguishable on the basis of its gray level being different from the surroundings) or virtual (in

1 Unless the idea of an "A" made up of "+"s has already been represented in some manner within the machine.
terms of a line of continuation of associated elements);

(ii) the idea of curved strokes (i.e., orientation changing gradually with distance) — there being an association between straight and curved strokes, in that a curved line can be considered as a straight line which has been transformed (curved);

(iii) strokes being adjacent, at an angle, touching, overlapping, coterminous, etc. — note that these relationships may be composed of even more basic relationships (e.g., there are clearly relationships enjoyed by pairs of strokes in Fig. 6-2d which depend on the strokes as wholes, and not only on the particular relationships obtaining between their ends); and

(iv) the particular configurations of strokes specific to each character in an idealized and basic form, including the normal orientation, locations, and sizes (with respect to a reference frame) of the individual strokes and the character as a whole.

In addition to the above information, the relationships between orientations (i.e., rotations), positions (translations), and sizes (dilations or contractions), are required for explaining unexpected differences in these attributes of either strokes or characters, in terms of the relevant transformations. As mentioned earlier in this section, characterization of smudges, spots, accidental versus deliberate gaps, serifs, superimposed and fused characters, deformation, etc., is also required.

The operation of the proposed model with each of the examples shown in Fig. 6-3 is now considered. With all the examples, the orientation of the presented image implies a reference frame for this property. Example (a) is the simplest case; the first phase of processing is selection of a potentially interesting part of the image on the basis of low resolution detail — once selected the whole of the character can probably be attended to at fine resolution. The black area is interpreted as being superimposed on the white area, by virtue of the boundary constraints and the continuation of the white ground after the interruption. Considering the left-most stroke first, this stroke is segmented (by detection of boundary constraints) from the other two strokes joined to it. Even at this early stage, the orientation of this stroke relative to the presumed reference frame has restricted the range of possible simple interpretations (i.e., not allowing for rotations of characters or
(a) Isolated well-formed character

(b) Ill-formed character

(c) Fused and overlapping characters

FIGURE 6-3  Images used as Processing Examples
strokes). With the segmentation of all strokes and recovery of their inter-relationships, interpretation "zeroes-in"on an "A" in preference, say, to an inverted "V" with an extraneous horizontal stroke. Note that the only possible comment regarding the location or size of this isolated character, is with respect to the size of the input image and the character's location relative to the image borders. No more-relevant reference frame previously exists for these attributes, but one has now been tentatively established for any other characters in the image. The complete description of this image is quite short, because the concept of an "A", with the addition of a few metrical properties, accounts for all the image detail.

Example (b) is fairly typical of the imperfections which might arise in printed characters. As has previously been stressed, accounting for (or 'recognizing') imperfections and unexpected differences, is an important process in the proposed model. The ease with which we can interpret images such as these, disguises the ambiguities present and the constructive nature of our perceptions. As with example (a), the black areas are soon selected for attention; possible initial segmentations are indicated in Fig. 6-3b. These segmentations are based on relationships between boundary segments, perhaps at a reasonably coarse resolution because of feathered or ragged edges. Most of the parts segmented have an inherent directionality or axis, and the next stage of processing involves a search for possible associations between these parts (taking account of mutually incompatible alternatives). Thus, the association of continuation between the elements composing the left-hand stroke, suggests that segmentation $\alpha$ is preferable to $\beta$. If there had been a suggestion of a blob on the bottom of the horizontal stroke, as occurs at its right-hand end, such an ambiguity may have already been resolved. An association of continuity between the bottom serifs is noted, but cannot at this stage be dismissed as subordinate rather than direct. The nature of the gap in the left-hand stroke suggests that it is accidental rather than deliberate (in that the bottom segment "fades out" rather than stops suddenly), and lends credence to the suggestion that it is intended to be a single stroke. Interpretation of the right-hand segments as a single stroke resolves the ambiguity of segmentation at the apex and suggests that segmentation $\gamma$ is preferable. Having employed only general rules of organization in the above processing, the possibility of interpreting according to
specific experience (i.e., as an "A") will already, no doubt, have been suggested. Such an interpretation accounts for most of the image, but leaves three short horizontal strokes (at the base and the apex) unaccounted for. They might be regarded simply as extraneous strokes, but the concept of serifs finishing off free ends of strokes affords a more satisfactory explanation.

The next examples considered are those of Fig. 6-3c, in which knowledge of specific configurations of individual strokes is required to help segment the fused characters, and to resolve ambiguous associations between strokes. For examples of previous work on identification of overlapping and running characters, see Kovalevsky (1968), Claydon et al. (1966), and Marill and Bloom (1966). Handwritten words such as those shown in Fig. 6-2c might also be considered here, but resolution of the manifold ambiguities of association and segmentation in typical handwriting, usually relies heavily on extra-pictorial information — the processing indicated below is relevant to, but not sufficient for, this case.

The objective is essentially that of applying knowledge of previously experienced character classes in describing the given groups. The problems exhibited in the examples are:— (i) characters running into one another, (ii) use of the one stroke as a component in more than one character, and (iii) superimposed characters, wherein the relationships between the strokes composing the overlapping characters are not always of direct significance, but may be concomitant with the relationships between the overlapping characters as wholes. It is remarkable that human perception of such characters is not hindered to any great extent by these problems.

Knowing that text is to be examined, horizontal bars (corresponding to lines of print or writing) noticed at a coarse resolution are then attended to at a finer resolution, starting at the left-hand end and scanning from left to right. The ease with which we can interpret the word "FETTER", emphasizes the importance of relationships between strokes as wholes, compared to local coincidence relationships. Gaps between strokes (as in "FE" versus "|E") are regarded as aids to, and not rules for, segmentation, with coincidence of strokes (as in "E" versus "|E") being similarly regarded with respect to grouping. Segmentation of strokes in the area attended to proceeds by noting boundary constraints and interruptions of the even white background. It is important to observe that the strokes do not have to be completely
isolated in the early stages of processing, because relationships may be perceived between only those parts of strokes visible within the current area of attention; these locally perceived relationships guide subsequent processing. Assuming that the area visible at fine resolution approximates the size of a single character, the vertical and horizontal strokes, and the relationships between these, will have already suggested interpretation as an "F". The right-hand vertical stroke does not fit into this interpretation and remains to be accounted for; the fact that this stroke is joined to the two horizontal strokes (which themselves continue to the right), hinders interpretation (in so far as segmentation has to await the application of specific knowledge), but does not prevent it. Interpretation of the other characters proceeds in the same manner, with the result that the long stroke along the top of the word "FETTER" is never seen as a single stroke, because by the time this whole stroke can be attended to at once, various parts of it have already been subsumed in local organizations.

Interpretation of the word "AND" proceeds without any problems until, having accounted for the strokes composing the A and the N, a left-facing curve remains and (apart from interpretation as a reversed C, say) cannot be explained in terms of known characters. The curve does, however, suggest a D, and the question of its missing vertical stroke arises. A stroke already accounted for as part of the N is otherwise satisfactory for this purpose, but the same element normally cannot be shared by two separate characters. This element has to be explained as two strokes superimposed or fused into one, because if it is accounted for as part of the D, no satisfactory interpretation of the earlier strokes arises. The explanation of fused strokes is supported if the stroke in question is thicker than the others.

With regard to the overlapping characters "A" and "H" in Fig. 6-3c, the first stage of processing results in the isolation of short black line-segments (on the basis of boundary constraints). Associations between these segments are then found, and various groupings are suggested. To the extent that there are many organization-suggestive associations between pairs of segments, there are already problems of ambiguity, but some associations are stronger than others and

1 Orientations are referred to that of the presented image.
play a large part in suggesting possible groupings. Continuity is a basic association which is important initially (e.g., "\( \backslash \)" is seen as two crossing strokes rather than as four strokes joined at a point); this type of grouping is relatively independent of the context and of specific prior experience, and is tentatively accepted.

Having grouped strokes according to continuity, the next phase of processing (i.e., the next cycle through the minor loop BCD in Fig. 4-3) concerns recovery of relationships between these larger strokes, and the suggestion of possible higher-level groupings (but not yet in terms of specific experience). At this stage, coincidence relationships become important — note the relative strengths of association between the strokes in "\( \backslash \)" , "\( \backslash / \)" , "\( \backslash / \)" and "\( \backslash / \)". It was stressed earlier that coincidence relationships are additional to more general relationships between strokes and are not the only relationships, but they do specify simple configurations of strokes and so represent preferred organizations.

Amongst many others, the relationships between pairs of strokes include those involved in the representations of an "A" and an "H". Through bidirectional associative links, there is a summation of implications in these representations, and interpretation as an "A" plus an "H" is suggested — interpretation as superimposed characters is not inconsistent with the image detail, and accounts for this detail simply (and is therefore preferred to interpretation in terms of "X's" and "I's", say). Relationships of strokes in the "A" to strokes in the "H" are accounted for as concomitant with the relationships between the "A" and the "H" as wholes.

6.2.4 Nuclear Tracks in Photographic Emulsion

The computer simulation experiments reported in Chapter 5 have been directed towards recovery of simple image organization, rather than towards the PI problem of discriminating and counting the number of tracks per unit area at different locations in the photographic emulsion. Recovery of image organization is clearly an important requirement in the track-counting problem, with association of isolated track segments into groups forming tracks, and segmentation of tracks from each other and from extraneous blobs, being needed.

The restrictions inherent in this image class (i.e., objects can effectively
be regarded as 2-D, specific prior experience does not play a large part in the interpretation process, and the image organization is relatively simple) are such that a simplified version of the conceptual model could be applied to the track-counting problem. In view of the restricted nature of this problem, the shortcomings of the computer simulation experiments (relative to the overall API problem) are not so serious, and these simulations provide a useful starting point for development of a track-counting system. A relationship which is equally important in suggesting that the black blobs are figures on a light background, as are constraints between the boundaries of these blobs, is that of repetition of the light background on the other side of a dark interruption; such repetitions have not been searched for in the simulation experiments. Similarly, repetitions of boundary curves after discontinuities or interruptions should also be searched for, and used for example, in segmenting blobs which adjoin tracks, from the tracks themselves.

The processing sequence envisaged for this application approximates that of the simulations as far as they went, except that an "attention" phase (see Section 4.2.1) would be included. Detail which is obvious at a coarse resolution is initially attended to, with subsequent processing concerning interpretation of this detail; small individual spots remote from more obvious detail may not be attended to except to the extent that they are eventually seen as part of a spotty background. From the point where the current simulations cease, repetitions of the background and boundary curves (in the vicinity of discontinuities) are searched for, and these, together with boundary constraints such as "correspondence", suggest simple figural components. In the next stage of processing, straight lines are distinguished from arcs and segments with changing curvature. Any image elements formed (described in terms of gray level, perceived constraints, orientations, locations, etc.) are compared with other nearby elements, to see if any groupings (particularly into "tracks") are suggested. After resolving ambiguous associations and effecting further segmentations and groupings (e.g., of crossing tracks, or tracks which are physically remote relative to the length of the individual segments, but not relative to the lengths of the groups formed from these), clearly distinguished tracks are counted.
The practical experiments suggest that such a track-counting system could not, at present, compete economically with a human photo-interpreter. After training, female laboratory assistants (\(<\$2\text{Aust. per hour}\) can count tracks at a rate of many per minute; the computer simulations indicate that an IBM 360/50 (\(\approx \$80\text{ per hour}\)) could take several minutes of processing for each track. With hardware implementations, techniques such as coherent optical processing (e.g., for correlating edge-detector masks with the input image) could reduce this disparity, but it is interesting to note that the information processing capacity of the human retinal neural structure alone (in which the information from \(10^8\) receptors is channeled into \(10^6\) output paths in the optic nerve) is comparable to the capacity of the largest present digital computers. It is possible that other, less sophisticated and more economical API schemes (such as template matching), could perform satisfactorily when applied to track-counting, but this is unlikely in view of the frequent occurrence of (i) dense groups of crossing tracks, (ii) very faint and gapped tracks, and (iii) extraneous detail adjacent to or superimposed on tracks.

The above application is in some respects rather similar to track recognition in bubble- and spark-chamber photographs, and the processes outlined are also relevant to this problem.

6.3 Recommendations for Further Research

The present research is a long way short of the overall goal of implementing reasonably general API systems, and has the status of an exploratory study. Progress in this field over the last decade has been slow and hesitant, and the present study has exposed many difficulties. It must be concluded that the API problem is solvable, in so far as the human perceptual system is a living existence proof (unless it is held that human cognitive processing cannot, in principle, be simulated — a view not held by the author). Whether or not practical solutions will be achieved in the foreseeable future is another question.

The current study suggests that solution of the overall API problem should be regarded as a long-term objective; the nature and extent of the sub-problems involved make it unlikely that there will be any sudden breakthrough resulting from
a simple but ingenious insight. A number of observations relevant to further development, and suggested avenues of attack on the remaining problems, have been presented, but substantial progress will probably be achieved only as a result of much further patient research. Useful lines of study include:

(i) further broad-based research concerning the present conceptual model (or another similar in scope and motivation), with the objective of refining, clarifying, and elaborating this model (to the extent practicable without specific study of the proposed components);

(ii) revision of and extensions to the simulations reported in Chapter 5;

(iii) development of computer-assisted photo-interpretation systems which are applied to reasonably difficult image analysis problems, with the aim of gradually increasing the role of the computer relative to that of the man;

(iv) investigations of restricted image analysis tasks which emphasize certain aspects of the processing assumed in the conceptual model, with a view to devising solutions for these tasks and also contributing to the overall API problem; and

(v) development of ad hoc solutions for practical image analysis problems, but through an awareness of the limitations of the techniques employed, modifying the images and fields portrayed so that as far as possible, these limitations are circumvented (e.g., by ensuring that the objects of interest are always presented in a standard orientation and are not differentially pigmented). It is not suggested that this line of research will contribute substantially to the overall API problem, but rather that immediately useful practical results may be achieved.

Regarding (i) above, the conceptual model requires further development, particularly with respect to the question of information representation, problems related to the use of decision thresholds, the process of choosing among alternative organizations on the basis of a simplicity measure, specification of procedures for recovering important associations and constraints, and integration of detail at various levels of resolution into a consistent description.
Much useful work could be performed by extending the scope of computer simulation experiments and bringing them more into line with the conceptual model (as currently specified). Some of the modifications and extensions required are indicated in Sections 6.1 and 6.2.4.

The dangers inherent in investigating restricted PI problems and expecting the results to generalize, have already been emphasized (Section 1.2). To the extent that the conceptual model for API described herein is valid, then by working in the context of this model and avoiding ad hoc solutions to the restricted problems, these dangers are reduced. It is anticipated that the results of such investigations will lead to modifications to the conceptual model as well as to elaboration and clarification of it.

Any of the potential applications discussed in Section 6.2 are suitable vehicles for further research as well as for the development of practical systems, in so far as there is a two-way interaction between the conceptual model and its application to specific problems.

A problem of considerable practical interest is that of character recognition. Character images are restricted in the sense that there is usually no concept of 3-D objects, and that binary rather than gray scale images are intended, but the interpretation of characters emphasizes the role of prior experience in resolving ambiguities and guiding interpretation. The results of the present study could, with some further research, be applied to the problems of segmenting character strokes from each other, and of recovering organization-suggestive relationships between strokes (as discussed in Section 6.2.3).

Encoding and decoding of pictorial information, as discussed in Section 6.2.1, are also suitable vehicles for further research, particularly with regard to the questions of development of appropriate properties for the description of image areas. Regeneration of pictorial information from an abstract representation is an interesting study in itself, but in addition, helps indicate (by means of the differences between the original and the regenerated pictures) whether the abstract representations specify all the important relationships, constraints and properties. The problem of counting nuclear particle tracks in photographic emulsion (discussed in Section 6.2.4) is also of interest, particularly with regard to coping
with image imperfections and extraneous detail.

An important process in the proposed model is that of interpretation in 3-D terms. Unfortunately, this process is intimately involved with many of the other processes outlined and, as such, cannot be effectively studied except in the context of the model as a whole. Nevertheless, further elucidation of image features which suggest simple 3-D arrangements, development of suitable representations for organized 3-D detail, and experiments with 3-D interpretations of images of simple 3-D fields (not restricted to only plane-faced objects), would be valuable.

Integration of 2-D or 3-D detail at various levels of resolution into a coherent description is required, but further research is needed. An apparently simple problem, in which consideration of several levels of resolution and simplicity of description is indicated, is that of positioning the boundary between two areas such as those shown in Fig. 6-4. A coarse resolution contour could readily be detected, but the accuracy of location of this contour corresponds to the resolution level employed, and does not match the perceived sharpness of the straight boundary between the two areas. Rosenfeld et al. (1969) have studied the problem of detecting and accurately locating this contour; they compute correlations with equally-weighted horizontal edge-detectors of several resolutions, and form a product of the correlations, the idea being that an edge will be detected only where there is a gradient at both coarse and fine resolutions, thereby locating the edge with an accuracy corresponding to the fine resolution. Isolated "noise" responses are filtered out by a horizontal line detector, and a line corresponding to the horizontal boundary is produced.

The approach suggested by the current study differs from that employed by Rosenfeld et al.; when applied to this problem, a set of variable-weight edge-detectors covering a range of resolutions and orientations is correlated with the given image, but a response at any of the resolutions (rather than all) has to be accounted for (but not necessarily by attending to every small area of the image in fine detail —

\[\text{1 If this image is encoded as a matrix of gray levels of the individual spots, then some elements in the initial representation correspond to actual pictorial entities; note that this is not the usual case.}\]
FIGURE 6-4 Adjacent "Noisy" Areas (after Rosenfeld et al., 1969)

Probability of black = 0.20

Probability of black = 0.80
At a coarse resolution, two areas are perceived, but at a finer resolution many small areas (i.e., the "noise" spots) would be noticed. The characteristics of these spots are determined by attending to several isolated spots (or to groups which are subsequently seen as being composed of joined individual spots). Two types are detected, white spots on a black background and black spots on a white background, with an ambiguous assignment of small areas in the middle of the image, as either spots or background. Resolution of these ambiguities depends on the neighbouring spots and also on the preference for a simple (e.g., straight) boundary between the two regions. Thus, a description of a white-spotted black area joined via a horizontal straight boundary to a black-spotted white area, would eventually be formed.

Despite the apparent simplicity of the above problem, it can be seen that its solution in the context of the proposed model, rather than in an ad hoc manner, will contribute substantially towards the overall API problem.

The interest in the pollen-interpretation task, used as an example throughout the present study, has been in its usefulness with respect to development of concepts, rather than with respect to the design of immediately-useful API systems (refer to Section 1.2). Pollen-interpretation was chosen for this purpose because of its non-trivial nature, the delicate and rich organization of pollen images, and the clear 3-D nature of the fields portrayed; this task could well be a suitable vehicle for further conceptual development, based on some of the suggestions made in Section 4.2 with regard to the model presented therein.

Pollen-interpretation is a reasonably "general" PI task, and its mechanization therefore depends on overall progress in the API field. Although there are other PI tasks of more immediate concern (e.g., bubble-chamber and fingerprint interpretation), a system capable of effective pollen-interpretation would automatically be capable of a wide range of real-world image-interpretation tasks, and development of such a system is a worthwhile (but necessarily long-term) practical objective.

In summary, progress towards solution of the API problem has been made in the present study, in so far as inherent difficulties have been identified, a concept-
ual model has been presented as a basis for discussion, certain practical techniques have been developed, and observations and suggestions relevant to future research have been made.

In the present study of the API problem, the first phase has involved an identification and clarification of major inherent difficulties, e.g., those related to (i) 3-D fields, (ii) object isolation, (iii) the quantity, organization, and variability of detail, (iv) ambiguity, and (v) information representation. A survey of previous work on API has been made, with the reported approaches being grouped into four major topics: template matching, property line, particular analysis, and picture parsing. An evaluation of these approaches with respect to the difficulties identified has uncovered many inadequacies. As a result of this evaluation, together with computer simulation experiments and consideration of human perception of many examples taken from real-world images, a conceptual model for API, particularly addressed to the difficulties identified, has been developed, and a number of observations and suggestions relevant to further research have been made within the framework of this model.

The simulation experiments have concerned the discovery of simple pictorial organization from realistic image data, which are digitized in a simple picture scanner, and correlated with improved FFT and in-echo parameterization algorithms with variable-weight edge-detection or sets of four orientations. Edge-responses are associated into boundary curves, which are divided into simpler segments in any sudden changes of orientation. These segments are fitted for several parameters, and compared to see if any "objects" approximately coincidental boundaries, may be formed. The results are sometimes very promising by means of an improved technique for generating connected outlines with a line printer. These simulations have indicated a substantial parameterizing effect and are rather promising of computer-time and storage.

The results, implications, and possible applications of the present study have been discussed in the previous chapter. The main conclusions drawn being:

1. The overall API problem is extremely difficult and complex, and the solution is seen as being a long-term objective. Given sufficient further research and development, national API systems will
Chapter 7

CONCLUSIONS

In the present study of the API problem, the first phase has involved an identification and clarification of major inherent difficulties, viz., those related to (i) 3-D fields, (ii) object isolation, (iii) the quantity, organization, and variability of detail, (iv) ambiguity, and (v) information representation. A survey of previous work in API has been made, with the reported approaches being grouped into four areas (template matching, property list, articulation analysis, and picture parsing). An evaluation of these approaches with respect to the difficulties identified, has uncovered many inadequacies. As a result of this evaluation, together with computer simulation experiments and consideration of human perception of many examples taken from real-world images, a conceptual model for API, particularly addressed to the difficulties identified, has been developed, and a number of observations and suggestions relevant to further research have been made within the framework of this model.

The simulation experiments have concerned the recovery of simple pictorial organization from nuclear-track images, which are digitized in a simple picture scanner, and correlated (via improved FFT and in-situ permutation algorithms) with variable-weight edge-detectors at each of four orientations. Edge-responses are associated into boundary curves, which are divided into simpler segments at any sudden changes of orientation. These segments are fitted for several parameters, and compared to see if any "objects" suggested by constrained boundaries, may be formed. The results of processing are displayed by means of an improved technique for generating pictorial output with a line printer. These simulations have involved a substantial programming effort and are rather demanding of computer time and storage.

The results, implications, and possible applications of the present study have been discussed in the previous chapter, the main conclusions being:-

1) The overall API problem is extremely difficult and complex, and its solution is seen as being a long term objective. Given sufficient further research and development, adequate API systems will
eventually be formulated, but the outlook for the near future is sobering — in view of the nature and extent of the problems identified, it is unlikely that there will be many substantial short-term results. Nevertheless, the possible applications of general API systems are stimulating, and it is expected that results from API research will facilitate man-machine communication (to the extent that the human and mechanical interpretations of visual data are alike), and will also contribute to a better understanding of human visual perception — continued research is therefore adequately justified.

2) Just as much further research is required with respect to understanding the processes involved in API, so (when these processes are sufficiently understood) a great deal of technological research will be required to implement various components (such as a memory with distributed processing and "active" links between associated information, as indicated in Section 4.2.8), and to develop techniques which can provide the formidable processing and information storage capacity indicated by the present study. Eventual implementations must be economical with respect to time and cost when compared to human photo-interpreters.

3) Of the current approaches to API which have been examined, the template matching and property list frameworks are too restricted and fundamentally inadequate to have much potential as complete API systems, although some of the techniques developed within these frameworks may be useful tools in more-sophisticated systems. Although the articular analysis approach (outlined in Section 3.3), is more a collection of techniques and concepts than a clearly formulated coherent approach, it nevertheless appears to be the most suitable candidate for further development and application to complex images. The picture parsing approach is more clearly defined and delimited, but (as currently formulated) faces major conceptual difficulties with regard to object and image variability, to the extent that "closed" descriptive formalisms are employed for "open" sets
of images. (Given the ability to generate (from a "closed" basic set) new attributes and relationships appropriate to the situation being described, these difficulties may be mitigated).

4) Effective interpretation of images which depict 3-D fields, must necessarily be based on recovery of 3-D information from 2-D images, because the objects of ultimate interest are 3-D and will therefore be characterized mainly in this domain. Derivation of a reasonable 3-D description from an image, is one of the most difficult problems encountered in API, unless there are drastic restrictions on the nature of the fields portrayed.

5) Global interpretations of complex images should be derived by merging interpretations of local areas, with these local interpretations being suggested by an interaction of image information with stored knowledge, rather than by an exhaustive test of possibilities. In particular, the range of possible organizations is effectively unbounded, and it is a question of discovering organization, rather than of detecting whether a pre-specified organization exists.

6) The sub-problems inherent in API are interdependent, and research concerning any of these should be performed in the context of a plausible overall model or approach. Despite its limitations, it is claimed that the conceptual model presented herein is suitable for this purpose. It is tempting to try to avoid the complexity evident in this model (even in its incompletely formulated state), but in order to cope with the problems identified, such complexity seems unavoidable.

7) Image and object organization plays a crucial role in the interpretation process¹, as does a knowledge of the relationships between objects, their images, and the initial machine encoding of these images.

¹This conclusion is at variance with the view of pattern recognition (and thus photo-interpretation) implicit in Rutovitz's (1966, p. 526) contention that "pattern recognition is a domain where deterministic schemes are mere camouflage for an underlying balance of probabilities... it is a field where the statistician should rule".
Important forms of organization include:

(i) hierarchical composition from simple components;
(ii) constraints within, and relationships between, boundary curves and surfaces; and
(iii) superimposition of components (e.g., a shadow on a surface, or summation of detail at several levels of resolution).

Coincidence relationships between components, which have been the prime concern in previous work, are additional to relationships between components as wholes (rather than at individual points). The "correspondence" relationship between boundary curves is an example of this latter class, but further research towards identifying (and specifying recovery procedures for) such relationships is required.

Within the context of complex images, the utility of the notion of "primitive" objects (whose internal structure is not of interest) is questionable — it seems that such objects are still organized, in that they can usually be defined in terms of simpler parts or of boundary constraints, so that if this organization is not captured, problems of generalization over changes in the size, orientation, location, etc., of these primitives may arise. The alternative view of primitives as the basic elements in the machine's initial encoding of the input image (e.g., as cells in a matrix of gray levels (Clowes, 1969a)), is awkward in the sense that these elements are usually not pictorially identifiable, and therefore should not form part of the input image's description (although they will, of course, be intimately involved in the process of deriving this description). Relationships and attributes are seen as being "prior" to objects.

The ability to learn new objects (i.e., to form and retain characterizations of given objects for use in subsequent processing) follows fairly readily, once an API system is able to form appropriate descriptions of unfamiliar objects by the application of general
rules of organization, and is also able to compare such descriptions, noting similarities and differences — there is a close relationship between comparing descriptions of objects in separate images, and comparing descriptions of components of the same image for the purpose of suggesting groupings on the basis of similarity. Formation of class characterizations arises from the two-way interaction between stored descriptions and given images, but there are problems with regard to abstraction of significant class characteristics, and the possibility of alternative taxonomies.

10) In view of the current state-of-the-art in API, much remains to be done before effective systems for tasks as difficult as pollen-interpretation can be developed (refer to (1) above). Nevertheless, PI tasks such as these are worth studying (even if suitable systems are not readily implementable), in so far as they enable a broader view to be taken of the API problem than do more-restricted tasks.

Among the recommendations for further research it is suggested that (in addition to broad study), the concepts exhibited in the proposed model should be applied to restricted PI and pattern recognition problems, whose solutions in the context of the model will thereby contribute to its refinement and clarification.

Important problems for future investigation include:-

(i) the development of effective information representations;
(ii) further identification of the features which suggest inter- or intra-domain organizations (e.g., 2-D features suggesting an underlying 3-D field);
(iii) integration of 2-D and 3-D detail at various levels of resolution;
(iv) determination of the necessity or otherwise of decision thresholds, and if these are unavoidable, of guidelines to their correct usage; and
(v) development of techniques for determining the relative simplicities of alternative descriptions, as an aid to choosing the best local or global interpretations.

It is suggested that this study has contributed insights into the nature and
extent of the API problem. An understanding of the formidable difficulties which lie ahead, however, is a first step towards coping with them. It is only when we realize the problems in mechanization of photo-interpretation, that we gain a fuller insight into the power, sophistication, and (under the circumstances) remarkable success, of the human perceptual processes that we so readily take for granted.
REFERENCES


CLOWES, M. B., (1968): "Transformational grammars and the organisation of pictures"; Lecture notes for the NATO Summer School on Automatic Classification and Interpretation of Images, Pisa.


EVANS, T. G., (1968a): "Descriptive pattern analysis techniques"; Lecture notes for the NATO Summer School on Automatic Classification and Interpretation of Images, Pisa.


HAKE, H. W., (1957): Contributions of psychology to the study of pattern vision; WADC Tech. Rept 57-621 (AD 142035).


MILLER, G.A., (1956): "The magical number seven, plus or minus two; some limits on our capacity for processing information"; Psychol. Rev., 63, pp.81-97.


NATHAN, R., (1968): "Picture enhancement for the moon, mars and man"; pp. 239-266 in Cheng et al.


O'NEILL, J. P., (1965): Letter regarding the implementation of SLIP; Co-op News (CDC Users' Group Newsletter), No. 54, pp. 7-14.


POLLOCK, D. K., KOESTER, C. J., and TIPPETT, J. T., (Eds.), (1963); Optical Processing of Information; Spartan, Baltimore.


IEEE Spectrum, 6, 3, pp.66-78.

SEBESTYEN, G.S., (1961): "Recognition of membership in classes";
Trans. IRE, IT-7, 1, pp.44-50.


SEBESTYEN, G., (1963): Adaptive memory research; Tech. Rept. AD427825,CFSTI.


SHAW, A.C., (1968): The formal description and parsing of pictures;


SOWERS, H.W., (1963): Star pattern recognition — a survey of the literature; Tech. Rept. AD 421473, CFSTI.


TEITELMAN, W., (1964): "Real time recognition of hand-drawn characters"; AFIPS, 26, pp. 559-575.


WOLTERS, A.W.P., (1933): The Evidence of Our Senses; Methuen, London.


Picture Digitisation Via A Modified X-Y Plotter

By I. D. G. Macleod *

Pictures may be digitised via a modified X-Y plotter which is interfaced to a computer by means of ADC's and DAC's. The picture to be scanned is placed on the bed of the plotter, whose normal pen assembly is replaced by a specially built scanning head. A simple scanner constructed in this manner had a gray-scale resolution of at least 64 levels, a spatial resolution of .010in. over an area of 10in. by 15in., and a maximum sampling rate of 200 points per second.

I. Introduction:
As a result of automatic photo-interpretation and digital picture processing experiments being carried out in the Department of Engineering Physics at the Australian National University, a method for digitising half-tone photographs was required. At least 32 repeatable gray-scale values and a spatial resolution of .010in. were necessary for adequate representation of the input pictures, which could range in size up to 2½in. square. The expense of obtaining and interfacing a standard picture scanning system to the ANU's IBM 360/50 computer could not be justified in view of the rather limited usage anticipated.

It was decided to construct a simple picture scanner and interface it to the computer through existing analogue to digital converters (ADC's) and digital to analogue converters (DAC's). Several approaches, including a drum-type scanner (Jessup and Wallace, 1968) and a CRT-based scanner (Ledley et al., 1965) were considered. The solution finally adopted (because of its simplicity, versatility, and low cost) was to place the pictures to be scanned on the bed of an existing analogue X-Y recorder, whose pen assembly was replaced by a specially constructed scanning head.

II. The Scanning Head and Circuits:
The scanning head consists of a 2½in. long piece of 3½in. diameter thin-walled tubing, mounted on the pen carriage perpendicular to the plotter's bed, with a convex lens inserted in the bottom and two photo-field-effect transistors (photo-FET's) mounted at the top. The tubing is lined with black felt to minimize internal reflections. The general configuration is illustrated in Fig. 1, the scanning head being connected to the rest of the circuit by a flexible shielded cable.

The picture being scanned is illuminated by two small light bulbs positioned at the base of the scanning head. A sampled area of .010in. diameter on the picture was chosen as being consistent with the X-Y recorder's repeatability specification of 0.1% for both the 10in. X-axis and 15in. Y-axis. The lens projects light reflected from the picture through a sampling window onto photo-FET P₁ and generates a photo-current which produces a signal voltage at P₁'s source terminal. P₂ is mounted in a similar thermal environment to P₁ but is kept in darkness. By suitable adjustment of R₁, the temperature dependent drifts of the dark currents in P₃ and P₁ cancel in the differential signal voltage V₄. This voltage is amplified and applied to one of the ADC inputs. A variable frequency pulse generator PG may be used to pulse the ADC START input, thus initiating sampling and subsequent conversion in the ADC.

III. Method of Operation:
The scanner is quite versatile, but the best method for acquiring digitised samples of the picture being scanned depends on the particular application. A TV-type scan is currently employed in acquiring a matrix of digitised picture points for later off-line processing. Each scan line is acquired by sweeping the scanning head at a constant velocity in the X direction, and using the pulse generator to control the sampling rate. The Y-axis position is decremented for each successive scan line.

IV. Performance:
Experience with the scanner and the results of tests carried out with a standard gray scale suggest that at least 64 repeatable gray levels and reasonable linearity are obtained. As far as could be determined, the response time of the system to a sudden black-white transition was of the order of one millisecond, this figure probably resulting from the RC time constant of resistor R₁ and stray capacitances. A faster response could be obtained by reducing the value of R₁, but to maintain the other performance specifications, the illumination intensity might have to be increased. With .010in. diameter sampled areas taken at .010in. intervals at a rate of 200 samples/second, a sample independence of better than 90% was estimated.

The maximum internal sweep rate of the recorder is 2in./second, which gives a maximum sampling rate for non-overlapping .010in. dia. sampled areas of 200 points/second. With allowance for carriage return after each scan, the average sampling rate is limited to about 150 points/second.

* Department of Engineering Physics, Research School of Physical Sciences, The Australian National University, Canberra, A.C.T., 2600. Manuscript received August, 1969.
V. Discussion:
Jessup and Wallace have recently reported an ingenious modification to an inexpensive facsimile stencil cutter which enables its use as a graphic input device to Basser Computing Department's KDF9 computer (Jessup and Wallace, 1968). A comparison with the present scanner is called for. The Basser scanner is a drum type and can handle pictures up to 9in. x 12in. which are wrapped around the drum and scanned in a helical manner. A spatial resolution of .005in., two gray levels, and a scanning rate of 10,000 points/second for a full size picture are reported. Modifications to Jessup and Wallace's scanner have been made at the CSIRO Division of Computing Research in Canberra, where four gray levels are obtained, and also at the University of Adelaide’s Computing Centre, where 16 gray levels are obtained. The ANU scanner can handle pictures up to 10in. x 15in., has a spatial resolution of .010in., at least 64 repeatable gray levels, and a maximum scanning rate of 200 points/second. The characteristics of the two scanners indicate rather different applications, the ANU scanner being slower, but offering greater gray-scale resolution and much more versatility with regard to the choice of a scanning technique. In the application for which the ANU scanner was designed, fidelity and resolution of gray-scale value were of rather more importance than scanning speed. When operated in a time-shared mode on the 360/50 computer, only 6 seconds of central processor time was used during the 12 minutes necessary to scan a 2½in. square picture on a .010in. grid.

As a result of successful use in the experiments being carried out in the Department of Engineering Physics, the scanner is also being used for digitising photographs of interference fringes obtained during experiments with the high-velocity shock tube in the ANU’s Department of Physics, School of General Studies. The reflectance values obtained are used to compute fringe locations and displacements, manual acquisition of reflectance values via a microdensitometer being avoided by use of the scanner. Another application, which is currently being investigated is the use of the scanner as an on-line curve follower. It is anticipated that graphs and other curves could be digitised both rapidly and accurately in this manner. When not in use as a scanner, the X-Y recorder can be used as a computer controlled plotter.

If care was taken to minimize the weight of the scanning head, a digital incremental plotter could be used instead of an analogue X-Y recorder. In this case, somewhat higher sampling rates could probably be achieved, and analogue output would not be required.

By constructing a small amount of special purpose hardware, and using facilities which are available in many computing installations, an inexpensive yet versatile picture scanner can be assembled.

References:


Pictorial Output with a Line Printer

I. D. G. MACLEOD

Abstract—An improved method for the production of pictorial output on a line printer is described. A reasonable black-white contrast ratio is obtained by overprinting up to eight characters, and pseudorandom noise is used to smooth out discontinuities in the range of print densities.

Index Terms—Linear interpolation, line printers, overprinting, pictorial output, picture coding, picture output.

The possible character positions on line printer output may be regarded as cells in a two-dimensional array. By choosing the character (or combination of overprinted characters) printed in each cell on the basis of average print density, a pictorial representation of any desired two-dimensional data may be generated. The maximum number of printed characters in a line is a restriction to the horizontal size of the picture unless it is assembled from several strips. Pictures produced in this manner will be inferior to those produced by special-purpose hardware, but the convenience and ready availability of a line printer will in many cases outweigh any loss of quality.

Perry and Mendelsohn [1] describe such a method of picture generation with a line printer. They use pairs of adjacent character positions in each printed line as basic density cells, and overprint a maximum of two characters. We have found that a greater degree of overprinting yields a worthwhile increase in maximum density (and thus contrast) without an undue penalty in printout time on the IBM 1403 line printer (1100 lpm max.) attached to the Australian National University's IBM 360/50 computer. Using a maximum of eight overprinted characters, the output shown in Fig. 2(b) took approximately one minute to generate on this printer. The use of individual (rather than pairs of) character positions as basic density cells results in finer resolution and allows more information to be represented on each line. With the 1403 printer set to ten characters per inch horizontally and eight lines per inch vertically, noticeable distortion results from the difference in horizontal and vertical scales. This distortion is reduced by linearly interpolating output lines between rows of the data array, printing only eight lines for every ten rows in the data array.

A difficulty in determining the density codes (i.e., overprinted character combinations) to be employed is that it is not possible, given a conventional character set, to choose combinations such that there is a smooth transition from white to black, without perceptible discontinuities. The sudden transition from white (blank) to the next lightest combination (comma) is noticeable in the examples of pictorial output given in [1]. The problem of representing numerical data with density
values chosen from a large but unevenly distributed set, e.g., the set of overprinted character combinations, is related to that of coding pictorial input with a small but evenly distributed set of numerical values. Following Roberts' technique for encoding pictures with a small set of symbols [2], the particular density code to be printed in each cell is chosen statistically, such that the average density of several neighboring cells is usually closer to the desired density than are any of the individual density codes. Roberts' examples illustrate the effectiveness of this technique.

Referring to Fig. 1, the following method is used to decide which density code is to be printed in each position.

1) The closest available density codes both below and above the desired density are determined.
2) A pseudorandom number is chosen from a uniform distribution with zero mean and a range equal to the density difference of the closest available codes, and added to the desired density.
3) The next code above the desired density is printed if the perturbed density is greater than the average value of the two closest codes, otherwise the next code below is printed. For the example given in Fig. 1, code $D_i$ would be printed.

The probability of a particular code being printed in a given cell thus depends on how close its density value is to the desired density. Schroeder [3] reports the use of this technique in the computer production of pictorial output on a microfilm plotter.

The quality of output which may be obtained is illustrated in Fig. 2. A scanner was used to convert the original photograph shown in Fig. 2(a) into a 160 x 128 array of density values. The pictorial output shown in Fig. 2(b) was generated from this array. The output is 128 lines (interpolated on the 160 rows of data) deep by 128 density cells (i.e., character positions) wide. The twenty-one basic codes employed, with their approximate density values on a normalized scale, are given in Table I. These values were obtained with medium print pressure and a well-inked ribbon, using a type QV print train in our IBM 1403 line printer. Intermediate codes, together with their normalized density values, were
TABLE I
DENSITY CODES EMPLOYED

<table>
<thead>
<tr>
<th>Overprinted Character Combinations</th>
<th>Estimated Density Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>blank</td>
<td>0.0</td>
</tr>
<tr>
<td>-</td>
<td>0.15</td>
</tr>
<tr>
<td>=</td>
<td>0.22</td>
</tr>
<tr>
<td>+</td>
<td>0.25</td>
</tr>
<tr>
<td>)</td>
<td>0.29</td>
</tr>
<tr>
<td>1</td>
<td>0.33</td>
</tr>
<tr>
<td>Z</td>
<td>0.37</td>
</tr>
<tr>
<td>X</td>
<td>0.40</td>
</tr>
<tr>
<td>A</td>
<td>0.42</td>
</tr>
<tr>
<td>M</td>
<td>0.45</td>
</tr>
<tr>
<td>O -</td>
<td>0.53</td>
</tr>
<tr>
<td>O =</td>
<td>0.56</td>
</tr>
<tr>
<td>O +</td>
<td>0.60</td>
</tr>
<tr>
<td>O + '</td>
<td>0.64</td>
</tr>
<tr>
<td>O + ' .</td>
<td>0.67</td>
</tr>
<tr>
<td>O + ' . =</td>
<td>0.79</td>
</tr>
<tr>
<td>O X ' . -</td>
<td>0.85</td>
</tr>
<tr>
<td>O X ' . H C</td>
<td>0.89</td>
</tr>
<tr>
<td>O X ' . H B</td>
<td>0.93</td>
</tr>
<tr>
<td>O X ' . H B V</td>
<td>0.97</td>
</tr>
<tr>
<td>O X ' . H B V A</td>
<td>1.00</td>
</tr>
</tbody>
</table>

determined by visual comparison with other codes rather than by measurements of inked-in areas. As a result, the values given in Table I tend to be linear on a scale of perceived rather than measured density.

REFERENCES
An Algorithm For In-Situ Permutation

By I. D. G. Macleod *

In-situ transposition of a vector-stored matrix is an example of the permutation of a set of elements. Any such permutation is composed of a unique set of disjoint cycles, and is most efficiently performed by tracing around cyclic paths and moving elements to their new positions. This paper discusses the problem of choosing "leader" elements for previously untraced cycles. It is shown that while it is usually not possible to say, without an exhaustive search, that an element has not already been moved, it is comparatively easy to identify, and hence reject as leaders, most of those which have. This reduction in the number of exhaustive searches required, together with the ease of recognising cycles of length 1 and length 2, leads to an effective practical algorithm which is described in detail.

I. Introduction

The operation of rearranging N distinct objects amongst themselves is called a permutation of degree N [Ledermann, 1953]. Certain computations require the permutation of array elements, the new position of any element being determined by its old position, rather than by its value. Matrix transposition, and unscrambling of spectral values in the fast Fourier transform, are examples of such computations. When a given element is shifted from its old to its new position, the value of any element already in the new position must be saved. This requirement may be avoided by shifting elements from their old positions in the original array to positions in an auxiliary array corresponding to the new positions in the original. On completion of this process, the values in the auxiliary array may, if required, be copied back into the original array. With large arrays, storage for an auxiliary array may not be available and in-situ permutation becomes desirable. The problem of in-situ permutation of array elements is, however, "so poorly understood that even the in-place transpose of a non-square matrix is impossible to do anywhere near efficiently today" [Gentleman, 1969]. The algorithm described below is believed to be the most efficient yet devised, but its efficiency varies from reasonable to poor, depending on certain characteristics of the desired permutation. The problem of devising an algorithm which is reasonably efficient for all permutations remains to be solved.

II. Permutation via Cycle-Tracing

Any permutation is composed of a unique set of disjoint cycles of one or more elements [Ledermann, 1953]. It is assumed that there are NA elements in the set to be permuted, and that these elements occupy contiguous positions in a storage vector A, elements being referenced by 1-origin indices on A (i.e. indices ranging from 1 to NA). Each element has two distinct attributes—its position and its value. In practice it is the element values which are permuted.

In-situ transposition of a vector-stored matrix satisfies our definition of a permutation, and the following discussion will use this process as an example. Fig. 1 shows the rearrangement of elements required when an 8x2 matrix, stored in row-major order in a vector A (of NA=16 elements), is transposed in-situ. Also shown are the cycles which compose this permutation; each element has to be moved to the position originally occupied by its anti-clockwise cyclic neighbour. Thus, for the second cycle of Fig. 1, we could in turn enter at element 2 and save its value, replace element 2 by element 5, element 5 by element 3, element 3 by element 9, and element 9 by the saved value of element 2. By tracing out all cycles in this manner, the desired permutation is achieved with the use of only one position of auxiliary storage. Note that no action is necessary for cycles of length 1 (i.e., only one element).

Any selected element has two cyclic neighbours. In cycles of length 1 and length 2, these neighbours will not be distinct. We refer to the clockwise and anti-clockwise cyclic neighbours as the "predecessor" and "successor" elements and say that they are in the "last" and "next" positions relative to the selected element, respectively. Given the form of the desired permutation, it will usually be possible to specify procedures for computing the indices of both cyclic neighbours as a function of the index of the selected element; these procedures will thus define the permutation to be executed.

For an in-situ M×N matrix transposition, it can be shown that if K is the index of the selected element, then

\[ \text{NEXT}(K) = [(K-1) \times N \mod (M \times N - 1)] + 1 \]

and

\[ \text{LAST}(K) = [(K-1) \times M \mod (M \times N - 1)] + 1 \]

For the 8×2 matrix example, the successor element for element 4 is

\[ \text{NEXT}(4) = [(4-1) \times 2 \mod (8 \times 2 - 1)] + 1 = \text{element 7} \]

and the predecessor element is

\[ \text{LAST}(4) = [(4-1) \times 8 \mod (8 \times 2 - 1)] + 1 = \text{element 10} \]

* Department of Engineering Physics, Research School of Physical Sciences, The Australian National University, Canberra, A.C.T 2600. Manuscript received July, 1969, in revised form October, 1969.
There are several possible approaches to executing permutations by means of cycle-tracing. One approach is to completely specify the cycles of length 2 or greater which compose the desired permutation, and to use this specification in executing the permutation. This approach minimizes execution time (if the time necessary to determine the complete specification is not taken into account), but the specification requires approximately \( NA/2 \) storage elements.

A second approach is to supply the means for tracing around cycles (i.e., the procedures for computing \( \text{NEXT} \) and/or \( \text{LAST} \)), and a table of representative or "leader" elements, one for each cycle of length 2 or greater. Pall and Seiden describe the use of this approach for in-situ transposition [Pall and Seiden, 1960]. The size of the leader table is usually not predictable and could require up to \( NA/2 \) elements (if all cycles were of length 2).

A third approach is to determine leaders during the execution of the desired permutation. This approach is logically complete and minimizes storage requirements, but it is slower than either of the above approaches. The problem of determining leaders in a reasonably efficient manner has not yet been satisfactorily solved; much more time may be spent in searching for leaders than in moving elements to their new positions. The algorithm \text{PERMUTE}, described in section III below, follows this third approach and employs a technique for determining leaders which is believed to be the most efficient yet devised.

It is not necessary to trace out each cycle in one step. Boothroyd employs a technique of exchanging pairs of elements lying on the same cycle, in an algorithm for in-situ transposition which does not rely upon leaders being found [Boothroyd, 1967]. Both Boothroyd's and Pall and Seiden's algorithms could readily be modified to accommodate more general permutations than in-situ transposition.

III. The Algorithm \text{PERMUTE}

In this algorithm, the element with the smallest vector index in each cycle is designated the leader element. The input vector is stepped through one element at a time, with cycles being traced and elements moved to their new positions as leader elements are encountered.

Several advantages accrue from the chosen definition of leader elements. Firstly, this is a simple and unambiguous definition. Secondly, if either the predecessor or successor element has a smaller vector index than the selected "candidate" element, it cannot be a leader. It is intuitively expected, and has been experimentally verified, that approximately three-quarters of all candidate elements will be rejected as leaders by this simple initial test. To pass the initial test a candidate need not be a genuine leader, and an exhaustive test may be required. No further testing is required, however, in either of the following special cases:

1. The current cycle is of length 2 or length 1, as indicated by the successor and predecessor indices being equal.
2. No cycles of length greater than 2 have previously been traced.

There are other less easily detectable special cases which may be worthwhile checking for, but these are beyond the scope of the present discussion.

The exhaustive test consists of tracing around the current cycle (without moving elements) until (i) the cycle "breaks" with an element being encountered whose index is smaller than the candidate's, thus proving that the candidate is not a leader, or (ii) the cycle closes without such an element being encountered, thus proving that the candidate is a leader. In this latter case, the current cycle can be retraced, this time moving elements to their new positions. By adopting the above approach, any elements shown to be leaders will lie on cycles which have not previously been traced, whereas elements shown not to be leaders will have already been moved to their new positions.

It is not uncommon for a permutation to be composed of a relatively small number of long cycles, such that the permutation will actually have been completed after only a few of the elements have been examined as possible leaders. For the example of Fig. 1, the permutation will have been completed after the eighth element has been processed, because the only remaining cycle is of length 1. Except in the case of there being more than one cycle of length 1 remaining, it is possible to detect such "premature" completions of processing by keeping a tally of elements moved to their new positions, and comparing this at the end of each step with the number of elements in the permutation.

The algorithm \text{PERMUTE} is given as an ALGOL procedure in the appendix; its operation may be summarized as follows:

1. Set \( \text{TALLY} = 0 \), \( \text{CAND} = 0 \), \( \text{TRIVIAL} = \text{true} \).
2. \( \text{CAND} = \text{CAND} + 1 \); where \( \text{CAND} \) is the index of the next candidate element to be tested as a possible leader.
3. \( \text{SUC} = \text{NEXT} (\text{CAND}) \); where \( \text{SUC} \) is the index of \( \text{CAND} \)'s successor.
   If \( \text{SUC} < \text{CAND} \), \( \text{CAND} \) is not a leader; go to step 2.
4. \( \text{PRE} = \text{LAST} (\text{CAND}) \); where \( \text{PRE} \) is the index of \( \text{CAND} \)'s predecessor.
   If \( \text{PRE} < \text{CAND} \), \( \text{CAND} \) is not a leader; go to step 2.
5. \( \text{CAND} \) has passed the initial test. If \( \text{PRE} = \text{SUC} \) we have a short cycle and \( \text{CAND} \) is definitely a leader; go to step 6.
   If \( \text{PRE} \neq \text{SUC} \) we have a longer cycle, but if only trivial cycles have previously been traced (i.e., \( \text{TRIVIAL} \) is true) \( \text{CAND} \) must be a leader; go to step 7.
   If \( \text{TRIVIAL} \) is false, an exhaustive search is required; go to step 8.
6. If \( \text{SUC} \neq \text{CAND} \), there are two elements in the cycle; exchange these, increment \( \text{TALLY} \) by 2 and go to step 9. Otherwise, \( \text{SUC} = \text{CAND} = \text{PRE} \) and we have only one element in the cycle; increment \( \text{TALLY} \) by 1 and go to step 9.
7. Set \( \text{TRIVIAL} = \text{false} \) and trace out the current cycle, moving elements to their new positions and incrementing \( \text{TALLY} \); go to step 9.
(8) Trace around the current cycle without moving elements. If the cycle breaks, go to step 2. Otherwise, retrace the cycle after it closes, this time moving elements to their new positions and incrementing TALLY.

(9) If TALLY < NA—1 go back to step 2 and continue, otherwise no further processing is required. Note that if TALLY = NA — 1, there is one cycle of length 1 remaining, but no action is required for this.

IV. Discussion

PERMUTE is a quite general algorithm and for this reason it will usually be possible to modify it to obtain improved efficiency for a restricted class of permutations. With in-situ transposition, for example, improved efficiency may be obtained in the following ways:

(i) The procedures NEXT and LAST can be coded in-line, thus saving the overhead associated with invocation of separate procedures.

(ii) NEXT (1) = 1, and NEXT (M x N) = M x N, for all M and N. The first and last elements of A therefore lie on cycles of length 1 and do not have to be moved. As a consequence, the processing can commence at the second elements with TALLY set to 1, and cease when TALLY ≥ NA—2.

(iii) The indices of the successor and predecessor elements for element K — 1 can be used to simplify the computation of the corresponding indices for element K [Boothroyd, 1967].

Timing tests for in-situ matrix transposition were carried out using (i) PERMUTE (modified as above), (ii) Boothroyd's TRANSPOSE, and (iii) a routine TABLE which employed the second approach outlined in section II above and required a table of leaders. An extended version of PERMUTE stored the indices of elements found to be leaders in this table, prior to the execution of TABLE. All tests were performed using FORTRAN IV on the ANU's IBM 360/50. Of the approximately 40 matrix transpositions timed, PERMUTE was relatively most efficient for a 90 x 122 matrix and least efficient for a 163 x 139 matrix. The 90 x 122 matrix took PERMUTE TRANSPOSE, and TABLE 1.9, 11.9, and 1.8 seconds respectively. The corresponding times for the 163 x 139 matrix were 20.3, 19.9, and 3.7 seconds.

PERMUTE is most efficient for transpositions which are composed of a small number of relatively long cycles, and which complete after only a few elements have been examined as possible leaders. PERMUTE is least efficient for transpositions which do not complete until most elements have been examined as possible leaders, and which contain a large number of "false" leaders (which pass the initial test and are rejected during the exhaustive test).

Several possible extensions to PERMUTE have been attempted, with the aim of making its rather variable efficiency more uniform, but except in special cases, no marked improvements have been obtained. The main cause of inefficiency is the time wasted in exhaustive testing of candidates which ultimately fail.

In summary, PERMUTE is not a final answer for the problem of in-situ permutation, but it is sometimes much faster than comparable algorithms.

**ORIGINAL MATRIX**

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>i</th>
<th>j</th>
<th>k</th>
<th>l</th>
<th>m</th>
<th>n</th>
<th>o</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TRANSPOSED MATRIX**

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>i</th>
<th>j</th>
<th>k</th>
<th>l</th>
<th>m</th>
<th>n</th>
<th>o</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**MAPPING INTO STORAGE VECTOR**

(ROW-MAJOR ORDER)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
<td>i</td>
<td>j</td>
<td>k</td>
<td>l</td>
<td>m</td>
<td>n</td>
<td>o</td>
<td>p</td>
</tr>
</tbody>
</table>

**CYCLES COMPOSING THE TRANSPOSITION**

![Figure 1: Matrix Transposition Example.](image)

Acknowledgement

The author is grateful to Mr. J. Boothroyd of the University of Tasmania Computing Centre for helpful discussions concerning this work.

References:


Appendix

procedure PERMUTE (A, NA);
value NA; integer NA; array A;
begin in teger SUC, PRE, CAND, CRRNT, LIMIT, TALLY;
real SAVE;
boolean TRIVIAL;
STEP1: TALLY := CAND := 0; TRIVIAL := true;
LIMIT := NA — 1;
STEP2: CAND := CAND + 1;
In-Situ Permutation

**Step 3:**
- `SUC := NEXT(CAND);`
- If `SUC < CAND` then go to Step 2;

**Step 4:**
- `PRE := LAST(CAND);`
- If `PRE < CAND` then go to Step 2;

**Step 5:**
- If `PRE ≠ SUC` then go to if TRIVIAL
  - Then Step 7 else Step 8;

**Step 6:**
- If `SUC ≠ CAND`
  - Then begin
    - `SAVE := A[CAND];`
    - `A[SUC] := SAVE;`
    - `TALLY := TALLY + 2;`
    - Go to Step 9
  - Else begin
    - `TALLY := TALLY - 1;`
    - Go to Step 9
  - End

**Step 7:**
- `TRIVIAL := false;` go to RETRACE;

**Step 8:**
- `CRRNT := SUC;`

**Trace:**
- `CRRNT := NEXT(CRRNT);`
- If `CRRNT = PRE` then go to RETRACE;
- If `CRRNT > CAND` then go to TRACE
  - Else go to Step 2;

**Retrace:**
- `SAVE := A[CAND];`
- `CRRNT := CAND;`

**Move:**
- `TALLY := TALLY + 1;`
- `CRRNT := SUC;`
- `SUC := NEXT(SUC);`
- If `SUC ≠ PRE` then go to MOVE;
- `A[PRE] := SAVE;`
- `TALLY := TALLY + 2;`

**Step 9:**
- If `TALLY < LIMIT` then go to Step 2
- End of permutation via cycle tracing;

**Note:** The publisher's errors corrected above are the subject of a forthcoming erratum.
Appendix 4  ON THE IMPLEMENTATION OF SLIP

In implementing Weizenbaum's (1963) Symmetric List Processor (SLIP) on the ANU's IBM 360/50 computer, several corrections, extensions, and improvements have been effected. At the time of these experiments, the only two SLIP implementations for the IBM/360 range of computers that the author was aware of, were an implementation in FORTRAN IV by Holly (1966), which uses System 360 decimal representation for linkages and does not include the recursive SLIP facilities, and an implementation in PL/I by Lamberth (1969), which is closer to the original SLIP, but uses PL/I recursion facilities rather than those specified by Weizenbaum.

SLIP is designed around a large number of often very short subroutines, and PL/I's overhead in the invocation of subroutines causes a marked degradation in performance. For this and other reasons, a System 360 FORTRAN IV (with ASSEMBLER) version of SLIP, more in the spirit of the original than Holly's, and similar in some respects to Lamberth's, has been implemented, complete with recursive facilities. In this implementation, System 360's 32-bit words are divided into a 2-bit identifier field and two 15-bit linkage fields, which are used to address 32-bit words relative to the start of the list of available space (LAVS) rather than 8-bit bytes relative to the beginning of core, as in System 360 machine architecture. This linkage system allows a maximum LAVS size of 16K double-word cells and also allows the current LAVS to be written to disc at any stage and to be used for restarting later execution, regardless of the new absolute address of the LAVS. Many errors and redundancies in Weizenbaum's original listing of the SLIP subroutines (1963, Appendix) have been reported in the letters section of the Communications of the ACM, (Yarborough, 1964; Elkin, 1964; Russell, 1965; Weizenbaum, 1964, 1965), and also by O'Neill (1965). Several additional errors, omissions, and redundancies have been discovered during the current implementation and application of the FORTRAN IV version of SLIP; among these are:-

1) In line 28 of the original listing,

CALL SETDIR (-1,-1, LNKR (CONT (M)), AVSL),

the invocation of LNKR is redundant because in filling in the fields
of AVSL, SETDIR uses only the LNKR portion of CONT(M) at any rate. This redundancy is repeated many times, e.g., lines 60, 85, and 86.

2) In the function INLSTL (lines 147-160) and INLSTR (lines 161-174), no check is made to see if the list being inserted is empty, in which case no insertion should take place. In fact, with the code given, the header cell of an empty list is inserted in the host list, which then contains two header cells.

3) Function NAMEDL does not return the name of the description list in NAME format.

4) In function RDLSTA, line 679 is redundant because the same statement is repeated at line 682.

5) The listing for function LSTMRK, although trivial, is not given.

In addition to correcting these shortcomings, it was noted that there were nearly 70 statements involving the sequences LNKR (CONT (............)), LNKL (CONT (............)), or ID (CONT (............)), so three new primitives — LNKRI, LNKLI, and IDI (i.e., indirect link-right, link-left, and identifier) — have been introduced, with a resultant simplification of many of the SLIP routines.
Appendix 5  COMPUTER PROGRAM LISTINGS

Selected programs which have been developed in the practical experiments described in Chapter 5, are listed below in alphabetical order. All the programs listed are coded in FORTRAN IV. There are frequent references to subprograms, particularly to SLIP routines (Weizenbaum, 1963) — familiarity with SLIP is presumed of the reader (refer to Appendix 4 for details of the present implementation and the added primitives LNKLI, LNKRI, and IDI). Other routines invoked by the listed programs but not included here are:-

(i)  NCELL(L) — an integer function which returns the number of cells in list L;
(ii) REVERS(L) — a subroutine which reverses the order of cells on list L;
(iii) SETB12(WORD, I1, I2), SETH1(WORD, I1), GETB34(WORD, I3, I4), etc. — ASSEMBLER subroutines which take 32-bit WORDS and set or get the 8-bit (Byte) or 16-bit (Halfword) fields indicated by the subroutine name, e.g., GETB3(WORD, I3) takes the third 8-bit byte of WORD, pads it with high-order zeros to 32 bits, and returns this value in I3;
(iv) SETBIT(A, M, N, I, J, K), GETBIT(A, M, N, I, J, K) — ASSEMBLER subroutines which enable the (I, J)th element in an array A(M,N) of 1-bit elements to be referenced, the element value (i.e., 0 or 1) being either set from K or returned in K, according as SETBIT or GETBIT is invoked.
(v) GETSTG(VECT, NE, LE, IOF), FRESTG(VECT, NE, LE, IOF) — ASSEMBLER subroutines which enable FORTRAN programs to acquire and free dynamic storage. A vector of one element, VECT(1), is defined in the program which requires dynamic storage, and the number (NE) and length (LE, in bytes) of elements required are set. On return from GETSTG, IOF is set to a positive or negative integer such that VECT(IOF + 1) addresses the Ith element of a block of dynamically acquired storage. This block is freed by a call to FRESTG.
SUBROUTINE BREAK1(A,B,BRK,M,THR1,FRACT1,DELTA,THR2,NBRK,NBRKN)
A IS VECTOR OF CURVATURES
B IS VECTOR OF ANGLES
M IS NO. OF ANGLES
SUGGESTED BREAK POINTS ARE RETURNED IN BRK
THR1 IS THRESHOLD FOR ABSOLUTE CURVATURE BELOW WHICH NO BREAK
POINYS ARE SUGGESTED
FRACT1 IS FRACTION OF CURVATURE AT PEAKS WHICH MUST NOT BE EXCEEDED
AT PLUS OR MINUS DELTA FROM THE PEAKS FOR PEAKS TO BE LABELLED AS
BREAK POINTS
THR2 IS FRACTION OF CURVATURE OF THE SMALLEST OF TWO ADJACENT PEAKS
WHICH MUST NOT BE GREATER THAN THE CURVATURE BETWEEN THE PEAKS
IF THE PEAKS ARE TO BE CONSIDERED AS SEPARATE
THE NO. OF BREAK POINTS WHICH ARE STEEP ON BOTH SIDES IS RETURNED IN NBRK
THE NO. OF BREAK POINTS WHICH ARE STEEP ON ONE SIDE IS RETURNED IN NBRKN
BREAK POINTS SUGGESTED ARE POSITIONED AT THE START OF A NEW CURVE

REAL A(M),BK(M),B(M)
INTEGER ILOC(25)
REAL BMN(25)
LOGICAL JOINED
LOGICAL TEST1,TST2
NBRK=0
NBRKN=0
NBRKTT=0
DO 10 I=1,M
  IF(M.LE.3) RETURN
  M=M-1
  DO 20 J=1,M
    ABSAI=ABS(A(J))
    IF(ABSAI.LT.THR1)GO TO 20
    SIGNAL=SIGN(1.0,A(J))
    IF(ABSAL.GE.ABS(A(J-1)) .OR. SIGNAL.NE.SIGNAL)TEST1=.TRUE.
    IF(ABSAL.GE.ABS(A(J+1)) .OR. SIGNAL.NE.SIGNAL)TEST2=.TRUE.
    IF(.NOT.TEST1.AND..NOT.TEST2)GO TO 20
    TEST1=.FALSE.
    TEST2=.FALSE.
    NBRK=NBRK+1
  20 CONTINUE
  IF(ABSAL.GE.0.5)GO TO 21
  21 CONTINUE
  IF(IHI.LE.M)GO TO 22
  22 CONTINUE
  AH=A(IHI)*DIFFH+A(IHH)*DIFFH
  AL=A(IHI)*DIFFL+A(IHL)*DIFFL
  AH=AL/FRACT1
  AL=AH/FRACT1
  SIGNAL=SIGN(1.0,AL)
  SIGNALH=SIGN(1.0,AH)
  TEST1=.TRUE.
  TEST2=.TRUE.
  IF(ABSAL.GE.0.5)GO TO 20
  NBRK=NBRKTT+1
  IF(.NOT.TEST1.AND..NOT.TEST2)GO TO 20
  NBRK=NBRK+1

REAL AIM) ,8RK(,B((
INTEGER ILOC(25)
REAL BMN(25)
LOGICAL JOINED
LOGICAL TEST1,TST2
NBRK=0
NBRKN=0
NBRKTT=0
DO 10 I=1,M
  IF(M.LE.3) RETURN
  M=M-1
  DO 20 J=1,M
    ABSAI=ABS(A(J))
    IF(ABSAI.LT.THR1)GO TO 20
    SIGNAL=SIGN(1.0,A(J))
    IF(ABSAL.GE.ABS(A(J-1)) .OR. SIGNAL.NE.SIGNAL)TEST1=.TRUE.
    IF(ABSAL.GE.ABS(A(J+1)) .OR. SIGNAL.NE.SIGNAL)TEST2=.TRUE.
    IF(.NOT.TEST1.AND..NOT.TEST2)GO TO 20
    TEST1=.FALSE.
    TEST2=.FALSE.
    NBRK=NBRK+1
  20 CONTINUE
  IF(ABSAL.GE.0.5)GO TO 21
  21 CONTINUE
  IF(IHI.LE.M)GO TO 22
  22 CONTINUE
  AH=A(IHI)*DIFFH+A(IHH)*DIFFH
  AL=A(IHI)*DIFFL+A(IHL)*DIFFL
  AH=AL/FRACT1
  AL=AH/FRACT1
  SIGNAL=SIGN(1.0,AL)
  SIGNALH=SIGN(1.0,AH)
  TEST1=.TRUE.
  TEST2=.TRUE.
  IF(ABSAL.GE.0.5)GO TO 20
  NBRK=NBRKTT+1
  IF(.NOT.TEST1.AND..NOT.TEST2)GO TO 20
  NBRK=NBRK+1
C DOES MAXIMUM FALL AWAY STEEPLY ON ONLY ONE SIDE
IF (.NOT. TEST1 .AND. TEST2) BPK(I) = -ABS(A)
IF (.NOT. TEST2 .AND. TEST1) BPK(I) = -ABS(A)
20 CONTINUE
NBRK = NBRK - NBRK
IF (ABS(B(I)) = 0.0) RETURN
IF (ABS(B(I)) < 0.0) GO TO 50
SIGNL = 0.0
NL0C = 0

C FIND MINIMUM CURVATURE BETWEEN ADJACENT PEAKS
DO 24 I = 2, M1
IF (B(I) .LE. 0.0) GO TO 25
NL0C = NL0C + 1
IF (NL0C .LE. 25) CALL ABEND(1111, 'MORE THAN 25 BREAK POINTS')
BL0C(NL0C) = I
IF (SIGN1 .LT. 0.0) .AND. (NBRK .LT. 0.0) BMNMM = 0.0
IF (NL0C .LT. I) BMN(INL0C-1) = BMNMM
SIGNL = SIGN1 .LT. 0.0
BMNMM = A(I)
GO TO 24

25 CONTINUE
IF (NL0C .LT. 1) RETURN
IF (SIGNL .LT. 0.0) .AND. (A(I) .LT. BMNMM) BMNMM = A(I)
IF (SIGNL .GT. 0.0) .AND. (A(I) .GT. BMNMM) BMNMM = A(I)

C IS GROOVE BETWEEN PEAKS DEEP ENOUGH TO SEPARATE THEM
IF (MIN(I) .LT. THRH1 .AND. MIN(I) .GT. THRH2 .AND. SIGN1 .LT. 0.0) GO TO 40
IF (MIN(I) .GT. THRH1 .AND. MIN(I) .GT. THRH2 .AND. SIGN1 .LT. 0.0) GO TO 40
IF (SIGN1 .LT. 0.0) .AND. (NBRK .LT. 0.0) BMNMM = 0.0
JOIN = .FALSE.

C START OF JOINED PEAKS
ISTART = I
IEND = I + 1
IF (I .NE. NL0C) GO TO 40
GO TO 43

40 CONTINUE
JOIN = .FALSE.
SUM1 = 0.0
SUM2 = 0.0

C MERGE PEAKS AND FORM ONE BREAK POINT
DO 44 II = ISTART, IEND
BRK(I) = 0.0
SUM = A(BLOC(I)) .PLUS. BLOC(I) .PLUS. SUM1
44 SUM = A(BLOC(I)) .PLUS. SUM1
SUM2 = SUM1 .PLUS. SUM2
SUM1 = SUM2 .PLUS. (IEND - ISTART + 1)
JOIN = .FALSE.

C POSITION BREAK POINT MORE ACCURATELY BY SHIFTING AT MOST PLUS
OR MINUS ONE UNIT
C POSITION BREAK POINT SUCH THAT LARGEST CHANGE IN ANGLE IS BETWEEN
C START OF NEW CURVE AND END CF CLD CURVE
50 CONTINUE
JOIN = .FALSE.

C AVOID SHIFTING BREAK POINT BY MORE THAN ONE UNIT
IF (JOIN .LT. 0.0) GO TO 53
IF (JOIN .LT. 0.0) GO TO 51
JOIN = 0
BRK(I) = 0.0
ATL = ABS(B(I)) - B(I-1)
ALL = ATL
IF (ATL .LT. 0.0) ALL = -ABS(B(I-1)) - B(I-2)
ANT = ABS(B(I-1)) - B(I-2)
SUBROUTINE COOL(XX,NN)

PURPOSE
PERFORMS DISCRETE COMPLEX FOURIER TRANSFORMS ON A COMPLEX ARRAY

USAGE
CALL COOL(XX,NN)

DESCRIPTION OF PARAMETERS
XX  - AS INPUT, XX CONTAINS THE COMPLEX ARRAY TO BE TRANSFORMED.
     AS OUTPUT, XX CONTAINS THE COMPLEX FOURIER TRANSFORM.

NN  - MAGNITUDE OF NN = LOGARITHM TO THE BASE 2 OF THE NUMBER OF INPUT POINTS.
     SIGN OF NN INDICATES WHETHER A FORWARD OR REVERSE TRANSFORM IS TO BE PERFORMED.
     IF NN>0 THEN THE FORWARD TRANSFORM IS PERFORMED.
     OTHERWISE IF NN<0 THEN THE REVERSE TRANSFORM IS PERFORMED.
     THE MAGNITUDE OF NN MUST BE IN THE RANGE 3 TO 20 INCLUSIVE.

REMARKS
THE DEFINITION OF FORWARD AND INVERSE TRANSFORMS AGREES WITH THE USAGE IN IBM'S SSP SUBROUTINE FARM. THIS USAGE VARIES FROM THAT USED IN SOME OTHER FOURIER TRANSFORM Routines.

THE FORWARD TRANSFORM IS DEFINED AS
\[ F(J) = \sum_{K=0}^{N-1} A(K) \cdot W^{K \cdot J} \]
WHERE \( N = 2^N \)
AND \( W \) IS THE \( N \)TH ROOT OF UNITY

THE REVERSE TRANSFORM IS DEFINED AS
\[ A(K) = \frac{1}{N} \sum_{J=0}^{N-1} F(J) \cdot W^{-K \cdot J} \]

REAL XX(1),W(14)
INTEGER JNT(20),OFFSET,NPIT(20),KADD(14,6,10,2,12,4,8,0/ 
N=ABS(NN)
NXON4=2**N(N-2)
NXON2=NXON4+NXON4
NX=NXON2+NXON2
NX2=NX+NX
NX70N4=NX+NX0N2+NX0N2
CON1=6.28318530717959/NX
RT2INV0=0.707106781186548
NWI=0
ASSIGN 81 TO LABEL IF(NX*LT.32)ASSIGN 82 TO LABEL
DO 3 K3,N
NBIT(K)=0
3 JNT(K)=2**N(N-K)
IF(NX*GT.0)GO TO 2
DO 1 K1,NX2LS1,2

END
1 \(XX(K+1) = -XX(K+1)\)

2 \(\text{LSTART} = N - N/3 \times 3 + 1\)

3 IF(\(\text{LSTART} - 2\) \(\neq\) \(9,8,6\), \(5 \text{ DO } 6 \text{ K0} = 1, \text{NXCN2}, 2\))

4 \(K1 = K0 + \text{NXCN2}\)

5 \(K2 = K1 + \text{NXCN2}\)

6 \(K3 = K2 + \text{NXCN2}\)

7 \(A0R = XX(K0) + XX(K2)\)

8 \(A0I = XX(K0+1) + XX(K2+1)\)

9 \(A1R = XX(K0) - XX(K2)\)

10 \(A1I = XX(K0+1) - XX(K2+1)\)

11 \(A2R = XX(K1) + XX(K3)\)

12 \(A2I = XX(K1+1) + XX(K3+1)\)

13 \(A3R = XX(K1) - XX(K3)\)

14 \(A3I = XX(K1+1) - XX(K3+1)\)

15 \(XX(K0) = A0R + A2R\)

16 \(XX(K0+1) = A0I + A2I\)

17 \(XX(K1) = A1R + A3I\)

18 \(XX(K1+1) = A1I + A3R\)

19 \(XX(K2) = A1R - A3I\)

20 \(XX(K2+1) = A1I + A3R\)

21 \(XX(K3) = A1R + A3I\)

22 \(XX(K3+1) = A1I - A3R\)

23 \(6 \text{ XX(K3+1) = A1I - A3R} \)

24 \(5 \text{ XX(K0) = A0R + A2R}\)

25 \(4 \text{ XX(K0+1) = A0I + A2I}\)

26 \(3 \text{ XX(K1) = A1R + A3I}\)

27 \(2 \text{ XX(K1+1) = A1I + A3R}\)

28 \(1 \text{ XX(K2) = A1R - A3I}\)

29 \(\text{DO } 5 \text{ K0} = 1, \text{NXCN2}, 2\)

30 \(K1 = K0 + \text{NXCN2}\)

31 \(A0R = XX(K0) + XX(K2)\)

32 \(A0I = XX(K0+1) + XX(K2+1)\)

33 \(A1R = XX(K0) - XX(K2)\)

34 \(A1I = XX(K0+1) - XX(K2+1)\)

35 \(A2R = XX(K1) + XX(K3)\)

36 \(A2I = XX(K1+1) + XX(K3+1)\)

37 \(A3R = XX(K1) - XX(K3)\)

38 \(A3I = XX(K1+1) - XX(K3+1)\)

39 \(XX(K0) = A0R + A2R\)

40 \(XX(K0+1) = A0I + A2I\)

41 \(XX(K1) = A1R + A3I\)

42 \(XX(K1+1) = A1I + A3R\)

43 \(XX(K2) = A1R - A3I\)

44 \(XX(K2+1) = A1I + A3R\)

45 \(XX(K3) = A1R + A3I\)

46 \(XX(K3+1) = A1I - A3R\)

47 \(5 \text{ XX(K3+1) = A1I - A3R}\)

48 \(4 \text{ XX(K0) = A0R + A2R}\)

49 \(3 \text{ XX(K0+1) = A0I + A2I}\)

50 \(2 \text{ XX(K1) = A1R + A3I}\)

51 \(1 \text{ XX(K1+1) = A1I + A3R}\)

52 \(\text{DO } 9 \text{ LAYER = LSTART + N, 3}\)

53 \(L8LK2 = 2^{(N - LAYER - 1)}\)

54 \(L8LK1 = L8LK2 - 2\)

55 \(L8LK8 = L8LK2 \times 8\)

56 \(\text{DO } 22 \text{ OFFSET} = 1, \text{NXCN2}, L8LK8\)

57 \(\text{IF(OFFSET NE. 0, 1) GO TO 52}\)

58 \(\text{ARG} = \text{CON1} \times \text{W1}\)

59 \(w(1) = \cos(\text{ARG})\)

60 \(w(2) = \sin(\text{ARG})\)

61 \(\cos\theta = w(1) \times w(1) + w(2) \times w(2)\)

62 \(\sin\theta = 2 \times w(1) \times w(2)\)

63 \(w(3) = \cos(\theta - 1)\)

64 \(w(4) = \cos(\theta)\)

65 \(w(5) = \cos(\theta + 3)\)

66 \(w(6) = \cos(\theta + 1)\)

67 \(\cos\theta = w(1) \times w(1) + w(2) \times w(2)\)

68 \(\sin\theta = 2 \times w(1) \times w(2)\)

69 \(w(7) = \cos(\theta - 1)\)

70 \(w(8) = \cos(\theta)\)

71 \(w(9) = \cos(\theta + 3)\)

72 \(w(10) = \cos(\theta + 1)\)

73 \(w(11) = \cos(\theta - 1)\)

74 \(w(12) = \cos(\theta)\)

75 \(w(13) = \cos(\theta + 3)\)

76 \(w(14) = \cos(\theta + 1)\)

77 \(\text{L8LKO = OFFSET + L8LK1}\)

78 \(\text{DO } 25 \text{ K0 = OFFSET + L8LKO, 2}\)

79 \(K1 = K0 + L8LK2\)

80 \(K2 = K1 + L8LK2\)

81 \(K3 = K2 + L8LK2\)

82 \(K4 = K3 + L8LK2\)

83 \(K5 = K4 + L8LK2\)

84 \(K6 = K5 + L8LK2\)

85 \(K7 = K6 + L8LK2\)

86 \(K0 = XX(K0)\)

87 \(XX(K0+1) = XX(K0)\)

88 \(\text{IF(OFFSET NE. 0, 1) GO TO 21}\)

89 \(\text{XX(K0+1) = XX(K0)}\)

90 \(\text{XX(K1) = XX(K1+1)}\)

91 \(\text{XX(K2) = XX(K2+1)}\)

92 \(\text{XX(K3) = XX(K3+1)}\)

93 \(\text{XX(K4) = XX(K4+1)}\)

94 \(\text{XX(K5) = XX(K5+1)}\)

95 \(\text{XX(K6) = XX(K6)}\)
COOL A5-6

21 XK1 WR = XX(K1) * W(1) - XX(K1 + 1) * W(2)
   XK2 WR = XX(K2) * W(3) - XX(K2 + 1) * W(4)
   XK3 WR = XX(K3) * W(5) - XX(K3 + 1) * W(6)
   XK4 WR = XX(K4) * W(7) - XX(K4 + 1) * W(8)
   XK5 WR = XX(K5) * W(9) - XX(K5 + 1) * W(10)
   XK6 WR = XX(K6) * W(11) - XX(K6 + 1) * W(12)
   XK7 WR = XX(K7) * W(13) - XX(K7 + 1) * W(14)

   AR = XK0 WR + XK4 WR
   A01 = XK0 WI + XK4 WI
   A1R = XK1 WR + XK5 WR
   A1I = XK1 WI + XK5 WI
   A2R = XK2 WR + XK6 WR
   A2I = XK2 WI + XK6 WI
   A3R = XK3 WR + XK7 WR
   A3I = XK3 WI + XK7 WI
   A4R = A0R + A2R
   A4I = A0I + A2I
   A5R = A0R - A2R
   A5I = A0I - A2I
   A6R = (A1R - A3R) * RT2 INV
   A6I = (A1I + A3I) * RT2 INV
   A7R = (A3R - A1R) * RT2 INV
   A7I = (A3I + A1I) * RT2 INV
   XX(K4) = A4R + A6R
   XX(K4 + 1) = A4I + A6I
   XX(K5) = A4R - A6R
   XX(K5 + 1) = A4I - A6I

25 XX(K7 + 1) = A5I - A7I

D0 200 K=4,N
IN(BIT(K)*NE.0) GO TO 259
N61 = NL1 + JNT(K)
G0 TO 22

259 NB1T(K) = 0
260 NL1 = NL1 - JNT(K)
22 CONTINUE

D0 200 K=4,N

221 NB1T(K) = C
10 CONTINUE
D0 80 K=0,NX2LS1,16
NL1 = NL1 + NX7ON4
SUBROUTINE LSTPLT(A,M,N,LST,ICFF,JOFF,NEXT,OCBOX)

C ENTERS CHARACTERS INTO A LOGICAL*1 ARRAY A(M,N) TO REPRESENT STRINGS
C OF CELLS ENCODED IN SLIP LISTS; THIS CHARACTER ARRAY HAS TO BE
C PRINTED OUT BY ANOTHER SUBROUTINE. THE STARTS OF LISTS ARE INDICATED
C BY AN ASTERISK, OTHER CELLS BEING INDICATED BY A CHARACTER
C CHOSEN SO THAT NO ADJACENT CHARACTERS BELONGING TO SEPARATE LISTS
C ARE THE SAME. SINGLE-CELL LISTS ARE PRINTED AS A SINGLE CHARACTER,
C AND NOT AS AN ASTERISK. THE CHARACTER SELECTED FOR TESTING TO
C SEE IF IT IS SUITABLE, IS INITIALLY DETERMINED BY NEXT. THEREAFTER
C THE NEXT CHARACTER IN SEQUENCE TO THE ONE PREVIOUSLY EMPLOYED IS
C TESTED. IF AN ADJACENT CHARACTER WHICH IS THE SAME AS THE SELECTED
C CHARACTER IS DISCOVERED, THEN THE NEXT CHARACTER IN THE SEQUENCE
C IS SELECTED AND TESTED.
C LST IS THE NAME OF THE LIST TO BE PLOTTED.
C IOFF AND JOFF ARE USED TO OFFSET THE PLOTTED LIST IN THE ARRAY A
C IF REQUIRED.
C OCBOX INDICATES WHETHER OR NOT THE LIST TO BE PLOTTED HAS AN
C OCCUPANCY BOX AS THE FIRST CELL.
C LOGICAL*1 A(M,N),CHARTH,USEO(36),STAR/'**'/,NUMB/'#'/
C LOGICAL CHARA0,OCBOX
C LOGICAL*1 AIJ,AI3J3,FIRST
C LESSI=-1
C IF(NEXT.LT.1.OR.NEXT.GT.36)NEXT=1
C SELECT INITIAL CHARACTER TO BE TRIED.
C S=SEQD(LST)
DO 5 K=1,36
5 USEO(K)=.FALSE.
C IF(OCBOX)WORD=SEQR(S,IND)
C GET NEXT CELL OFF THE CURRENT LIST AND FIND THE I AND J LOCATIONS.
10 WORD=SEQR(S,IND)
IF(IND.GT.0)GO TO 20
CALL GETB34(WPRED,WORD,I,J)
C FIND ANY CHARACTERS ADJOINING THE CURRENT POINT (I,J).
DO 11 II=LESSI,I
11 I3=II-IOFF
IF(I3.LT.1)GO TO 11
IF(I3.GT.M)GO TO 10
DO 12 JJ=LESSJ,J
12 J3=JJ-JOFF
IF(J3.LT.1)GO TO 12
IF(J3.GT.N)GO TO 11
II=AIJ JJ=AJJ
C CHAREQ IS A FUNCTION RETURNING .TRUE. OR .FALSE. ACCORDING AS THE
C TWO CHARACTERS PASSED AS PARAMETERS ARE EQUAL OR NOT.
IF(CHAREQ(AIJ,''))GO TO 12
IF(.NOT.CHAREQ(AIJ,**))GO TO 14
I=II
14 J=JJ
DO 15 I2=LESSI,I
15 I3=I2+1
IF(I3.LT.1)GO TO 15
IF(I3.GT.M)GO TO 14
DO 16 J2=LESSJ,J
16 J3=J2+1
IF(J3.LT.1)GO TO 16
IF(J3.GT.N)GO TO 14
C
LSTPLT/NEAR

J3 = J4 + J2
IF (J3 .LT. 1) GO TO 16
IF (J3 .GT. N) GO TO 15
AI3J3 = [1, J3)
IF (CHAREQ(AI3J3, ' ')) GO TO 16
DO 17 K = 1, 36
IF (.NOT. CHAREQ(AI3J3, CHARS(K))) GO TO 17
C FLA W: ADJACENT CHARACTER IN USE.
USE(K) = .TRUE.
GO TO 16
17 CONTINUE
16 CONTINUE
15 CONTINUE
14 CONTINUE
DO 13 K = 1, 36
IF (.NOT. CHAREQ(AI3J3, CHARS(K))) GO TO 13
USE(K) = .TRUE.
GO TO 12
13 CONTINUE
12 CONTINUE
11 CONTINUE
GO TO 10
C SELECT NEXT CHARACTER IN SEQUENCE WHICH IS NOT ADJACENT TO CURRENT
C LIST OF CELLS.
20 DO 21 K = NEXT, 36
IF (USE(K)) GO TO 21
CHARH = CHARS(K)
NEXT = K + 1
GO TO 22
21 CONTINUE
C IF NO SUITABLE CHARACTER WAS FOUND IN THE RANGE NEXT TO 36, THEN
C SEE IF ONE CAN BE FOUND IN THE RANGE 1 TO NEXT.
DO 25 K = 1, NEXT
IF (USE(K)) GO TO 25
CHARH = CHARS(K)
NEXT = K + 1
GO TO 22
25 CONTINUE
C GETTING DESPERATE, USE SPECIAL CHARACTER RESERVED FOR EMERGENCIES.
CHARH = NUM3
NEXT = 1
22 CONTINUE
C SET CELLS IN ARRAY TO SELECTED CHARACTER.
WORD = SEQLR(S, IND)
FIRST = .TRUE.
23 CONTINUE
WORD = SEQLR(S, IND)
IF (IND .GT. 0) GO TO 24
CALL GETB34(WORD, I, J)
II = I + 10FF
JJ = J + J0FF
IF (II .LT. 1) GO TO 31
IF (II .GT. M) GO TO 31
IF (JJ .LT. 1) GO TO 31
IF (JJ .GT. N) GO TO 31
AI3JJ = CHARH
C IF THIS IS THE FIRST CHARACTER TO BE PRINTED FOR THE CURRENT LIST,
C AND THERE ARE MORE CELLS TO FOLLOW, THEN PRINT AN ASTERISK TO INDICATE
C THE START OF THE LIST.
IF (FIRST .AND. IND .NE. 2) A(I1, JJ) = STAR
31 CONTINUE
FIRST = .FALSE.
GO TO 23
24 RETURN
END

SUBROUTINE NEAR(LIST1, MIN, RMAX, ACMAX, ALMAX, LOCS)
C TRACES ALONG A BOUNDARY STRING AND COMPA RES DISTANCES ALONG THE STRING
C BETWEEN POINTS WITH EUCLIDIAN DISTANCE, A TENDENCY FOR THE BOUNDARY
C STRING TO CLOSE BEING MANIFESTED BY A LARGE RATIO OF DISTANCE ALONG THE
C STRING TO STRAIGHT-LINE DISTANCE.
C LIST1 IS USED TO LOAD VECT
C VECT CONTAINS THE NEAREST LOCATION TO EN points OF UNIT VECTORS DRAWN
C ALONG THE BOUNDARY STRING. I IS IN THE FIRST BYTE AND J IS IN THE SECOND.
C THE OTHER TWO BYTES ARE NOT EXAMINED.
C MIN SPECIFIES THE MINIMUM DISTANCE OF INTEREST (ALONG THE STRING).
C THE PAIR OF POINTS WHICH GIVE MAXIMUM RATIO OF DISTANCES ARE RETURNED
C IN LOCS. I1, J1, I2, J2 ARE RETURNED IN SUCCESSIVE BYTES OF LOCS.
INTEGER VECT(1)
RMAX=1.0
ALMAX=1.0
ACMAX=1.0
LOC5=0
NC=NCELL(LIST1)-1
IF(NC LE MIN)RETURN
CALL GETSTG(VECT,NC,4,ICF)
CALL LOADVR(LIST1,VECT(ICF+1),NC2)
CALL GETB12(VECT(ICF+1),I2MAX,J2MAX)
DO 10 I=2,NC
CALL GETB12(VECT(ICF+J),I2,J2)
IF(I.EQ.I2.AND.J1.EQ.J2)GOTO 20
ACROSS=(I1-I2)**2+(J1-J2)**2
ALONG=(I1-J1)**2
IF(ALONG.GT.12 MAX)GOTO 20
IF(RATIO.GT.RMAX)GOTO 20
IF(RATIO.LT.1.0.AN.D ALONG.LT.ALMAX)GOTO 20
RMAX=RATIO
ACMAX=ACMAX
ALMAX=ALMAX
I2MAX=I2
J2MAX=J2
20 CONTINUE
CONTINUE
IF(RMAX.GT.1.0)CALL GETB12(VECT(ICF+NC),I1MAX,J1MAX)
CALL SETB14(L0CS,I1MAX,J1MAX)
ACMAX=SQRT(ACMAX)
ALMAX=SQRT(ALMAX)
RMAX=SQRT(RMAX)
CALL FRESSTG(VECT,NC,4,ICF)
RETURN

END

SUBROUTINE OVLAP(AUX,WORK1,WORK2,PARM1,PARM2,N1,N2,OVLPL2,OVLPL21)

C CHECKS A PAIR OF BOUNDARY CURVES DESCRIBED BY PARAMETERS SUPPLIED
C IN PARM1 AND PARM2, TO SEE IF FOR EACH OF THE EQUALLY SPACED POINTS
C ON ONE CURVE, THERE IS A "CORRESPONDING" POINT ON THE OTHER CURVE
C SUCH THAT A CHORD JOINING THESE POINTS SUBTENDS APPROXIMATELY EQUAL
C ANGLES ON THE INSIDE OF THE TWO BOUNDARY CURVES. THE NUMBER OF
C CORRESPONDING POINTS FOR EACH CURVE IS RETURNED IN OVLPL2 AND OVLPL21
C RESPECTIVELY. OVLPL2 IS NOT NECESSARILY EQUAL TO OVLPL21 BECAUSE OF
C DIFFERENT CURVATURES OF CURVE1 AND CURVE2. CURVE1 AND CURVE2 ARE
C N1 AND N2 UNITS LONG RESPECTIVELY. AUX, WORK1, AND WORK2 ARE
C AUXILIARY STORAGE ARRAYS.
C SFUNC AND CFUNC COMPUTE THE I AND J DISPLACEMENTS RESPECTIVELY,
C AS A FUNCTION OF THE DISTANCE ALONG A CURVE, WITH INITIAL ANGLE=ANG,
C INITIAL CURVATURE=CURV, AND INITIAL DERIVATIVE OF CURVATURE=DCURV.
C REAL AUX(20),WORK1(3,N1),WORK2(3,N2)
C REAL PARM1(5),PARM2(5)
C PARMN(1)=I LOCATION OF START OF CURVE1 (WITH N=1 OR 2).
C PARMN(2)=J LOCATION OF START OF CURVE1.
C PARMN(3)=INITIAL ANGLE AT START OF CURVE1.
C PARMN(4)=INITIAL CURVATURE AT START OF CURVE1.
C PARMN(5)=INITIAL DERIVATIVE OF CURVATURE AT START OF CURVE1.
C NIM=20
C EPS=1.0E-3
C NLESS=N1-1
C NLESS=N2-1
C OVLPL2=0.0
C OVLPL21=0.0
C AP0S=PARM1(1)
C APOS=PARM1(2)
C ANG=PARM1(3)
C CURV=PARM1(4)
C DDCURV=PARM1(5)
C DO 10 I=1,NLESS
C FILL IN WORK1 ARRAY WITH ANGLE AND I AND J LOCATIONS OF POINTS
C SPACED 1 UNIT APART ON CURVE1.
AI = I - 1
W0RK1(1, I) = ANG + AI * CURV + AI + AI * 0CURV / 2.0
C WQRK1(1, I) = ANGLE OF CURVE1 AT DISTANCE I UNITS FROM THE START.
XL = AI
XU = AI + 1.0
W0RK1(2, I) = AIPOS
W0RK1(3, I) = AJPCS
C FIND AND LOCATION OF POINTS 1 UNIT APART ON CURVE1 BY NUMERICAL
C INTERGRATION OF SFUNCT AND CFUNCT.
CALL QATR(XL, XL, EPS, NDIM, SFUNCT, XADD, IER1, AUX)
CALL QATR(XL, XL, EPS, NDIM, CFUNCT, YADD, IER2, AUX)
AIPOS = AIPOS + XADD
AJPOS = AJPOS + YADD
10 CONTINUE
AI = N1 LESS
I = N1
W0RK1(1, I) = ANG + AI * CURV + AI * AI + CURV / 2.0
W0RK1(2, I) = AIPOS
W0RK1(3, I) = AJPCS
C REPEAT PROCESS WITH WORK2 AND CURVE2.
AIPOS = PARM2(1)
AJPOS = PARM2(2)
ANG = PARM2(4)
CURV = PARM2(5)
DO 20 I = 1, N2 LESS
AI = I - 1
W0RK2(1, I) = ANG + AI * CURV + AI + AI * CURV / 2.0
XL = AI
XU = AI + 1.0
W0RK2(2, I) = AIPOS
W0RK2(3, I) = AJPCS
C CHECK EACH POINT ALONG CURVE2 TO SEE IF IT "CORRESPONDS" (WITHIN A
C CALCULATED ERROR LIMIT) TO THE CURRENT POINT ON CURVE1.
DO 40 J = 1, N2
ABSOTH = ABS(W0RK2(2, J) * SINA - W0RK2(3, J) * COSA - CONST)
IF (ABSOTH .GE. DMIN) GO TO 40
DMIN = ABSOTH
AIMIN = W0RK2(2, J)
AJMIN = W0RK2(3, J)
40 CONTINUE
IF (AIMIN .LE. 0.5) OVLPL2 = OVLPL2 + 1.0
30 CONTINUE
C FIND OVLPL2.
DO 50 I = 1, N2
ANGNOW = W0RK2(1, I)
SINA = SIN(ANGNOW)
COSA = COS(ANGNOW)
AINOW = W0RK2(2, I)
AJNOW = W0RK2(3, I)
CONST = AJNOW * COSA - AINOW * SINA
DMIN = 1.0E50
DO 60 J = 1, N1
ABSOTH = ABS(W0RK1(2, J) * SINA - W0RK1(3, J) * COSA - CONST)
IF (ABSOTH .GE. DMIN) GO TO 60
DMIN = ABSOTH
AIMIN = W0RK1(2, J)
AJMIN = W0RK1(3, J)
60 CONTINUE
IF (DMIN .LE. 0.5) OVLPL2 = OVLPL2 + 1.0
50 CONTINUE
RETURN
END
SUBROUTINE PICOUT(A,M,N,L,DMIN,DMAX)

PRODUCES A PICTORIAL REPRESENTATION OF DATA SUPPLIED IN MATRIX
A(M,N,L) BY MEANS OF PRINTING (OR OVERPRINTING) CHARACTERS SELECTED
FOR THE VARIOUS PRINT DENSITY. THE VALUES SUPPLIED IN DMIN AND
DMAX ARE USED AS THE WHITE AND BLACK LEVELS RESPECTIVELY; DMIN MAY
BE LESS THAN, EQUAL TO, OR GREATER THAN DMAX. LINES OF PRINT
CORRESPOND TO ROWS OF DATA. NOTE THAT A MAXIMUM OF 128 COLUMNS IS
PRINTED.

REAL*8 RLINE(16,8),BLANKS/
DIMENSION A(M,N,L),LINE(12B,3)
LOGICAL*1 LINE,IA,IB,IC,II,IM,IO,IV,IX,IZ,IBLK,IMNS,IEQL,
1 IPLS,IQTE,ISTP
EQUIVALENCE (RLINE,LINE)
DATA IA,IB,IC,II,IM,IO,IV,IX,IZ,IBLK,IMNS,IEQL,IPLS,IEQL,
1 IPLS,IQTE,ISTP,JX
! A 'B,'C,'D,'E,'F,'G,'H,'I,'J,'K,'L,'M,'N,'O,'P,'Q,'R,'S,'T,'U,'V,'W,'X,'Y,'Z',
90 1,153714239/
N=N1
IF (N.GT.128) N=128
DRANGE=DMAX-DMIN
LMAX=8
DO 200 I=1,M
DO 230 K=1,LMAX
RLINE(K,L)=BLANKS
230 CONTINUE
220 CONTINUE
LMAX=1
DO 100 J=1,N
C GET A NEW UNIFORM RANDOM NUMBER IN THE RANGE 0.0 TO 1.0.
CALL RANDU(JX,JY,RNUM)
JX=JY
RNUM1=0.01925*(RNUM-0.5)
RNUM2=RNUM1+RNUM
DVAL1=(A(I,J)-DMIN)/DRANGE+RNUM1
DVAL2=DVAL1+RNUM1
DVAL4=DVAL2+RNUM2
DVAL6=DVAL4+RNUM2
DVAL8=DVAL6+RNUM2
IF (DVAL8.LT.0.073) GO TO 100
IF (DVAL4.LT.0.192) GO TO 72
IF (DVAL2.LT.0.236) GO TO 73
IF (DVAL2.LT.0.273) GO TO 74
IF (DVAL2.LT.0.309) GO TO 75
IF (DVAL2.LT.0.346) GO TO 76
IF (DVAL2.LT.0.382) GO TO 77
IF (DVAL2.LT.0.437) GO TO 78
IF (DVAL4.LT.0.491) GO TO 80
IF (LMAX.LT.2) LMAX=2
IF (DVAL2.LT.0.546) GO TO 81
IF (DVAL2.LT.0.592) GO TO 82
IF (DVAL2.LT.0.619) GO TO 83
IF (LMAX.LT.3) LMAX=3
IF (DVAL2.LT.0.655) GO TO 84
IF (LMAX.LT.4) LMAX=4
IF (DVAL6.LT.0.729) GO TO 85
IF (LMAX.LT.5) LMAX=5
IF (DVAL2.LT.0.820) GO TO 86
IF (DVAL2.LT.0.874) GO TO 90
IF (LMAX.LT.6) LMAX=6
IF (DVAL2.LT.0.910) GO TO 88
IF (DVAL2.LT.0.946) GO TO 89
IF (LMAX.LT.7) LMAX=7
IF (DVAL2.LT.0.982) GO TO 90
IF (LMAX.LT.8) LMAX=8
GO TO 91
20 IF (A(I,J).LE.DMIN) GO TO 100
LMAX=8
GO TO 91
72 LINE(J,1)=IMNS
GO TO 100
73 LINE(J,1)=IEQL
GO TO 100
74 LINE(J,1)=IPLS
GO TO 100
75 LINE(J,1)=IBKT
GO TO 100
76 LINE(J,1)=IONE
GO TO 100
SUBROUTINE RATND(A, POSBIT, NEGBIT, M, N, ATHR, BTHR, CTHR, DTHR, ETHR)

ATHR IS THRESHOLD FOR MIN ABSOLUTE MAGNITUDE OF CORRELATION
BTHR IS THRESHOLD FOR MIN MAGNITUDE OF SECOND DERIVATIVE RELATIVE TO
MAGNITUDE OF CORRELATION
CTHR IS THRESHOLD FOR MAX RELATIVE MAGNITUDE OF SECOND DERIVATIVE
ALONG DIRECTION PERPENDICULAR TO THAT OF THE MAXIMUM SECOND
DERIVATIVE
DTHR IS THRESHOLD FOR MAX RELATIVE MAGNITUDE OF DERIVATIVE IN THE
PERPENDICULAR DIRECTION
ETHR IS THRESHOLD FOR MAX RELATIVE MAGNITUDE OF DERIVATIVE IN DIRECTION OF
MAXIMUM SECOND DERIVATIVE

REAL A(M,N),DD2(4),DD(4),B(9),RT2/1.414214/,RT2BY2/2.828428/
INTEGER POSBIT(1),NEGBIT(1)
MLESS1=M-1
MESS1=N-1

NEGLECT EDGE RESPONSES AT BOUNDARY OF PICTURE
DO 10 I=2,MLESS1
1PLUS1=I+1
JLESS1=I-1
DO 20 J=2,JLESS1
JPLUSJ=J+1
B(5)=A(I,J)
ABSB5=ABS(B(5))

C IS MAGNITUDE OF CORRELATION LESS THAN ATHR
1F(ABSB5.LT.ATHR)GO TO 20
B(1)=A(JLESS1,J)
B(2)=A(JLESS1,JPLUSJ)
B(3)=A(JLESS1,J)
B(4)=A(J,JLESS1)
B(5)=A(J,JPLUSJ)
B(6)=A(JPLUSJ,J)
B(7)=A(JPLUSJ,JLESS1)
B(8)=A(JPLUSJ,J)
B(9)=A(JPLUSJ,JPLUSJ)
B5BY2=B(5)+B(5)
DD2(1)=B(4)+B(4)-B5BY2
DD2(2)=(B(3)+B(7)-B5BY2)/RT2
DD2(3)=B(2)+B(8)-B5BY2

END
SUBROUTINE RECODE(LIST1, LIST2, LIST7, SWITCH)

C TRACES AROUND THE BOUNDARY STRING SUPPLIED IN LIST1 AND SCRIBES UNIT VECTORS. THE ANGLES OF SUCCESSIVE VECTORS ARE RETURNED IN LIST2 AND THE BOUNDARY CELL(S) CLOSEST TO THE START OF EACH VECTOR ARE RETURNED IN LIST7. A CHECK IS MADE TO ENSURE THAT THE BOUNDARY IS TRACED IN THE DIRECTION OF BLACK ON THE LEFT. IF THIS TEST FAILS, SWITCH IS SET TO TRUE, THE BOUNDARY STRING IS REVERSED AND PROCESSING RECOMMENCES.

LOGICAL FIRST, LAST
LOGICAL SWITCH
SWITCH = .FALSE.

CONTINUE
LAST = .FALSE.
FIRST = .TRUE.
ANGCOR = 0.0
ANGGLO = 0.0
ANGLST = 0.0
IOUT = 0
DDFT01 = 0.0
CALL LIST(LIST2, 0)
CALL LIST(LIST3, 0)
CALL LIST(LIST6, 0)
CALL LIST(LIST7, 0)
S = SEORDR(LIST1)
CL0G1 = SELWR4(S, IFLAG)
IF(FLAG = .T.) GO TO 230
CALL GETBL4(CL0G1, IDIRN1, ICCRN1, I1, J1)
ITOP = I1
JTOP = J1
D0R1 = IDIRN1 / 4.0 + 6
CALL GETH2(CL0G1, IH2)
CALL SETH2(IDIRN1, IH2, 0)
CALL NEWBUT(LIST6)
CL0G1 = SELWR4(S, IFLAG)
IF(FLAG = .F.) GO TO 6
CALL IRAList(LIST7)
LIST7 = LIST6
CALL NEWBUT(IDIRN1, LIST2)
GO TO 231
CONTINUE
AI0LD = I1
AJ0LD = J1

DD2(4) = (B(1) + B(9) - B5B(2))/RT2
I MAX = 1
DD 14 IM = 2.4
IF(ABS/DD2 (IM)) .GT. ABS/DD2 (IMAX)) IMAX = IM
CONTINUE
C IS MAXIMUM SECOND DERIVATIVE LESS THAN RELATIVE THRESHOLD
IF(ABS/DD2 (IMAX)) .GT. BTHR*ABS(B5)) GO TO 20
C IS A TRUFFLE RATHER THAN A RIDGE
IF(SIGN(1.0, DD2(IMAX)) .EQ. SIGN(1.0, B(5))) GO TO 20
IRECT = IMAX + 2
C IS A SADDLE POINT
IF(SIGN(1.0, DD2(IMAX)) .EQ. SIGN(1.0, DD2(IRECT))) AND
1 ABS/DD2 (IRECT)). GT. CTHER*ABS(B5)) GO TO 20
DD(1) = (B(6) - B(4))/2.0
DD(2) = (B(5) - B(7))/RT2/B2
DD(3) = (B(4) - B(8))/2.0
DD(4) = (B(1) - B(9))/RT2/B2
C IS DERIVATIVE ALONG RIDGE GREATER THAN RELATIVE THRESHOLD
IF(ABS/DD2 (IMAX)) .GT. FTHR*ABS(B5)) GO TO 20
C IS DERIVATIVE ACROSS RIDGE GREATER THAN RELATIVE THRESHOLD
IF(ABS/DD2 (IMAX)) .GT. GTHR*ABS(B5)) GO TO 20
C THIS POINT IS A LOCAL MAXIMUM ACROSS THE RIDGE
IF(B(5) .LT. 0.0) GO TO 15
IF(B(5) .LT. B(5 + IMAX)) GO TO 20
IF(B(5) .LT. B(5 - IMAX)) GO TO 20
GO TO 16
CONTINUE
IF(B(5) .LT. B(5 - IMAX)) GO TO 20
IF(B(5) .LT. B(5 + IMAX)) GO TO 20
CONTINUE
IF/DD2(IMAX))/GT.0.0) CALL SETBIT(POSBIT, M, N, I, J, 1)
IF/DD2(IMAX)).LT.0.0) CALL SETBIT(NEGBIT, M, N, I, J, 1)
CONTINUE
10 CONTINUE
RETURN
END
GO TO 5
10 CONTINUE
  CL0G1=SEQLR(S,IFLAG)
  IF(IFLAG.GT.0)GO TO 80
5 CONTINUE
  CALL GETBL14(CL0G1,DIRN2,ICORR2,I2,J2)
  DIRN2=DIRN2/40.6
  CALL GETH2(CL0G1,IH2)
  CALL SETH2(IL0G3,IH2,0)
  I4=I2
  J4=J2
  IDIST2=(I2-I1)**2+(J2-J1)**2
  IF(IDIST2.GT.2)GO TO 311
  SN0W2=S
  CLOG1=SEQLR(S,IFLAG)
  IF(IFLAG.GT.0)GO TO 39
  CALL GETBL14(CL0G1,DIRN3,ICORR3,I3,J3)
  DIRN3=DIRN3/40.6
  IDIST3=(I3-I1)**2+(J3-J1)**2
  IF(IDIST3.GT.2)GO TO 40
  CALL GETH2(CL0G1,IH2)
  CALL SETH2(IL0G3,IH2)
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  IF(IDIST3.GT.14159)DIRN2=0
  IF(IDIST3.GT.14159)DIRN3=0
  IDIST=(I3-I1)**2+(J3-J1)**2
  IDIST=(I3-I1)**2+(J3-J1)**2
  IF(IDIST.GT.2)GO TO 41
  SN0W2=S
  CLOG1=SEQLR(S,IFLAG)
  IF(FOOT.0)GO TO 80
  CALL GETBL14(CL0G1,DIRN2,ICORR2,0)
  CALL GETH2(CL0G1,IH2)
  CALL SETH2(IL0G3,IH2)
  I4=I2
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  IF(IDIST.GT.2)GO TO 41
  SN0W2=S
  CLOG1=SEQLR(S,IFLAG)
  IF(FOOT.0)GO TO 80
  CALL GETBL14(CL0G1,DIRN2,ICORR2,0)
  CALL GETH2(CL0G1,IH2)
  CALL SETH2(IL0G3,IH2)
  I4=I2
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
  SN0W2=S
  I4=I3
  J4=J3
  AINEW=(I2*ICORR2+I3*ICORR3)/FLOAT(ICORR2+ICORR3)
  AJNEW=(J2*ICORR2+J3*ICORR3)/FLOAT(ICORR2+ICORR3)
RECODE/SQPIC

ATHS = SEQLR(S6, IFLAG)
CALL NEWBJT(ALST, LIST7)
PARTA = 0.0
DISTA = SEQLR(S7, IFLAG)
DISTB = DISTA
IF(IFLAG GT 0) GO TO 200
ANGLEA = SEQLR(S2, IFLAG)
ANGLB = ANGLEA
FIRST = TRUE.
210 CONTINUE
REMA = DISTA - PARTA
IF(REMA GE 1.0) GO TO 220
DISTB = SEQLR(S7, IFLAG)
IF(FIRST AND IFLAG GT 0) GO TO 250
IF(IFLAG GT 0) GO TO 200
ANGLEB = SEQLR(S2, IFLAG)
ALST = ATHS
ATHS = SEQLR(S6, IFLAG)
BTA = ANGLEB - ANGLEA
RCBTA = REMA * COS(BTA)
FACT = BCTA * RCSBTA - REMA * REMA + 1.0
IF(FACT LT 0.0) FACT = 0.0
BITOFII = -RC SQTA + SQRT(FACT)
BTSNB T = BITOFII * SIN(BTA)
IF(BTSNBT LT 1.0) BTSNBT = 1.0
IF(BTSNBT LT -1.0) BTSNBT = -1.0
GAMMA = ARSIN(BTSNBT) + ANGLEA
CALL NEWBOT(GAMMA, LIST2)
PARTA = BITOFII
DISTA = DISTB
ANGLEA = ANGLEB
CALL NEWBOT(ATHS, LIST7)
FIRST = FALSE.
GO TO 210
220 CONTINUE
GAMMA = ANGLEA
PARTA = PARTA + 1.0
CALL NEWBOT(GAMMA, LIST2)
CALL NEWBOT(ATHS, LIST7)
FIRST = FALSE.
GO TO 210
250 CONTINUE
CALL NEWBOT(ANGLEB, LIST2)
CALL NEWBOT(ATHS, LIST7)
REMA = 0.0
200 CONTINUE
IF(REMA LE 0.5) GO TO 230
CALL NEWBOT(ANGLEA, LIST2)
CALL NEWBOT(ATHS, LIST7)
230 CONTINUE
CALL IRLST(LIST6)
231 CONTINUE
CALL IRLST(LIST3)
RETURN
311 WRITE(3, 112) JTCP, JTOP, I1, J1, I2, J2, IDIST2
CALL EXIT
112 FORMAT('JUMP IN LIST OF CELLS'/
113 FORMAT('POSSIBLE ERROR IN DIRECTION OF LINE'/
1 ' JTCP=', I5, ' JTOP=', I5, ' AVDDIF=', G15.6, ' NOTOT=', I5, '/0')
114 FORMAT('AMBIGLOUS DIRECTION OF TRACE IN RECODE')
115 FORMAT('TOP1=', I5, ' BOTTOM=', I5)
END

SUBROUTINE SQPIC(A, M, N, MODE, DMINI, DMAXI, SKIP)

PRODUCES A PICTORIAL REPRESENTATION OF THE DATA SUPPLIED IN MATRIX A(M,N), BUT LINEARLY INTERPOLATES LINES OF PRINT ON ROWS OF DATA TO OVERCOME THE ASPECT DISTORTION RESULTING FROM THE IMP. 4403 LINE PRINTER'S 10 CHARACTERS/INCH VERSUS 8 LINES/INCH.


SKIP IS A LOGICAL VARIABLE WHICH DETERMINES WHETHER OR NOT THE PRINT CONTINUES OVER THE PERFORATIONS BETWEEN PAGES. IF SKIP = .TRUE., A NEW PAGE IS STARTED EVERY 64 LINES (TO SAVE TEARING THE PAPER), OTHERWISE IF SKIP = .FALSE., PRINTING IS CONTINUOUS.

NOTE THAT A MAXIMUM OF 128 COLUMNS IS PRINTED.
C

REAL A(M,N), B(123)
LOGICAL SKIP
IMAX=0.5
N=N1
IF(N.GT.128)N=128
DMIN=MIN
MAX=MAX
IF(MODE.EQ.-1)GO TO 2
IF(MODE.EQ.+1)GO TO 2
DMIN=MIN
MAX=MAX
GO TO 3
2 CONTINUE
AI=0.25
DO 10 I=1, IMAX
AI=AI+1.25
ILO=AI
IHI=ILO+1
IF([H1I.GT.M) IHI=M
IF([ILO.GT.M-1) ILO=M-1
FLODIFF=AI-ILO
CLUDIFF=IHI-AI
DO 20 J=1, N
DTEMP=A(IHI,J)*FLODIFF+A(ILO,J)*CLUDIFF
IF(DTEMP.LT.MIN)DMIN=DTEMP
IF(DTEMP.GT.MAX) MAX=DTEMP
20 CONTINUE
10 CONTINUE
IF(MODE.EQ.-1)GO TO 4
DTEMP=DMIN
DMIN=MAX
MAX=DTEMP
4 CONTINUE
DMIN=DMIN
MAX=MAX
3 CONTINUE
WRITE(3, 100)DMIN, MAX
AI=0.25
DO 30 I=1, IMAX
AI=AI+1.25
ILO=AI
IHI=ILO+1
IF([H1I.GT.M) IHI=M
IF([ILO.GT.M-1) ILO=M-1
FLODIFF=AI-ILO
CLUDIFF=IHI-AI
DO 40 J=1, N
B(J)=A(IHI,J)*FLODIFF+A(ILO,J)*CLUDIFF
IF((I-I)/64.EQ.-1.AND.I.NE.1.AND.SKIP)WRITE(3, 101)
40 CONTINUE
101 FORMAT('**')
100 FORMAT('**','DMIN = ',G13.5, ' MAX = ',G13.5)
END

SUBROUTINE STRTN(LIST1, IND)

PURPOSE

EXAMINES ADJOINING CELLS ON LIST1 TO SEE IF THEY ARE WITHIN SORT(2,0)
OF EACH OTHER. AND IF THIS IS NOT THE CASE, TO SEE IF IT IS POSSIBLE
TO REARRANGE THE ORDER OF THE CELLS SUCH THAT OVERALL ADJACENCY IS
OBTAINED. A FINAL PASS THROUGH THE LIST OF CELLS SWAPS PAIRS OF CELLS
WHEN THIS EXCHANGE HELPS TO STRAIGHTEN OUT THE PATH TRACED.

IF NO GAPS IN LIST IND=0 ON RETURN, OTHERWISE IND= +OR-1000*I+J
WHERE (I,J) IS LOCATION OF GAP

LOGICAL NOLI, NCN, TEST
IND=0
NC=NC1ELIST1)-1
IF ONLY ONE CELL NO ACTION REQUIRED
IF (NC.LE.1) RETURN
TEST=FALSE
1 CONTINUE
S=SEQDR1(LIST1)
AII2=SEQLR(S,IFLAG)
CALL GETB34(AII2, ILast, JLast)
GO TO 5
CONTINUE
ILAST=THIS
JLAST=JTHIS

CONTINUE
SLAST=S
AII2=SEQLR(S,IFLAG)
IF(IFLAG.LE.0)GO TO 6
IF(IST)GO TO 200
TEST=.TRUE.
GO TO 1

CONTINUE
STHIS=S
CALL GETB34(AII2,THIS,JTHIS)
IDIST=(THIS-ILAST)**2+(JTHIS-JLAST)**2
C IF DISTANCE BETWEEN THIS AND LAST IS <= SQRT(2.0) THEN NO GAP HERE
IF(IDIST.LE.2)GO TO 10
IF(IST)GO TO 330
ILIAD=LNKLIL(NLKL(S))+1
ILAD=LNKL(S)+1
ITAD=LNKRIL(NLKL(S))+1
INAD=LNKK(S)+1
ALAST1=CONT(ILIAD)
ALAST=CONT(ILAD)
THIS=CONT(ITAAD)\nNEXT=CONT(INAD)
S=LAST
AII2=SEQLR(S,IFLAG)
IF(IFLAG.LE.0)GO TO 21
NBACK=1
GO TO 24

CONTINUE
CALL GETB34(AII2,ILAST1,JLAST1)
AII2=SEQLR(S,IFLAG)
IF(IFLAG.LE.0)GO TO 23
NBACK=2
GO TO 24

CONTINUE
CALL GETB34(AII2,ILAST?,JLAST?)
NBACK=3

CONTINUE
S=THIS
AII2=SEQLR(S,IFLAG)
IF(IFLAG.LE.0)GO TO 25
NFWD=0
GO TO 28

CONTINUE
CALL GETB34(AII2,INEXT,JNEXT)

CONTINUE
IF(NBACK.EQ.1)GO TO 31
ILI=(ILAST1-THIS)**2+(JLAST1-JTHIS)**2
IF(ILI.GT.2)GO TO 31
IF(NBACK.EQ.2)GO TO 41
ILI2=(ILAST-ILAST2)**2+(JLAST-JLAST2)**2
IF(ILI2.LE.2)GO TO 41

CONTINUE
IF(NFWD.EQ.0)GO TO 33
ILI=(ILAST-INEXT)**2+(JLAST-JNEXT)**2
IF(ILN.GT.2)GO TO 31
IF(NBACK.EQ.2)GO TO 43
IF(ILI2.LE.2)GO TO 43

CONTINUE
INO=1000*ITIBS+JTHIS
RETURN

CONTINUE
SWAP LAST AND LAST BUT ONE
CALL STRING(ILAST,ILIAD)
CALL STRING(ILAST1,ILAAD)
ILAST=THIS
JLAST=JTHIS
S=THIS
GO TO 5

CONTINUE
SWAP THIS AND NEXT
CALL STRING(THIS,INAD)
CALL STRING(INEXT,ITAAD)
ILAST=INEXT
JLAST=JNEXT
S=THIS
GO TO 5

CONTINUE
SWAP LAST AND LAST BUT ONE THEN SWAP THIS AND NEXT
CALL STRIND(ALAST, ILIAO)
CALL STRIND(ALAST1, ILAD)
CALL STRINO(ATHIS, INAO)
CALL STRINO(ANEXT, ITAD)
ILAST = INEXT
JLAST = JNEXT
S = STHIS
GO TO 5
200 CONTINUE
C IF ONLY TWO CELLS NO EXCHANGES REQUIRED
IF(NC .EQ. 2) RETURN
S = SEQRK(#1, S, IFLAG)
201 AI2 = SEQRK(S, IFLAG)
IF(IFLAG .EQ. 0) RETURN
CALL GETB34(AII2, ITHIS, JTHIS)
ILAD = LNKL(S)
ILAD = LNKL(ILAD)
IIAD = LNKL(ILAD)
INAO = LNKL(S)
CALL GETB34(CONT(ILAD + 1), ILAST, JLAST)
NOLL = .TRUE.
IF(I11(I11AD) .EQ. 2) GO TO 202
NOLL = .FALSE.
CALL GETB34(CONT(ILAD + 1), ILAST, JLAST)
W = SEQRK(S, IFLAG)
202 CONTINUE
C WOULD AN EXCHANGE REDUCE KINKS IN LIST
IF(I11(I11AD) .EQ. 2) GO TO 203
NOLL = .TRUE.
CALL GETB34(CONT(I11AD + 1), INEXT, JNEXT)
IF(NOLL) GO TO 204
ILN = (ILAST - INEXT)**2 + (JLAST - JNEXT)**2
ITN = (ITHIS - INEXT)**2 + (JTHIS - JNEXT)**2
IL1 = (ILAST1 - ITHIS)**2 + (JLAST1 - JTHIS)**2
I11 = (ILAST1 - ILAST)**2 + (JLAST1 - JLAST)**2
GO TO 210
203 CONTINUE
IL11 = (ILAST1 - ITHIS)**2 + (JLAST1 - JTHIS)**2
I11 = (ILAST1 - ILAST)**2 + (JLAST1 - JLAST)**2
IF(I11 .EQ. 1) GO TO 210
GO TO 201
204 CONTINUE
ILN = (ILAST - INEXT)**2 + (JLAST - JNEXT)**2
ITN = (ITHIS - INEXT)**2 + (JTHIS - JNEXT)**2
IF(I11 = 1) GO TO 210
GO TO 201
210 CONTINUE
C SWAP THIS AND LAST
ALAST = CONT(ILAD + 1)
ATHIS = CONT(ITAD + 1)
CALL STRIND(ATHIS, ILAD + 1)
CALL STRIND(ALAST, ITAD + 1)
GO TO 201
330 CONTINUE
IND = -(1000 # THIS + JTHIS)
RETURN
END

SUBROUTINE TRACE(POSBIT, PICBIT, PCLIST, A, EDGES, M2, N2, EP, ER, 1, IL, IH, JL, JH)
C TRACES AROUND EDGE RESPONSES FLAGGED AT EACH OF EIGHT DIRECTIONS IN POSBIT
AND FORMS LISTS OF CELLS FOR A CELL TO BE ASSIGNED AS A SUCCESSOR
TO THE CURRENT CELL IT MUST ADJOIN THE CURRENT CELL AND HAVE A SIMILAR
DIRECTION OF EDGE RESPONSE. ALSO, IT MUST NOT ALREADY BE ACCOUNTED FOR ON
THIS OR ANY OTHER BOUNDARY STRING, THIS BEING INDICATED BY THE
CORRESPONDING BIT BEING FLAGGED IN PICBIT. THE DIRECTION OF TRACE IS THAT
WITH BLACK ON THE LEFT. THE SLIP ADDRESS OF THE HEADER CELL FOR THE
BOUNDARY STRING CONTAINING ANY CELL IS PLACED AS A HALF-WORD IN THE
CORRESPONDING LOCATION OF VMEM.
C REAL CCNF(3)
REAL GCNF(3)
INTEGER IADD(8) = 0, -1, -1, 0, 1, 1, 1, 1, JADD(8) = 1, 1, 0, -1, -1, -1, 0, 1/
INTEGER POSBIT(512, 3), PCLIST, PICBIT(512), IP(8)
REAL EDGES(M2, N2, 5), CON61(0:785399)/
LOGICAL A(128, 128)
C A IS AN ARRAY OF 8-BIT ELEMENTS OF THE PSYCHIC-PHYSICALLY SCALED
C INPUT IMAGE
LOGICAL DROPI, DROPP
INTEGER ICNF(3), JCNF(3), LESS1/-1/
INTEGER*2 VMEM,HWORD(2)
EQUIVALENCE(HWORD,FWORD)
COMMON/VISMEM/VMEM(128,128)

C VMEM = A 128 X 128 ARRAY OF HALFWORDS WITH ELEMENTS POINTING TO
C THE LIST ON WHICH THE CORRESPONDING EDGE-RESPONSE HAS BEEN INCLUDED.
C VMEM IS TO BE USED LATER FOR FACILITATING THE RECOVERY OF
C RELATIONSHIPS BASED ON THE PROXIMITY OF EDGE-RESPONSES OR BOUNDARY
C CURVES, BY PROVIDING POINTERS FROM THE POINTS THEMSELVES TO THE
C HIGHER-LEVEL CONSTRUCTS IN WHICH THEY ARE INVOLVED.

DO 2 I=1,128
DO 2 J=1,128
2 VMEM(I,J)=0
CALL LIST(PCLIST,0)

DO 1 I=1,512
DO 20 J=0,127
IF(GETBIT(IPICBIT,128,128,I,J,OLDVAL).NE.O.O)GO TO 20
IPOS=0
IF(GETBIT(IPICBIT,128,128,I,J,OLDVAL).EQ.O.O)GO TO 11
IPOS=IPOS+1
IPOS=IP0S
11 CONTINUE
IF(IPOS.EQ.O.O)GO TO 20
CALL SETBIT(IPICBIT,128,128,I,J)
CALL LIST(NULIST,0)
CALL SETDIR(0,NULIST,FWORD)
DIRN=1
CALL INFU(LA,EDGES,M2,N2,EP,ER,I,J,GAMMA,CMAX,KMAX,A1,A2,A3,AMAG)
INFO IS A ROUTINE FOR SUPPLYING INFORMATION ABOUT CORRELATIONS WITH
EDGE-DETECTORS OF ANY ORIENTATION AT THE POINT (I,J) IN ARRAY A.
EDGE-DETECTORS AT FOUR ORIENTATIONS ARE SUPPLIED IN EDGES.
CMAX=MAXIMUM ABSOLUTE CORRELATION AT ANY OF 8 ORIENTATIONS
CORRESPONDING TO POSITIVE OR NEGATIVE CORRELATIONS WITH THE FOUR
EDGE-DETECTORS SUPPLIED.
KMAX=ORIENTATION AT WHICH CMAX WAS NOTED, EXPRESSED AS AN INTEGER
IN THE RANGE 1 TO 8.
IF THE POINT (I,J) LIES TO ONE SIDE OF A CONTOUR, THE MAGNITUDE
OF THE CORRELATION MAY BE A MAXIMUM FOR EDGE-DETECTORS ORIENTED
SLIGHTLY AWAY FROM PARALLEL TO THE CONTOUR, SUCH THAT THE MAGNITUDE
OF CORRELATION VERSUS ANGLE HAS A DOUBLE-HUMPED PEAK. IN SUCH
CASES THE TRUE ORIENTATION OF THE CONTOUR DOES NOT CORRESPOND TO
THE PEAK CORRELATION BUT RATHER TO THE PHASE OF THE FUNDAMENTAL
COMPONENT IN THE CORRELATION MAGNITUDE VERSUS ANGLE.
THE MAGNITUDE AND PHASE OF THIS FUNDAMENTAL IS COMPUTED VIA AN
8-POINT FOURIER TRANSFORM USING THE 8 OBSERVED CORRELATIONS.
A1=MAGNITUDE OF IN-PHASE COMPONENT.
A2=MAGNITUDE OF IMAGINARY COMPONENT.
AMAG=SQRT(A1*A1+A2*A2)
GAMMA=PHASE OF FUNDAMENTAL
=GAMMA+ANGCOR/CONST1+1.5
260 CONTINUE
IF(IANG.GT.8)IANG=IANG-8
NCNF=0
DO 200 II=LESS1,1
IF(ISUB.IE.ISUB+8)GO TO 200
IF(ISUB.GT.ISUB+8)GO TO 200
JCON=JNOW+JADD(ISUB)
IF(JCON.NE.JCON+JADD(ISUB).NE.O.O)GO TO 200
IF(JCON.JCON+JADD(ISUB).NE.O.O)GO TO 200
IF(GETBIT(IPICBIT,128,128,JCON,JCON,OLDVAL).NE.O.O)GO TO 200
DO 210 III=1,128
IANG2=IANG+III

IF(ANGCOR.EQ.0.0) IANG2 = IANG2 - 4
IF(IANG2.LT.1) IANG2 = IANG2 + 8
IF(IANG2.GT.8) IANG2 = IANG2 - 8
GO TO 210
NCNF = NCNF + 1
ICNF(NCNF) = I
ICNF(NCNF) = JCON
JCNF(NCNF) = JCON
GO TO 211
210 CONTINUE
211 CONTINUE
IF(NCNF.EQ.0) GO TO 250
DROP1 = FALSE
DROP2 = FALSE
IF(NCNF.EQ.3) DROP1 = TRUE
A3ST = -255.0
A3ST2 = 255.0
DO 270 I = 1, NCNF
CALL INFOITA, EDGES, M2, N2, EP, ER, ICNF(II), JCNF(II), GCNF(II), GCNWF, AMAX, KMAX, A1, A2, GCNWF, AMAG)
GAMNW = GCNF(II) + GCNWF
GAMCLD = GAMCOR
GAMDF = GAMST - GAMNW
IF(GAMDF.GT.3.14159) GAMCCOR = GAMCOR + 6.283185
IF(GAMDF.LT.-3.14159) GAMCCOR = GAMCOR - 6.283185
GAMNW = GAMNEW
GAMCLD = GAMCOR
GAMDF = GAMST - GAMNW
IF(II.EQ.3) A3ST2 = CCNF(II)
IF(II.EQ.4) A3ST2 = CCNF(II)
IF(II.EQ.5) A3ST = CCNF(II)
IF(CCNF(II).GT.A3ST2) IINMIN = II
IF(CCNF(II).LT.A3ST) IINMAX = II
IF(CCNF(II).GT.A3ST) IINMIN = II
IF(CCNF(II).LT.A3ST) IINMAX = II
270 CONTINUE
IF(DROP1.AND.IINMIN.EQ.0) DROP2 = TRUE
IF(DROP2) WRITE(3, 101) IOW, JNOW, ICNF(I), JCNF(I), ICNF(II), JCNF(II), II = 1, NCNF,
1 IINMIN, IINMAX
IF(NCNF.EQ.2.AND.ICNF(2).GT.ICNF(1)) GT.1)
1 WRITE(3, 102) IOW, JNOW, ICNF(I), JCNF(I), II = 1, 2, IINMIN, IINMAX
IF(NCNF.EQ.2.AND.ICNF(2) .LT. IINMIN) GO TO 220
IINMIN = II
IINMAX = II
DO 220 II = 1, NCNF
IF(II .EQ. IINMIN) GO TO 220
IF(DROP2.AND.II.NE.IINMAX) GO TO 220
IINMIN = II
IINMAX = II
220 CONTINUE
IF(DROP1.AND.IINMIN.EQ.0) DROP2 = TRUE
IF(DROP2) WRITE(3, 100) IOW, JNOW, ICNF(I), JCNF(I), II = 1, 4,
1 IINMIN, IINMAX
END
ON THE BANDWIDTH OF CARRIER-TYPE DC AMPLIFIERS

ABSTRACT:

For a carrier-type amplifier (e.g., a chopper amplifier), it is well-known that the low-frequency response may extend to dc. The position regarding the high-frequency response is not as clear, and statements in the literature give various fractions of the fundamental carrier frequency as an upper limit to the input signal bandwidth which can be recovered from the modulated carrier. It is argued here that there is no simple well-defined theoretical upper limit, and that in practical circuits the fundamental carrier frequency may or may not be one of several factors determining the useful bandwidth.

The circuit of a novel wide-band dc amplifier, consisting of two synchronized chopper amplifiers operating in parallel, is described. The frequency response of this amplifier is independent of the chopping rate, and its method of operation may help clarify bandwidth limitations.

INDEX TERMS:

Frequency response, bandwidth, chopper amplifiers, dc amplifiers, modulation, demodulation.

I INTRODUCTION:

To reduce problems of drift and low-frequency noise in conventional dc amplifiers, the technique of carrier-type dc amplification outlined in Fig. 1 is often employed [1]. The low-level input signal modulates a carrier waveform which then passes through an ac-coupled amplifier and after suitable demodulation and filtering, yields an amplified version of the input signal, complete with any dc component. The question to be examined below is that of the practical and theoretical limitations imposed by the modulation process on the frequency spectrum of the input signal which may be recovered at the output of such an amplifier. This recoverable frequency spectrum will be referred to as the "bandwidth".

Many types of modulation and carrier waveforms could be employed in the circuit of Fig. 1. For certain modulation-carrier combinations, such as pulse-code modulation and pulse-position modulation, the knowledge which can be gained from the modulated carrier of the input signal's behaviour, is restricted to sampled values. The sampling theorem \[2\] is applicable in this case and since the theoretical bandwidth limit is readily obtained, this case will now be considered. Given an input signal strictly limited to a bandwidth \(W\), if all frequency components in this bandwidth are to be recovered, the average rate at which new samples are represented in the modulated carrier must be at least \(2W\). No practical signal is strictly bandlimited \[3\] and a sampling rate higher than \(2W\) is necessary in practice if a useful bandwidth \(W\) is to be obtained. The required increase beyond \(2W\) is a function of circuit complexity, acceptable reconstruction error, and spectral content of the input signal outside \(W\). The case of representation of the input in the modulated carrier by only sampled values will not be considered further.

An important limitation to the range of modulation-carrier combinations likely to be employed in practical amplifiers, is that of allowable circuit complexity, including the demodulator and output filter. For this reason, amplitude modulation of either a square wave (e.g., chopptype dc amplifier \[1\]) or a sinusoid (e.g., vibrating-capacitor modulator \[4\]) is a common approach.\(^1\) The following discussion will be primarily concerned with these two combinations, but the conclusions reached are fairly general.

To reduce noise and problems associated with unwanted high-frequency components in the input signal, the ac-coupled amplifier in Fig. 1 is sometimes tuned to the fundamental carrier frequency. In this case the bandwidth of the carrier-type dc amplifier is limited by that of the tuned amplifier. It is assumed below that the only restriction imposed by the ac-coupled amplifier is that dc and very low frequency components in the modulated carrier will be rejected.

Given details of the modulation method and the carrier waveform, the knowledge which can be gained from the modulated carrier of the input
signal's behaviour is, in many cases, continuous for some part(s) of each carrier period. The representation of the input signal in the modulated carrier is thus "partly-continuous", rather than sampled. Practical chopper and vibrating-capacitor modulators produce modulated carriers which, in essence, satisfy this criterion. Those part(s) of each carrier period during which the input signal is not accurately represented in the modulated carrier will be termed "gaps" (i.e., in our knowledge of the input). Examples of such gaps are:

(i) The times during which the mechanical contacts in a full-wave chopper amplifier are changing from one polarity to another, including the interval necessary for cessation of contact bounce and switching transients (if these transients affect the recoverability of the input);

(ii) Those intervals when the carrier waveform is effectively zero, such as the times during which the signal path in a half-wave chopper amplifier is broken, or when the amplitude of a sinusoidal carrier waveform approaches zero.

For partly-continuous modulated carriers, bandwidth limitations are not as clear as for the case where only sampled values of the input signal are involved. In most previous literature on this subject, the bandwidth is assumed to be limited to a fraction of the fundamental carrier frequency $[1, 5, 6, 7]$. Some statements regarding bandwidth limitations may have referred only to the useful bandwidth of particular practical amplifiers, but others give what appear to be theoretical limits. It will be argued below that there is no clear theoretical limit to the input signal bandwidth recoverable from such modulated carriers, and that in practical amplifiers the width of gaps is, given suitable demodulation, of greater significance in determining the useful bandwidth than the carrier period itself.

II THEORETICAL BANDWIDTH LIMITATIONS:

Given the nature of the representation of the input in a partly-continuous modulated carrier (i.e., given the modulation technique and carrier waveform employed), the input's behaviour may be deduced continuously except for the duration of any gaps. Fig. 2 illustrates the knowledge of the input
which may be gained from a typical modulated carrier. The question of bandwidth limitations thus involves the question of limitations arising from gaps in our knowledge of a continuous signal.

The sampling theorem may readily be applied in this case to give a lower bound for the bandwidth to which an input signal must be strictly limited if it is to be accurately reconstructed. We assume for this purpose that all gaps are of duration $t_{\text{gap}}$, and that the intervals between gaps are integral multiples of $t_{\text{gap}}$. In this case, it is possible to deduce equi-spaced samples of the input, as shown in Fig. 2. The lower bound to the input bandwidth is thus $\frac{1}{2} \frac{1}{t_{\text{gap}}}$ Hz. It is clear that we have more knowledge of the input than is represented in these equi-spaced samples, and it is natural to ask what effect this additional knowledge has on the allowable bandwidth. To reconstruct a function strictly limited to a bandwidth $W$ over an interval of $T$ seconds, $2TW$ independent numbers are required, but these numbers "need not be ... equally spaced samples ..., the samples can be unevenly spaced, although, if there is considerable bunching, the samples must be known very accurately to give a good reconstruction of the function" [2].

Thus, if samples are taken in bunches during the intervals when the input behaviour is accurately represented in the modulated carrier such that the average sampling rate is $2W$ per second, any input signal strictly limited to a bandwidth $W$ may be reconstructed. By taking more and more samples in these intervals, greater and greater bandwidths could be reconstructed. There is no clearly defined theoretical limit to the recoverable bandwidth. If circuit complexity were not a consideration, then it would be possible to relate the useful bandwidth to the acceptable interpolation error, given details of:-

(i) The widths of gaps and of the intervals between neighbouring gaps.

(ii) The spectral content of the input signal (after any filtering which may have to be introduced).

(iii) The likely errors in position and amplitude of deduced samples.
III PRACTICAL BANDWIDTH LIMITATIONS:

In discussing limitations to the bandwidth which can be usefully employed, it will be assumed that the input signal spectrum is fairly uniform over a broad band and that the objective is to recover as much of this spectrum as possible, consistent with various practical requirements and limitations. Some factors involved in determining the useful bandwidth of a carrier-type dc amplifier are:

(i) Maximum gap duration,
(ii) Acceptable interpolation error during gaps,
(iii) Allowable circuit complexity,
(iv) Requirements with regard to the frequency response in the amplifier's passband and the rate of attenuation of frequency components beyond the cut-off frequency,
(v) Synchronous errors.

The limitations imposed by synchronous errors are considered in Section IV as they are basically different from those imposed by the other factors listed.

The length of gaps will often be determined by the components and circuits employed in modulating the carrier. In the case of mechanical choppers, the contact changeover time may be reduced slightly by modifications to the driving waveforms, but is essentially fixed. For a sinusoidal carrier, noise voltages and errors in knowledge of the carrier phase will mean that the input is, in practice, irrecoverable whenever the carrier waveform's magnitude is less than some fraction of its peak value. In this case, the gap duration is a function of the carrier frequency. Generally speaking, the shorter the duration of gaps, then (given suitable demodulation and interpolation) the greater the useful bandwidth.

If the interpolation error during gaps were of no concern, as would be the case if a carrier-type dc amplifier were used as the vertical amplifier in a cathode-ray oscilloscope and any gaps occurred during the beam retrace, then the gaps would have no effect on the useful bandwidth. The more usual situation is that the behaviour of the frequency spectrum of interest in the input signal must be interpolated, with less than a specified error, during gaps. Holding other factors constant, the mean interpolation error increases in a roughly exponential manner as the bandwidth to be recovered increases.
Any increase in acceptable interpolation error would thus lead to only a small increase in the useful bandwidth which may be recovered.

Allowable circuit complexity is an important limitation to the bandwidth which can be obtained practically. Circuits employing a limited-bandwidth carrier-type dc amplifier to "drift-correct" a wide-band conventional dc amplifier, give a performance which is satisfactory for most applications \[8\]. If the circuits employed in extending the bandwidth of a carrier-type dc amplifier are too complex, a drift-corrected amplifier may provide a simpler solution.

The restriction on circuit complexity will generally be manifested in the use of non-ideal demodulation and interpolation techniques, with synchronous rectification followed by low-pass filtering being a common approach. For a full-wave chopper modulator, synchronous rectification approaches the ideal, but this is not true for a sinusoidal carrier where ideal demodulation would involve multiplication by a cosecant waveform (truncated in the vicinity of carrier wave zeros, i.e., during gaps). Simple low-pass filtering is not the ideal interpolation technique in either case. Reconstruction from samples would give better results, but would not usually be considered because of circuit complexity. Perhaps the simplest interpolation technique to implement is a crude Taylor's series approximation, where the amplitude, or amplitude and rate of change, etc., of the reconstructed input signal just prior to each gap is maintained across the gap.

Having chosen the input bandwidth to be recovered, any frequency components in the demodulated signal beyond this specified bandwidth corrupt the knowledge (of the chosen input bandwidth) which is to be used for interpolation during gaps. The input is taken to be broad-band, and the potential performance of the amplifier (in terms of the bandwidth recoverable for a given gap duration and interpolation error) can be improved by filtering out extraneous components. It is therefore assumed that the input signal will be low-pass filtered before being applied to the modulator. As the rate of attenuation beyond the cut-off frequency of the filter is increased, the magnitude of extraneous components decreases, and a smaller interpolation error may be achieved for the given bandwidth.
Alternatively, a wider bandwidth could be achieved for the same interpolation error. The effects of a sharp cut-off filter on transient response (and stability in feedback circuits) have to be considered, together with the resulting circuit complexity and possible frequency response variations in the filter's passband.

The preceding discussion is complicated by interrelationships between various factors involved in determining the practical bandwidth. There is a conflict between allowable circuit complexity, interpolation error, and transient response, and the obtainable bandwidth depends on the relative importance of these factors.

IV SYNCHRONOUS ERRORS:

The useful bandwidth of a carrier-type dc amplifier could be restricted by some unusual errors which are generally most pronounced when the input signal contains a large spectral component in the vicinity of the (fundamental) carrier frequency. These "synchronous" errors have traditionally been avoided by restricting the bandwidth of the input signal (by means of low-pass filtering) to a fraction of the carrier frequency. If a wider bandwidth is required, other methods must be sought. There are two distinct classes of synchronous errors, according to their origin.

The first class of synchronous errors results from the rejection of dc and low-frequency components in the modulated carrier during ac-coupled amplification. Low-frequency components in the modulated carrier arise, at least for amplitude modulation, when there are components in the signal and carrier waveforms of nearly equal frequency. Unless these components are negligible, or are somehow replaced, distortion will occur. Fig. 3 illustrates the distortion which results when a synchronous sinusoidal input is applied to an ideal full-wave chopper amplifier. The distortion does not arise immediately on the application of a synchronous input component, but with a rise time which depends on the low-frequency decay time of the ac-coupled circuits.

One possible approach to reducing this first type of synchronous error is to enrich the carrier spectrum so that the number of frequency
components is increased whilst their amplitude is decreased. There will thus be more frequencies which will give rise to distortion, but the magnitude of the distortion will be smaller, except in the unlikely event that the spectra of the input and the carrier waveforms are similar.

Another approach which can, if applicable, drastically reduce this type of error, is restoration of low-frequency components to the modulated carrier after ac-coupled amplification. If the carrier waveform is known to be zero during some portion of each carrier period, then the modulated carrier should also be zero during this period, and if it is not, the discrepancy represents the amplitude of a rejected low-frequency component. If a low-frequency component is added to the amplified modulated carrier in such a fashion that it is "reset" to zero whenever the carrier waveform is known to be zero, then the low-frequency component rejected during ac-coupled amplification will be largely replaced. A simple technique for achieving this compensation in a half-wave chopper amplifier is shown in Fig. 4. This technique, together with the use of an overlapped second amplifying channel, is employed in the "overlapped" amplifier discussed in Section V.

The second class of synchronous errors results from inaccurate interpolation during gaps. The output signal may be considered as a faithfully amplified version of the input signal, plus an error signal which will typically be of small magnitude except in the vicinity of gaps. Because the acceptable interpolation error, and thus the peak amplitude of the error signal, will usually be small, the spectral components of this signal will not normally cause trouble. This does not always follow, however; consider the case of a half-wave chopper amplifier in which an input signal bandwidth of only a small fraction of the chopping frequency is to be recovered, and interpolation consists of setting the output signal to zero during gaps, with the resultant large errors being reduced by drastic low-pass filtering. The effect on low-frequency components of the input signal is simply a reduction in magnitude which is proportional to the ratio of the gap width to the chopping period. If high-frequency input components are not filtered out before amplification, then a consistent positive or negative interpolation error may arise from a synchronous input component. The resulting dc offset may be unacceptable, and has been
referred to as "synchronous drift" [5]. This second type of error is not likely to be troublesome in carrier-type dc amplifiers where an attempt is made to improve the bandwidth by filtering out extraneous input components and accurately interpolating the input's behaviour during gaps.

V "OVERLAPPED" AMPLIFIER:

The input signal is still present during gaps, and provided that all gaps occupy less than half of each carrier period, it is possible to synchronize two carrier-type dc amplifiers so that the behaviour of the input signal is always available from at least one of their outputs. As is shown in Fig. 5, by combining the two outputs, the composite signal can be made to follow the input signal with a bandwidth determined by the ac-coupled amplifiers. Synchronous errors are negligible because both amplifiers are "reset" during each carrier period, and the interpolation during gaps is accurate. The frequency response of this "overlapped" amplifier is independent of the rate at which the choppers operate, and its method of operation may help clarify bandwidth limitations.

A circuit similar to that of Fig. 5 could be used as a stable wideband dc amplifier, but in most applications the conventional drift-corrected dc amplifier would be preferred, as it is a simpler alternative. Because of its mode of operation, the drift of the overlapped circuit depends mainly on the rate of change of drifts in the ac-coupled amplifiers. This fact may be of advantage in amplifiers required to operate over a very large (but slowly varying) temperature range, or which have to be relatively insensitive to large (but slow) supply voltage variations, or which are required to maintain low drifts over long periods without adjustment.

VI CONCLUSIONS:

Limits to the modulation bandwidth recoverable from a modulated carrier waveform have been examined. For pulse-type carriers, the theoretical and practical limits are fairly clear. This is not the case when the representation of the input (or modulation) in the modulated carrier
is partly-continuous. The implication by some authors that the recoverable modulation bandwidth is restricted to a clear upper limit, which is a function of the (fundamental) carrier frequency, has been rejected. It has been argued that there is no clear theoretical limit, and that in practical circuits the carrier frequency is one of several factors which may or may not be involved in determining the useful bandwidth. The other factors are gap width, allowable circuit complexity, acceptable interpolation and synchronous errors, and spectral content of the input signal outside the useful bandwidth.

Several ways in which the useful bandwidth may be related to the carrier frequency are:–

(i) The chopping rate employed with mechanical choppers is often a function of their contact changeover time, as is the width of gaps in the modulated carrier. As a result, the useful bandwidth can be indirectly related to the chopping rate.

(ii) With a sinusoidal carrier the gap width is related to the carrier frequency and the useful bandwidth can thus be a function of the carrier frequency. With this type of carrier, a more direct relationship is likely to arise through the use of non-ideal demodulation.

(iii) If synchronous errors are of an unacceptable magnitude, it may prove necessary to low-pass filter the input signal, with a filter whose cut-off frequency is somewhat lower than the carrier frequency.

A novel wide-band carrier-type dc amplifier, whose bandwidth is independent of the carrier-frequency employed, has been introduced for the purpose of clarifying bandwidth limitations. It is possible that some practical applications may be found for this circuit.

VII ACKNOWLEDGEMENT:

The author is grateful to Professor Gordon Newstead and Dr S. Kaneff for helpful discussions concerning this work.
REFERENCES:


FOOTNOTE:

1. The type of modulation usually employed is multiplication of the carrier waveform by the input signal, which is equivalent to double-sideband (suppressed-carrier) amplitude modulation [5].
FIGURE 1 - Carrier-type amplifier

Waveforms depicted are for an ideal half-wave chopper amplifier.
FIGURE 2 - Representation of input signal in modulated carrier
FIGURE 3  -  Example of errors arising from a synchronous input component

(a) Synchronous sinusoidal input signal.
(b) After full-wave chopping.
(c) After ac-coupled amplification.
(d) After full-wave synchronous rectification.
FIGURE 4 - Restoration of low-frequency components to modulated carrier
When the contacts are in the position shown, the input to amplifier A is zero and capacitor C charges to a voltage such that the output at point  is "reset".
FIGURE 5 (continued)

Overlapped amplifier
(a) Schematic diagram,
(b) Input signal.
(c) Waveform at point \( \alpha \).
(d) Waveform at point \( \beta \).
(e) Output signal.
FIGURE 5  - Overlapped amplifier
(a) Schematic diagram.
(b) Input signal.
(c) Waveform at point $\lambda$.
(d) Waveform at point $\beta$.
(e) Output signal.