AD NUMBER

AD880642

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited.

FROM:

**Distribution authorized to U.S. Gov't. agencies and their contractors;
Administrative/Operational Use; JAN 1971. Other requests shall be referred to Army Research Office, Durham, NC.**

AUTHORITY

AROD ltr 4 Aug 1971

ARO-D Report 71-1

PROCEEDINGS OF THE 1970 ARMY NUMERICAL ANALYSIS CONFERENCE

Sponsored by

The Army Mathematics Steering Committee

on Behalf of

THE OFFICE OF

THE CHIEF OF RESEARCH AND DEVELOPMENT

ARO-D Report No. 71-1

PROCEEDINGS OF THE 1970 NUMERICAL ANALYSIS CONFERENCE

U. S. Army Research Office-Durham

Report 71-1

January 1971

PROCEEDINGS OF THE 1970 ARMY NUMERICAL ANALYSIS CONFERENCE

Sponsored by the Army Mathematics Steering Committee

Host

Night Vision Laboratory

Fort Belvoir, Virginia

2-3 April 1970

U. S. Army Research Office-Durham
Box CM, Duke Station
Durham, North Carolina

INTRODUCTORY REMARKS

Jenny Bramley
Night Vision Laboratory
Fort Belvoir, Virginia

I want to welcome this audience to the 1970 Army Numerical Analysis Conference. As you all realize, this year one particular topic, namely pattern recognition, has been singled out for emphasis.

My opening remarks on pattern recognition may at first appear to be directed solely towards a narrow field, but I hope to make clear that they are much more general.

According to the latest unabridged Webster's Dictionary, the noun "pattern" has 16 different meanings, including "patron saint's day in Ireland." But even if we restrict ourselves to the more conventional meanings of "pattern," such as "representative instance" or "typical example," or "mechanical design," or "form or configuration," we must conclude that PATTERN RECOGNITION is the most encompassing of all disciplines. For example, in atomic theory, we try to recognize the pattern of atomic energy levels, in electron devices, we want to recognize the pattern of, say, the voltage-current characteristic, while in numerical analysis, we obviously try to recognize the pattern of number sequences.

Forgetting these claims to grandeur for our chosen topic, I should like to quote the formulation given by Brick and Owen in a book called "Computer and Information Sciences":

"Pattern recognition implies the assignment of a name or representation (be it simple or complex) to a class (or group of classes) of stimuli whose members have something in common. They are "similar" in a useful sense. One of the major problems in pattern recognition is the determination of a set of representative characteristics which

(a) exploit differences between pattern stimulus classes;

(b) are of minimal sensitivity to variations among members of the same class, i.e., they are most characteristic of the class as a whole, and

(c) satisfy (a) and (b) with maximum efficiency as measured over the ensemble of J pattern classes to be handled."

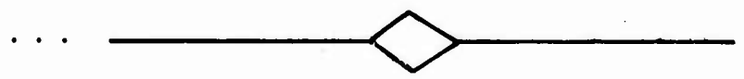
Problems in pattern recognition (in the narrower sense of the word) arise in a variety of fields, some of which are included among the papers at this conference. As I see it, once a set of representative characteristics has been specified, the types of solutions desired fall into two basic categories:

1. We want to clarify a pattern distorted during transmission or reception so as to make it more readily recognizable by a human observer.

2. Using a specified training data set, we want to design or train a recognition apparatus so that -- within prescribed limits of error -- it can recognize samples not included in the training set.

Sometimes, the human observer specified in the first category may be replaced by a recognition apparatus, i.e., the first approach acts as a preprocessor for the second.

In conclusion I want to say that we have tried to include both broad overviews and specific applications among the papers at this conference. We hope that you will recognize the pattern of our endeavors.



The Army Mathematics Steering Committee, the sponsor of these conferences, wished to thank the various speakers and chairmen for their help in the conduction of the 1970 Army Numerical Analysis Conference. Dr. John Giese, the Chairman of the conference, and the members of his Arrangements Committee--Jenny Bramley, Francis Dressel, Jack Hillard, and William Sacco--are to be congratulated on organizing such an informative program on the Theory and Foundations of Pattern Recognition.

TABLE OF CONTENTS

---

*The author was unable to present his paper at the conference.

**This paper was presented at the Conference. It does not appear in the Proceedings.

---

*The author was unable to present his paper at the conference.

**This paper was presented at the Conference. It does not appear in the Proceedings.

A G E N D A

1970 ARMY NUMERICAL ANALYSIS CONFERENCE
Night Vision Laboratory, Fort Belvoir, Virginia

Thursday, 2 April 1970

0800-0900 REGISTRATION - Lobby, Humphrey Hall, Building 247

0900-0915 OPENING OF THE CONFERENCE - Room 2B, Humphrey Hall

Jenny Bramley, Local Representative, Night Vision Laboratory, Fort Belvoir, Virginia

0915-1015 GENERAL SESSION 1 - Room 2B, Humphrey Hall

Chairman: Fred Frishman, Mathematics Branch, Office of the Chief of Research and Development

FEATURE EXTRACTION: A SURVEY
Martin D. Levine, Department of Electrical Engineering McGill University, Montreal, Canada

1015-1045 BREAK

1045-1300 TECHNICAL SESSION 1* - Room 2B, Humphrey Hall

Chairman: Raymond B. Schnell, U. S. Army Advanced Materiel Concepts Agency, Washington, D. C.

DIGITAL IMAGE PROCESSING FOR TERRAIN PATTERN RECOGNITION
Laurence P. Murphy, Engineering Topographic Laboratories, Fort Belvoir, Virginia

MULTI-IMAGE CLUSTERING
Robert M. Harlick, Engineering Topographic Laboratories, Fort Belvoir, Virginia

THE FACTORS ANALYSIS TECHNIQUE APPLIED TO MULTI-SPECTRAL DATA
Everard G. A. Barnes and Francis G. Capece, Engineering Topographic Laboratories, Fort Belvoir, Virginia

*Technical Sessions 1 and 2 will run concurrently in Humphrey Hall.

1045-1300 TECHNICAL SESSION 2 - Room 1A, Humphrey Hall

Chairman: Alan S. Galbraith, Mathematics Division, U. S. Army Research Office-Durham, Durham, N. C.

MIXTURE PROBLEMS IN PATTERN RECOGNITION
William J. Sacco, Applied Mathematics Division, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

IDENTIFICATION OF THRUST CHANGES
William D. Powers, Pacific Missile Range, Point Mugu, Calif.

FOUNDATION OF APPROXIMATE SOLUTIONS AND ITS USE IN NUMERICAL ANALYSIS
A. Sakurai, Waterways Experiment Station, Corps of Engineers, Vicksburg, Mississippi

1300-1400 LUNCH

1400-1500 TECHNICAL SESSION 1 (Continued)

Chairman: Raymond B. Schnell

ANALYSIS OF GEOMETRICAL MOMENT FEATURES EXTRACTED FROM DIGITIZED TANK PHOTOGRAPHS USING AN ON LINE PATTERN ANALYSIS AND RECOGNITION SYSTEM
James R. Rapp, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland; Donald Roberts and Jerry Milton, Rome Air Development Center, Griffiss Air Force Base, New York

1400-1500 TECHNICAL SESSION 2 (Continued)

Chairman: Alan S. Galbraith

A FAMILY OF CUBIC SPLINE SUBROUTINES
R. D. Scanlon, Maggs Research Center, Watervliet, New York

1500-1530 BREAK

1530-1630 GENERAL SESSION 2, Room 2B, Humphrey Hall

Chairman: J. Hiller, Night Vision Laboratory, Fort Belvoir, Va.

PATTERN IDENTIFICATION - A REVIEW OF PERCEPTION RESEARCH
Selby Evans, Institute for Study of Cognitive Systems, Texas Christian University, Fort Worth, Texas

Friday, 3 April 1970

0830-1030 TECHNICAL SESSION 3, Room 2B, Humphrey Hall

Chairman: Ben Noble, Mathematics Research Center, U. S. Army, University of Wisconsin, Madison, Wisconsin

THE ELEMENTARY CONCEPT OF THE KALMAN FILTERING TECHNIQUES
John Matsushino, USASTRACOM, SAFEGUARD, Fort Huachuca, Arizona

THE TWO-DIMENSIONAL POWER SPECTRUM REPRESENTATION OF AN IMAGE AND ITS RELATIONSHIP TO THE POWER SPECTRUM OF A VIDEO SCAN SIGNAL
George R. Jones, Night Vision Laboratory, Fort Belvoir, Va.

CLASSIFICATION TECHNIQUES FOR STRIP CHART RECORDINGS
Barry Rodin, Theodore Hlibka and William Sacco, Applied Mathematics Division, Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland

0830-1030 TECHNICAL SESSION 4 - Room 1A, Humphrey Hall

Chairman: Major Pasqual Perrino, Division of Medical Chemistry, Walter Reed Army Institute of Research, Walter Reed Army Medical Center, Washington, D. C.

A COMPUTER MODEL OF THE INITIAL STAGES OF MAMMALIAN PATTERN PROCESSING
Otto Zinser, Night Vision Laboratory, Fort Belvoir, Virginia and Selby H. Evans, Texas Christian University, Fort Worth, Texas

RESOLUTION AND NOISE LIMITATIONS OF NIGHT VISION DEVICES
Gertrude H. Kornfeld and Walter R. Lawson, Visionics Technical Area, Fort Belvoir, Virginia

PATTERN RECOGNITION ASPECTS OF EVOLUTION
Aurel N. DeHollan, Electronics Research Center, NASA, Cambridge, Massachusetts

1030-1100 BREAK

1100-1300 GENERAL SESSION 3 - Room 2B, Humphrey Hall

Chairman: Jenny Bramley, Night Vision Laboratory, Fort Belvoir, Va.

PROBLEMS OF DESIGNING A FLEXIBLE SCENE-ANALYSIS SYSTEM
Marvin L. Minsky, Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass.

1100-1300 GENERAL SESSION 3 (Continued)

INTEGRAL GEOMETRY AND THE RECOGNITION OF SHAPES
Eugene Wong, Department of Electrical Engineering,
University of California at Berkeley, Berkeley, Calif.

1300-1400 LUNCH

FEATURE EXTRACTION: A SURVEY

M. D. Levine
Department of Electrical Engineering
McGill University, Montreal, Quebec

I would like to thank the organizers of the 1970 Army Numerical Analysis Conference for inviting me here to present this talk today. The subject of my talk is primarily based on material appearing in a paper published in the Proceedings of the IEEE entitled "Feature Extraction: A Survey" and appearing in Volume 57, No. 8 of its 1969 issue [1]. For a complete set of references, I recommend that you refer to this paper. The talk will be a presentation of different slides, but because of the large number involved it will be impossible for me to give the individual credit for most of these.

Let me begin by asking the following question: "Do we as human observers know what we are looking for? Do we really know what the desired features are or should be?" I contend, after considerable study in this area, that in fact we do not. It is interesting to make a comparison between some of the really fantastic operations that the human mind is capable of achieving as compared with the rather mundane tasks that computers have been so far programmed to do. Slide (1) [all slides are at the end of this article] shows a sample of scanned letters A which was obtained in a character recognition project; it is obvious that the different A's vary in many aspects, however, it is quite easy for a person to recognize all six examples as a letter A. How are we doing this? Are we using template matching? Possibly, but in addition, human recognition is very dependent on context. For instance, if we examine slide (2), I think everyone would agree that most people would read the words shown as THE CAT even though the second letter in the word THE and the second letter in the word CAT are identical. It appears that we are using context in order to read this. Similarly, slide (3) shows an interesting phenomenon known as the "Peter-Paul Goblet" [2]. In this case, the appearance of a part depends on the whole in which it is embedded rather than the contrary, which is usually accepted as the conventional wisdom. Let us consider slide (4). "Is this an elliptical object or is it a circle in perspective?" We really cannot answer this question without the use of context until we examine slide (5) where it becomes obvious that in fact it is a circle in perspective representing the hoop being held by the little boy [3]. A similar example is shown in slide (6). What does this figure represent? If we examine slide (7) we see that is is a puddle and in fact this really stresses the importance of contextual information which is required for human

*Those slides for which figure captions (and references in the captions) appear, refer to this paper.

recognition [3]. Another question we may ask is the following: "Are familiar patterns recognizable no matter where or how they fall on the retina of the eye? Slide (8) shows that we can tolerate considerable ambiguity in this respect. If one tilts one's head slightly to the right, one can see that this figure represents a little puppy; if on the other hand one tilts one's head to the left, one can see that the figure could possibly represent the chef in a rather prestigious restaurant. Human recognition is very heavily dependent on verbal cues as well. For instance, if we examine slide (9), it is quite difficult for us to decide what this figure represents. However, if I present to you a very simple cue by saying that it belongs to a rather ill-defined category, such as for example musical instruments, I think most of your will agree that the figure is recognizable as a violin. Finally, let me show you slide (10) which represents the "RAT-MAN FIGURE" [2]. If we look at this in one way, we can see a little rat with a rather long tail scurrying around. On the other hand, with a little bit of imagination we can also see this figure as the portrait of a man (rather ugly, I guess) wearing large glasses, which in fact are quite the style these days. It has been shown that after viewing many animals most people will see this as a rat. In other words, the previous visual cues have conditioned the recognizer to see this as a rat; however, there are people who will still see this as a man. My intention in showing you these last few slides is to demonstrate to you that man's capabilities as far as feature extraction and pattern recognition are concerned are very complicated and are really not well understood. Therefore, when we set ourselves the task of programming a computer to perform quite similar operations, we are setting for ourselves a very difficult task indeed.

The main body of my talk will be concerned with two main headings: the first, micro-analysis and micro-operations, and the second, macro-analysis and macro-operations. This division is really only helpful from a conceptual point of view and in fact I think one could put strong arguments for taking any particular method and considering it under heading one rather than two and vice versa. I will briefly discuss under various headings some interesting, or what I consider the most interesting methods and philosophies. Most of the methods are unfortunately rather primitive and it will become obvious that a considerable amount of research is required to develop organized and theoretical methods for obtaining features. In my talk I will stress the importance of practical problems for the following reasons:

(i) I think that in engineering we should use practical problems to stimulate the discovery of new techniques and

(ii) I feel this is where the important real problems arise.

For example, it is not enough to write a program which will categorize chromosomes or cells; in an actual situation one finds overlapping

chromosomes or cells and a very important aspect of such a problem would be the writing of a program which could distinguish between, for example, two chromosomes that are overlapping and a large chromosome. The problem of isolating different parts of a picture in order to then recognize them is a difficult one for a computer, and therefore the solution to the problem of deriving geometrical relationships is not easily found. If we examine slide (11) we see that it is very easy for us to distinguish figure (a) submerged in figure (d) even though a rather complicated pattern has been superimposed on the former. Now I think that if we had to write a program to distinguish or isolate figure (a) we would be confronted with a difficult task indeed.

Let us examine the standard conceptual configuration which is typically used for pattern recognition problems (shown in slide (12). On the left, we assume that the image is obtained by some hardware device which could be a flying spot scanner, an image dissector, a mechanical device or some other digitizer such as for example, in the case of charts of graphs. In this way we are able to take a picture, a chart, or any other data and represent them internally in the computer as a list of data. This is referred to as the transducer stage. The next stage is the preprocessor or feature extractor followed by the classifier which classifies the different inputs to the whole system. Generally speaking, the feature extractor and preprocessor stage is not really very well defined and is usually accepted to mean anything that occurs between the hardware, transducing stage and the output classification stage.

Suppose we are confronted with a real problem. Do we have any rules or theory to guide us with respect to the feature extraction stage (slide 13)? Generally speaking, the following three points are relevant:

(i) What features are important? - in fact we have no theory to guide us.

(ii) The design of feature extractors has until now been empirical and uses many ad hoc strategies, and finally

(iii) We can usually get some guidance from biological prototypes and possibly maybe even psychology.

The first concept which I wish to examine is referred to as smoothing. Here, we are concerned with the averaging-out of noise much in the same way as the problem which arises in the case of electrical noise. The noise in the digitized image case can either be represented by the absence or the appearance of a signal. An

example of a noisy situation is demonstrated in slide (14) which is a x80 magnification of a sectioned lung. The structure that is visible here is the alveolar structure which is prevalent in the human lung. This slide was used in a project in which I am presently involved and one of whose main objectives is to obtain the mean linear intercept of the alveolar structure in the lung [4,5,6]. This measurement can be related to the internal surface area of the lung which is of interest to pathologists in their study of emphysema. In order to obtain the mean linear intercept it is desirable to neglect such things as small specks which are artifacts and the holes in the alveolar walls. From a physical point of view one knows that there are not actual holes in the walls but these can arise either because of the way in which the histological section was obtained or because of noise in the digitization process. Therefore, the first difficulty to us in this problem and in fact most such problems is the smoothing-out or elimination of this noise. One method of accomplishing this is demonstrated in slide (15). Normally the image matrix is rectangular and one superimposes on this matrix a 3 x 3 submatrix or window where each element in the submatrix can assume one of eight light density values, $Y_1, Y_2, \ldots, Y_8$, where Y_1 represents white and Y_8 represents black. Using the relationship that I will show you, we can map the original digitized matrix A into a new matrix B which has been smoothed or averaged. Slide (16) shows the algorithm that we used. If the sum of the light densities impinging on a given submatrix is greater than a threshold and the middle element is not white, then we set b_{ij} equal to a_{ij}. If on the other hand, the middle is blank we set b_{ij} equal to Y_2. If the sum is less than a threshold, then we set b_{ij} equal to Y_1, which is in fact just making it equal to white. There are many problems even with such a simple method. For example, how large an aperture should we use? Is 3 x 3 enough? Should we use 5 x 5 and so on? What kind of aperture? Is there any value in using a square aperture? Maybe we should use a hexagonal aperture as shown in slide (17). How do we choose the threshold for the algorithm that I just discussed? Should we choose it once and for all, or should we make it adaptive to the actual data? It can be easily shown with such an algorithm that it is quite possible for us to insert things or remove others that are of interest to us. Slide (18) shows a digitized version (matrix A) of the lung section that I showed in slide (14). Slide (19) is just a close-up of slide (18) showing the different grey levels represented by a number from 0 (blank) to 7. Slide (20) is the smoothed version, matrix B, of the original digitized matrix A and we can see that the lines are generally thicker and the islands or specks are generally bigger in size.

Another operation similar in nature to the one mentioned above is that related to filtering operations on lines. Mainly, these have been used for real time character recognition where one is concerned with stroke analysis in real time as a person writes on an instrument such as a RAND tablet. In fact the methods are

similar to those generally in use in numerical analysis. As opposed to the general averaging problem mentioned above, in this case the line for the most part is well defined. Usually a thinning operation is also performed in real time.

Contour tracing is generally used for obtaining line drawings. The contour or outline carries a significant amount of the information required for the recognition of objects and an additional advantage is that it is independent of translation, size and rotation. An example of the use of contour analysis is the experimental reading machine for the blind by Mason and Clemens [7]. The contour is determined before any coding which later can be used in the recognition of the characters. The next slide, slide (21) shows the simple algorithm for following the contour. The follower turns right after meeting a white element, turns left after meeting a black element, but after three similar moves is required to make the opposite choice. This is necessary since in view of the noise in the scanning process, it is quite possible to return to an element and find a different reading. Slide (22) shows a human fingerprint which was obtained during a project whose main objective was the determination of a digital method for generating a reference point in the fingerprint [8]. The contour was used to determine the ridge curvature, and then a standard gradient method was employed to find the direction of movement. In such a noisy situation, as can be seen from the slide, it is quite difficult to program a contour follower which will always successfully follow the desired contour.

At this point it is interesting to consider the general area of line drawings and sketches and indeed the intimate relationship that exists between the field of computer graphics and pattern recognition. After all, in both areas of endeavour one is concerned with visual images as programmed on the computer. An interesting system where graphics and pattern recognition were used to advantage is that suggested by Krull and Foote [9]. These authors were involved with a contour follower whose progress was monitored on a graphics console and the program so constructed that human intervention was possible when the follower failed. Another interesting problem in this area is the recognition of faces from photographs. In order to accomplish this, it is necessary to first determine line drawings and from these perform the necessary pattern recognition. In a similar vein is the problem of cartoon recognition, in particular political cartoons. It is well known that we are able to recognize faces of different political personages drawn by different cartoonists. How do we do this? It would be interesting to study this problem as it may give us some interesting insights into the pattern recognition in both humans and computers.

Related to contour followers are the programs or algorithms which perform edging operations. This procedure is equivalent to that of differentiation and results in a display of the sharp changes in a given image. The reason possibly for the interest in this type of procedure is that it is hypothesized that the visual cortex in the human tends to detect straight lines. What we are doing here is high frequency filtering as opposed to the averaging operation discussed previously. Slide (23) is an example of an image communication system which incorporates separate transmission of high and low frequency content. This is an example of a possible picture communication system and is really not suggested as a technique for pattern recognition. A possible advantage of this type of communication system is that fewer samples are required for the low frequency data than for the high frequency data which it is hypothesized contain the information obtained by an edging operation. Slide (24) shows a typical receptor configuration postulated by Deutsh [10] which was used as an integral part of his SLEN. One can see that the ON receptors in the middle have a weighting of two while the OFF receptors on either side have weightings of minus one. If we excite this array, for example, with an even excitation, then the output of this sort of a combination would be zero. On the other hand, if we convolve this array with long orthogonal lines, the output would also be zero. Slide (25) shows how a SLEN might be connected to give us a binary output. We can now arrange these SLENS in groups to cover a whole field and Deutch has used this concept to recognize digits. Since we recognized that edging is a process of differentiation, we may accomplish this in two dimensions as is shown in slide (26). If we apply an edging operation in two dimensions to a typical scene, for example an office scene (slide (27)), we can produce a line drawing of this rather complicated image. The next slide,(slide (28)) shows the lower half of the previous picture. The features are given by line segments at any one of eight different angles and the centers of the lines are plotted. Slide (29) shows a close-up of the previous image. It is quite obvious that this scene is far too complex for such a simplistic approach and one would have great difficulty in making any headway with only this type of operator.

Another interesting approach to pattern recognition and preprocessing is that using spatial frequency response methods as has been done at the Jet Propulsion Laboratory with respect to the image enhancement of lunar pictures [11, 12]. This is accomplished using a digital filtering or as it is otherwise known, spatial filtering, and is capable of producing some rather striking results. Slide (30) is an example of an unprocessed radioisotope scanner chest film. The horizontal lines that can be seen are rather distracting to the observer. Slide (31) shows the same film after the removal of the scan lines and slide (32), after a low pass filtering operation,

which has the effect of removing the disturbing non-random structure. A comparison of slides (30) and (32) shows the power of these methods. Slide (33) shows an unprocessed radiograph of bone while slide (34) shows the same radiograph after the background has been removed using a high pass filter and the contrast has been enhanced. A considerable amount of detail is now visible which was not available in the original film. Slide (35) shows an unprocessed retina photograph while slide (36) shows an enlargement of the section which will be enhanced. Slide (37) gives this area after some contrast enhancement and slide (38) shows the same scene after even further enhancement. Note the improved definition of the blood vessel wall in a picture which is quite striking.

Shape and curvature are two concepts which are of great concern in pattern recognition. It has been stated that "inflection points on a contour are its informationally richest part" [13]. Slide (39) shows a picture where the regions of curvature on the original photograph have been connected by straight lines. I think it is obvious to everyone that this line drawing represents a small kitten and thus all the pertinent information is represented by the curvature. Slide (40) shows a simple way of coding the curvature along a straight line where an eight level scale is used for quantizing the directions along a given curve. This method of coding, called chain encoding, has been postulated by Freeman [14] and has been used to describe pieces in an apictorial jigsaw puzzle. The problem in this case was to use a computer to solve jigsaw puzzles where only the shape of the pieces is of importance.

Another topic of interest can be referred to as correlation methods. These are often called either template matching or windowing techniques. In general, the operation can be performed either optically or digitally. Some of the problems associated with this type of approach relate to the fact that correlation methods are sensitive to translation magnification, brightness, contrast, orientation, and noise of any nature. Often one reads in the literature that by using prenormalization as an initial stage one can then successfully apply correlation methods. However, prenormalization is extremely difficult since even examples of one particular type of object may not have similar shape. If we do assume that the application of correlation methods to pattern recognition by computer is difficult, an interesing argument arises with respect to human recognition. I have often discussed this with psychologists who insist that the basic level of human recognition is premised on the use of templates. Is this in fact a valid argument? I am prepared to agree with them that at the lowest level of recognition, even in the computer, it is necessary to use templates. However, a question does arise as to how high in the hierarchy of pattern recognition can one go until one reaches a point at which templates are no longer useful.

A large number of templates have been postulated in the literature and some of these are shown in the following slides. Slide (41) shows an edge detector very similar to the SLEN previously discussed. Slide (42) demonstrates one of the problems that arises with such an edge detector where we desire to correlate with respect to an object exhibiting a certain radius of curvature. We can see the effect that orientation has by noting that the correlation increases monotonically with the radius of the object and decreases with the angular rotation of the edge. Slide (43) shows a window which could be used for detection of corners; slide (44) shows a line segment detector which would unfortunately be very sensitive to the width of the line, and would require extreme uniformity in the lines being detected to be of any use; slide (45) shows two interesting windows - one for detecting orthogonal intersections, and one for detecting spots, again of a certain uniform size. Slide (46) shows an interesting example of correlation techniques which have proven to be successful. The application is in the area of real time recognition of hand printed characters. As the person writes on a RAND tablet, the pattern recognition program in the computer is required to identify the printed characters. A property vector is constructed, which describes the position of the "pen" as a function of time, using the matrix shown in the slide. The next step is to correlate these vectors. Note that here we have an extremely valuable piece of additional information, that is, the temporal information which is usually not available in the cases mentioned above.

An important concept in preprocessing and feature extraction is that of connectivity. Let us consider a previous slide where I showed you the detailed alveolar structure in the lung. Here we were interested in the mean linear intercept which can then be related to the total internal surface area of the alveoli in the lung. In order to determine the mean linear intercept, it is necessary to detect intersections with the wall structure, which by the way, is known to form a connected set. It is extremely difficult to detect the intersections using, for example, the intersection detector I showed you a moment ago. For one thing, it is necessary to ignore all the islands and specks and other artifacts which are not part of the alveolar structure. To solve this problem we used the concept of connectivity which has the effect of labelling the wall structure with one label as shown in slide (47), and all the other extraneous material in the slide with other labels which could then later be ignored. Slide (48) shows a blow-up of the previous slide where the two islands which are not part of the desired structure are shown to have different labels than the wall structure itself. It is interesting, that connectivity is an important criterion for human visual recognition as well, but the big problem in its computer application is the effect of noise.

What kind of topological properties can one actually use as features? There is, of course, an infinite variety and the specifics will depend on the actual application. People have used lakes, loops, holes, arcs, arches, curves, etc., etc., etc. Two interesting properties which are slightly different from the usual ones appearing in literature are presented by Munson [15]. Slide (49) shows an example of how one might obtain what is called the concavity in a figure. First a contour follower is used to obtain the contour of the figure 7 and from this the convex hull is calculated. The concavity is then defined as the connected region adjacent to both the figure and the hull. Similarly, slide (50) shows an example of how one might obtain the enclosures in a figure. Here we define the enclosure as the connected regions of ground touching the figure but not the convex hull.

Let me now briefly discuss the concept of macro-processing, where I again would like to stress that this is really a conceptual categorization rather than, in most cases, a practical one. The first concept of importance here is that of parallel processing an approach that has been deified considerably in the literature. Everyone agrees that parallel processing is a good idea but not much has been produced of a practical nature along these lines. There are many theoretical ideas on how this could be incorporated into pattern recognition problems. An interesting application is that of Hawkins [16] who has used electro-optical techniques in the form of image intensifier tubes. He was able to perform parallel operations of addition, substraction, multiply sums, spatial filtering and threshold logic. As an example, let us consider slide (51) which shows a typical scene at one moment in time. Slide (52) shows the same scene a little later where now a man with a white shirt is standing behind the tree; notice his shadow in the foreground. Now if we difference the two scenes in parallel, we see on the resulting slide (53) white regions which represent the shirt and shadow of the man. Of course, we are still left with the problem of pattern recognition in this new scene which is by no means a trivial task.

A feature which is of considerable importance to human recognition is that of textural information. This type of global concept is easily recognized by people, but it is extremely difficult to achieve on a computer. Essentially, in this case, we are interested in the frequency content of the spatial data. For example, in slide (54) showing different lung sections, the pathologist is interested in grading the lung sections according to the degree of emphysema prevalent. We can see in the slide that the texture, at least in part of the sections, varies from section to section with the one at the top left hand corner being a normal lung and the one at the bottom right hand corner being a diseased lung. To program the computer to recognize these textural differences would be an extremely difficult task.

How do we as human observers describe a given object? This is usually done by using certain terms describing its shape, and of course we desire to do this in the computer as well. An interesting shape descriptor is the medial axis transform due to BLUM [17]. The gross properties of the MAT are related to the structural properties of the pattern and can therefore be used as a description of the shape. The MAT is generated by causing a given pattern to shrink down in size by allowing the area outside the pattern to propagate with uniform velocity into the pattern. The MAT is then defined by the locus of self-intersections of the propagating boundary area. We may also use a grass-fire analogy to describe the construction of the MAT. If we consider the pattern superimposed on a field of grass, and if we consider that a fire is lit on the boundary of the pattern, we may imagine the fire propagating to the center and extinguishing itself when the waves would meet up with each other. The lines of extinction can be considered to be the MAT. In fact this process has been shown to be reversible if the temporal information is retained. Slide (55) shows typical medial axis transforms for some simple objects. Slide (56) shows the MAT in terms of propagating wave as described previously, as does slide (57). Slide (58) is an interesting demonstration of the possible usefulness of the MAT. On the left is a simple sketch of a "stick man" and on the right is the same man, grossly distorted. Note however that the MAT's in both cases are similar and possibly this might be used as a powerful method of shape description which would overcome the usual problems in matching objects belonging to the same class. Finally slide (59) shows a CRT display of the MAT for a dog.

I now come to what I consider to be the most promising approach to the pattern recognition problem, that is, the use of articular analysis or a linguistic approach to produce a two dimensional pattern recognition language. This is analagous, I feel, to how a human perceives and recognizes patterns. An interesting book on this subject is the one by Neisser [2]. Some papers on this subject have appeared but a considerable amount of research has yet to be done. It is possible to make an analogy between pattern images and ordinary language as shown in slide (60). Here we may arbitrarily relate the parts of the image to the characters, the more complicated structure in the images to sentences, and the overall pictures or subpictures to paragraphs in a language. This powerful concept may be organized as shown in slide (61) which shows a hierarchical configuration for pattern recognition using articular analysis. At the lower level is the feature extractor which would use methods and algorithms of the type I have already discussed. The output of this stage would be fed to a block where names would be assigned to classes of property lists as described in the previous stage. Next the program would, using a given pattern recognition language, make statements about relationships between the objects, the classes and the properties previously recognized. On the basis of these statements the algorithm would then produce a classification which

obviously would be quite different from the normal classification procedures now found in the literature. I further feel that for this type of approach to show results one will have to perform the design of the language in a man-machine interaction environment as shown in slide (62). Parenthetically, I might add that we are presently developing such a system in the Department of Electrical Engineering at McGill University. At the bottom is shown a person interacting with a hybrid computer system via either a typewriter or preferably a graphics display and light pen. The graphics display unit would be capable of displaying grey tone pictures of quite large matrices in order that the effect of computer algorithms on images and scenes can be evaluated properly. In addition, via the display, the person is capable of interacting with the image scanner, requesting various scans as desired. This is not such an unusual environment, since it should be accepted that pattern recognition is intimately related to images, as is graphics. In this context, an interesting study would be the relationship between such a postulated two dimensional pattern recognition language and certain graphics languages now appearing in the literature.

Finally to end this talk let us examine the tremendous power of the human brain to perform recognition when the clues presented by the eye are minimal. Consider slide (63). This is not a set of meaningless random lines but represents solid configurations: we all can recognize that this is a washerwoman on her knees next to her bucket of water!

REFERENCES

1. Levine, M. D., Feature Extraction: A Survey, Proc. IEEE, Vol. 57, No. 8, August, 1969, pp. 1391-1407.

2. Neisser, U., Cognitive Psychology, New York: Appleton-Century-Crofts, 1967.

3. Gregory, R. L., Eye and Brain, World University Library, McGraw-Hill Book Company, 1966.

4. Levine, M. D., Reisch, M. L., Thurlbeck, W. M., A Computer Method for Measuring the Internal Surface Area of the Human Lung, Proc. 8th International COnference on Medical and Biological Engineering, July 20th-25th, 1969, Chicago, Illinois, Session 12-9.

5. Levine, M. D., Reisch, M. L., Thurlbeck, W. M., Image Processing and the Morphology of the Human Lung, Quarterly Bulletin of the Canadian Information Processing Society, Vol. 9, No. 5, Fall-Winter, 1969, pp. 6-14.

6. Levine, M. D., Reisch, M. L., Thurlbeck, W. M., Automated Measurement of the Internal Surface Area of the Human Lung, IEEE Trans. on Biomedical Engineering, July, 1970 (to appear).

7. Mason, S. J., Clemens, J. K., "Character recognition in an experimental reading machine for the blind," in Recognizing Patterns, P. A. Kolers and M. Eden, Eds., Cambridge, Mass.: M.I.T. Press, 1968, pp. 155-167.

8. Karasik, R. P., "A digital method for generating a reference point in a fingerprint," M Eng. Thesis, Dept. of Electrical Engineering, McGill University, Montreal, Canada, July, 1969.

9. Krull, F. N., Foote, J. E., "A line scanning system controlled from an on-line console," 1964 Fall Joint Computer Conf., AFIPS, Proc. Vol. 26 (San Francisco, Calif.) pp. 397-410.

10. Deutsch, S., "Conjectures on mammalian neuron networks for visual pattern recognition," IEEE Trans. Systems Science and Cybernetics, Vol. SSC-2, pp. 81-85, December, 1966.

11. Selzer, R. H., Improving biomedical image quality with computers, Technical Report 32-1336, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, October 1, 1968.

12. Nathan, R., Digital Video-Data Handling, Technical Report 32-877, Jet Propulsion Laboratory, California Institute of Technology Pasadena, California, January 5, 1966.

13. Kolers, P. A., "Some psychological aspects of pattern recognition," Notes of M.I.T. Special Summer Program Course 6.59s on Pattern Recognition M. I. T., Cambridge, Mass., July 5-15, 1966.

14. Freeman, H., "On the encoding of arbitrary geometric configuartions," IEEE Trans. Electronic Computers, Vol. EC-10, pp. 260-268, June, 1961.

15. Munson, J. H., "The recognition of hand-printed text," IEEE Pattern Recognition Workshop, Puerto Rico, October, 1966.

16. Hawkins, J. K., "Parallel electro-optical picture processing," in Pictorial Pattern Recognition, G.C. Cheng, R. S. Ledley, D. K. Pollock, and A. Rosenfeld, Eds Washington, D. C.: Thompson, 1968, pp. 373-385.

17. Blum, H., "A transformation for extracting new descriptors of shape," in Models for the Perception of Speech and Visual Form. Boston, Mass: M.I.T. Press, 1967, pp. 362-380.

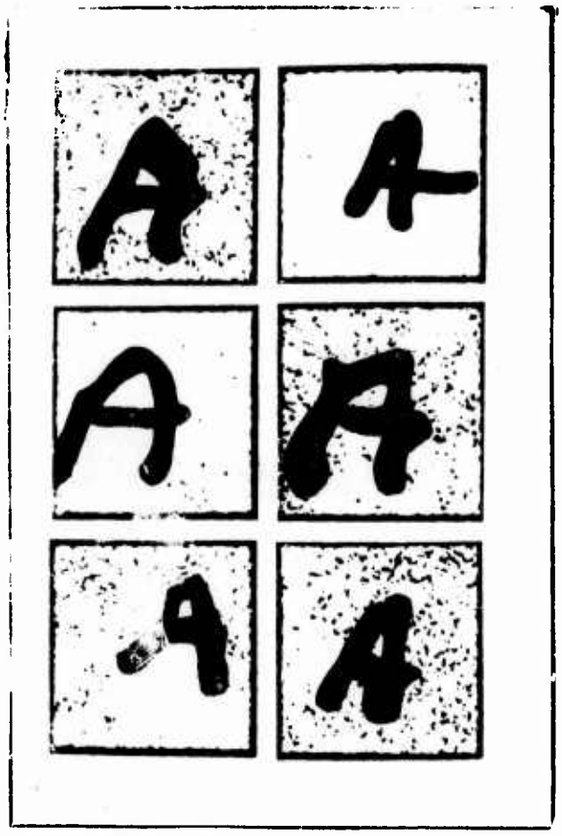


Figure 1

Figure 2

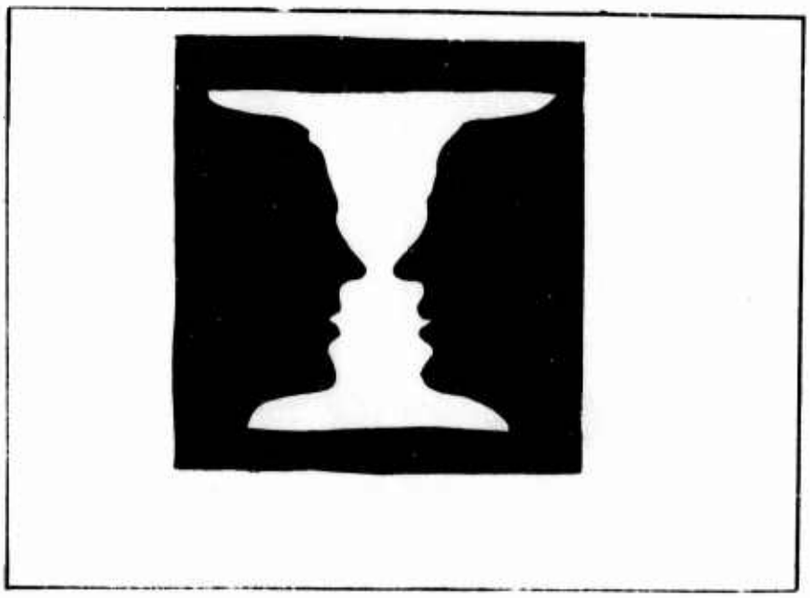


Figure 3

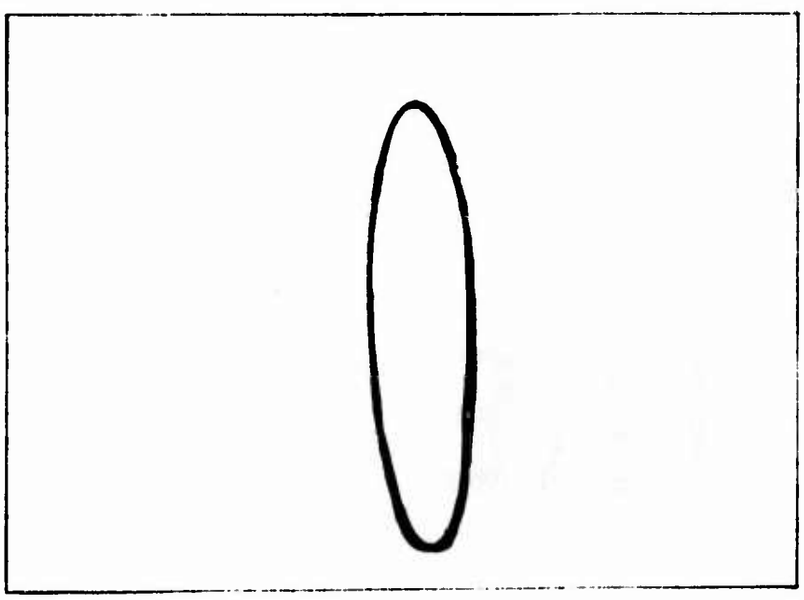


Figure 4

Figure 5

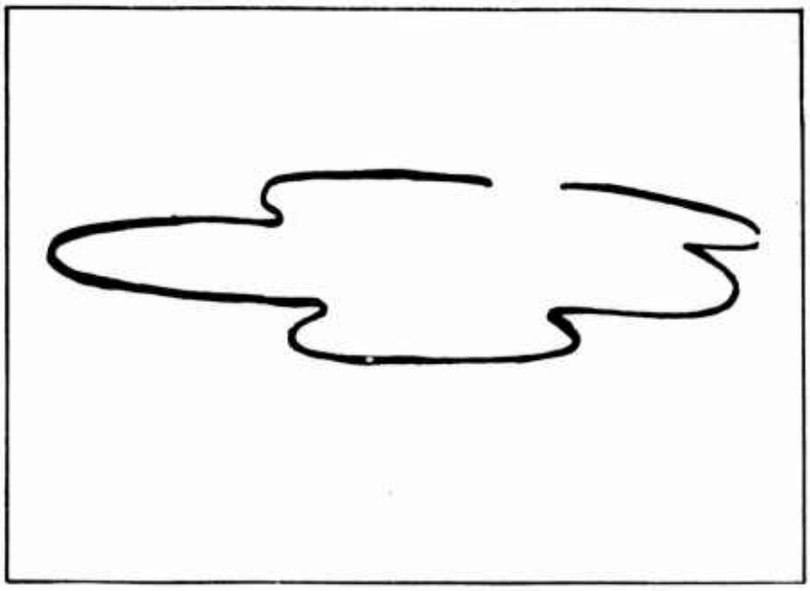


Figure 6

Figure 7

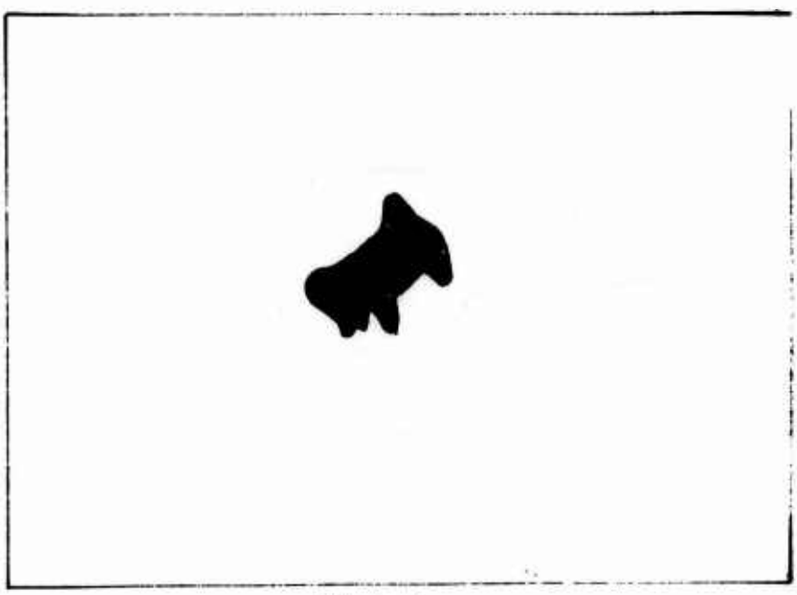


Figure 8

Figure 9

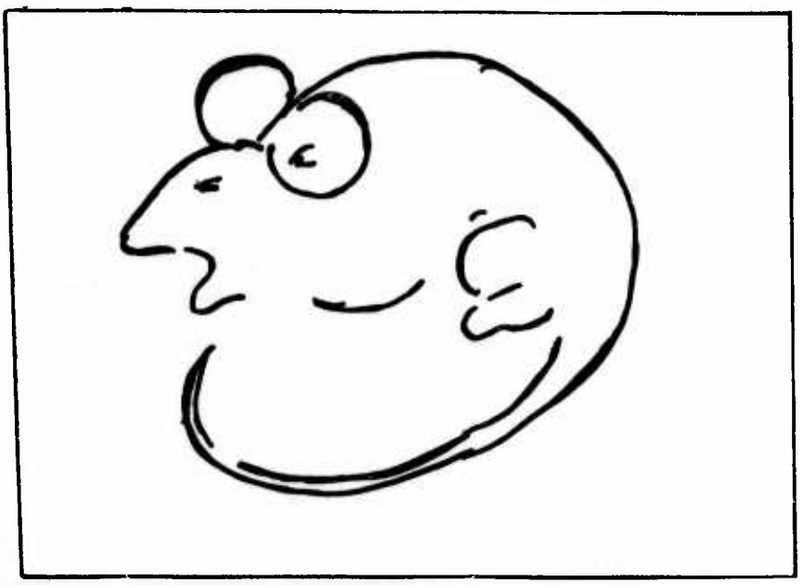


Figure 10

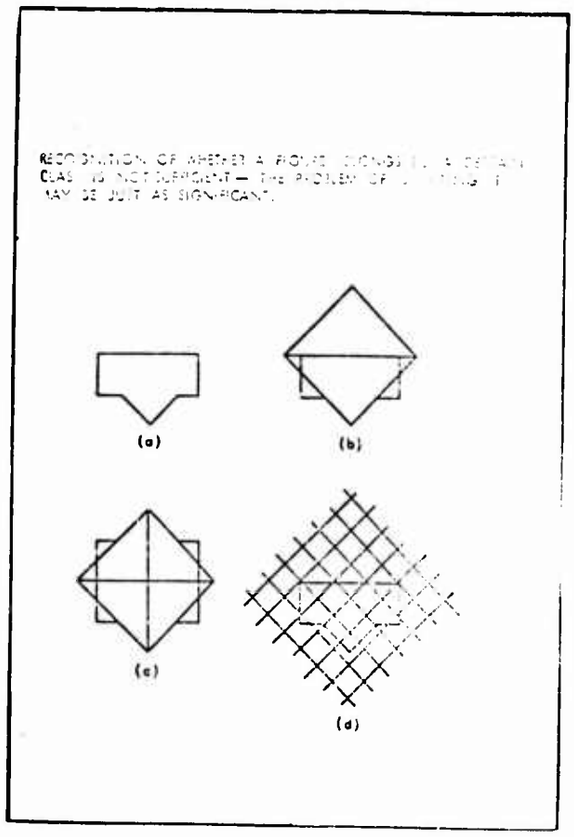


Figure 11

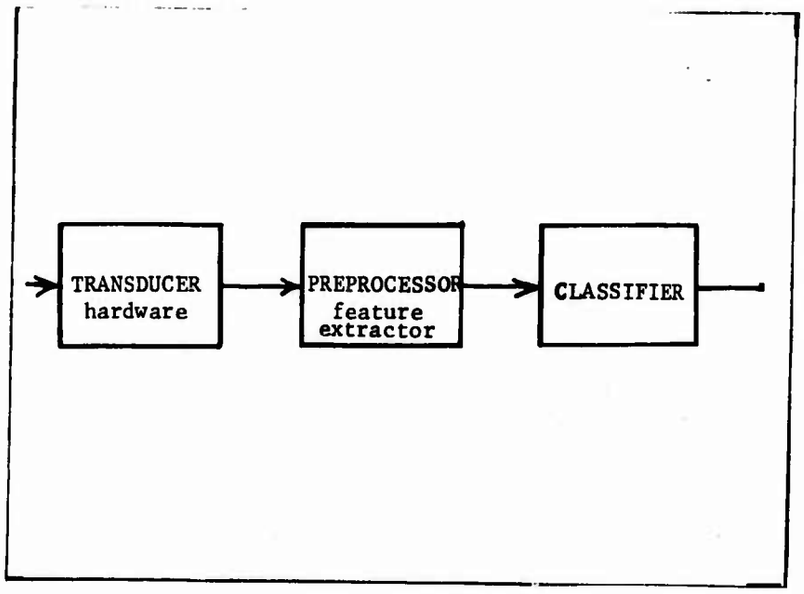


Figure 12

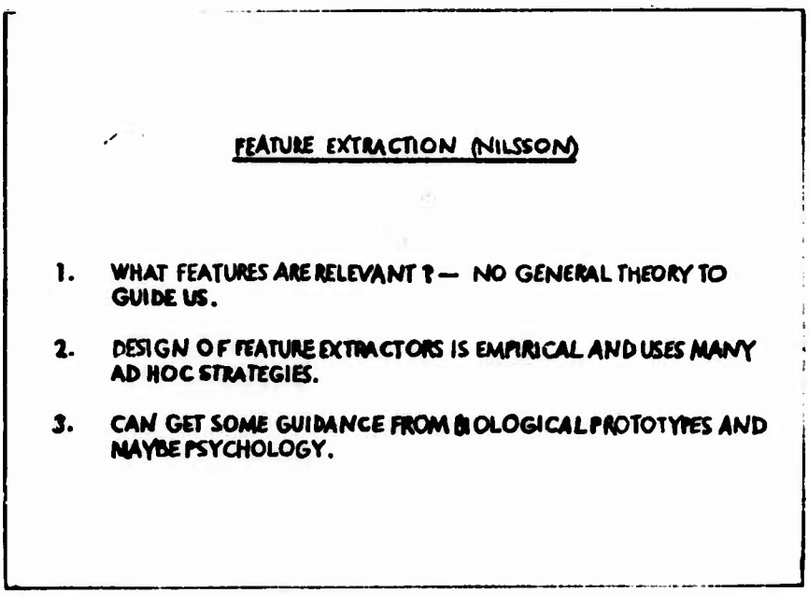

FEATURE EXTRACTION (NILSSON)
1. WHAT FEATURES ARE RELEVANT ? — NO GENERAL THEORY TO GUIDE US.
2. DESIGN OF FEATURE EXTRACTORS IS EMPIRICAL AND USES MANY AD HOC STRATEGIES.
3. CAN GET SOME GUIDANCE FROM BIOLOGICAL PROTOTYPES AND MAYBE PSYCHOLOGY.

Figure 13

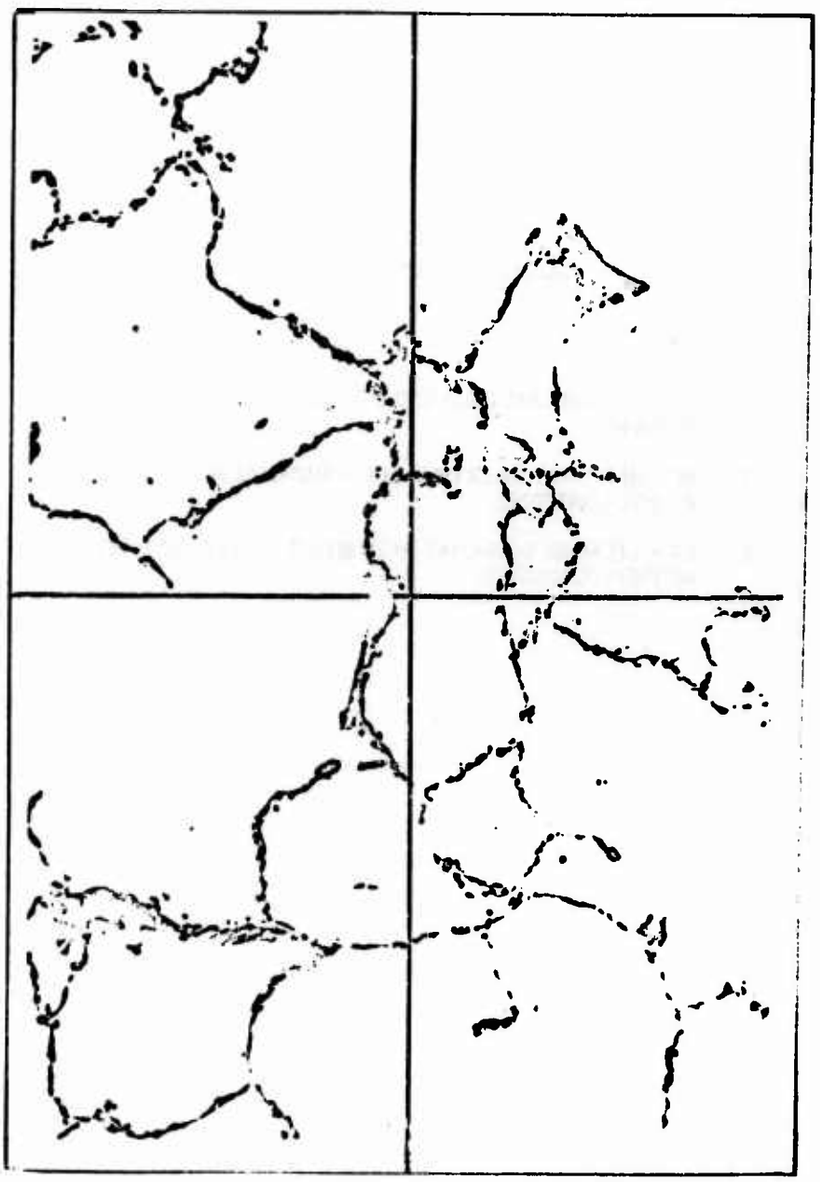


Figure 14

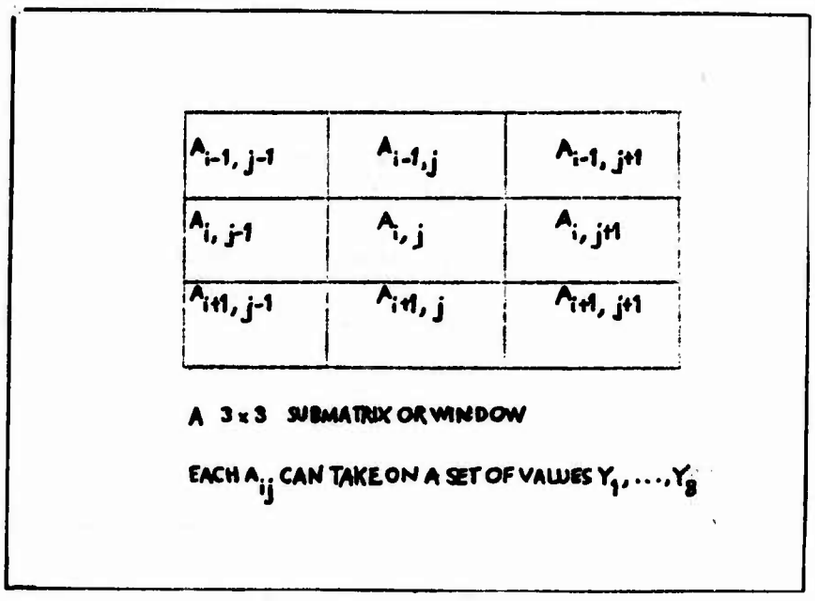

$A_{i-1, j-1}$
$A_{i-1, j}$
$A_{i-1, j+1}$
$A_{i, j-1}$
$A_{i, j}$
$A_{i, j+1}$
$A_{i+1, j-1}$
$A_{i+1, j}$
$A_{i+1, j+1}$
A 3 x 3 SUBMATRIX OR WINDOW
EACH A_{ij} CAN TAKE ON A SET OF VALUES $Y_1, \ldots, Y_g$

Figure 15

LET B_{ij} BE THE ENTRY IN THE MAPPED MATRIX B CORRESPONDING TO A_{ij}

(i) IF $\sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} A_{m,n} \geq \theta$

AND $A_{i,j} \geq Y_2$, THEN SET $B_{i,j} = A_{i,j}$

$A_{i,j} < Y_2$, THEN SET $B_{i,j} = Y_2$

(ii) IF $\sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} A_{m,n} < \theta$

THEN SET $B_{i,j} = Y_1$ REGARDLESS OF THE VALUE OF $A_{i,j}$.

Figure 16

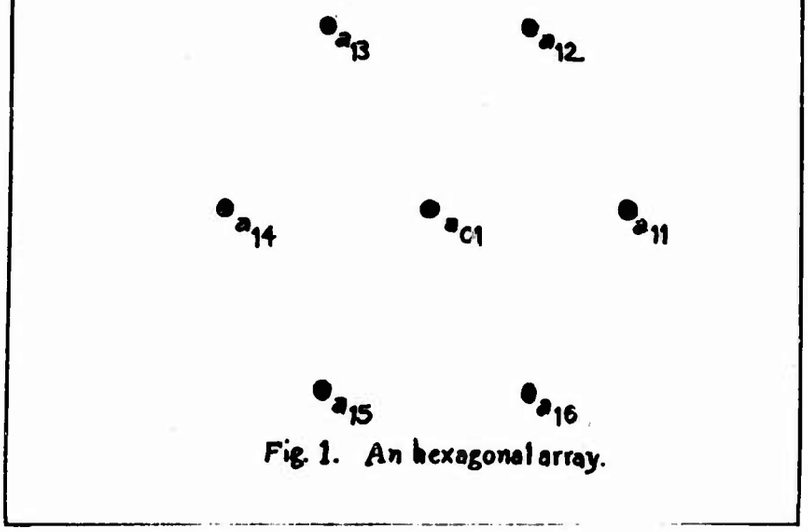


Fig. 1. An hexagonal array.

Figure 17

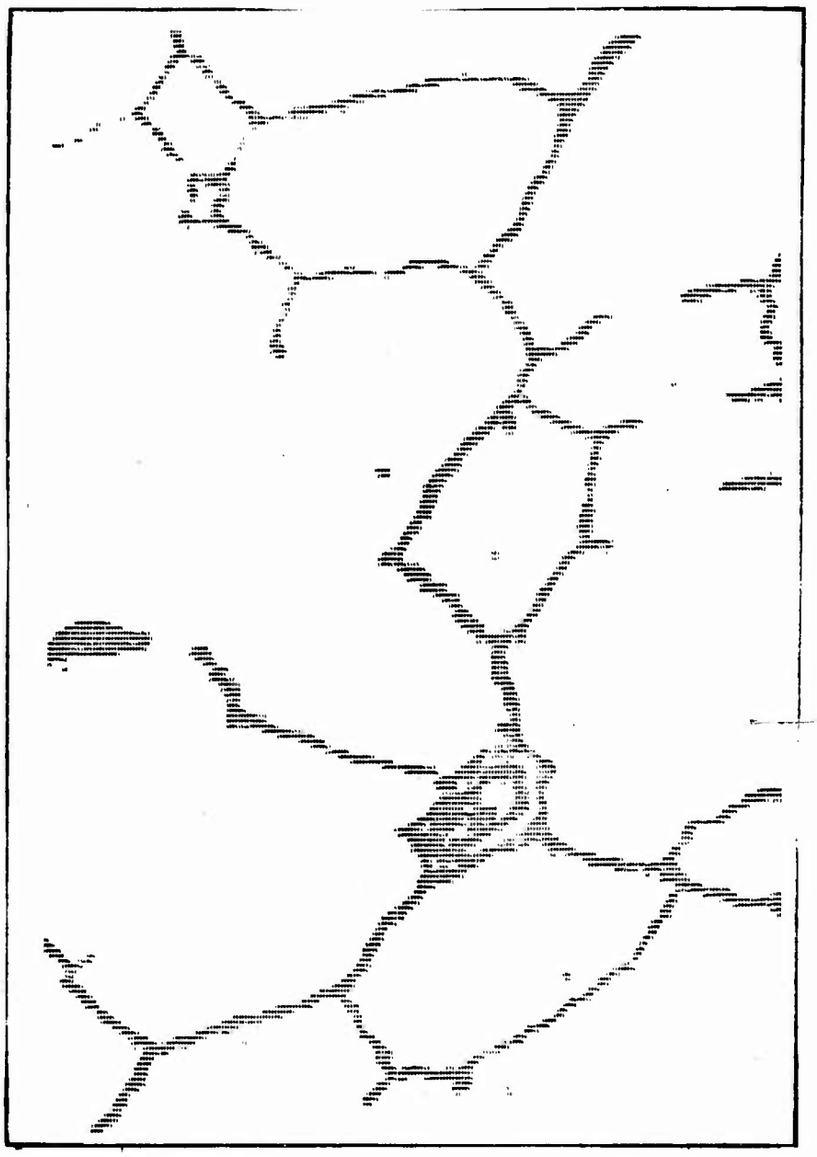


Figure 18

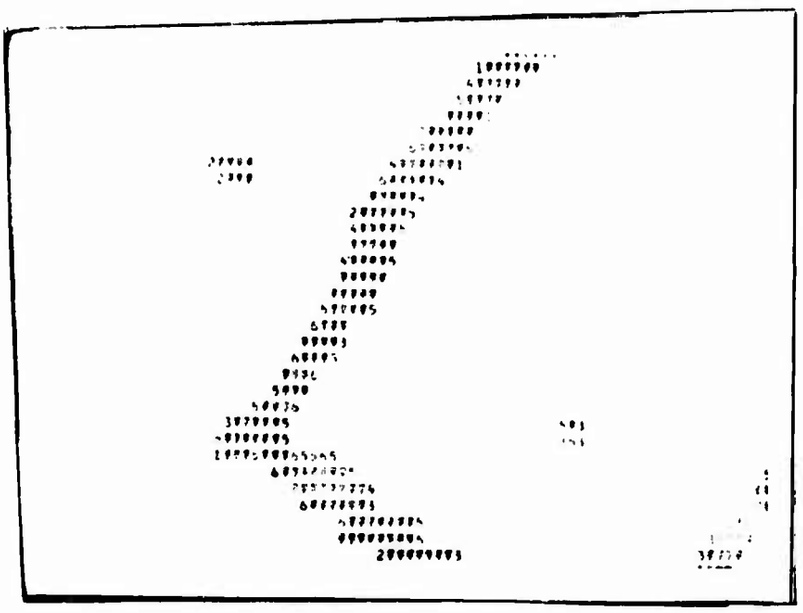


Figure 1

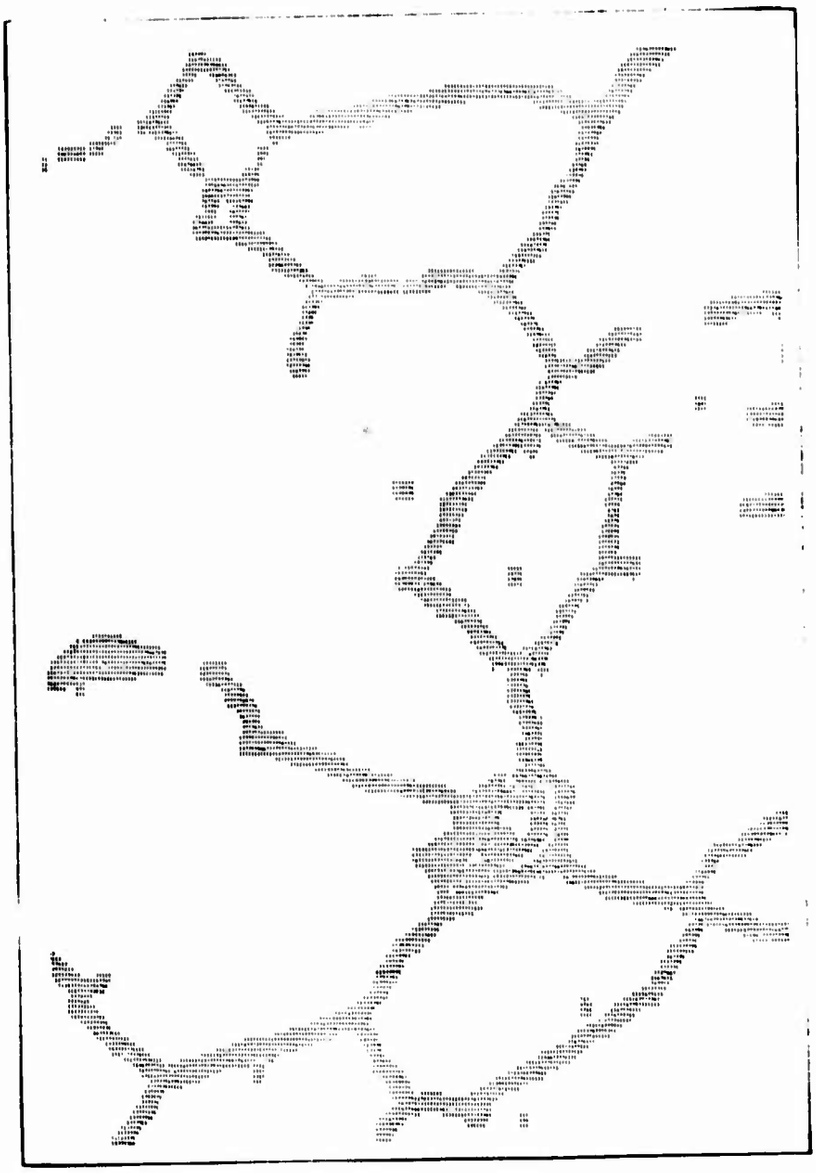


Figure 20

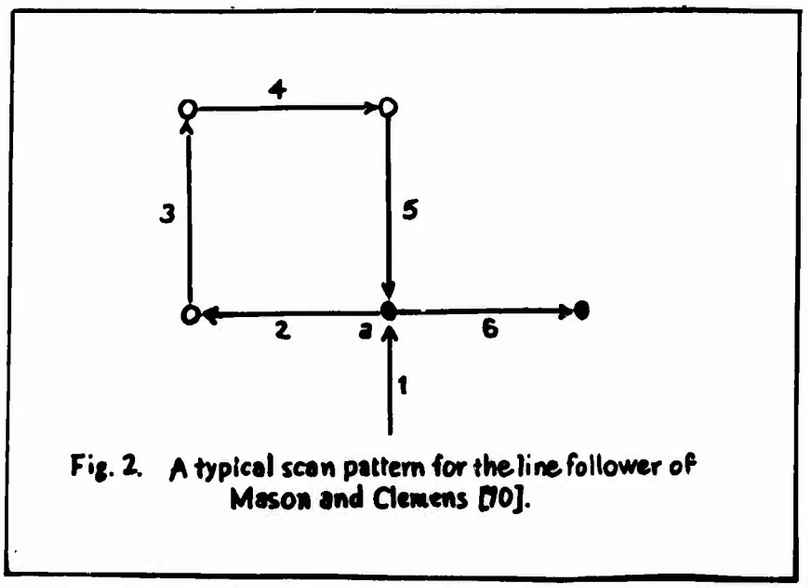


Fig. 2. A typical scan pattern for the line follower of Mason and Clemens [70].

Figure 21

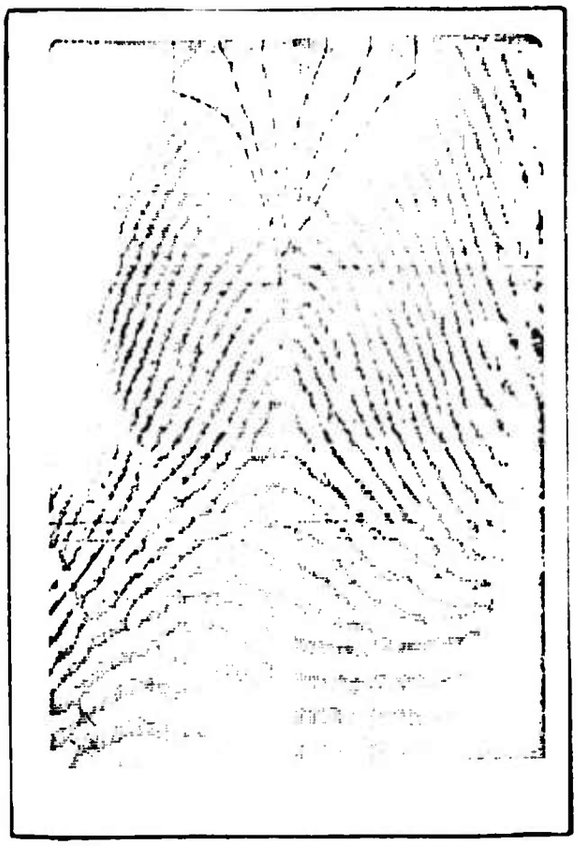


Figure 22

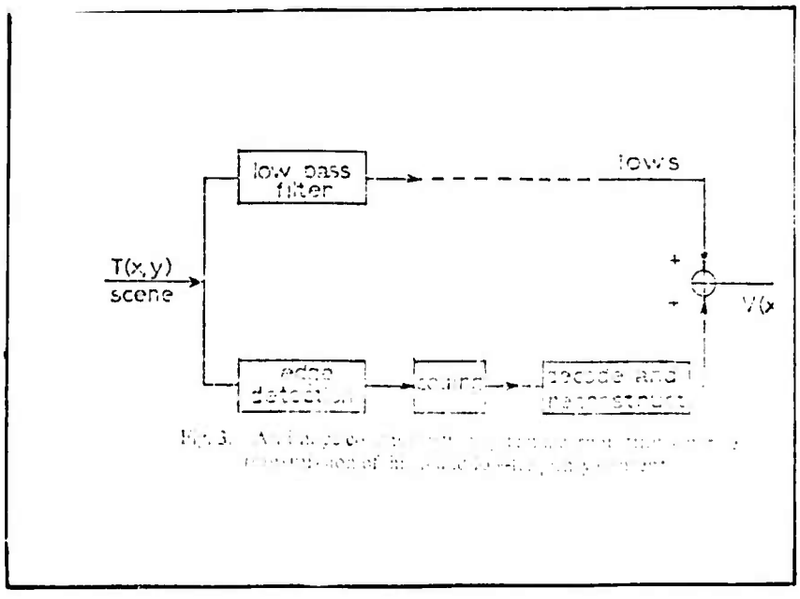


Figure 23

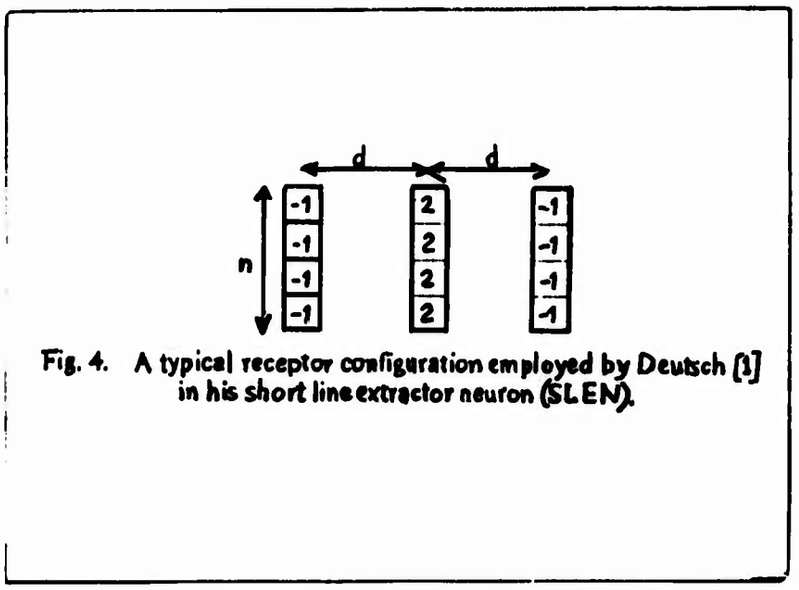


Fig. 4. A typical receptor configuration employed by Deutsch [1] in his short line extractor neuron (SLEN).

Figure 24

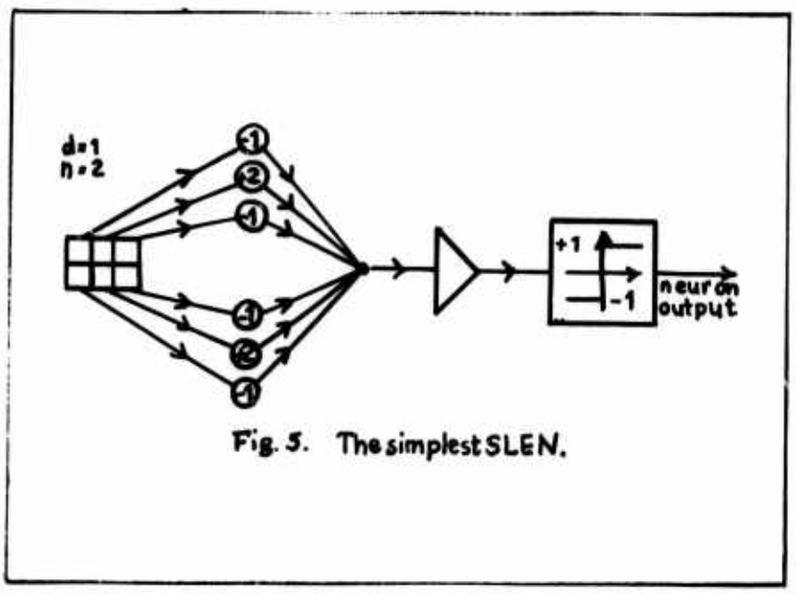


Fig. 5. The simplest SLEN.

Figure 25

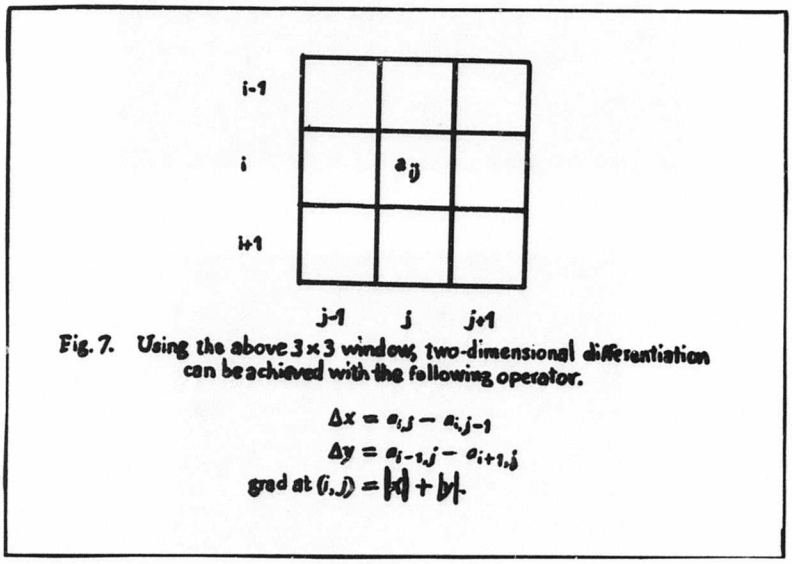


Fig. 7. Using the above 3×3 window, two-dimensional differentiation can be achieved with the following operator.

$$\Delta x = a_{i,j} - a_{i,j-1}$$

$$\Delta y = a_{i-1,j} - a_{i+1,j}$$

$$\text{grad at } (i,j) = |x| + |y|.$$

Figure 26

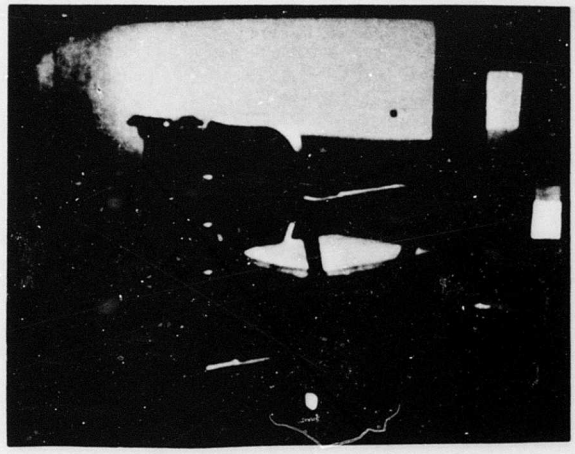


Figure 27

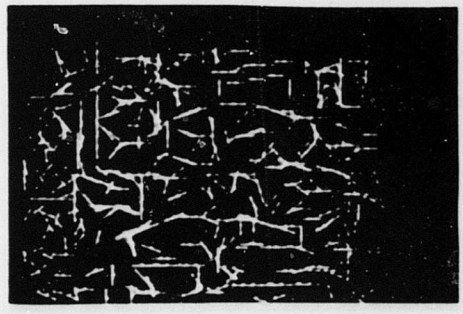


Figure 28

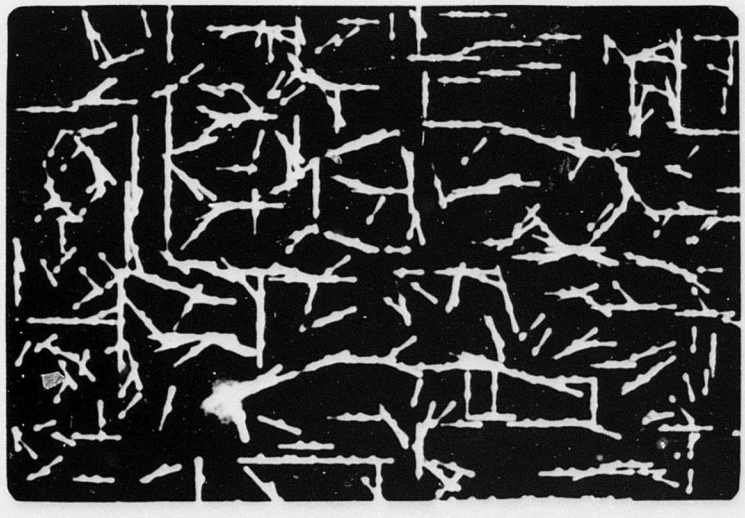


Figure 29

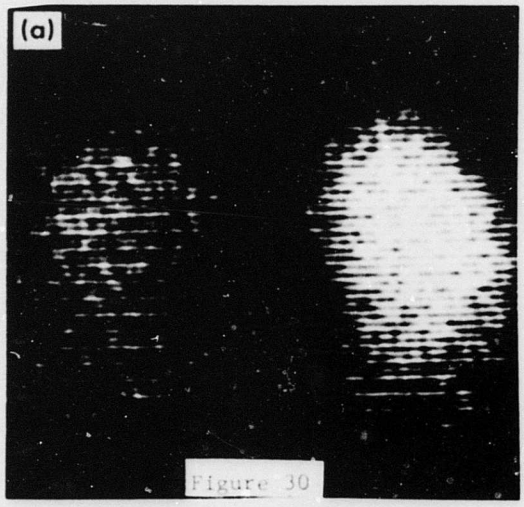


Figure 30

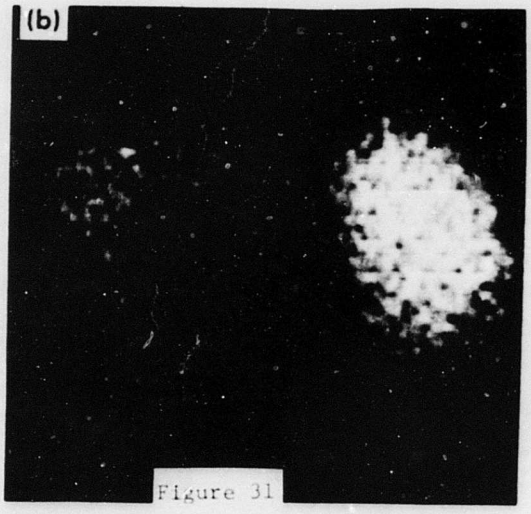


Figure 31

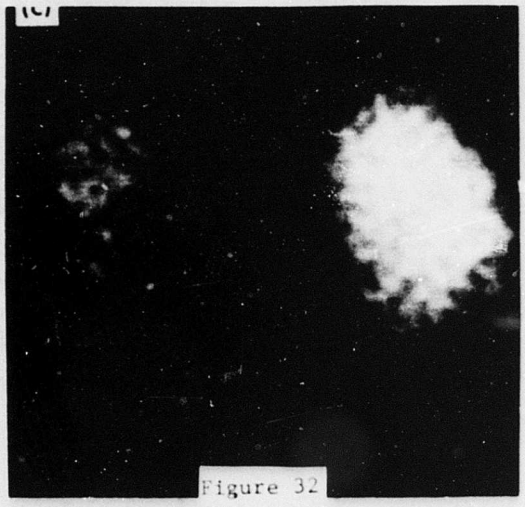


Figure 32

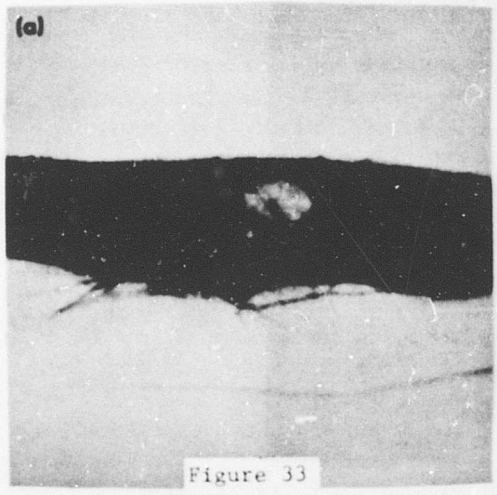


Figure 33

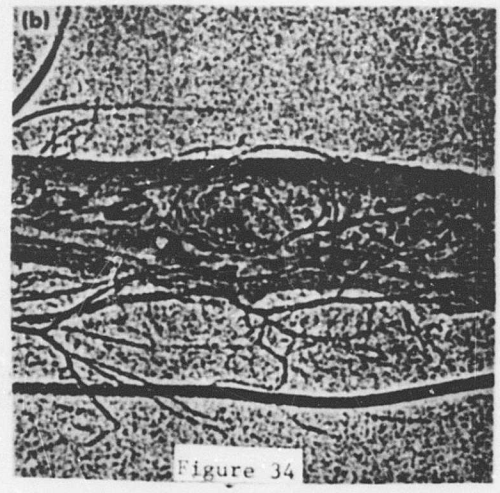


Figure 34

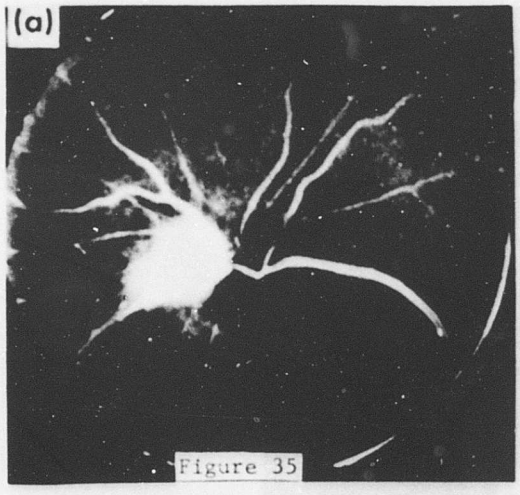


Figure 35

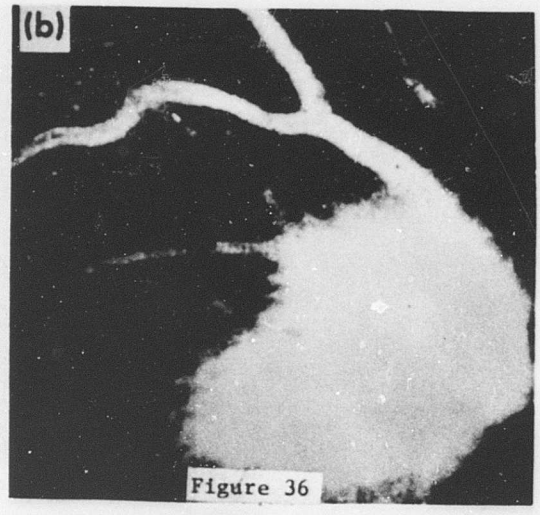


Figure 36

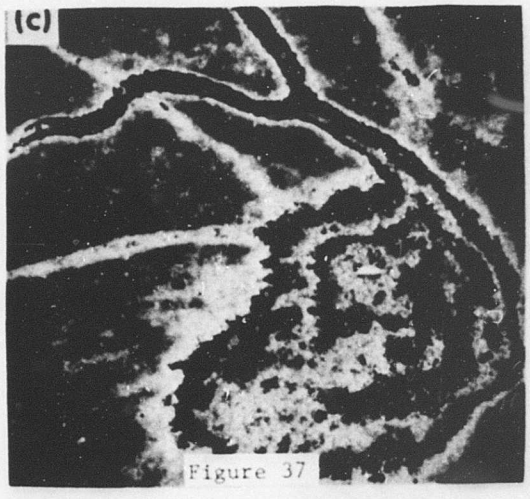


Figure 37

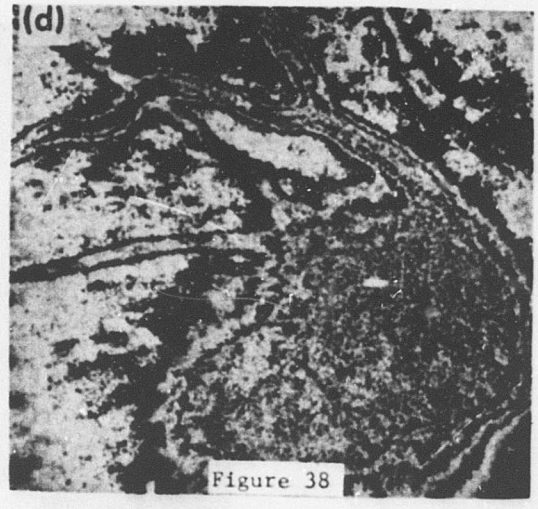


Figure 38

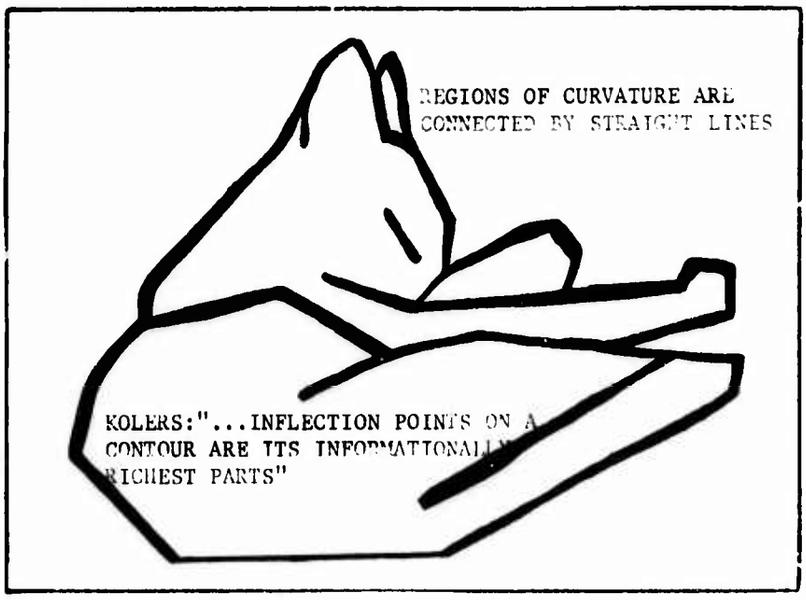

REGIONS OF CURVATURE ARE
CONNECTED BY STRAIGHT LINES
KOLERS:"...INFLECTION POINTS ON A
CONTOUR ARE ITS INFORMATIONALLY
RICHEST PARTS"

Figure 39

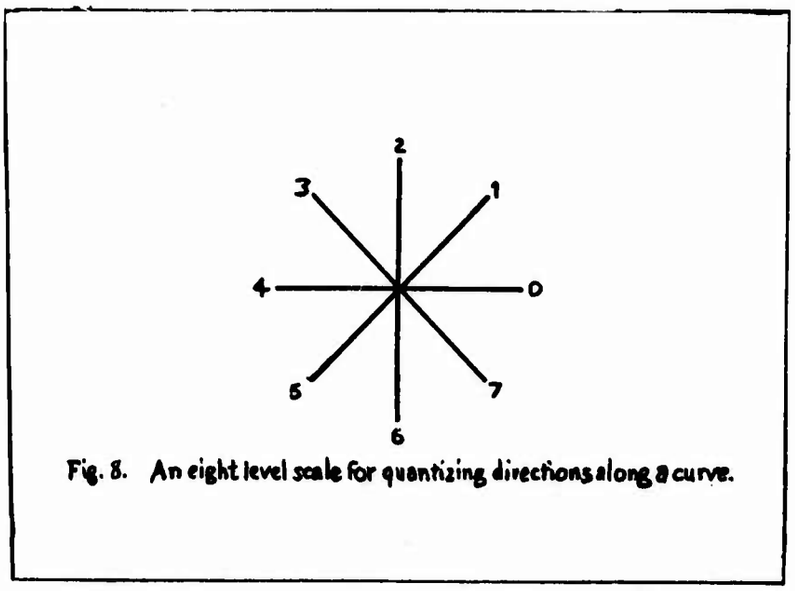


Fig. 8. An eight level scale for quantizing directions along a curve.

Figure 40

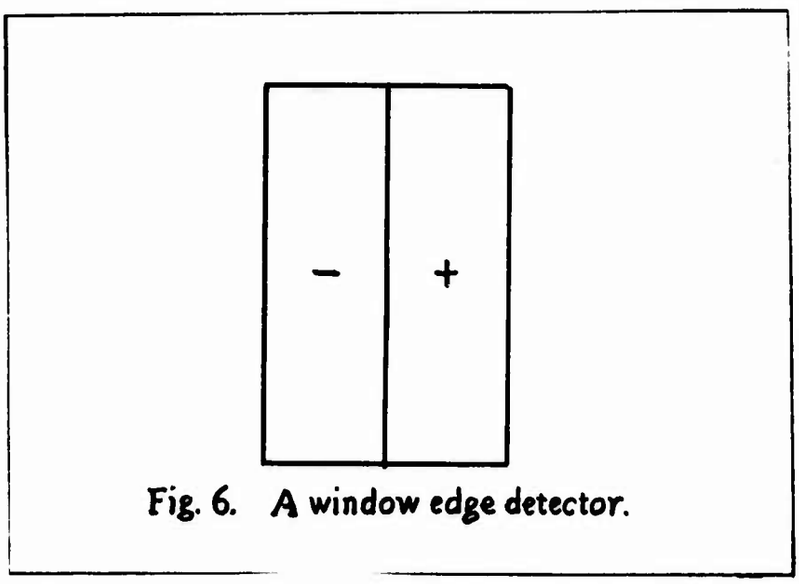


Fig. 6. A window edge detector.

Figure 41

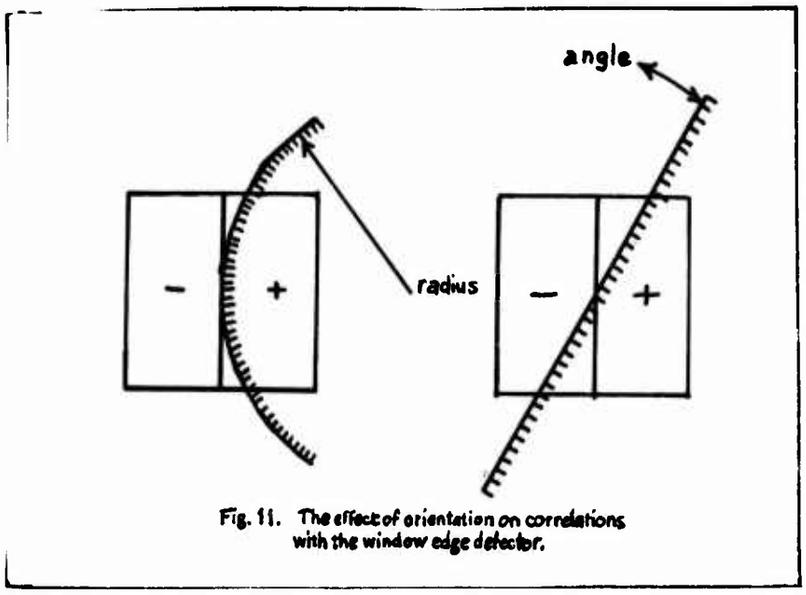


Fig. 11. The effect of orientation on correlations with the window edge detector.

Figure 42

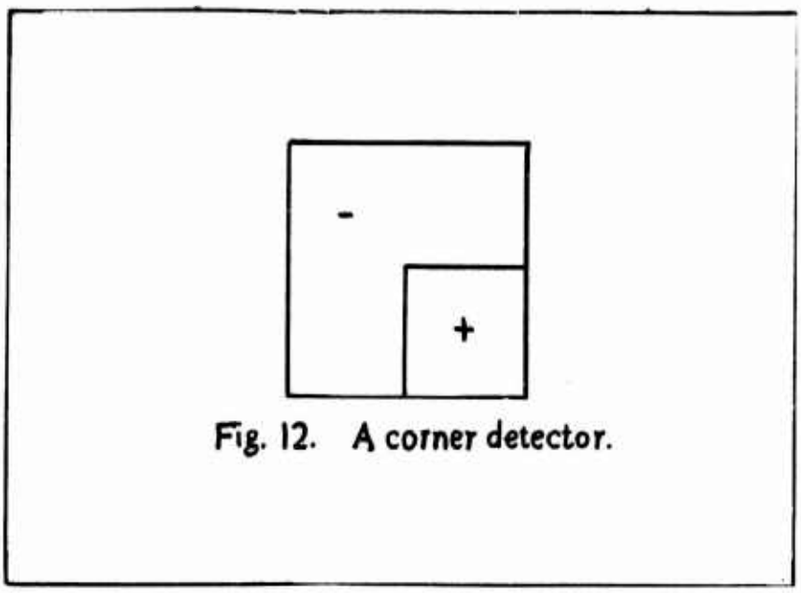


Fig. 12. A corner detector.

Figure 43

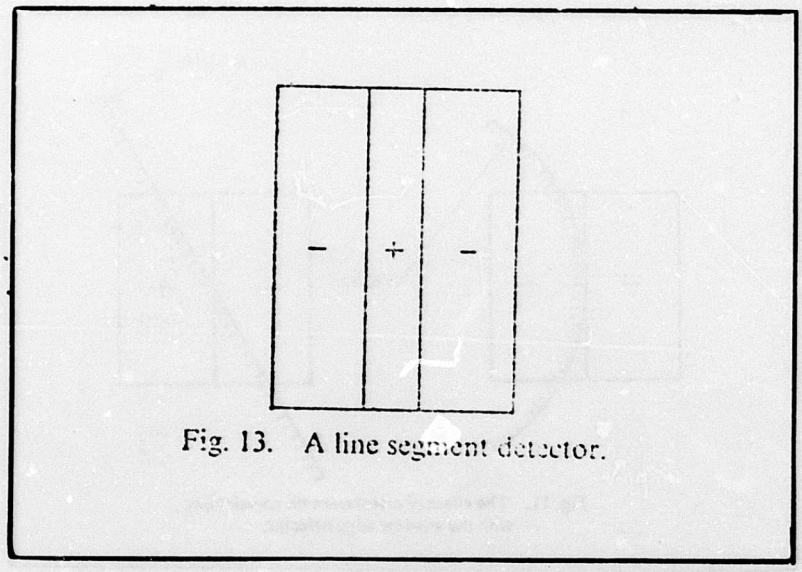


Fig. 13. A line segment detector.

Figure 44

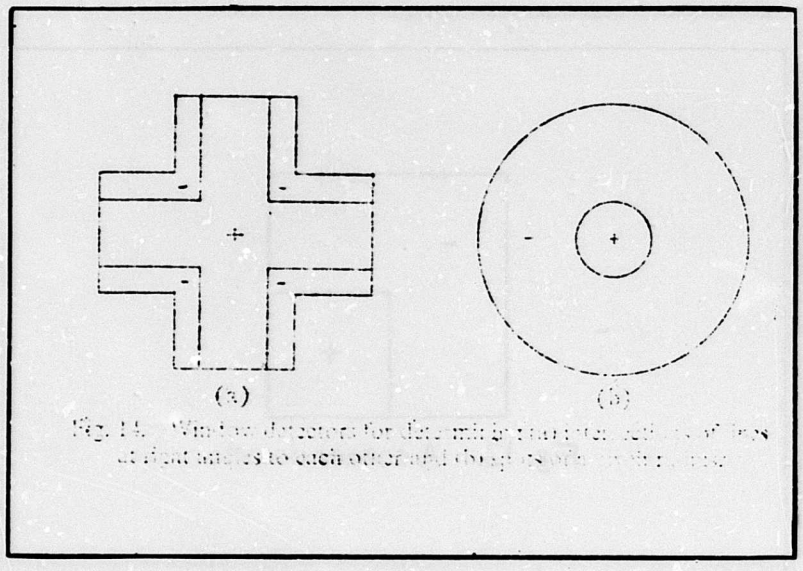
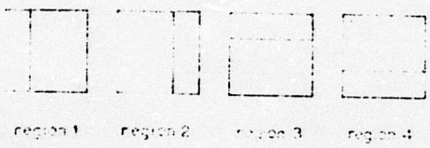


Fig. 14. Window detectors for detecting certain features of lines at fixed angles to each other and of specified lengths.

Figure 45



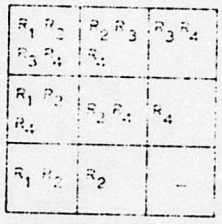


Fig. 15. Regions used to construct a property vector for the recognition in real time of handwritten characters [115].

Figure 46

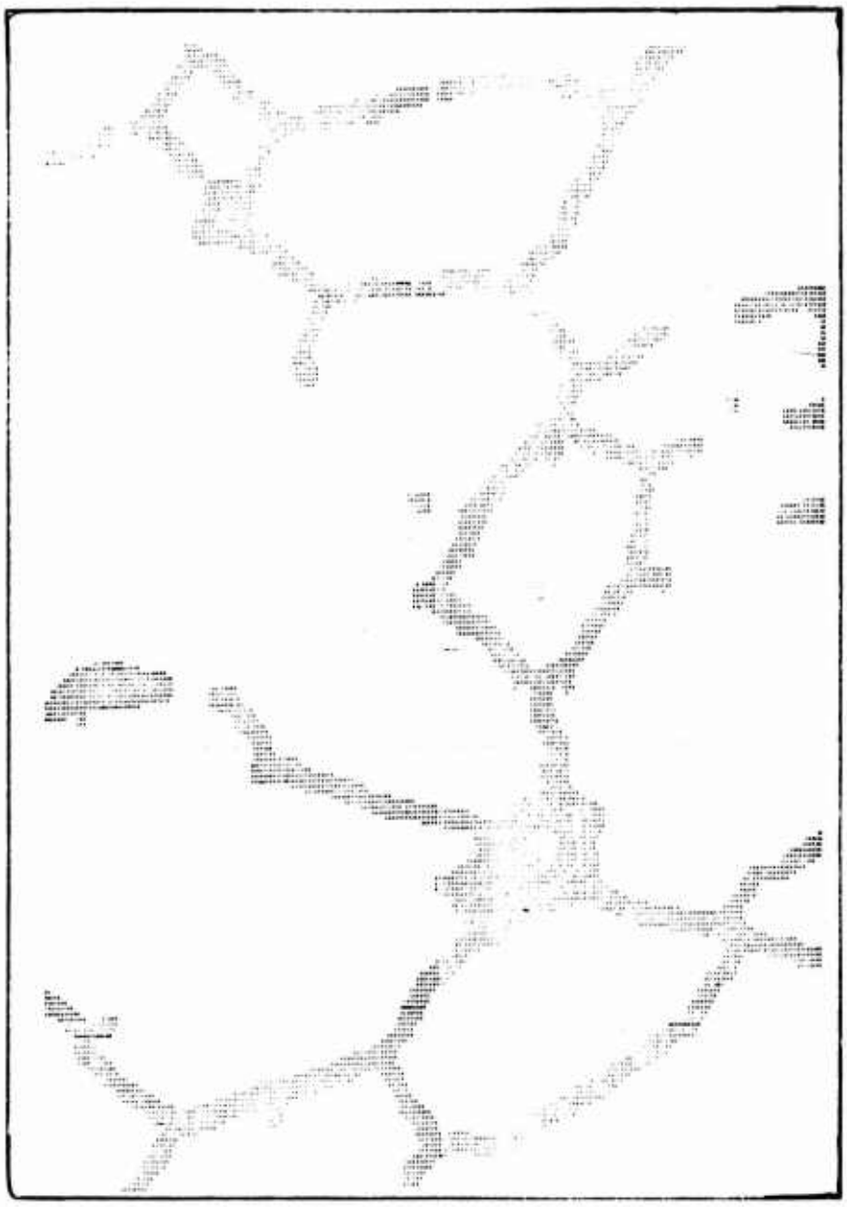


Figure 47

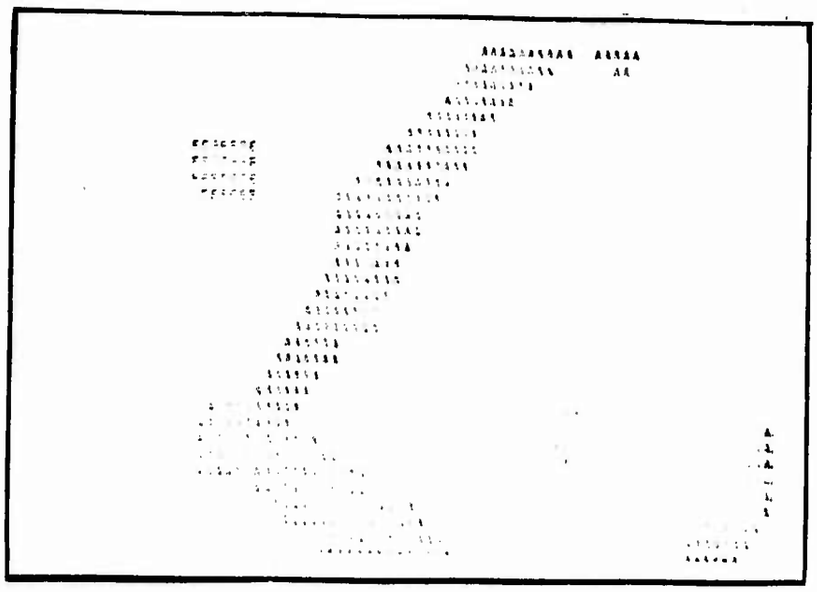


Figure 48

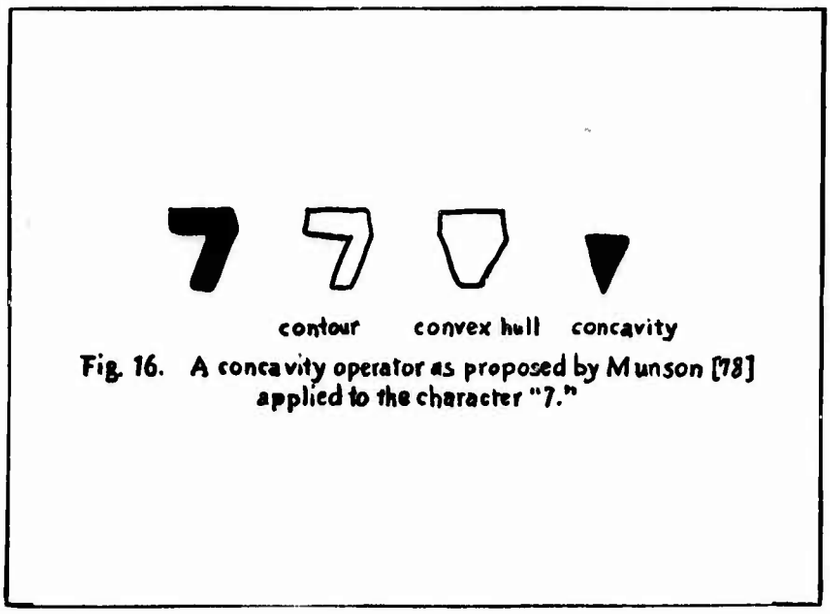


Fig. 16. A concavity operator as proposed by Munson [78] applied to the character "7."

Figure 49

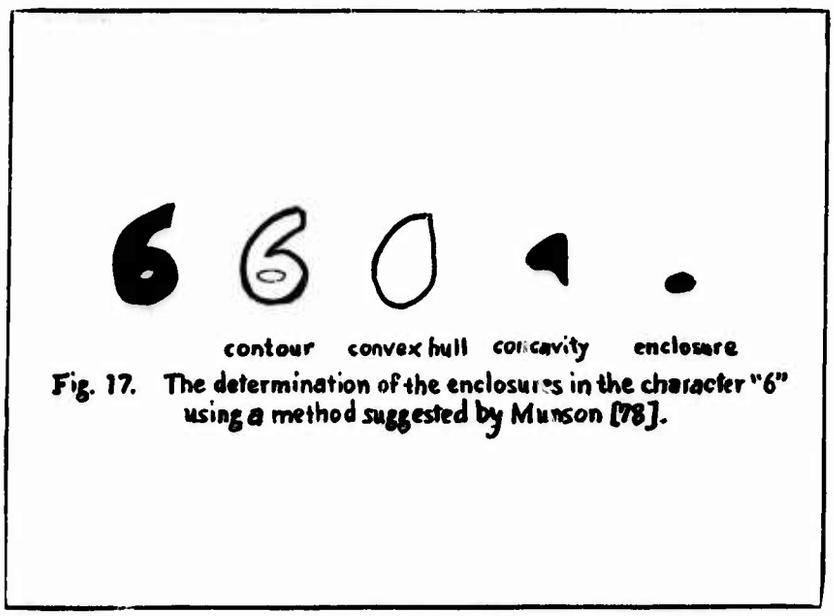


Fig. 17. The determination of the enclosures in the character "6" using a method suggested by Munson [78].

Figure 50

Figure 51

Figure 52

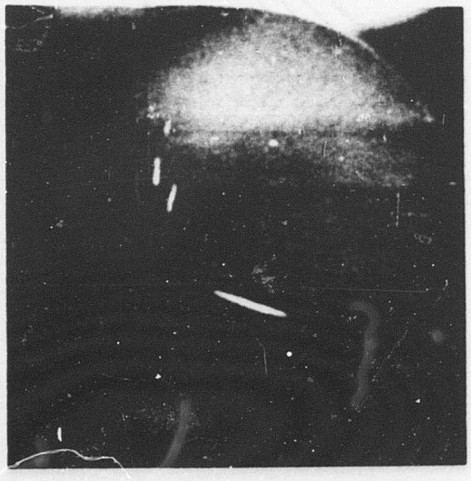


Figure 53

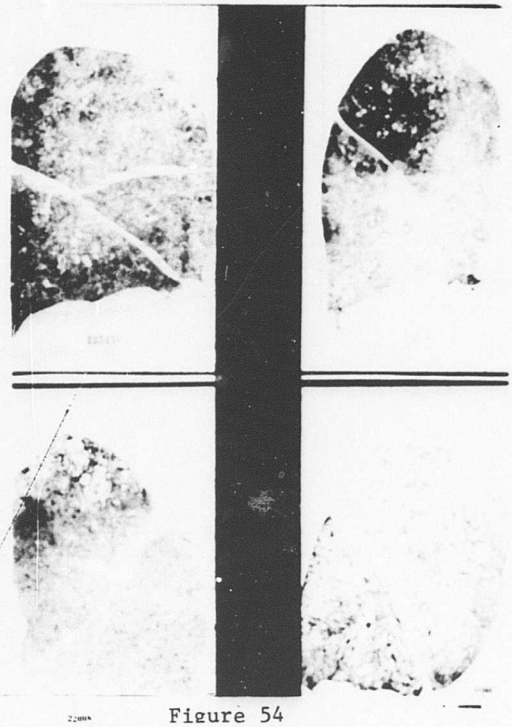


Figure 54

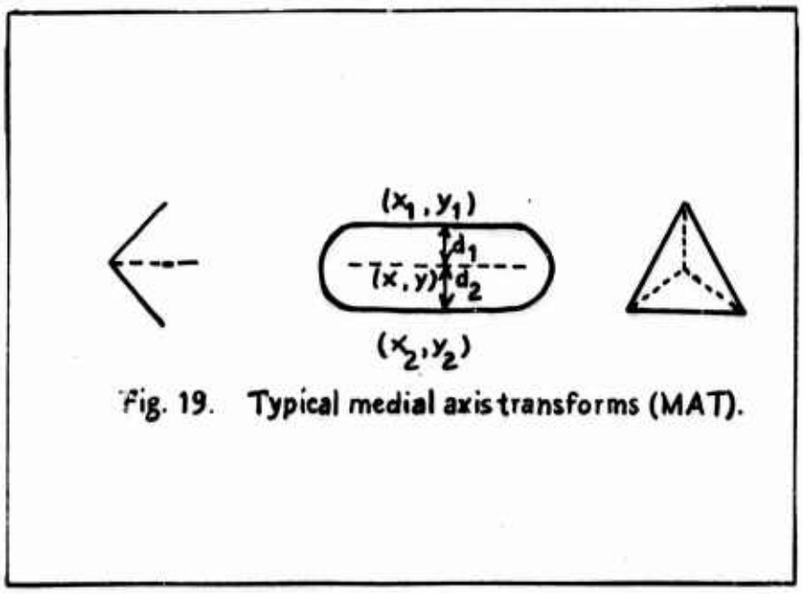


Fig. 19. Typical medial axis transforms (MAT).

Figure 55

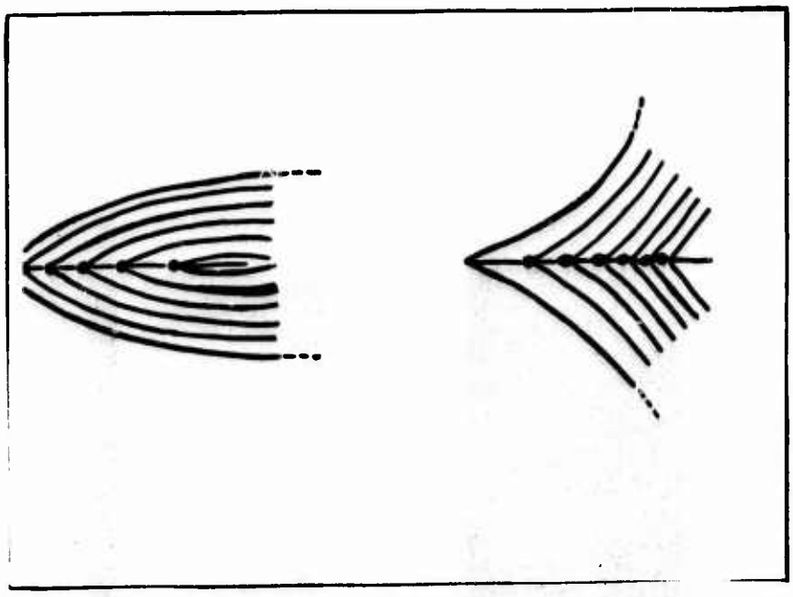


Figure 56

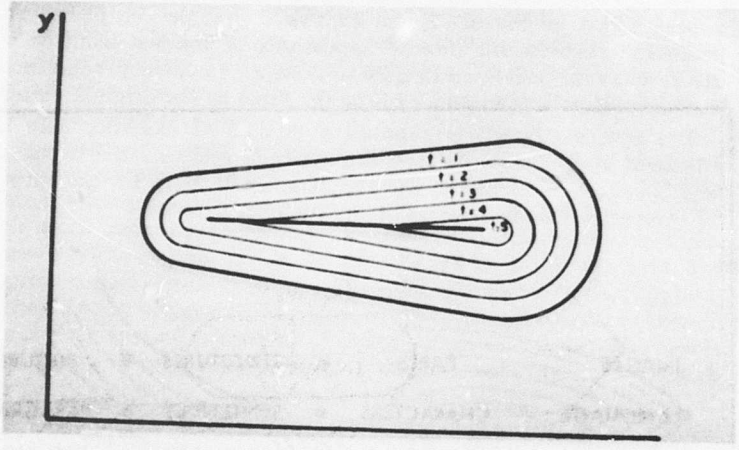


Figure 57

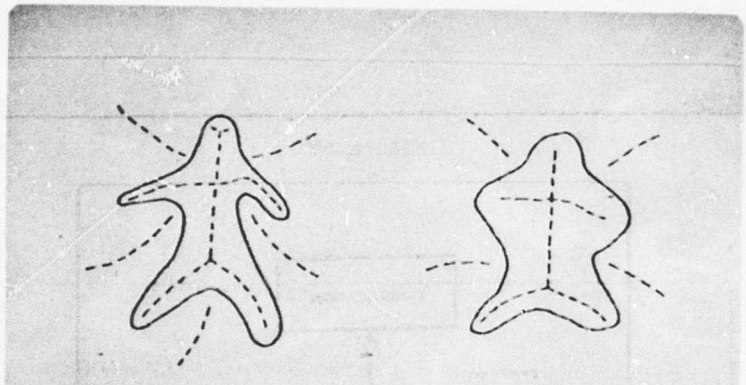


Fig. 20. The MAT for both a simple sketch and a distorted version of a man.

Figure 58

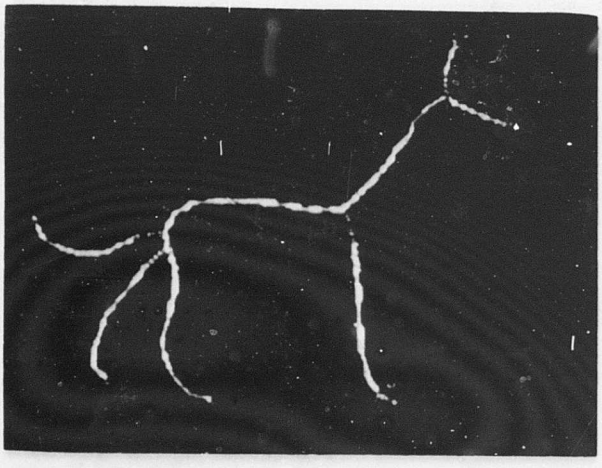


Figure 59

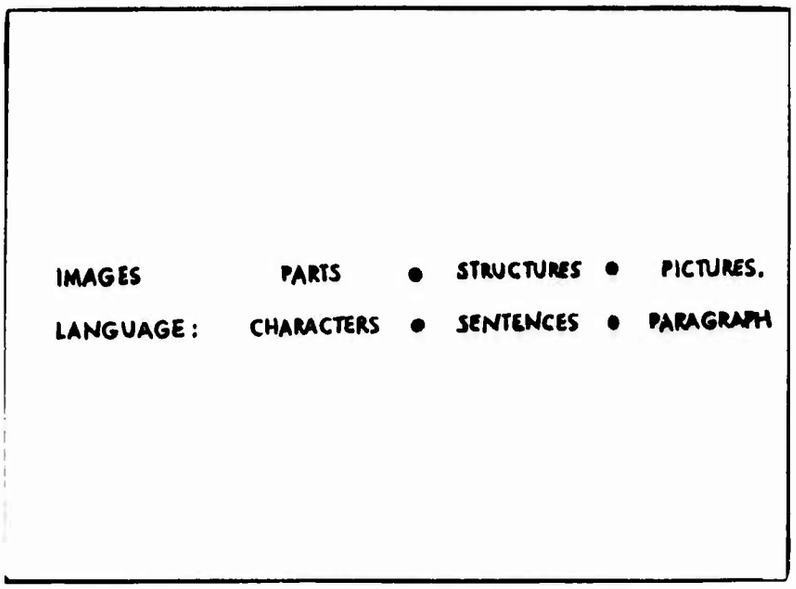


Figure 60

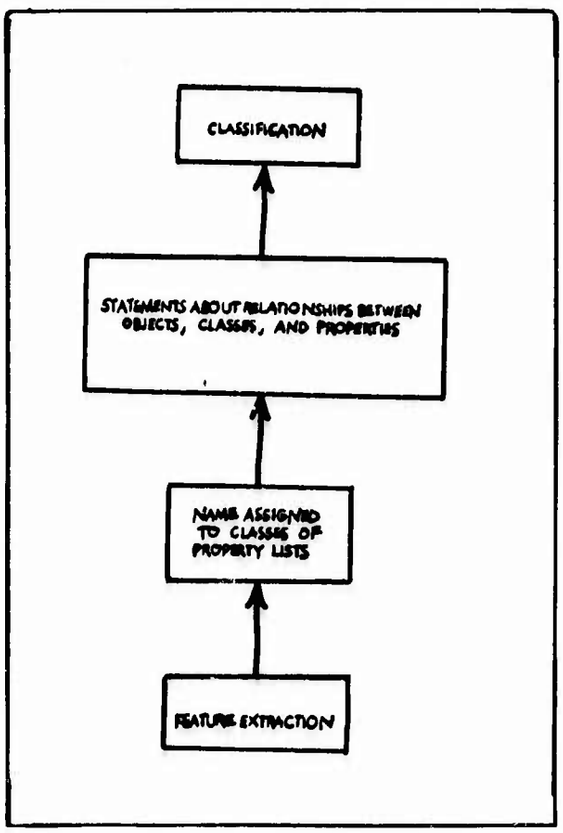


Figure 61

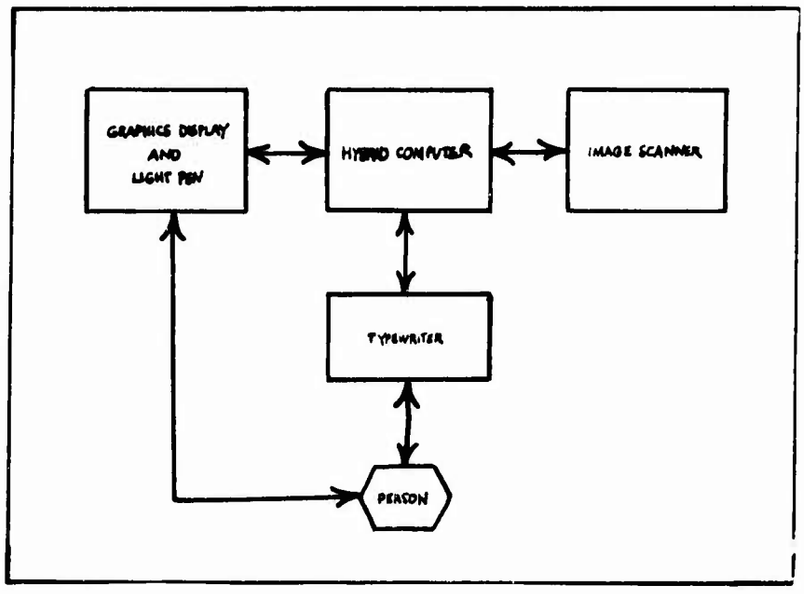


Figure 62

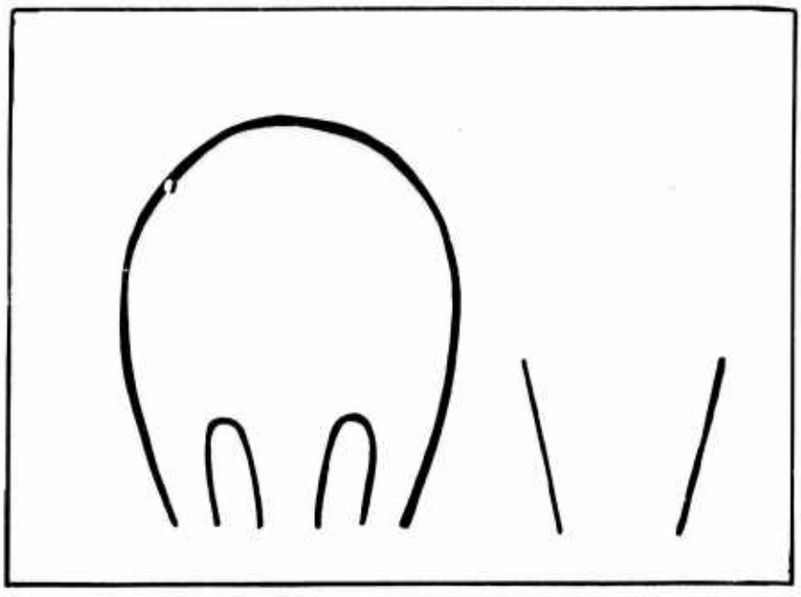


Figure 63

BLANK PAGE

DIGITAL IMAGE PROCESSING FOR TERRAIN PATTERN RECOGNITION

Lawrence P. Murphy

Geographic Information Systems Branch, Geographic Sciences Division,
U. S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia

ABSTRACT. The U. S. Army Engineer Topographic Laboratories have evaluated and tested a prototype digital image processing system for performing pattern recognition experiments. Tests and procedures are described for evaluating the Natural Image Computer (NIC) with its systems software, in addition to the conclusions reached from the analysis. The prototype system was designed as an exploratory laboratory device for pattern recognition studies and limited feature delineation capability using vertical aerial photography as an input. The structure of the recognition algorithms is based upon recognition and correlation to basic feature shapes and statistical characteristics of the grey scale distribution of Military Geographic Intelligence. The NIC is a versatile laboratory device capable of accepting cut film photographs and producing grey scale digitization of the imagery at 16, 32 and 64 grey scale levels.

INTRODUCTION. The mapping and intelligence community has developed photographic collection systems that acquire data at a rate far in excess of man's ability to utilize or analyze. However, to date little progress has been made in automating the processes of detection, recognition, extraction, and symbolization of geographic intelligence and mapping information contained on vertical aerial photography.

In considering the eventual backlog of mapping and intelligence photography, the U. S. Army Engineer Topographic Laboratories (USAETL) of the U. S. Army Topographic Command (TOPOCOM) initiated a pattern recognition research program for mapping and military intelligence in 1963. To implement this pattern recognition research, USAETL contracted with the Aeronutronics Division of Philco-Ford in late 1964 to design, fabricate, test and produce a system study and analysis for a prototype pattern recognition system, the Natural Image Computer (NIC). This contract included contractor testing of the system and development of software to identify and extract orchards, woods, lakes, oil tanks and railroad yards. The prototype NIC system was delivered to USAETL in mid-1967. In-house testing and evaluation of the NIC concluded in mid-1969 with the completion of an evaluation report. Though in-house work is continuing with the system, this paper will describe the NIC system hardware and software and the tests and results developed during the evaluation period at USAETL.

SYSTEMS DESCRIPTION. A block diagram of the NIC system with systems flow is shown in Figure 1. The NIC is a modular system consisting of PDP-7 Digital Equipment Corporation computer, paper tape and small dual magnetic tape units, teletypewriter, image processor, monitor scope and plotter.

Nine by nine inch film transparencies are mounted on the X-Y film table which moves in increments of .005 inches under computer control. The flying spot scanner (CRT) produces a spot of .001 inch at the film plane, which can

be positioned to anyone of 256 by 256 points. When the film table is in a fixed position, the scanner field of view is therefore approximately 1/4 inch square at the film. Film density values are digitzed at 16, 32 and 64 grey level steps for 0.00 to 2.00 diffuse density. The NIC controller is unable to store this large an array (256 x 256) of multi-level grey level information. Consequently, a field of 64 x 64 points, or one sixteenth of the 256 x 256 array, is stored in the controller.

During the film scanning operation, the processor can be instructed to perform a linear-threshold operation at a given X-Y location within the scanner field of view.

This operation is defined as

$$D = 1, \; S \geq \theta$$

$$D = 0, \; S < \theta$$

$$S = \sum_{\Delta x, \Delta y} C_{\Delta x, \Delta y} \, I_{\Delta x, \Delta y}$$

Where D = binary decision

S = pre-decision sum

θ = threshold value

$C_{\Delta x, \Delta y}$ = weighting function

$I_{\Delta x, \Delta y}$ = image values

The processor performs this operation by deflecting the scanner beam $\Delta x, \Delta y$, and multiplying the results $I_{\Delta x, \Delta y}$ by the corresponding stored weighting function value $C_{\Delta x, \Delta y}$.

The weighting function, a two-dimensional array of numbers, is designed to enhance selected local image shapes. Typical weighting functions "masks", designed by the contractor are shown in Figure 2. All of the masks are designed to produce S = 0 when applied to a field of uniform grey level imagery. A dot in each mask indicates where the mask sum is accumulated. A generalized masking routine is available for the operator to quickly develop new mask designs

The sum S can be generated by using either the 0-63 image values (6 bits) or a binary version of the image value. The binary value of $I_{\Delta x, \Delta y}$ is developed by reading the film and comparing the results to a "clipping level" from 0 through 63 when processing in the 64 grey level digitization mode. If $I_{\Delta x, \Delta y}$ exceeds or equals the clipping level, the grey level code value is replaced with a 1 in forming S; otherwise it is replaced with 0. The main purpose of this process is to emphasize transitions in grey level imagery and to provide enhancement of the resulting data.

Up to 1,024 mask points may be stored in the processor representing any mask points within that limit. The processor may be directed to output the binary results of the mask application to the plotter, the monitor CRT or transmit the results to the controller. Once the processor has received and stored the mask data and instructions from the controller it performs the

operation independently of the computer. The processor can expand the mask by factors of 2 and 4, and coarsely rotate the mask about the central point of application.

FUNCTIONAL TESTING. Three types of recognition processes and series of linear feature tracking tests were performed in evaluating the prototype NIC. The recognition tests were based on (1) feature recognition by analysis of image grey scale statistics, (2) orchard recognition by masking techniques and grey scale analysis, and (3) railroad yard recognition by masking techniques only.

The recognition testing of woods, lakes, orchards, railroad yards, urban areas, rivers and bays was based solely on the imagery grey scale spatial distribution statistics. Sample areas of each test feature were selected at a nominal photo scale of 1:50,000. A total of 576 test images were used. The statistical (STA) program was used to develop grey scale statistics for each of the image types by field scan which is a 64 x 64 point array at 64 grey levels. Twenty-two types of grey scale distribution statistics were developed for each image tested. An example of computer printout (Figure 3) is identified as follows:

a. Mean-average grey level of the 64 x 64 field (4096 points)

b. Variance (σ^2) - the averaged square deviation from the mean grey level.

c. Distribution - the distribution of the 4096 points in the coded grey levels.

d. Total texture 1, 2 - a measure of texture of the field. A count (summation) of the first and second nearest neighbors that are a grey level from the given point.

e. and f. Difference No. 1 and No. 2 - these are detailed lists of the number of first and second neighbor differing grey level that were found in the texture calculations.

g. Run length X and Y - continuous tone column or row grey level runs in the digitized grey level 64 x 64 array.

From the statistical program output data, twenty-two sets of limit values were extracted from each of the images processed. The upper and lower limit (range) of each grey scale distribution parameter was determined for each image type. A sample of the tabulated data is shown in Table I. This figure illustrates the properties that were extracted to form simple decision criteria limits for determining whether a feature existed in the test imagery. As an example, in order for a feature to qualify as a wooded area it must have a mean grey level of 30.48 to 42.87, a variance about the mean of 0.18 to 6.21 and so on, as extracted from statistical data derived from the set of test imagery. If all of the 22 statistical measures for any image sample met these simple decision criteria, it was classified as that image type defined by the decision criteria limits.

The next slide (Table II) shows the resultant false alarms (false identification) produced when using this simple decision criterion for each test image type. Column two of the table shows the false alarm rates produced by accepting the wooded area (120 images) statistical range for a zero woods error miss decision rule. In reading this column of the table, 19 lakes of the 48 tested provided grey scale statistical properties which fell within the range of the wooded area decision rule. Consequently, these 19 lake images were falsely "identified" as woods. The remainder of the data for column two can be interpreted in like manner. If all of the non-wood test samples were combined into a class called "other", 26 other images from the test set of 456 images would have been falsely classified as woods. Consequently, a false alarm rate of 5 percent would result when processing for wooded areas.

The same testing technique was performed for each of the features listed in column one of the table. The range of the 22 statistical parameters for each test image listed in column one was used to develop a zero error (miss) decision rule for that image type. The results show that high false alarm rates occur between many of the tested features.

ORCHARD RECOGNITION BY MASKING TECHNIQUES AND ANALYSIS OF GREY SCALE STATISTICS. The second series of tests performed on the NIC were in the use of an algorithm as a combination of iterative masking techniques and statistical calculations. It is designed to detect the regular spacing of trees in an orchard, shrink each tree to a single point and determine the mean and variance of the sets of distances from every point to its nearest neighbor. The decision criteria are based on the idea that orchards exhibit uniformity in the distance of each point to its nearest neighbor.

Orchard feature extraction routines were performed on 96 fields of known orchards and 186 fields of non-orchard imagery as selected from photography of 1:50,000 scale. Data from 15 fields of known orchards were used to develop the ETL orchard recognition decision criteria. The decision criteria as recommended by the contractor were also tested. The next illustration (Figure 4) shows a 64 grey level printout of an orchard field with the most representative grey levels outlined for each orchard tree. This is a sample of the densitometric information that is analyzed by means of the Orchard Algorithm. In the first step of the process the NIC determine which grey level coding in the grey level imagery array accounts for 25 percent of the 4096 array points. This grey level is then grey level clipped, assigning ones to the tree-sized clusters and zeros to the non-qualifying points. The resultant binary array is then further processed by the orchard algorithm. A printout resulting from this process is shown in Figure 5. The elements in this printout are identified as follows:

a. Orchard Algorithm - The four columns of 3-digit numbers list the "closest distance" for each point on the processed image. The clusters of ones in the orchard binary imagery array are successively shrank by a series of group detection and shrinking masks until the center of each tree sized object is located. The closest distance is the resolution distance from a point to its nearest neighbor.

b. Mean, Variance, Number - The mean distance (MD) between tree centers, the distance variance from the mean, and the number of points (NP) in the orchard algorithm results.

c. SR - A "shift right" is performed using the unmasked original clipped binary data array. Each value is the number of points resulting after each one resolution element right shift and a logical "and" operation. Twenty-five one resolution element shifts are performed.

d. SD - Shift down is the same as "shift right" except the matrix shift is downward (90^{o}).

e. 4D - A diagonally downward shift at -45^{o} through 20 one-resolution elements with a logical "and" operation after each shift and a point count.

f. 4U - A diagonally upward shift at 45^{o} processed as in 4D above.

The point counts from the shifting routines result in high and low values referred to as peaks and valleys in the illustration. The next slide (Figure 6) shows a plot of these data from the orchard algorithm point out. This graph shows the periodicities that may be expected by shifting and logically "anding" a clipped image, element at a time over the original clipped image. This result occurs only with very regular spacing of discrete features in rows and columns.

From the orchard algorithm, five values were extracted and two values calculated for use in forming the orchard decision criteria limits. The parameters used and decision logic are shown in the next illustration (Figure 7). The decision parameters used are identified as follows:

1. Mean distance between trees.
2. Distance variance (σ^2) about the mean.
3. Number of points (trees).
4. Ratio of Mean Distance to Variance Calculations.
5. Number of peaks.
6. Number of valleys.
7. Total point count difference (correlation difference) between peaks and proceeding valleys in the shifting routine.

Before a field is identified as an orchard, all seven of the extracted statistical values must fall within the orchard Decision Criteria Limits set for 1:50,000 scale photography. Two sets of Orchard Decision Criteria Limits

were used in the testing, the decision criteria limits determined by contractor and a modification of these limits developed by ETL. The ETL modification was based on a sample set of 15 test images where the upper limit of tree spacing criteria was raised to approximately 50 feet at ground scale, the variance criteria lower limit was reduced and the correlation difference lower limit was increased. Ninety-six known orchard images were processed by the orchard algorithm. The data from this processing were subjected to the contractor's recognition criteria and the ETL modified decision criteria limits. Table III shows a summary of the results of this analysis. Using the contractor's recognition criteria, 70 images were correctly identified from the test set. Twenty-six (26) images or 27 percent were falsely rejected (missed). Eighty-four images were correctly identified using the ETL modified decision criteria and, thus, produced a false reject or miss rate of 12.5 percent.

One hundred and eighty-six (186) non-orchard images were subjected to the orchard algorithm and decision processing to determine the approximate false alarm rate for orchards in non-orchard areas. Table IV show the results of this processing where 12 different map feature images are identified in the test imagery. The major part of the imagery falls in the southwest area of the U. S. as was the imagery used in developing the recognition decision parameters. Thirty-eight images were falsely identified as orchards using the contractor's recognition criteria limits producing a false alarm rate of 20.4 percent. Five test images were falsely identified as orchards using ETL modified recognition criteria producing a false alarm rate of 2.7 percent.

RAILROAD YARD RECOGNITION BY MASKING TECHNIQUES. The railroad yard recognition program is based on line and edge mask application to detect the parallel striations appearing in the imagery. Although not reliably resolved by the NIC scanner, this appearance is caused by the tracks in the railroad yard image. Masks shown in Figure 2 of the type represented by "a" through "g" are used to detect these thin parallel streaks of density differences. The masks are rotated to each of eight orientations separated by approximately 22 degrees. These masks are applied over a 128 x 128 element area to detect edges and lines. Distributions of the mask sum values are accumulated and then printed on the teletypewriter. The printout specifies the angle of application for the line or edge mask and the mask sum values. A typical example of a plot of these quantities for a typical railroad yard is shown in Figure 8. The number of mask sum points accumulated at each angular application is shown in this figure. As expected, a high mask response is noted when the masks are applied in the general direction of the railroad yard. In this illustration, the railroad yard was mounted in the NIC scanner with the railroad yard paralleling the x scan motion.

A series of tests were performed using the railroad yard program as applied to 25 test images. Ten of the images were known railroad yards and 15 of the images were non-railroad yards. An example of plots for line and edge mask results for non-railroad yard and railroad yard features is shown in the next Figure 9. It is evident from these plots that poor data separation exists at the output. The line mask sums do not produce significantly different plots. The edge mask sums produce greater separation in

the plots of the data. The most distinguishing event for railroad yards in either plot is the occurrence of a single high peak mask sum value. Based on this observation and other angular relationships, the contractor devised a railroad yard decision logic which ETL used to analyze the results of the railroad yard program process data.

The decision criteria used in the analysis of the masking data to determine the existence or non-existence of a railroad yard is based on the number of edge and line peaks, angular relationship of the peaks, and first and second peak ratio values. Twenty-five test images were processed with the railroad yard program and the resultant data were subjected to the railroad yard decision criteria. The results of this processing are shown in Table V. This type of processing and logic performs badly as evidenced by the miss rate of 20 percent, false alarm rate of 40 percent and the combined error rate of 32 percent where all 25 test images were used.

FEATURE TRACKING. A linear feature tracking program was developed to investigate the procedures and operational problems associated with the extraction and delineation of mapping and military geographic intelligence features. The tracking program was designed to detect the edges of linear features and output the results on the plotter and teletypewriter. The program operates mainly in two modes, prediction and search. In the first mode, the system is given the X-Y location of the linear feature edge to be delineated and instructed to predict the next probable location of the edge. In the second mode, if the feature edge is not detected, what search pattern should be entered to locate the edge? The actual program consists of four major sections as follows:

a. Tracking Routine - Predicts the most probable next position of an edge based on the direction of the present edge point and the last direction moved in arriving at the present point.

b. Test Routine - Applied the appropriate mask as predicted in "a" above and signals success or failure.

c. Search Routine - When the predicted point fails, jumps to exploratory routines and searches the immediate area.

d. Operator Assistance Routine - Permits the operator to initialize the program and assists during the operation of the program. Allows display of the area, mask scale factor change, printout of 64 x 64 grey level area over the last good point, move scanner table, apply any mask and print result, reenter starting parameters.

The tracking of program basically entails the application of a long (approximately 10 resolution elements in length) edge detection mask (see Figure 10). Sixteen masks representing different angles of applications are used. To operate the program, the operator specifies, through the teletypewriter, the grey level operation mode, size of the area to be processed,

center X and Y coordinates of the initial search area, initial search direction, mask scale factor, threshold response. Additionally, the teletypewriter prints the X and Y table movements, and "X" when in the search routine, or the word "HELP", along with the current mask number, each time the program needs help.

Testing of the tracking program consisted of tracking roads on two sets of photographs at 16, 32 and 64 grey levels. The results of test runs at 16 and 64 grey levels on the first photograph were unacceptable. Using the same photograph, the tracking program ran at 32 grey levels but, on its best run, required operator assistance twelve times.

A second photograph (Figure 11) was used for tracking tests. This photograph is twice the scale of the first test photograph or approximately 1:25,000 scale. At first, this second photograph could not be road tracked at any grey level coding. The photograph was subjected to two stages of photographic enhancement using orthochromatic emulsions in photo lab processing. Figure 12 shows the tracking run results using the enhanced photograph. Although this road would not track satisfactorily at 16 grey level input coding, the tracking programs perform well at 32 and 64 grey levels. At 32 grey levels of input, eight breaks occurred. At 64 grey levels of input, six breaks occurred. The two tests were repeated with approximately the same results. Even though the test at 32 grey levels produced more breaks, the resultant plot was smoother and appeared to be a better cartographic representation of the road.

DISCUSSION - EVALUATION OF TEST RESULTS. In evaluating the data and test results developed during the testing of the NIC, many limiting factors became apparent to performing automated analysis of aerial imagery. In the use of imagery grey scale information, the statistical outputs are not only non-gaussian but are of great range. Within the range of many of the statistical outputs gross discontinuities occur. Limited separation of features is detected by comparing the range of each statistical output. Separation of features is not detectable by comparing the peak number of occurrences in each statistical output.

The results of the grey scale spatial distribution analysis provide little supporting evidence of the separability of the major map feature classes or subclasses using the statistical techniques developed by the contractor. There is evidence that the broad classes of cultural and non-cultural features may be separated. Other parameters such as shape correlation, feature height and spectral response will be required in the recognition algorithms. A considerable effort will be required to collect controlled imagery by region, season and exposure. Other statistical properties and transformations should be sought in developing feature recognition algorithms.

A limited amount of testing was performed using 16 and 32 grey scale levels as input from the scanner for orchard algorithm processing. It was apparent that there is an increase in the orchard recognition error rate with less than 64 grey levels of input to the system. Testing performed was insufficient for determining the actual performance fall-off rates due to grey scale resolution. It should also be determined if other types of

orchard processing and analysis can be developed. The current orchard processing routine has been packaged as an automated program at ETL. This program requires approximately 2.5 minutes of processing by the NIC for each 64 x 64 element image array before it determines the existence or non-existence of an orchard.

The results of the railroad yard recognition program testing did not produce the recognition accuracies reported by the contractor. The high false alarm rates shown from the ETL testing indicate that other parameters are definitely required in the recognition algorithm. Indications are that grey scale statistical parameters will need to be added to the railroad yard recognition criteria. Modification of the program, recognition parameters and logic will be required.

The results of the tracking tests show that the present program is easily mislead by road intersections, grey scale structure edges intersections, grey scale structure edges intersecting with roads and breaks in the road image. There is a definite indication that fine grey scale information, such as presented at 64 grey levels, causes the system to be mislead when edge tracking. The tests showed that tracking at 32 grey levels will produce a smoother and more cartographic representation of the road.

During the testing, it was observed from the CRT monitor that it is possible to clip a very narrow range of grey scale values from the photograph representing the road or roads. In this case, much of the edge noise about the road is eliminated. If the tracking program were modified to accommodate clipped grey levels and apply edge or line masks to this binary image rather than to the 16, 32, or 64 grey level images, it would perform with fewer breaks in the plotter road track output.

CONCLUSIONS. Based on the in-house testing of the prototype NIC at ETL, it is concluded that the NIC is satisfactory for performing experiments in pattern recognition and delineation from a wide variety of photographic input materials. However, software and hardware improvements to the NIC should be made for future work to increase its efficiency and accuracy. This will improve simulation and modeling and reduce eventual cost of any systems developed. The statistical criteria limits used in the recognition algorithms are questionable because of the small sample sizes, the narrow regionality and seasonality of the test image data base used. Broader contextual clues and more optimal program strategies must be pursued. The lack of automated data analysis programming for the system causes excessively complex and time consuming analysis of the output data by the system's user. Some means for rapidly analyzing the data that the NIC can produce must be developed.

The development of an operational system based on the NIC concept could evolve only through long range development. Even so, image variability on a world basis and on the basis of film and processing variability will restrict the number of automatic pattern recognition operations that may be implemented. It will be necessary to prepare regional versions of certain programs as well as scale and seasonal adaptations. Color and height information will be required as inputs to increase the confidence level of the recognition algorithms. A more realistic first goal would be to develop

an interim operational system where the man makes the majority of the decisions in the system. This system is envisioned as providing for the merging of data from more than one channel of input, a color display with light pen editing capability, and automatic tracking and output of most linear and area features. Pattern recognition and delineation experimentation and development should continue in the meantime with the NIC.

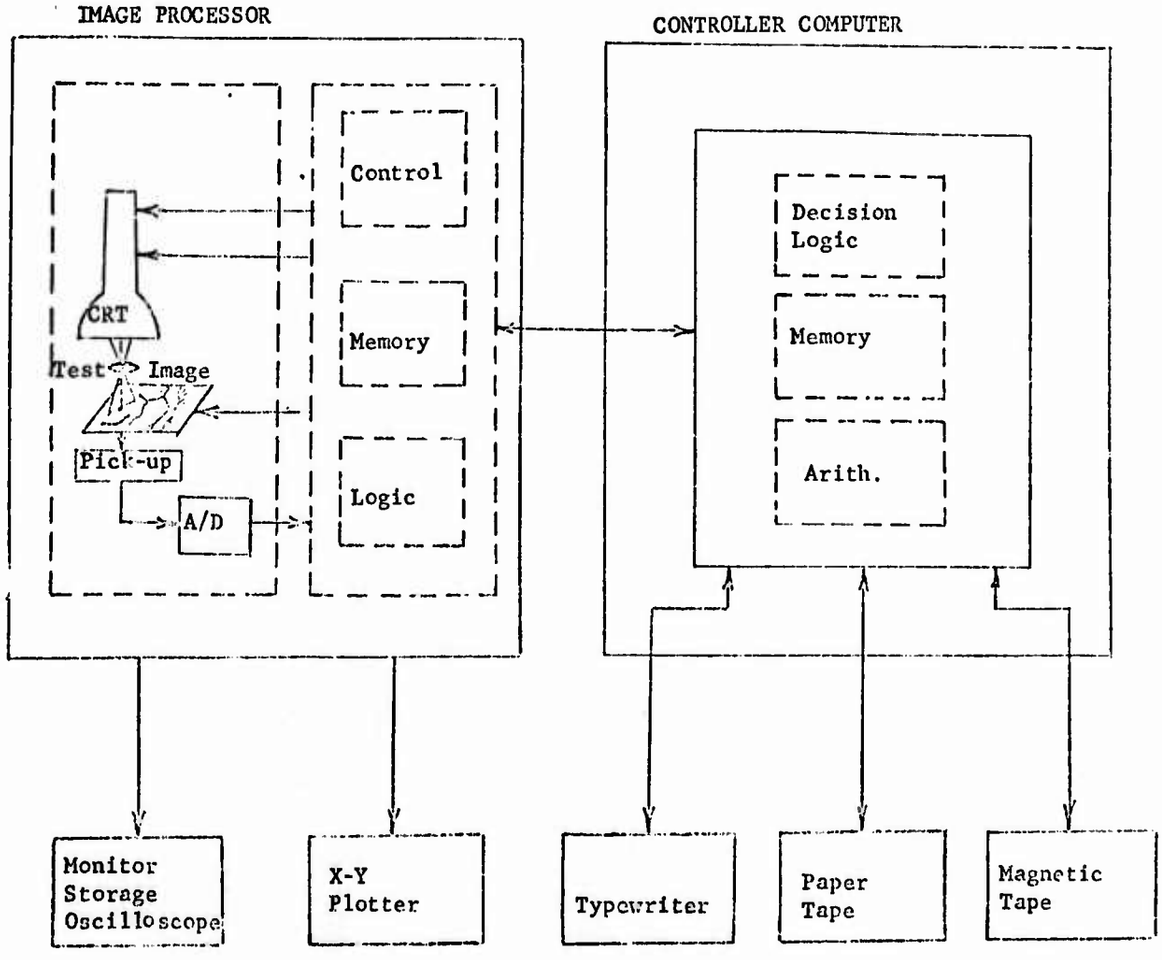


FIGURE 1. NATURAL IMAGE COMPUTER SYSTEM BLOCK DIAGRAM

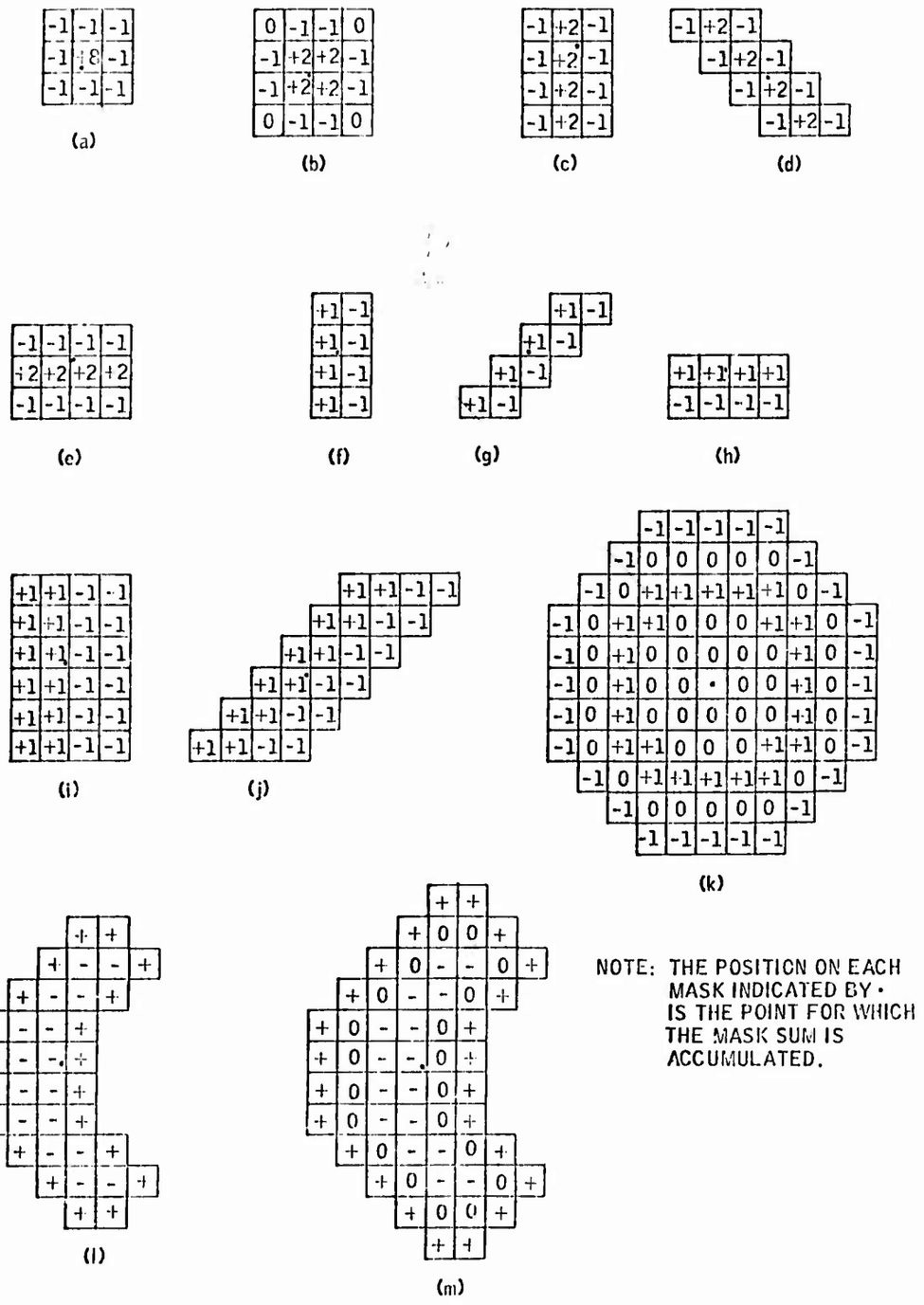


FIGURE 2. IMAGE PROCESSING MASKS

MEAN, VARIANCE
40.79, 6.35 (A), (B)

DISTRIBUTION

36	7	
37	118	
38	551	
39	865	
40	773	
41	531	
42	366	(C)
43	88	
44	121	
45	412	
46	261	
47	3	

TOTAL TEXTURE 1, 2
13865, 34031 (D)

DIFFERENCE NO. 1

1	12484	
2	1307	(E)
3	74	

DIFFERENCE NO. 2

1	25715	
2	6639	(F)
3	1552	
4	125	

RUN LENGTH X

1	703	
2	390	
3	181	
4	105	
5	78	
6	51	
7	25	
8	22	
9	13	
10	8	
11	7	
12	2	
13	1	
14	5	
15	3	
16	3	(G)
17	1	
26	1	
28	2	

RUN LENGTH Y

1	1054	
2	447	
3	243	
4	149	
5	59	
6	32	
7	20	
8	8	
9	4	
10	1	
12	5	
13	2	

FIGURE 3. STATISTICAL PROGRAM PRINT-OUT

STATISTICAL PARAMETER	WOOD LIMITS		LAKE LIMITS		ORCHARD LIMITS		RAIL YARD LIMITS		URBAN LIMITS		RIVERS & BAYS	
	LOW	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH
MEAN	30.48	42.87	30.23	38.54	34.75	45.82	36.06	41.90	36.81	44.73	32.54	45.03
VARIANCE	0.18	6.21	0.18	0.64	0.46	7.43	0.46	3.73	1.41	9.32	0.25	13.320
TEXTURE 1 ($\times 10^3$)	8.457	12.624	8.447	13.399	12.587	18.443	11.178	14.079	11.598	16.158	10.987	14.749
TEXTURE 2 ($\times 10^3$)	17.781	28.269	18.065	28.637	29.277	46.964	25.717	34.852	29.477	39.312	23.241	34.947
DIFFERENCE 1 - 1 ($\times 10^3$)	8.304	12.470	8.158	12.983	11.928	14.074	10.836	12.998	11.234	13.262	10.588	12.951
DIFFERENCE 1 - 2 ($\times 10^3$)	0.095	0.370	0.090	0.801	0.605	5.450	0.262	1.289	0.350	2.559	0.205	1.353
DIFFERENCE 1 - 3 ($\times 10^3$)	0.000	0.018	0.000	0.050	0.002	1.231	0.000	0.111	0.000	0.271	0.000	0.281
DIFFERENCE I > 3 ($\times 10^3$)	0.000	0.016	0.000	0.028	0.000	0.221	0.000	0.154	0.000	0.260	0.000	0.038
DIFFERENCE 2 - 1 ($\times 10^3$)	17.293	25.978	16.982	27.235	19.375	28.492	23.465	27.425	24.639	28.047	22.074	27.230
DIFFERENCE 2 - 2 ($\times 10^3$)	0.253	2.223	0.341	2.558	3.468	15.122	1.383	6.972	2.550	10.159	0.864	7.666
DIFFERENCE 2 - 3 ($\times 10^3$)	0.000	0.075	0.000	0.528	0.080	8.867	0.015	1.091	0.130	3.289	0.001	2.738
DIFFERENCE 2 > 3 ($\times 10^3$)	0.000	0.032	0.000	0.112	0.000	3.600	0.000	0.509	0.000	1.139	0.000	1.116
RUN LENGTH 1 X	617	927	603	978	777	1616	669	1006	711	1152	762	1162
RUN LENGTH 2 X	180	385	193	456	337	706	254	550	325	518	270	487
RUN LENGTH 3 X	91	240	93	223	180	291	134	285	173	252	148	244
RUN LENGTH > 3 X	302	399	293	466	99	332	168	368	242	360	235	357
RL3/RL > 3 X	.257	.681	.291	.709	.548	2.308	.364	1.04	.507	1.14	.384	1.000
RUN LENGTH 1 Y	586	918	586	932	774	1.989	668	975	698	1284	768	1067
RUN LENGTH 2 Y	190	381	167	439	274	695	283	485	324	530	272	446
RUN LENGTH 3 Y	86	201	97	225	158	263	156	277	175	281	137	233
RUN LENGTH > 3 Y	309	403	302	359	62	347	220	377	239	372	266	372
RL 3/RL > 3 Y	.252	.619	.281	.700	.500	3.274	.423	.981	.508	1.106	.383	.864
	120 images		48 images		99 images		74 images		155 images		80 images	

RANGE OF STATISTICAL PROPERTIES FOR IMAGES TESTED

TABLE I

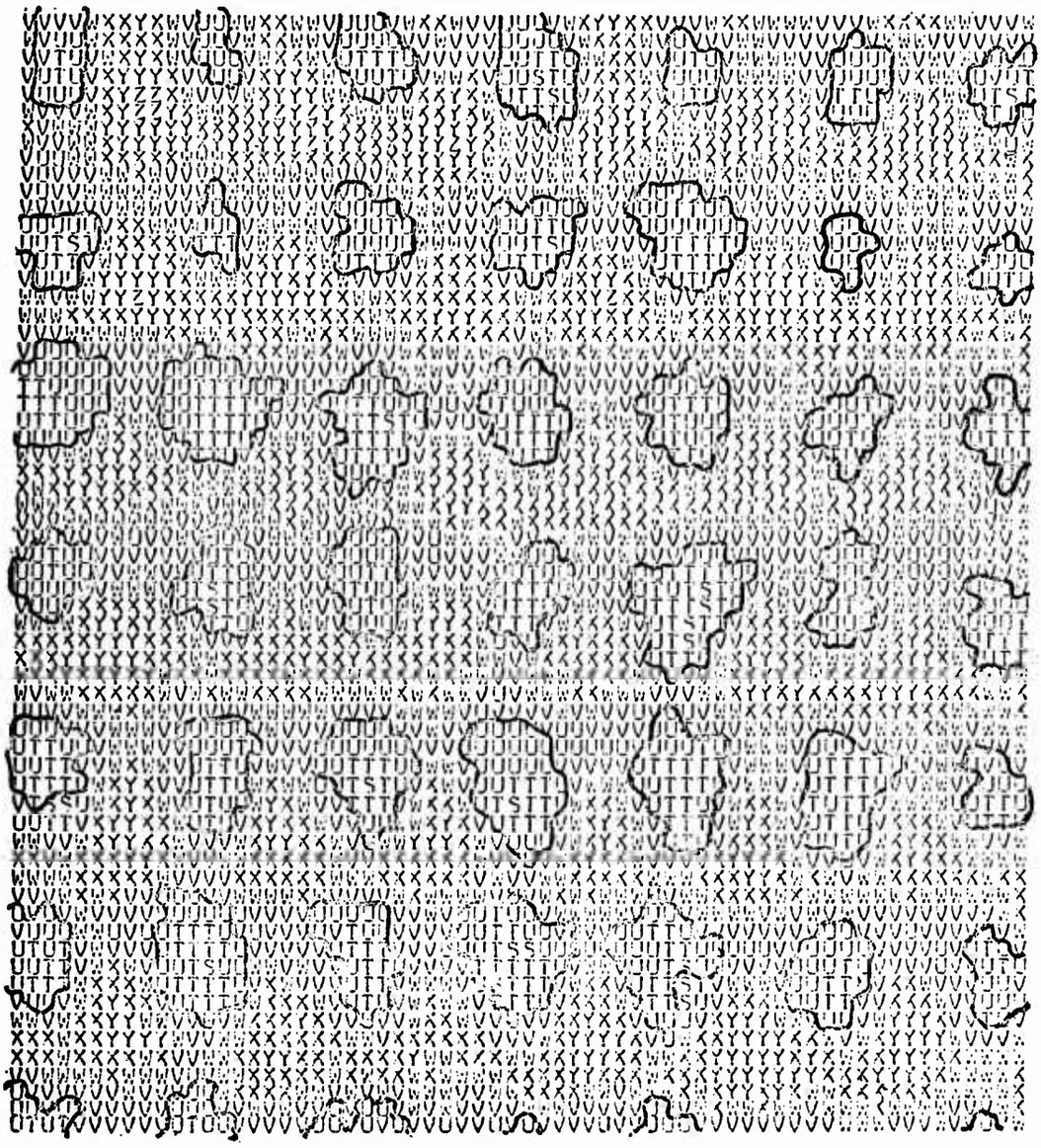


FIGURE 4. ORCHARD GREY LEVEL IMAGE

	WOOD FALSE ALARM RATE	LAKE FALSE ALARM RATE	ORCHARD FALSE ALARM RATE	RAIL FALSE ALARM RATE	URBAN FALSE ALARM RATE	RIVER FALSE ALARM RATE
WOOD		11/120=.091 9.1%	0/120= 0 0%	0/120=0 0%	0/120=0 0%	2/120=.017 1.7%
LAKE	19/48=.396 39.6%		0/48=0 0%	2/48=.042 4.2%	0/48=0 0%	12/48=.25 25%
ORCHARD	0/99=0 0%	0/99=0 0%		6/99=.06 6%	17/99=.171 17.1%	8/99=.08 8%
RAIL	1/74=.014 1.4%	2/74=.027 2.7%	31/74=.419 41.9%		23/74=.31 31%	26/74=.351 35.1%
URBAN	0/155=0 0%	0/155=0 0%	114/155=.735 73.5%	62/155=.40 40%		34/155=.219 21.9%
RIVER	6/80=.075 7.5%	25/80=.312 31.2%	3/80=.038 3.8%	16/80=.20 20%	6/80=.075 7.5%	
TOTAL	26/456=.05 5%	38/528=.072 7.2%	148/477=.31 31%	86/502=.171 17.1%	40/441=.09 9%	82/496=.165 16.5%

RECOGNITION FALSE ALARM RATES BASED
ON GREY SCALE STATISTICS

TABLE II

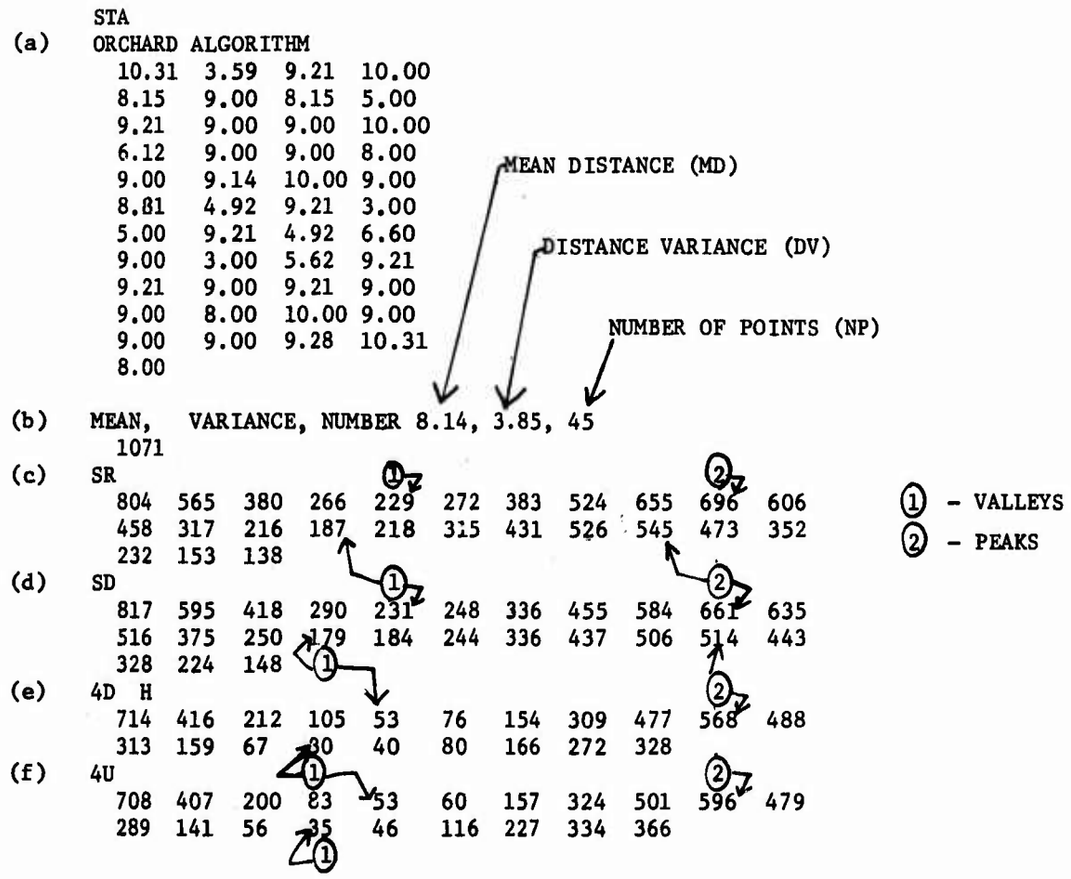


FIGURE 5. ORCHARD ALGORITHM PRINT-OUT

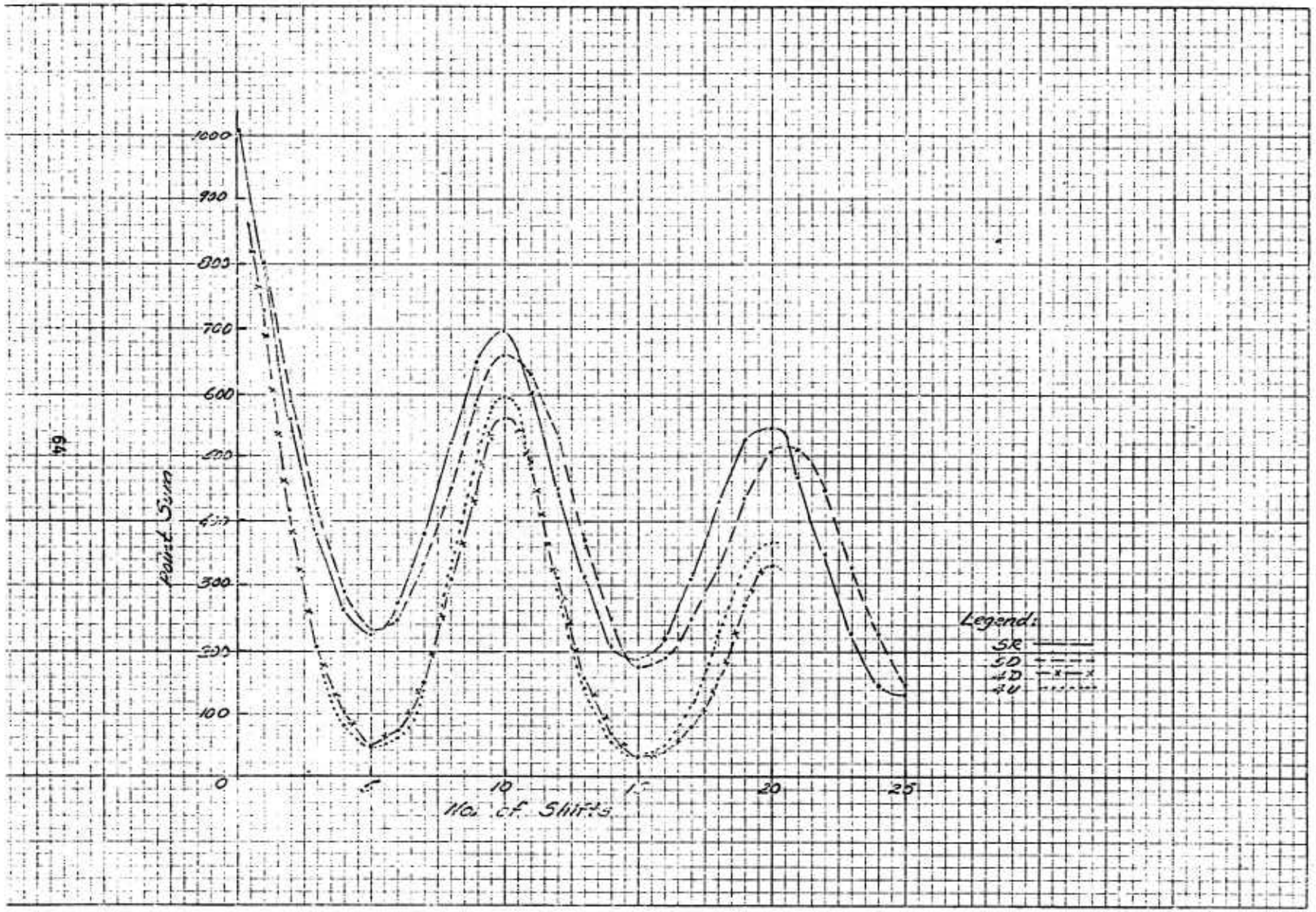
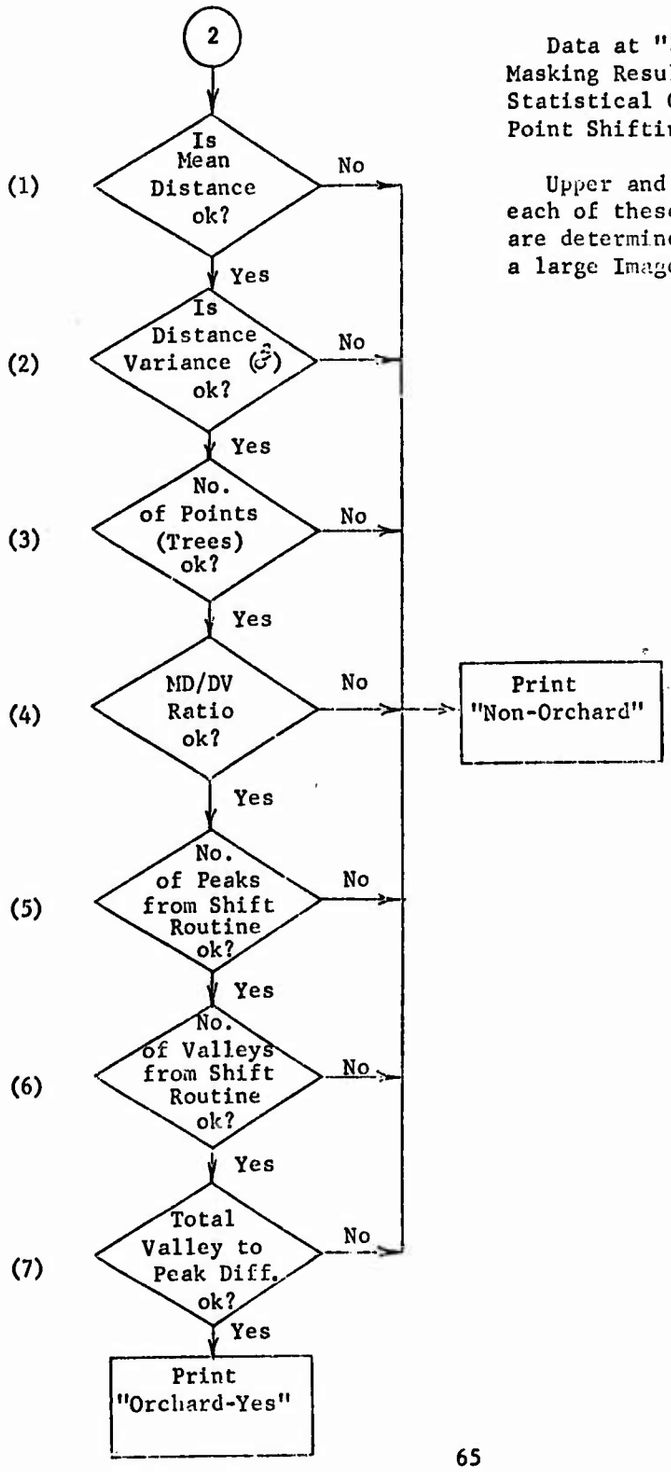


FIGURE 6. PLOT OF ORCHARD SHIFTING - ANDING



Data at "2" available from Masking Results, Distance and Statistical Calculations and Point Shifting Routines.

Upper and Lower Limits for each of these Decision Points are determined empirically from a large Imagery Data Base.

FIGURE 7. ORCHARD DECISION LOGIC

PATCH RESULTS
KNOWN ORCHARD FIELDS

TEST PATCH NO.	POSITIVE ORCHARD RESPONSES		NO. OF FIELDS TESTED	REMARKS
	ORCHARD CRITERIA			
	CONTRACTOR	MODIFIED		
1	4	2	7	Poor separation of Orch trees.
2	4	3	4	
3	10	10	12	
4	3	9	10	Variance of MD very small.
5	4	15	16	Variance of MD very small.
6	15	15	16	45% of missed field is non-orchard.
7	16	15	16	
TABLE VII, FIELDS	14	15	15	
TOTALS	70	84	96	
FALSE REJECT RATE (MISS RATE)	27.1%	12.5%		

TABLE III

PATCH RESULTS
NON-ORCHARD FIELDS

TEST PATCH - TYPE	POSITIVE ORCHARD RESPONSES		NO. OF FIELDS TESTED	REMARKS
	ORCHARD CRITERIA			
	CONTRACTOR	MODIFIED		
WATER & PIERS	0	0	8	
LIGHT INDUSTRY	0	0	8	
HEAVY INDUSTRY	1	1	16	
RURAL HOUSING DEVELOPMENT	6	0	39	Includes some rural streets but no houses.
RURAL AREA - MAIN ROADS	2	0	9	Some rural houses in each field but no large town major roads.
DRY WASH	1	0	16	
VINEYARD (LARGE ROW CROP)	7	4	16	Crop tends to line up in rows although dim.
URBAN AREA - ROW HOUSES	13	0	16	Includes streets and landscaping.
CULIVATED AREA - FALLOW	0	0	16	
NEWLY CULIVATED AREA	0	0	10	
MEADOW/PASTURE	0	0	16	
WOODED	8	0	16	October (Fall) photos; many contrasting trees.
TOTALS	38	5	186	
FALSE ALARM RATE	20.4%	2.7%		

TABLE IV

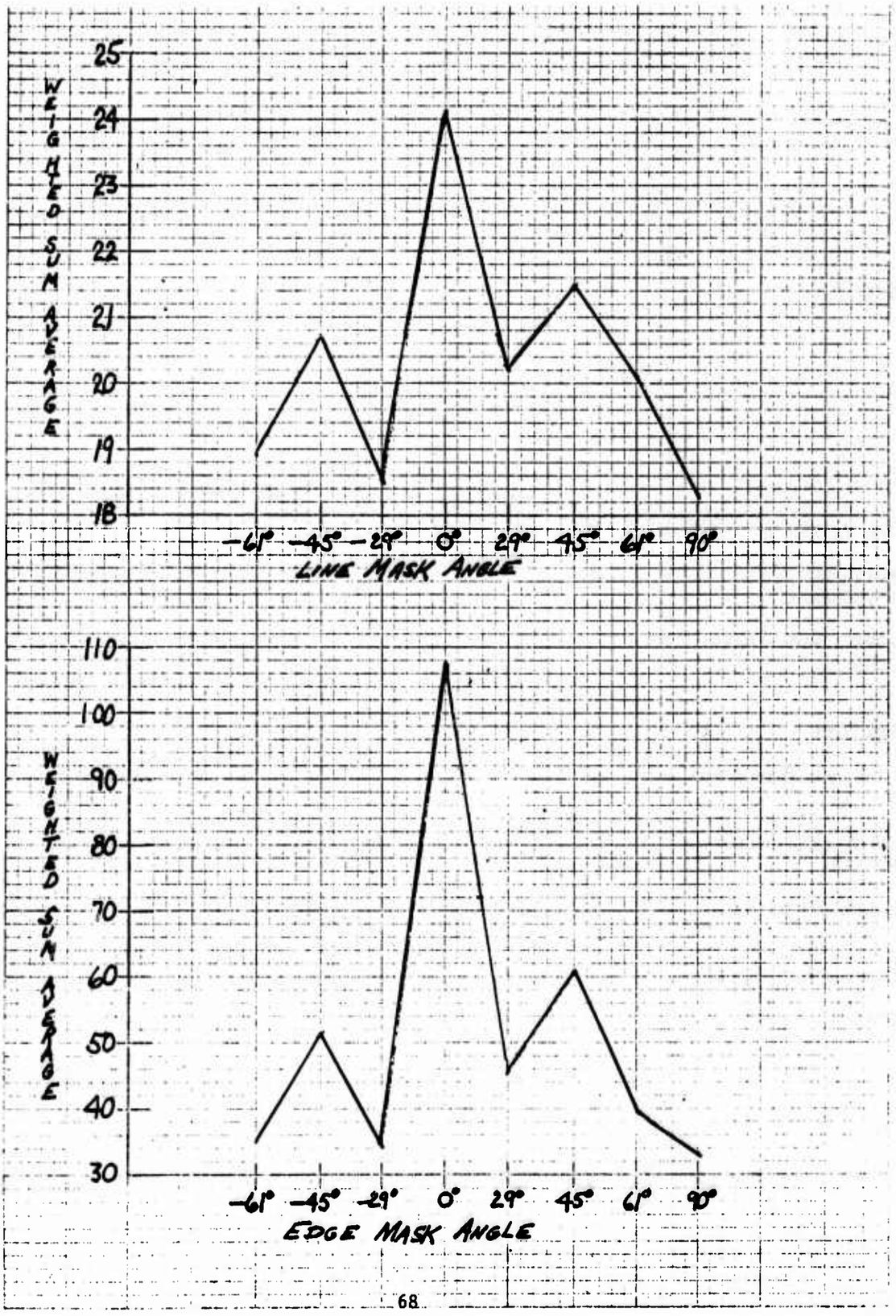


FIGURE 8 RR YARD LINE AND EDGE MARKING PLOT

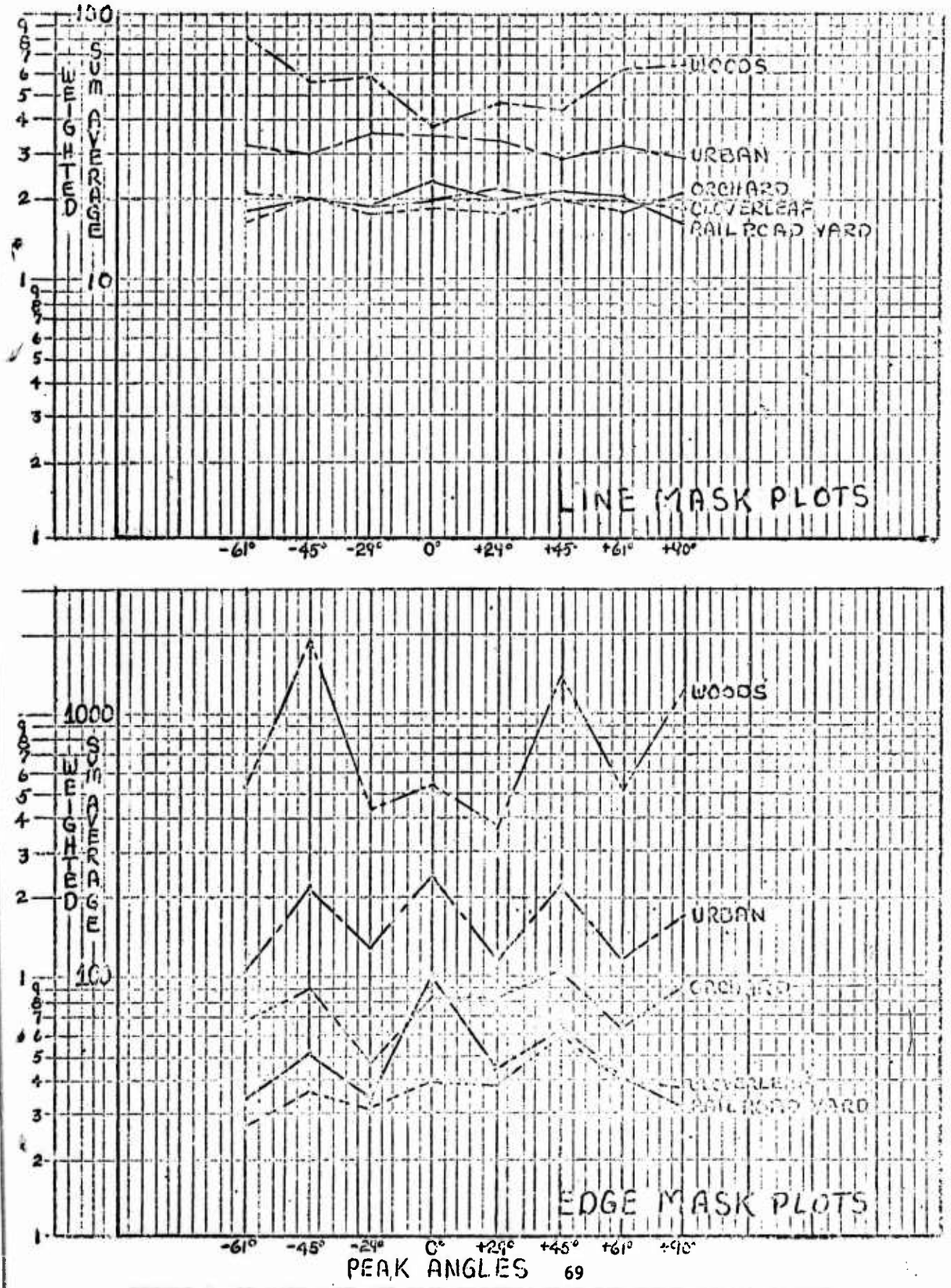


FIGURE 9. RR YARD LINE AND EDGE MARKING PLOT FOR NON-RAILROAD IMAGES

RAILROAD YARD RECOGNITION PROCESSING

1:50,000 Scale Criteria

MISS RATE	2/10 = 20%
FALSE ALARM RATE	6/15 = 40%
COMBINED ERROR RATE	8/25 = 32%

TABLE V

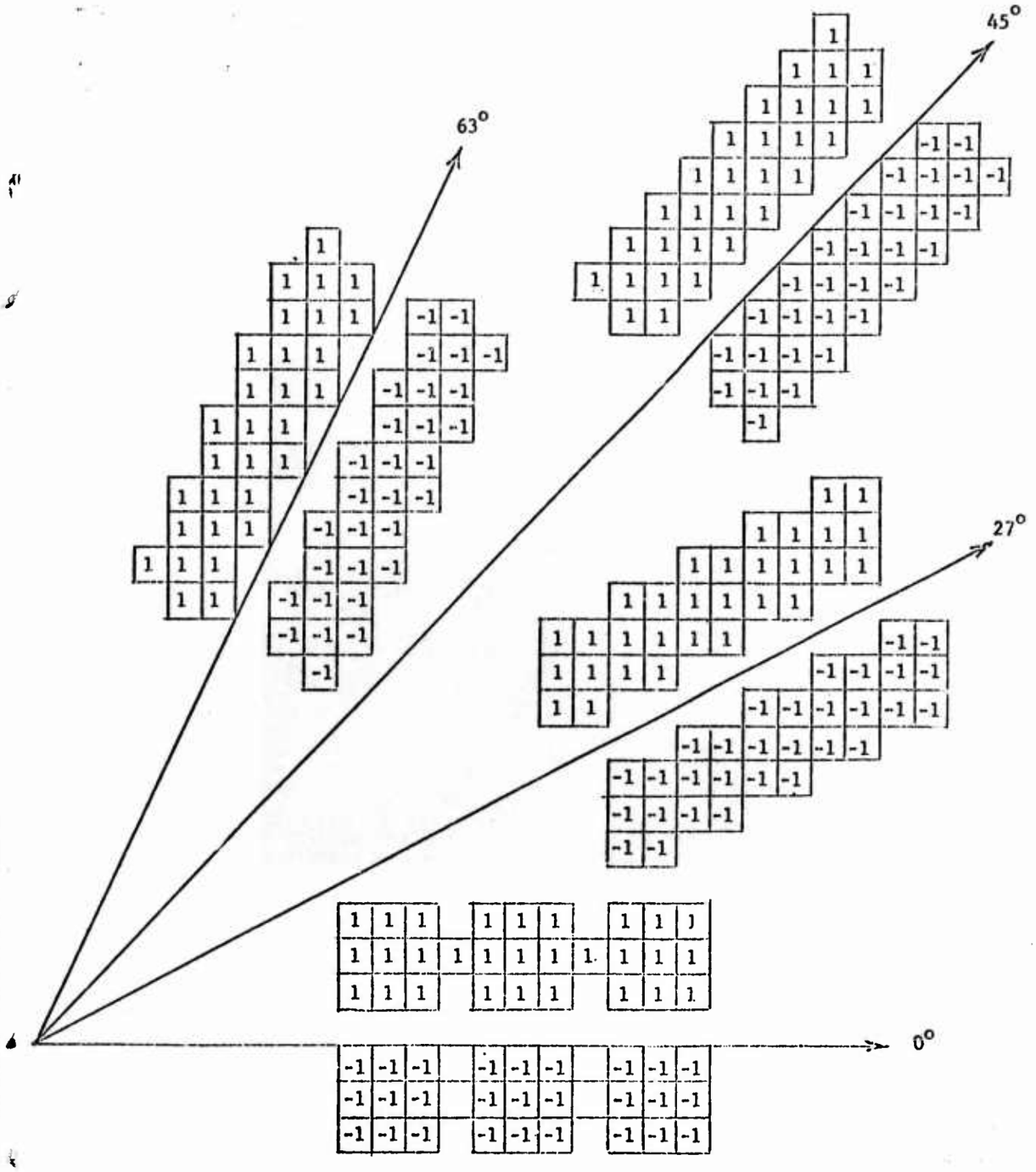
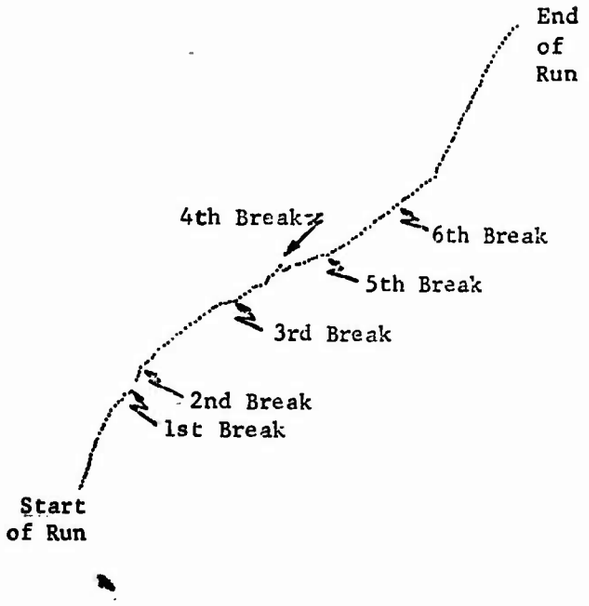
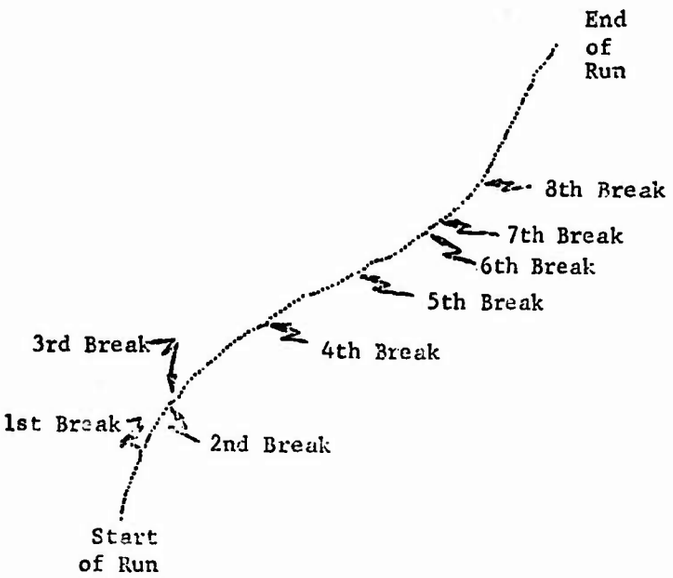


FIGURE 10. BASIC TRACKING MASKS

FIGURE 11. TEST IMAGE USED FOR TRACKING RUN EXAMPLE



64 GREY LEVEL RUN OF ENHANCED IMAGE



32 GREY LEVEL RUN OF ENHANCED IMAGE

FIGURE 12. TRACKING TESTS USING PHOTOGRAPHICALLY ENHANCED IMAGE

BLANK PAGE

MULTI-IMAGE CLUSTERING

Robert M. Harlick
Engineering Topographic Laboratories
Fort Belvoir, Virginia

ABSTRACT. One technique used for image data analysis is clustering. Clustering procedures group the data points or resolution cells into blocks having members which are highly interrelated or similar. The distinct blocks are hypothesized as having been produced by different environmental entities or processes. Hence, examination of the blocks yields interpretation concerning the nature and diversity of the environment from which the data were gathered. Since the blocks reflect the "natural" structure of the environment as seen by the instruments which collected the data, clustering techniques can help the researcher formulate the concepts needed to deal with the instrument-environment system.

A measurement space and spatial clustering procedure is utilized on a multi-image data set obtained from the northern part of Yellowstone Park with the Michigan 12 channel scanner system. The results indicate that about a half-dozen distinct environmental categories can be recognized without any *a priori* ground truth information.

I. MULTI-IMAGE CLUSTERING. Clustering is a way to automatically extract information from data, in our case multi-image data (1-10). The traditional way to extract information from imagery is to present the imagery to skilled photo-interpreters for analysis. But this is a slow process intended only for limited amounts of imagery. To analyze the large amounts of data which today's remote sensors gather by airplane and satellite, automatic processing techniques must be used.

The automatic processing techniques fall into two classes: supervised and unsupervised. Clustering is an unsupervised technique. With supervised techniques, one gathers a training set of data for which one knows the correct identification of each distinct entity in the data. One then estimates the necessary conditional probability distribution and determines a decision rule from them. The decision rule can then be employed to identify any other data set gathered under similar conditions. With unsupervised techniques there is no training data set or decision rule. One attempts to determine the structures in a data set. Distinct structures are then interpreted as corresponding to distinct objects or environmental processes.

The advantage of the supervised techniques is that the scientist is able to decide what types of environmental categories among which he wishes to distinguish. The decision rule then determines to which such environmental category an arbitrary data entity belongs. The disadvantage of the supervised techniques is that they are sensitive to mis-calibrations. Any slight difference between the sensor calibrations or state of environment for the training data and the new data will cause error. For instance, if one were analyzing multi-images to determine vegetation or crop type on the basis of multi-spectral image grey tone, then one would find that the grey tone associated with a crop when it is sunny is not the same grey tone associated with it when it is cloudy. Perhaps it had rained and the ground was wet or perhaps it was hot and the ground was dry perhaps the irrigation and fertilizer were different; these differences too would make different grey tones. Perhaps one corn field was planted a few weeks before another or the hybrid of corn was different. Perhaps the film emulsions came from different emulsion batches or the developers were not all of equal strength. It is obvious that one could compile a long list of the intervening variables, to measure them (if he could), which affect the environment-image system calibration.

The advantage of the unsupervised techniques is that they are not sensitive to calibration problems. Two small-area patches of corn growing in the same field are going to be detected as being similar because they have similar grey tone associated with them. The disadvantage of the unsupervised techniques is that they are not able to identify what are the distinct environmental structures they determine.

II. POTENTIAL APPLICATION OF THE MULTI-IMAGE CLUSTERING TECHNIQUE. The use of the potential information contained in images is well documented in the remote sensing literature, a selected sample of which is given in reference 10-24. One can potentially determine vegetation species distribution, location of mineral deposits, worldwide distribution of geomorphic features, sea surface temperatures, location of fish schools and icebergs, spread of pollutants, cloud cover distribution and movement, developing storm systems, water movement, soil and vegetation moisture content, transportation networks, and extent of urbanization to name a few.

What is not clear is what sensor or set of sensors is best for what jobs. By employing multi-image clustering to imagery obtained from a set of sensors, one can determine the distinct environmental structures as seen through the sensor's eyes. By then comparing these structures to what is actually in the environment, as determined

from field work, one can tell what categories among which the set of sensors will distinguish best. This procedure is more efficient than one based on supervised techniques since it only has to be performed once while the supervised technique will have to be evaluated for each a priori set of categories chosen.

Once the multi-image clustering technique is developed so that it can be cheaply implemented in real time, another use for it will become feasible: it can be used to pre-process the data performing a feature extraction function. The features obtained will not be sensitive to the calibration problem. Identification decisions are then made as usual on the basis of the extracted features. The result will be a multi-image processing system which can work well despite the calibration difficulties.

III. DATA ANALYSIS. Twelve images taken by the Michigan scanner system were the multi-image data set. These images were taken of a 2 mile by 6 mile area in the northern part of Yellowstone Park at approximate coordinates $100^{\circ}30'$ by $44^{\circ}57'$ on September 19, 1967. Figure 1 shows an old panchromatic photograph of the area taken in 1954. Each image of the multi-image set was, in effect, a picture taken with a different narrow-band filter where the filters passed light in narrow bandwidths from the near infrared band part of the spectrum through the ultraviolet portion of the spectrum. Table 1 tabulates these bands. The images were digitized to 256 levels on a grid of 220 x 1260 for a total of about 270,000 resolution cells for each image. Each resolution cell contains the returns from 12 spectral bands coming from a 20 ft. x 20 ft. small-area ground patch. Successive resolution cells contain returns from small-area ground patches separated by a gap of 20 ft.

In order to reduce computer time, the original twelve images were processed to yield four smaller images, but with most of the statistical and spatial structure preserved. The first part of the pre-processing consisted of a principal components analysis. A principal components analysis may be considered in the following fashion. In any image, some grey tones occur more frequently than others. We may consider the relative frequence of the grey tones on the image as defining a one-dimensional probability distribution. This probability distribution has a mean and variance. Similarly in the twelve image multi-image set each resolution cell has a 12-tuple of grey tones. Some of these 12-tuples occur more frequently than others, and we may consider the relative frequency of the 12-tuple grey tones as defining a twelve-dimensional probability distribution. This probability distribution too has a mean and variance. The variance is called the total variance. A principal components

Figure 1. 1954 Photographic Image Taken of Area

Image	Wavelength band
1	800-1000 milli-microns
2	720-800
3	660-720
4	620-660
5	580-620
6	550-580
7	520-550
8	500-520
9	480-500
10	460-480
11	440-460
12	400-440

Table 1 Tabulates the wavelength bands for each of the images

analysis, as might be expected, determines principal components. The first principal component image is obtained by taking that linear combination or weighted average of the original twelve images such that the variance of the probability distribution of grey tones on the first principal component image is higher than it could be for any other linear combination. The second principal component is obtained by taking that linear combination, orthogonal to the first, such that the variance of the probability distribution of grey tones on the second principal component image is higher than it could be for any other linear combination orthogonal to the first. In general, the kth principal component image is obtained by taking that linear combination, orthogonal to the earlier 1st, 2nd, ...,(k-1)th linear combinations of the original twelve images such that the variance of the probability distribution of grey tones on the kth principal component image is higher than it could be for any other linear combination orthogonal to the earlier 1st, 2nd,...,(k-1)th ones. Because the sum of the variances of all the principal components equals the total variance of the original twelve-dimensional probability distribution, the ratio of the variance of the kth principal component to the total variance is called the variance accounted for by the kth principal component. The variance accounted for by the kth principal component is an indicator of how much statistical structure from the original twelve images is preserved by the kth principal component image.

The principal components provided the following results.It was determined that the first component accounted for 97.4% of the variance, the second component 1.6% of the variance, and the third component .9% of the variance. The respective weights used in the linear combination are listed in Table 2.

Table 2 has an interesting interpretation. The first linear combination has weights which are all positive and which are about the same magnitude. The first principal component image is then very close to what a panchromatic image of the area would be. This should not be surprising since most photo interpreters will prefer a panchromatic image over any narrow-band image because they see more structure in it. The second linear combination weights the infra-red part of the spectrum negatively, the middle of the spectrum hardly at all and the ultra-violet part of the spectrum positively. This weighting trend from the infra-red to the ultra-violet is almost a linear one. It is indicative of the fact that the spectral reflectance curve for most natural objects shows that when infra-red reflectance is high, then the ultra-violet reflectance is low and when the ultra-violet reflectance is high then the infra-red is low. Hence, the weighting done by the second principal component will enhance the difference between those objects with high infra-red and high ultra-violet reflectance. The third linear combination is perhaps indicative of the spectral reflectance difference between vegetation and rock. The spectral reflectance curve for most vegetation slopes positively at the ultra-violet end of the spectrum

	1st	2nd	3rd
	Principal Component		
	% Variance Accounted for		
Wavelength Band	97.4%	1.6%	.9%
800-1000	.15702	-.33153	.27651
720-800	.22758	-.33529	.26796
660-720	.28342	-.31332	.13188
620-660	.23184	-.23019	.11530
580-620	.20199	-.16154	.010509
550-580	.17486	-.094258	.025696
520-550	.32559	-.070442	.25499
500-520	.24756	.031677	.046381
480-500	.42727	.09246	-.12730
460-480	.52815	.22196	-.60118
440-460	.25916	.30966	.043766
400-440	.14878	.65714	.61121

Table 2 Table of weights used to obtain the linear combinations for the 1st, 2nd, and 3rd principal components.

while that for ryolite (a volcanic rock known to be prevalent in the area photographed) is almost flat in that very same region. Hence, weighting the 460-480 milli-micron part of the spectrum negatively and the 400-440 milli-micron end of the spectrum positively will enhance the difference between vegetation and ryolite.

Since the first three principal components accounted for about 99% of the variance, there is no need to cluster twelve images. Almost all the information is contained in 3 linear combinations (principal components) of the original twelve. However, because the clustering procedure treats each image equally, the second and third principal component images, which together only account for a few percent of the total variance, would be unduly emphasized. Therefore, a new set of images, called the principal components dispersed images, was prepared by taking linear combinations of the first three principal components. The linear combinations are mutually orthogonal and result in distributing the variance almost equally among the dispersed images. Table 3 shows the weights used in these linear combinations.

The next step in the preprocessing consisted of reducing the size of the three dispersed principal component images. Each was 220 resolution cells horizontally by 1220 resolution cells vertically. They were reduced in size to 73 resolution cells horizontally by 406 resolution cells vertically by taking every third row of resolution cells and every third resolution cell on each such row taken. To compensate for the roughening effect by such a reduction, a fourth image was prepared by reducing the size of the first principal component image by averaging the grey tones of nonoverlapping 3 x 3 blocks of resolution cells.

The final step in the preprocessing consisted of quantizing the grey tones of the four images to 13 quantized grey tone classes. This was done in two parts: the images were first quantized to 64 grey tone classes by a folded-tail linear quantizing procedure (see Figure 2). The folded tails quantizing is essentially a linear quantizing procedure modified to ignore extreme wild points on the tails of the distribution. In other words, instead of determining the highest grey tone and the smallest grey tone and then equally divide the resulting interval up into 64 pieces as the linear quantizing does, the folded tails linear quantizing determines a "high" grey tone less than the highest and a "small" grey tone greater than the smallest and equally divides that resulting interval up into 64 quantized classes. Of course, grey tones higher than the determined "high" and smaller than the determined "small" get put in the highest and smallest quantized class respectively.

	1st	2nd	3rd
	Dispersed Components		
1st Principal Component	$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{3}$
2nd Principal Component	$1/\sqrt{6}$	$-2/\sqrt{6}$	$1/\sqrt{6}$
3rd Principal Component	$1/\sqrt{2}$	0	$-1/\sqrt{2}$

Table 3 Table of weights for the mutually orthogonal linear com combinations used to disperse the variance of the first three principal components.

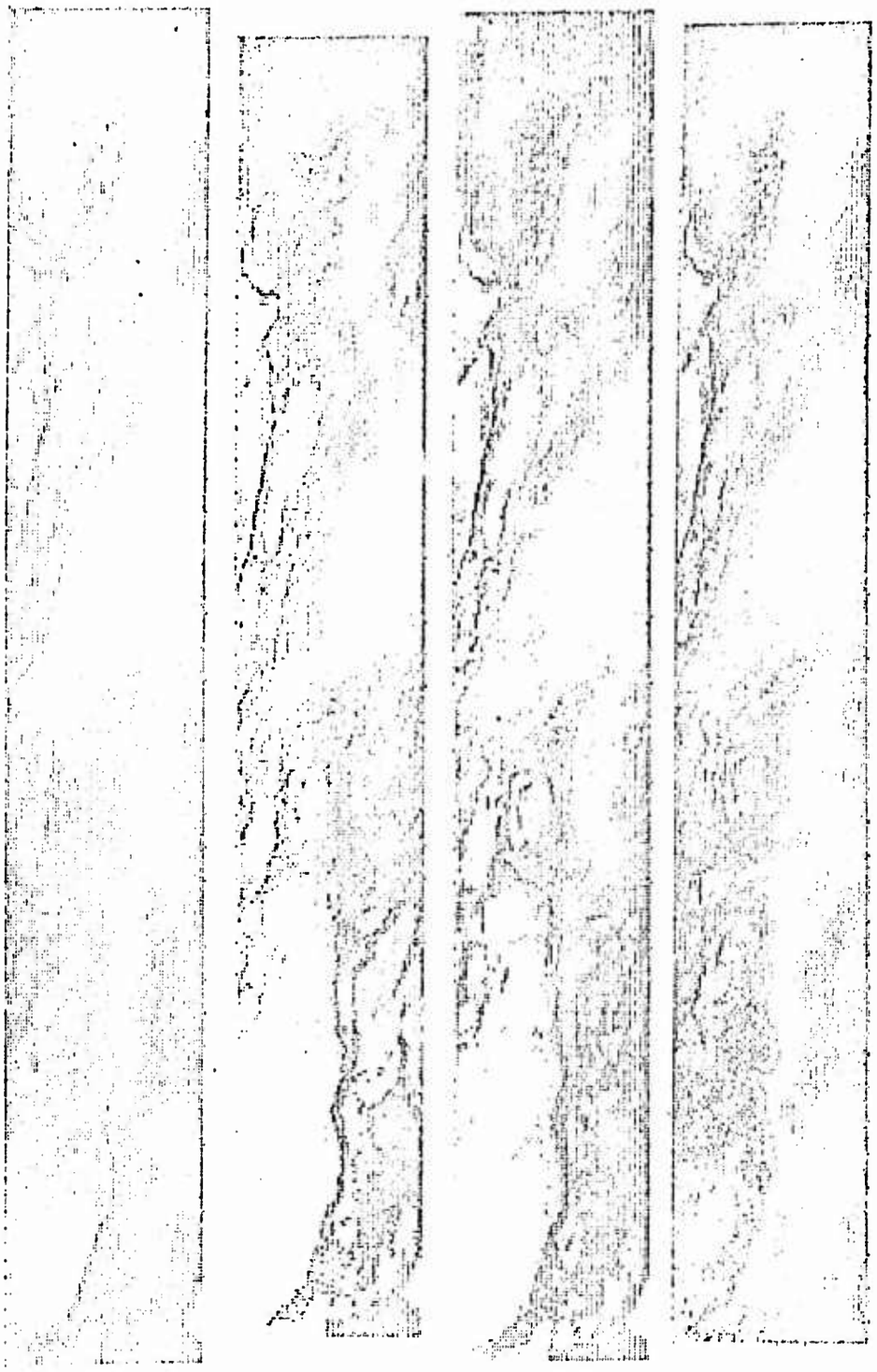


Figure 2. Quantized Images Used for the Input of the Spatial Clustering Procedure

The spatial quantizing procedure further reduces the number of quantized classes from 64 to 13 in a way which capitalizes on the spatial dependence of the 64 quantized grey tone classes. A 64 x 64 Markov transition probability matrix is set up where the (i,j)th element is the probability that a resolution cell having a grey tone in the ith grey tone class will be next to a resolution cell having a grey tone in the jth grey tone class. The spatial quantizing uses the information in this matrix to form quantized classes whose grey tones have a high probability of occurring next to each other. Hence, spatial continuity of grey tones tends to be preserved. These quantizing procedures are fully described in reference 18.

After quantizing the four images to 13 grey tone classes, the images were clustered.

IV. DESCRIPTIVE OUTLINE OF SPATIAL CLUSTERING. The clustering procedure starts by finding sets of resolution cells called center sets. Center sets are good places to build a cluster around. Resolution cells whose grey tone N-tuples are similar enough are sequentially added to the existing center set. When there are no more resolution cells similar enough, the existing cluster is complete, and a new cluster is begun from a new center set.

We assume that each object on the ground produces grey tones which are similar and homogeneous. There can be any spatial distribution of objects; one object may only occur once and another hundreds of times or each object may occur approximately the same number of times. It would be intuitively reasonable to form center sets from those spatial locations which have fairly homogeneous measurement space coordinates and which are representative measurements of a class of objects. However, since the location and extent of objects are unknown to the clustering procedure, it must try to induce this information from the data structure. Since we assume that the set of measurements recorded from any object form a homogeneous set of highly similar grey tones, and the location of these measurements in the image sequence is a small, more or less, spatially connected region, perhaps by breaking up the image sequence into a set of spatially connected subsequences and examining the measurements in each subsequence we can obtain the necessary information. Thus we make each spatially connected subsequence: (1) large enough to include within it a substantial proportion of the measurements recorded from at least one object; and (2) small enough so that the substantial proportion of the measurements recorded from the object make up a large proportion of the measurements in the subsequence. If we can form subsequences in this way, then the empirically observed probability distribution of the grey tone N-tuples in the sequence

will be dominated by the substantial proportion of grey tone N-tuples in the sequence recorded from some particular type of object. Thus, if a particular object occurs only once then there will be one subsequence dominated by it. By picking out the kind of measurements which typify that subsequence (i.e. those which have high probability or high self-association in the subsequence), then the set of all the spatial locations containing these measurements is a good center set.

A grey tone N-tuple has high self-association if there is high probability that a resolution cell having that grey tone or a grey tone similar to it will be contiguous to a resolution with the given grey tone N-tuple. Two grey tone N-tuples are said to have high cross-association if there is high probability that a resolution cell having one of the grey tone N-tuples, or a grey tone similar to one of them, will be contiguous to a resolution cell having the other grey tone N-tuple or a grey to similar to the other.

Since the clustering procedure we have proposed starts with center set one, build on it until no more similar measurements can be found, and then starts building on center set two, etc., we must specify how the order is determined for center sets. We should naturally start with the most important center set and here importance can be correlated with self-association. That center set is most important which has the highest self-association of all center sets in the subsequence form which it originates.

Next we must consider exactly how each center set grows. If there exists a resolution cell outside but next to some resolution cell in the center set and if the cross-association of the grey tone N-tuples of these resolution cells are sufficiently high and the self-association of the grey tone in the original center set is not too different from the self-association of the grey tone in the outside resolution cell, then the outside resolution cell is added to the center set. When no such outside resolution cell exists, the growing is terminated and a new center set is defined. A precise description of the clustering procedure can be found in reference (8).

V. RESULTS. A map of the distinct environmental objects as determined from the clustering procedure is displayed in Figure 3. The computer took .3 hour processor time and 1.8 hour wall time to produce these results. Some of the regions on the map can immediately be seen as representing some of the dominant types of categories in the area. A detailed comparison of the map with the ground truth for the region is required to make a good evaluation of these results.

HOMOGENEOUS REGIONS PRODUCED BY THE SPATIAL CLUSTERING METHOD

Figure 3

This is the first time that such a clustering algorithm has been tested on a multi-image and there are some problems with the clustering procedure itself which have to be solved. One problem concerns the specification of the clustering parameters. This first time they had been specified on a trial and error basis and it resulted in much wasted computer time. It should be possible to determine them from appropriate probability distributions which the multi-image generates. A second problem concerns boundary delineation. It appears that some boundaries extend too much while others not enough. The problem may be related to either optimal determination of clustering parameters or to the way the clusters themselves grow. A third problem is concerned with the large amount of computer time needed to implement the algorithm. A careful study of the growth of the clusters can probably lead to a new algorithm for which the computer time is cut by an order of magnitude.

REFERENCES

1. Bonner, R. E., "On Some Clustering Techniques," IBM Journal of Research and Development, Jan. 1964.

2. Mattson, R. L., and J. E. Dammann, "A Technique for Determining and Coding Subclasses in Pattern Recognition Problems," IBM Journal of Research and Development, July 1965.

3. Patrick, E. A., and J. C. Hancock, "Nonsupervised Sequential Classification and Recognition of Patterns," IEEE Transaction on Information Theory, Vol, IT-12, No. 3, 1966.

4. Cooper, D. B., and P. W. Cooper, "Nonsupervised Adaptive Signal Detection and Pattern Recognition," Information and Control, Vol. 7, 1964.

5. Ward, J. H., "Hierarchial Grouping to Optimize an Objective Function," American Statistical Association Journal, March 1963.

6. Lance, G. N., and W. T. Williams, "A General Theory of Classificatory Sorting Strategies II. Clustering Systems," Computer Journal, Vol. 10, No. 3, Nov. 1967.

7. Haralick, R. M. and G. L. Kelly, "Pattern Recognition with Measurement Space and Spatial Clustering for Multiple Images," Proceedings of the IEEE, Vol. 57, No. 4, April 1967.

8. Haralick, R. M. Multi-Image Pattern Recognition: Ideas and Results, Ph.D. Dissertation, University of Kansas, 1969.

9. Fralick, S. C., "Learning to Recognize Patterns without a Teacher," IEEE Transactions on Information Theory, Vol. IT-13, No. 1. pp. 57-64, January 1967; Stanford Technical Report 6103-10, SEL-65-011, March 1965.

10. Spragins, J. D., "Learning without a Teacher," IEEE Transactions on Information Theory, IT-12, No. 2, April 1966.

11. Colwell, R. N., et al,. The Usefulness of Thermal Infrared and Related Imagery in the Evaluation of Agricultural Resources, Vol. 1, "Cultivated Crops," NASA Contract No. 12-12-100-8316(20), Berkeley, University of California, 1966, 137pp.

12. Gerlach, A. C., "Geographic Applications of Remote Sensor Data from Aircraft and Spacecraft," Proceedings of the International Geographical Union Regional Conference, Mexico City, Mexico, 1966.

13. LARS, Remote Multispectral Sensing in Agriculture, Annual Report Vol. 3, LARS Publication No. 844, Lafayette, Purdue University Agricultural Experiment Station, September 1968, 175 pp.

14. Menzer, F. E., and C. R. Cooper, "Aerial Photographic Methods of Potato Disease Detection," University of Maine, Agricultural Experiment Station Bulletin 646, 1967, 14pp.

15. Meyer, M. P., and L. Calpouzos, "Detection of Crop Diseases," Photogrammetric Engineering, XXXIV, 6, 1968, 554-557.

16. Simonett, D. S., "Land Evaluation with Remote Sensors in the Infrared and Radar Regions," in Land Evaluation, edited by G. A. Stewart, Canberra, Australia: CSIRO and MacMillan, 1968, 349-366.

17. Thorley, G. A., "Some Uses of Color Aerial Photography in Forestry," in ASP, Manual of Color Aerial Photography, 1968, 392-393.

18. Weiss, M., et al., "Evaluation of an Infrared Sensor as a Clear Air Turbulence Detector," Sixth Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, October 1969.

19. Carneggie, D. M., "Analysis of Remote Sensing Data for Range Resource Management," Forestry Remote Sensing Laboratory, Berkeley, University of California, 1968.

20. Garrison, W. L., et al., "Five Papers on Remote Sensing and Urban Information Systems," Technical Report No. 1, Urban and Transportation Information Systems, Department of Geography, Northwestern University, Evanston, Illinois, 1966.

21. Lewis, T., "Evaluation of Multiple Polarized Radar Imagery for the Detection of Selected Cultural Features," U.S. Department of Interior, Geological Survey, Technical Letter NASA-130, NASA Contract No. W-12589, 1968.

22. Marble, D. F. and F. E. Hortin, "Remote Sensing and the Study and Planning of Urban Areas, NASA, Earth Resources Aircraft Program Status Review Vol. I, Manned Spacecraft Center, Houston, Texas, 1968.

23. Morain, S. A., and D. S. Simonett, "K-Band Radar in Vegetation Mapping," Photogrammetric Engineering, July 1967, 730-740.

24. Schwarz, D. E., D. S. Simonett, G. F. Jenks, and J. R. Ratzlaff, "The Construction of Thematic Land Use Maps with Spacecraft Photography," Department of Geography, Center for Research in Engineering Science, University of Kansas, U.S.G.S. Contract USGS 14-08-0001-10848, 1968.

FACTOR ANALYSIS APPLIED TO MULTISPECTRAL DATA

Everard G. A. Barnes and Francis G. Capece
Geographic Information Systems Branch
Geographic Sciences Division
U. S. Army Engineer Topographic Laboratories
Fort Belvoir, Virginia

ABSTRACT. Several techniques are used by the Bendix Aerospace Corporation under contract to the U. S. Army Engineer Topographic Laboratories (USAETL) to analyze remote sensing data collected with the Bendix multispectral scanner. The primary technique used by Bendix was factor analysis. This technique separates the underlying physical phenomena which affect the signals recorded by the multispectral scanner. After these "factors" are separated and enhanced images can be produced which show the individual and/or collective effects of each isolated factor on the original data. The manner in which a factor affects the original data may be used to enhance or emphasize those features which are determined to be of interest. Scatter diagrams are also formed which show whether two factors, when plotted against each other, separate or cluster the data points. If clustering occurs then additional enhanced imagery termed window imagery can be produced. This technique has useful application in analyzing multisensor data.

Results of the contract study, plus examples of original imagery, enhanced imagery and scatter diagrams are included.

I. INTRODUCTION. A contract for the exploitation of multispectral techniques for the location of engineer construction materials was awarded in March 1969 to the Bendix Corporation. Multisensor flights were carried out in March 1969 over an area of SW Louisiana (Figure 1). The data discussed in this paper were acquired using a Bendix Multispectral Scanner, Bendix Thermal Mapper, and a 70 mm Hulcher camera mounted in a Beechcraft D18S aircraft.

Analysis techniques were applied solely to the multispectral scanner data. Data samples were stored, during analysis, in an IBM 360/50 computer. The bulk of the data analyzed were obtained over a small area of the Delta south of Lake Charles, Louisiana, which contains many beach ridges. The portion of terrain selected for the detailed analysis is known as Pumpkin Ridge (Figure 2).

II. INVESTIGATION.

1. Data Acquisition Equipment. The airborne components of the Bendix system include a multispectral scanner, an electronic subsystem and an analog tape recorder. (Figure 3). The scanner has eight channels operating in the spectral range from 0.38 (ultraviolet) to 1.0 microns (near IR). Tapes recorded inflight have eight channels of video data and one channel of synchronized signal. Some processing of the multispectral scanner signals are performed inflight. The video signal produced by the scanner has a nominal bandwidth so FM recording of the signal was necessary to preserve information over the entire frequency range. Since the AR 1600 recorder was not an FM tape recorder frequency modulation of the viedo signal was accomplished on the electronic subsystem and then passed to the AR 1600 for analog recording.

Once collected, the video signals are displayed in the laboratory by playing the tape back through a wide band FM system. (Figure 4A). An Ampex FR-1800H tape recorder is utilized for this purpose. At this point single channel imagery may be produced by directing the video signal from the desired channel to a film recorder. Digitized samples may also be taken from the analog tapes. (Figure 4B). This was accomplished by taking a 5-μ sec sample simultaneously from each channel at the output of the recorder. This sampling point occurs at a constant, adjustable, time delay from the leading edge of the video signal. When the video signals are film recorded side by side, these sampling points appear as a straight line in the 120^{o} field of view of the scanner and running the total length of the flight line. These samples are recorded on a digital tape transport and subjected to analysis techniques. Thus the analyzed data contained sampling points from both the important features and their background and should qualify as a representative sample of the imagery.

2. Analysis Techniques. The technique chosen by Bendix in this investigation was a form of factor analysis known as principal components analysis. The technique was chosen for the following reasons:

(1) With factor analysis no a priori knowledge was required and ground truth is used solely to infer conclusions about the derived variables. Thus data of an unknown origin may be analyzed and enhanced imagery produced without information about the source of the data.

(2) The eight channels of the MS possess high first order correlations; they are not independent. In addition, an examination of the bandpass of the eight channels reveals a certain amount of overlap.

(3) Due to some properties of the approach which will be mentioned later, the application of the technique usually results in fewer "meaningful" variables than the original number of variables (channels).

3. Derivation. Assume that we have a data matrix X consisting of m rows each representing an MS channel and n columns each representing an observation. This matrix has correlation between elements as we have stated earlier. The object of factor analysis is to find that matrix which will transform the data matrix X into a new matrix with uncorrelated elements. The new matrix we will call the factor score matrix and the matrix that does the transforming we will call the factor coefficient matrix.

In order to work with the data it must be in standard form. The means of the rows of X must be computed and subtracted from appropriate elements in X. The matrix must then be multiplied by the inverse of the standard deviation matrix related to the matrix X. This new matrix now in standard form we shall call our standardized data matrix X_S. Multiplying X_S by its transpose X_S^T and divided through by n to normalize the matrix we get the correlation matrix R

$$R = \frac{1}{n} X_S X_S^T$$

We now find the eigenvalues and eigenvectors associated with R. Since R is real and symmetric, the eigenvalues are positive definite and the matrix of eigenvectors U is orthogonal ($UU^T = I = UU^{-1}$). By definition $URU^{-1} = D$ where D is a diagonal matrix of eigenvalues of R.

$D = \text{diag}\ (\lambda_1, \lambda_2, \ldots, \lambda_8)$ where λ_i's are the eigenvalues of R

Since U is orthogonal

$URU^T = D$

Defining $E = \text{diag}\ (\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_8})$ then

$EE = D$

$URU^T = EE$

Pre- and post-multiplying by E^{-1} we get

$E^{-1}URU^TE^{-1} = I$

Substituting for R we get

$$* \qquad E^{-1}UX_S X_S^T U^T E^{-1} = nI$$

If we define the factor score matrix to be $F = E^{-1}UX_S$ then it must have the property of independence between elements. By definition

$$F^T = X_S^T U^T (E^{-1})^T = X_S^T U^T E^{-1}$$

If the elements in F are uncorrelated then the off-diagonal elements of its correlation matrix must be 0's. Referring back to equation * and substituting we get

$$FF^T = E^{-1}UX_S X_S^T U^T E^{-1} = nI$$

The off-diagonal elements are 0 so the elements of F are uncorrelated. The new matrix F has m rows and n columns but now the rows represent factors rather than MS channels. Geometrically, factor analysis defines a transformation from one set of axes (channels) to a new set (factors). The factor coefficient matrix that does the transforming of X_S is $P = E^{-1}U$ an 8x8 matrix whose rows represent factors and columns represent MS channels. The set of measurements in a row of P represent a vector which assigns weights to the different channels for a particular factor.

This anlysis associates direction of maximum data variation with the first variable. This occurs because eigenvalues of R represent an indication of amount of variance contained in the factors. The higher the eigenvalue, the greater the variance. With eight eigenvalues their sum = 8. If the sum of the first three eigenvalues equal 7 then $\frac{7}{8}$ or 87.5% of total data variation is contained in the first three factors. These may possibly be the only variables worth considering. The underlying trends or factors cannot be defined except possibly with ground truth information.

Rotated analyses are also derived from the initial factor coefficient matrix. A straightforward rotation of axes following an exact analytic criterion results in a new set of factors. Two rotating procedures, Varimax and Oblimin rotations, were applied. The Varimax procedure attempts to simplify factors by extracting components that tend toward unity or zero while maintaining zero correlations of factors. The Oblimin procedure minimizes a similar expressions without the requirement of maintaining zero correlation. 1/

1/ H. Harman, Modern Factor Analysis

4. Manipulation. After the factor coefficient matrix and factor score matrix have been determined these matrices can be used several ways. Using the factor coefficient matrix a set of enhanced images possessing a continuous scale of grey levels can be produced which show the effect the individual factors had on the original image. A row of the factor coefficient matrix yields processing constants that when slightly modified to fit the analog system are used to yield enhanced signals. Each channel of the original recorded data is delivered to a separate processor and multiplied by the corresponding element of the factor coefficient matrix. The modified channels are summed forming a new signal which is a linear combination of the eight channels. In vector language multiplying each channel and their summation is comparable to a dot product. Imagery is then produced using a fiber optics cathode ray tube (CRT). Black and white areas on the imagery indicate correlations (+ or -) with the factor; while grey areas in the imagery indicate little correlation.

Scatter diagrams can also be produced by using the factor score matrix (the data matrix multiplied by the factor coefficient matrix). Two rows of this matrix each representing an individual factor are plotted in pairs. If by plotting these factors the distributions of several targets don't appreciably overlap, clusters are formed which indicate that certain features are separated by these two factors and hence some type of enhanced imagery may be producible. Usually the first few factors yield the most distinct target clusters, especially when two adjacent factors are related. By establishing boundaries around clusters corresponding video signals representing these boundaries may be applied to an analog system as threshold voltages. With the present system only two processed video signals may be thresholded simultaneously. The imagery resulting from the two dimensional decision process is termed window imagery. This window imagery will be a binary image (black versus white) enhancing those points within the threshold limits, preferably the target desired.

If the threshold of target A includes no other target spectra, the technique will work perfectly. If not, the threshold levels must be set to either minimize incorrect decisions or guarantee the inclusion of all of Target A. The choice of approach would depend upon whether the problem posed required minimizing the false alarm rate or the miss rate.

III. DISCUSSION. The objective of applying factor analysis was to produce enhanced imagery which would contribute to the ability to distinguish between marshes, beach ridges, and open water. The area we will analyze will be Pumpkin Ridge (Figure 5).

An example of imagery produced from the scanner data before analysis techniques were applied is shown in Figure 6.

The eight channels appear from left to right with the ninth film strip representing the sum of the eight channels. The average spectral properties of the three terrain types are represented in Figure 7 on a quantizing scale of 0-256. Note that in considering standard deviation no single channel could effectively separate the three terrain features. The lower graph shows the same results on a voltage scale.

Figure 8 shows the results of application of principal components analysis to the scanner data. Images representing the effects of the individual factors (1-8) are shown from left to right. The matrix of weights used to produce the imagery of Figure 8 is shown in Figure 9. This is a form of the factor coefficient matrix produced in factor analysis. The original factor coefficient matrix is shown above. The values had to be transformed to fit the analog system.

Figure 10 shows the mean and standard deviation of the three features now plotted on a voltage scale indicating that factors 1 and 2 separate the features considerably and that factors 3-8 do not. The latter factors describe more subtle differences in the target spectra and may be useful for more difficult discrimination problems.

Concentrating further on Figure 9, several conclusions about the nature of the factors could be drawn. The eigenvalues related to factor 1 consistently ranged between 4 and 6.5 accounting for 50-80% of the variance in the collected data. Also the coefficient vector of factor 1 consistently gave a maximum weight to channel 1 (UV) and decreased to Channel 8 (Near IR). This factor was determined to be a measurement of the albedo.

The eigenvalues for factor 2 invariably ranged between 1 and 2.5 accounting for from 12 to 30% of the total variance of the collected data. The coefficient vector associated with factor 1 usually weighted the UV and blue-green channels against red IR (high negative values versus high positive values). Thus on the average 70-90% of the variance of data was contained in the first 2 factors. It would seem likely then that scatter diagrams (previously discussed) would result in maximum clustering using

these two factors. A scatter diagram that resulted from plotting these two factors in pairs against each other is shown in Figure 11.

Figure 12 contains scatter diagrams ralating other factors to each other. Figure 11 definitely shows the greater clustering of the data.

The scatter diagrams strongly indicate that the windowing technique mentioned earlier can result in an enhanced image with the important features enhanced. Finding the approximate threshold values from the scatter diagram and applying window processing, each feature emerged distinctly (Figure 13).

A color enhanced photo of Pumpkin Ridge was created from the three enhanced photos (Figure 14). The example given demonstrates the application of factor analysis and window processing. Most often two features would be uniquely distinguished, but this example shows the possible merit of the technique.

IV. CONCLUSION. The purpose of this project was to develop procedures for automatic recognition of certain geologic features. Since the final decision of the photointerpreter concerning the appearance of a feature includes information about its surrounding elements, it would be up to the interpreter to decide whether or not the area selected by factor analysis actually is the designated feature. It seems evident that principal components analysis would have limitations caused by seasonality, regionality, moisture content, etc., which would limit the ability to distinguish features with a unique set of processing constants. More data would have to be collected over the same area at different times in the year and over different areas (both adjacent and distinct) under varying moisture conditions to determine the seriousness of the effects of seasonality and regionality on the processing constants so as to determine the generality of the solutions.

This analysis demonstrated that imagery which enhanced drainage patterns and manmade objects could be produced. It was also shown that byproducts of factor analysis, scatter diagrams and the window process, could sufficiently separate local marshes from beach ridges and open water as was illustrated in the Pumpkin Ridge slide. This, however, depended on how separable the spectra of the desired features became after factor analysis had been applied.

Thus the feasibility of such an approach has been shown even though the analysis was hindered because of data acquisition over a limited area at one point in time.

Analysis of multispectral data could be improved by:

1. Refining existing techniques including the development of special purpose hardware for implementing the techniques.

2. Applying additional techniques such as the recently popular Bayesian decision theory to the factor analyzed data.

3. Acquiring more data over the same area at different times and different areas under varying moisture conditions to study the generality of the procedure.

4. Improving existing hardware to allow more factors to be used simultaneously in the window process.

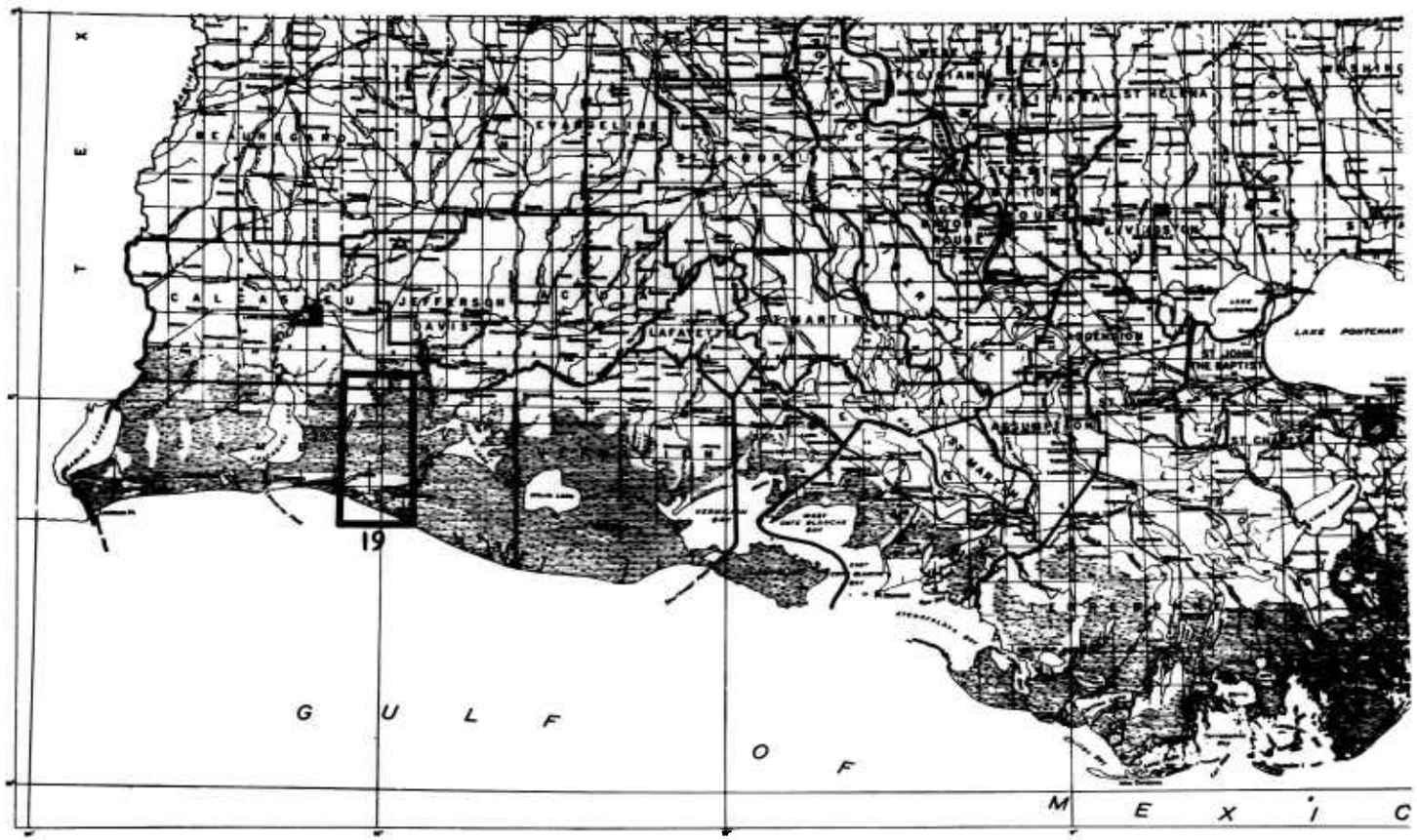
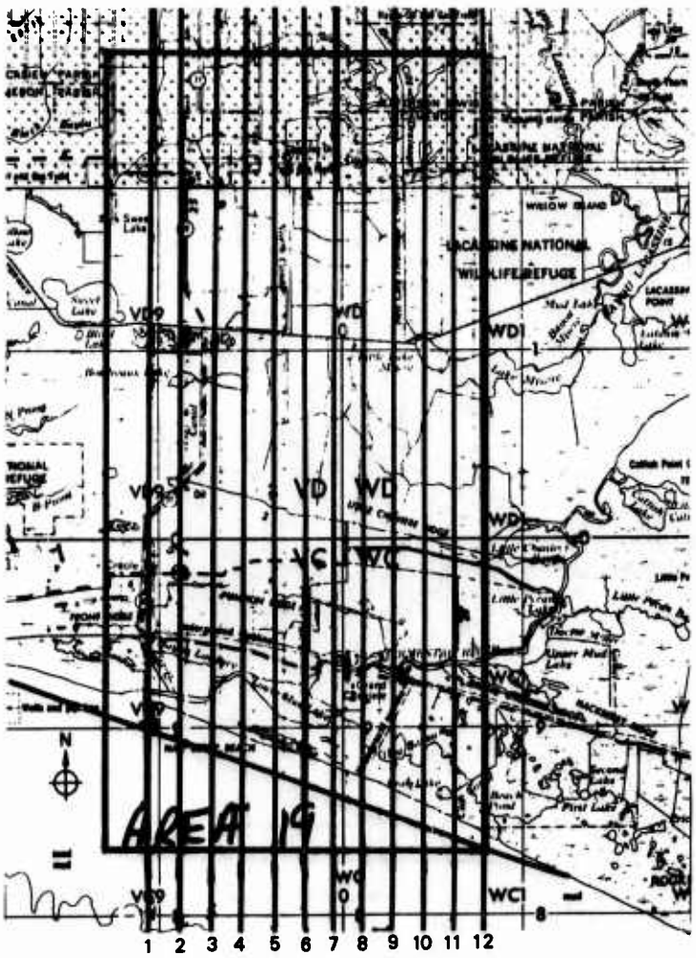
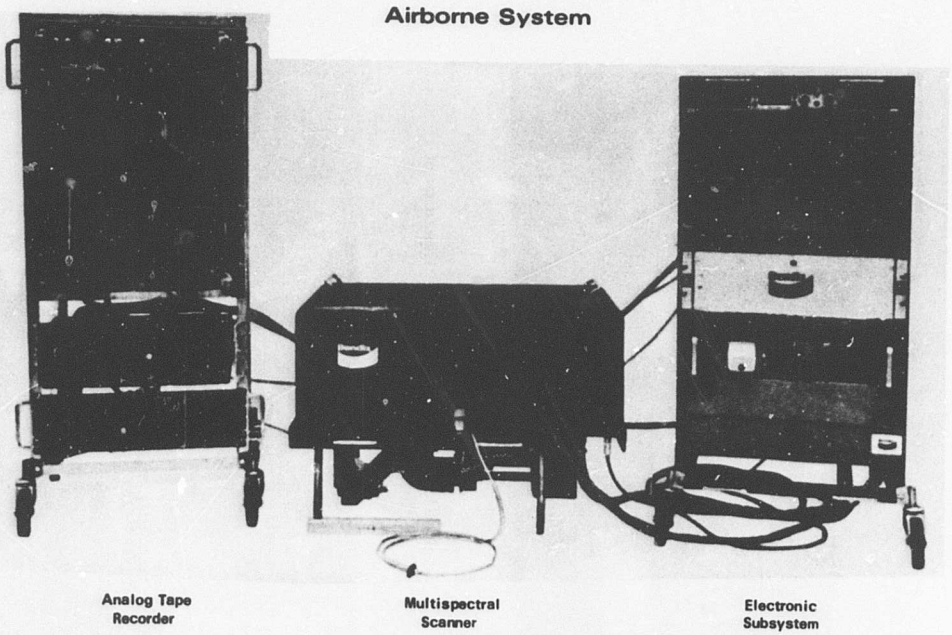

T E X
CALCASIEU
JEFFERSON
DAVIS
LAFAYETTE
ST MARTIN
ST HELENA
LAKE POINTCHART
ST JOHN THE BAPTIST
ASSUMPTION
ST CHARLES
19
G U L F O F M E X I C

FIG. 1



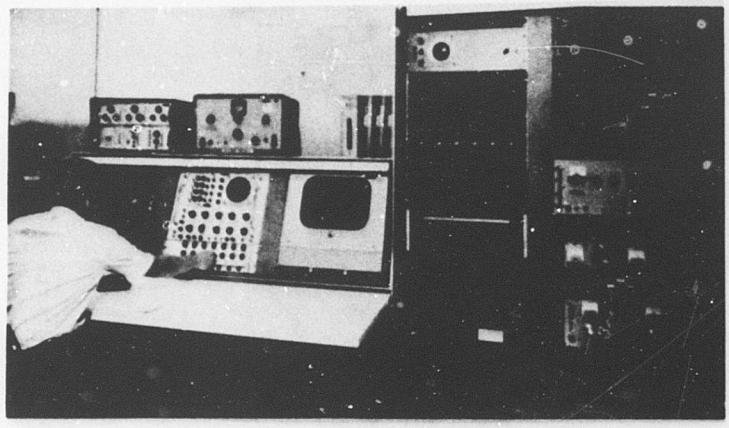
Area 19 of Lake Charles, La.

FIG. 2



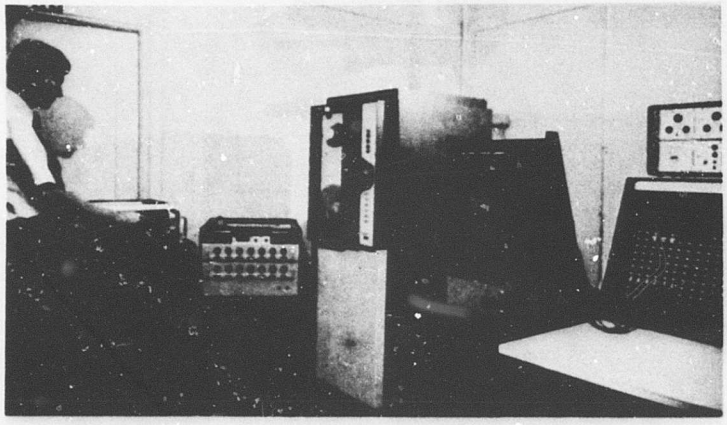
Multispectral Scanner Airborne System

FIG. 3



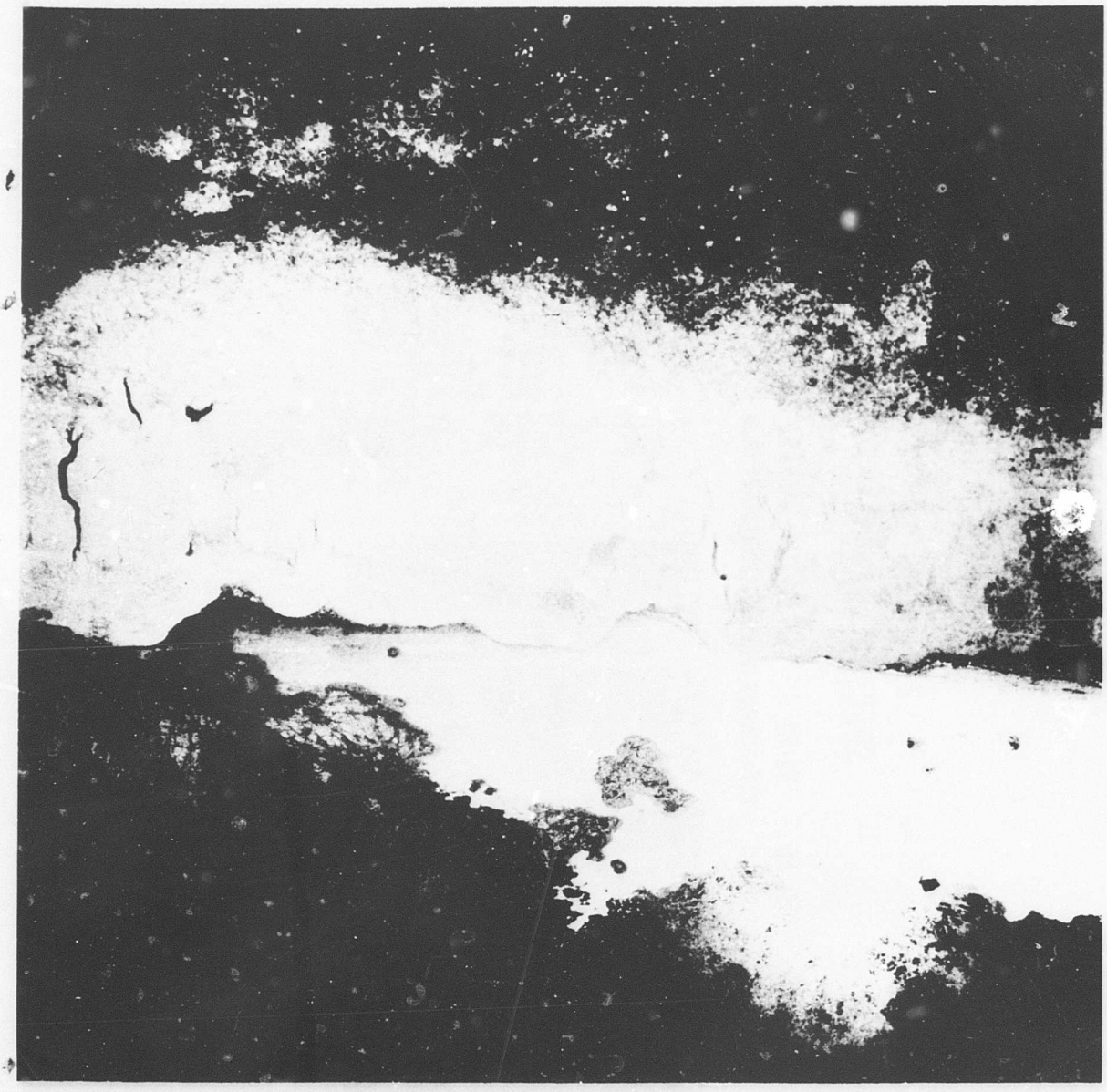
Analog Data Processing Equipment

FIG. 4A



Digital Data Sampling

FIG. 4B



AERIAL PHOTO OF PUMPKIN RIDGE

FIG. 5

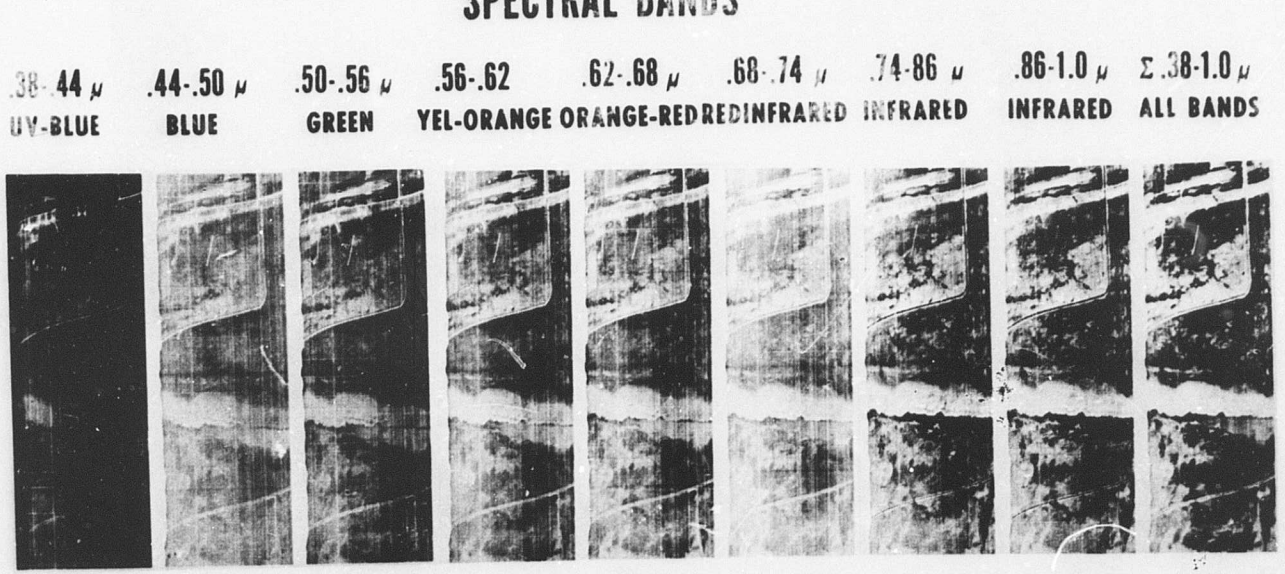
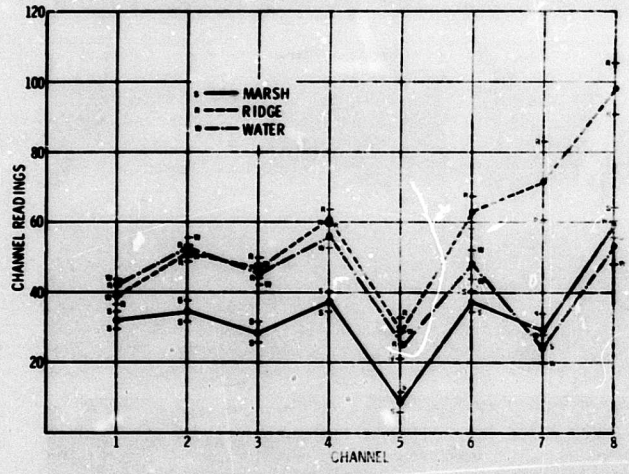
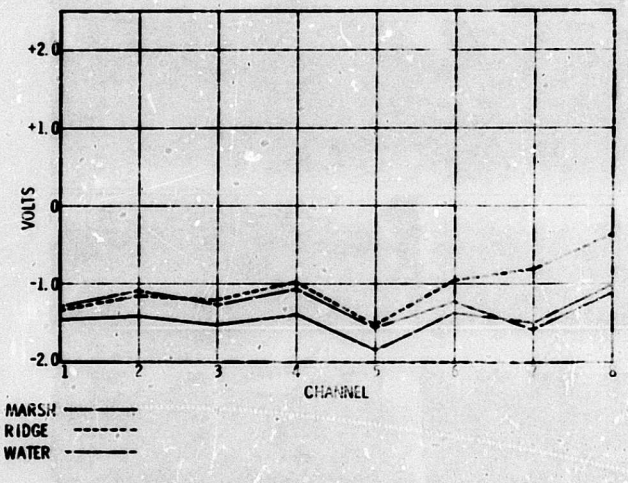

SPECTRAL BANDS
.38-.44 μ
UV-BLUE
.44-.50 μ
BLUE
.50-.56 μ
GREEN
.56-.62
YEL-ORANGE
.62-.68 μ
ORANGE-RED
.68-.74 μ
REDINFRARED
.74-86 μ
INFRARED
.86-1.0 μ
INFRARED
Σ .38-1.0 μ
ALL BANDS

FIG. 6

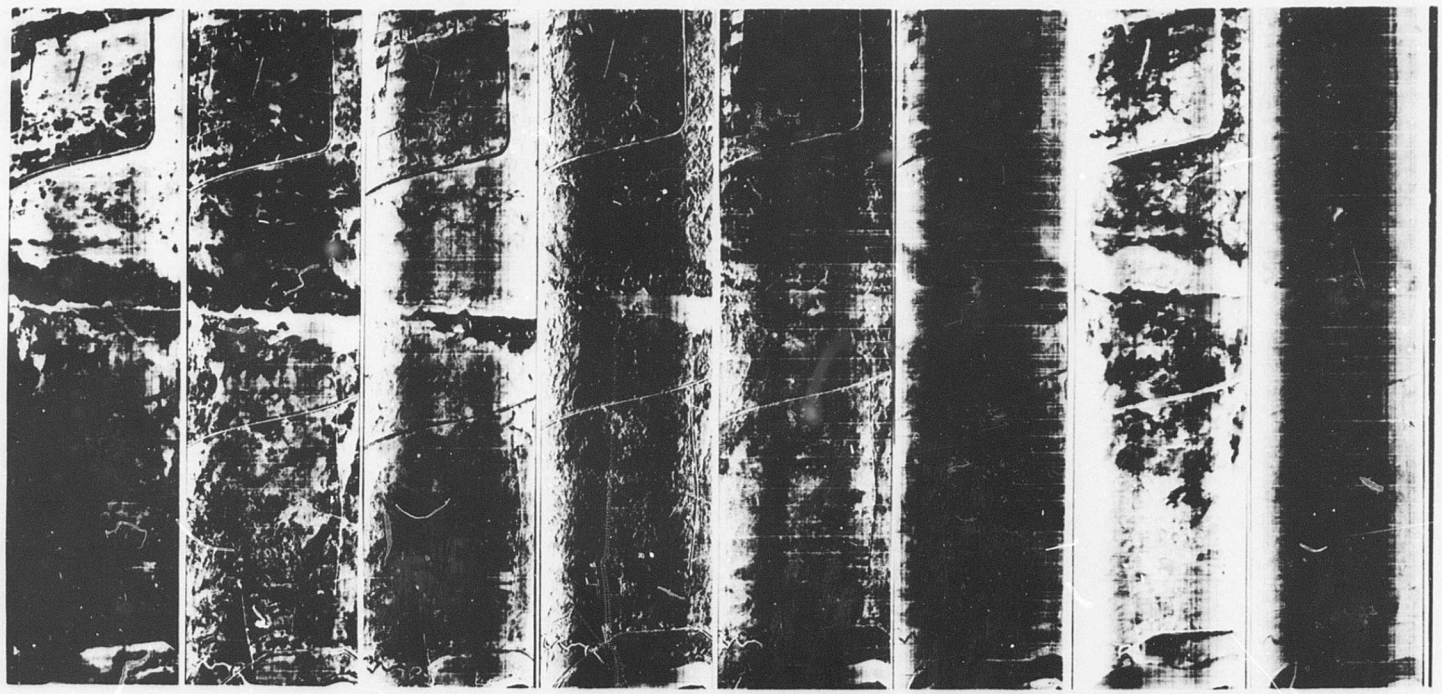


Reflectance Spectrum of Three Terrain Types



Reflectance Spectrum of Three Terrain Types

FIG. 7



FACTOR 1 FACTOR 2 FACTOR 3 FACTOR 4 FACTOR 5 FACTOR 6 FACTOR 7 FACTOR 8

FIG. 8 ENHANCED IMAGERY

FACTOR SCORE COEFFICIENT MATRIX : F = $[R_{yy}]^{-1/2}$ $[U_S]$ x NON-STANDARD DATA
BEFORE ROTATION

FACTOR								
1	0.0e311525	C.05018C4C	C.04805221	0.03869818	0.04301363	0.02661941	-0.00033496	-0.00145523
2	-0.03035871	-3.01340371	0.00054889	-0.00077515	-0.00232242	0.05632480	0.02609513	0.03658922
3	-0.32083c61	-C.07480233	0.C7998471	-0.03495803	C.13372675	0.09116446	-0.01939069	-C.02232248
4	-0.33221261	2.09522453	0.09C54[illegible]7	0.17870401	0.00517C21	-0.21671682	0.00667132	0.04756227
5	-0.252C9C3	0.23657463	-C.17C6R	C.12C62447	-0.21326976	0.20233593	-0.00792485	-0.05735499
6	C.02574732	3.14273535	-3.45505.	-0.01563437	0.33C69999	-0.01030115	-0.02851002	0.05776538
7	0.14253416	-3.48711943	-0.14923741	C.34086339	C.02392145	0.07346974	-0.00531413	-C.03025044
8	0.05558575	-3.03053382	0.08424301	0.01506477	-0.15874576	0.10899829	-0.13324678	0.15456613

CHANNEL COEFFICIENT SETTINGS
BEFORE ROTATION

FACTOR								
1	10.00000000	7.95059756	7.61340658	6.13135153	6.81509273	4.21758827	-0.05307114	-0.23056664
2	-5.38993717	-2.37971666	0.09745166	-0.13762179	-C.41232560	10.00000000	4.63297333	6.49611243
3	-1C.0CCC0000	-2.33148165	2.49300902	-1.08959183	4.16807180	2.84146586	-0.60438010	-0.69575981
4	-1C.0C00000C	2.56537643	1.52151596	5.37923857	C.15563058	-6.52347674	0.20081598	1.43169015
5	-1C.00000000	9.36804763	-6.77317452	9.79462129	-8.46308490	8.02920267	-0.31447810	-2.27599139
6	0.56541132	3.13632766	-10.00000000	-0.34796896	7.26731031	-0.22637330	-0.62652293	1.26942539
7	3.05134302	-1C.00003000	-3.19484483	7.42558267	0.51210558	1.57282568	-0.11376393	-0.64759560
8	3.75373544	-2.30140453	5.30678824	C.94898744	-1C.00000000	6.86621714	-8.39372223	9.73670915

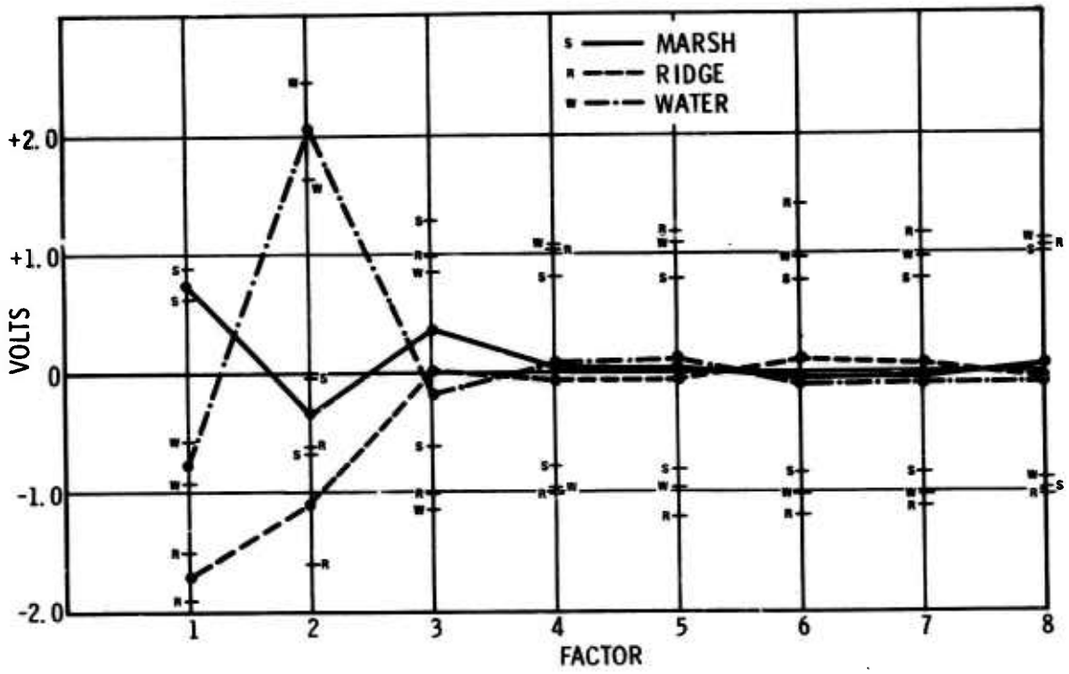
EIGENVALUES

4.33614188	2.30072E15	C.52255243	C.27732164	0.19615682	C.13141152	0.11891965	0.11676792

CUMMULATIVE PROPORTION OF TOTAL VARIANCE

3.54201773	C.82560E75	C.89492781	0.92459301	0.954112e1	0.97053505	0.98540401	1.00000000

FIG. 9



Factor Score Distribution of Three Terrain Types

FIG. 10

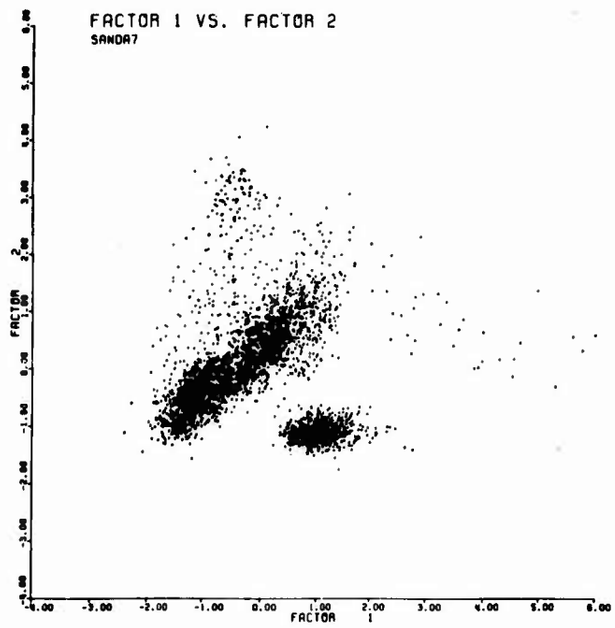
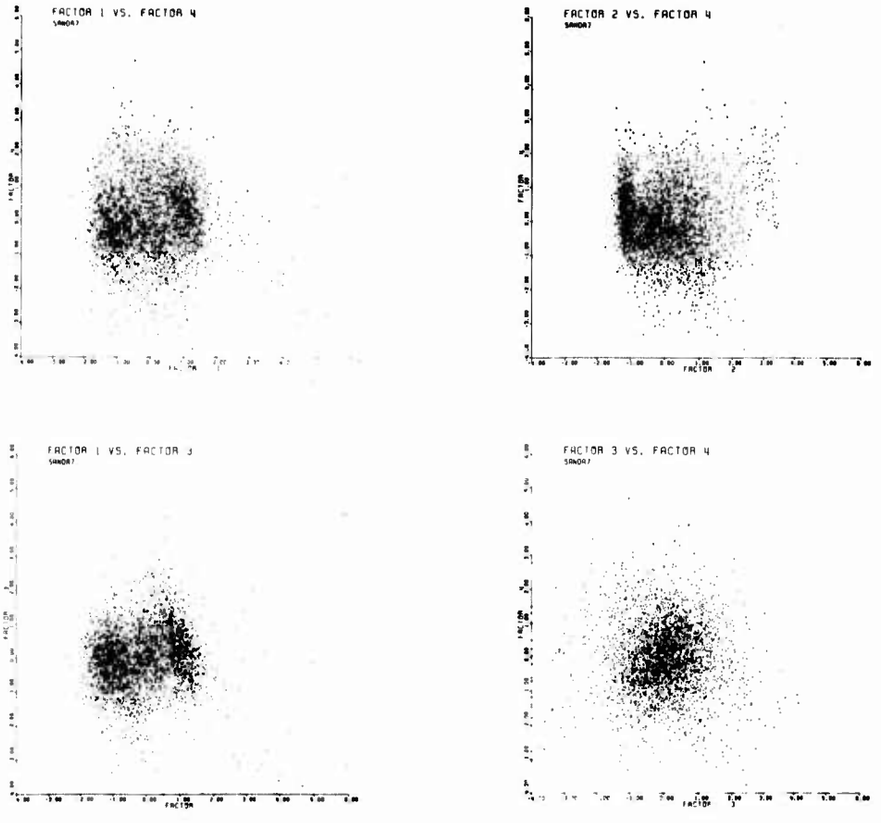
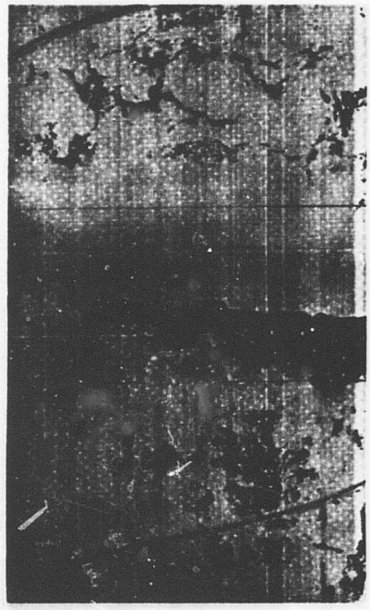


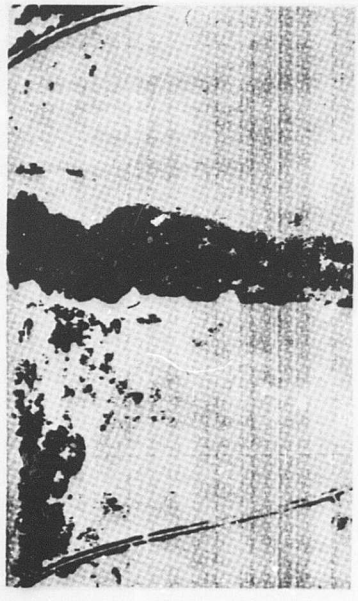
FIG. 11

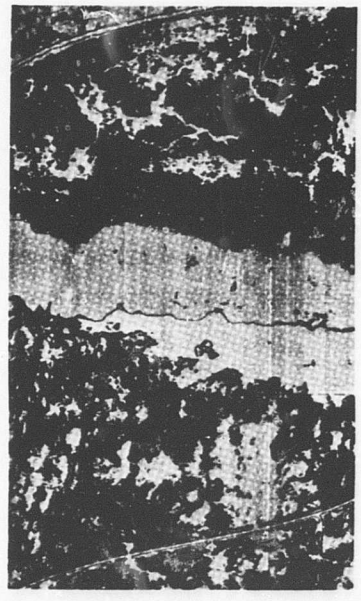


Scatter Diagrams of Factor Pairs

FIG. 12



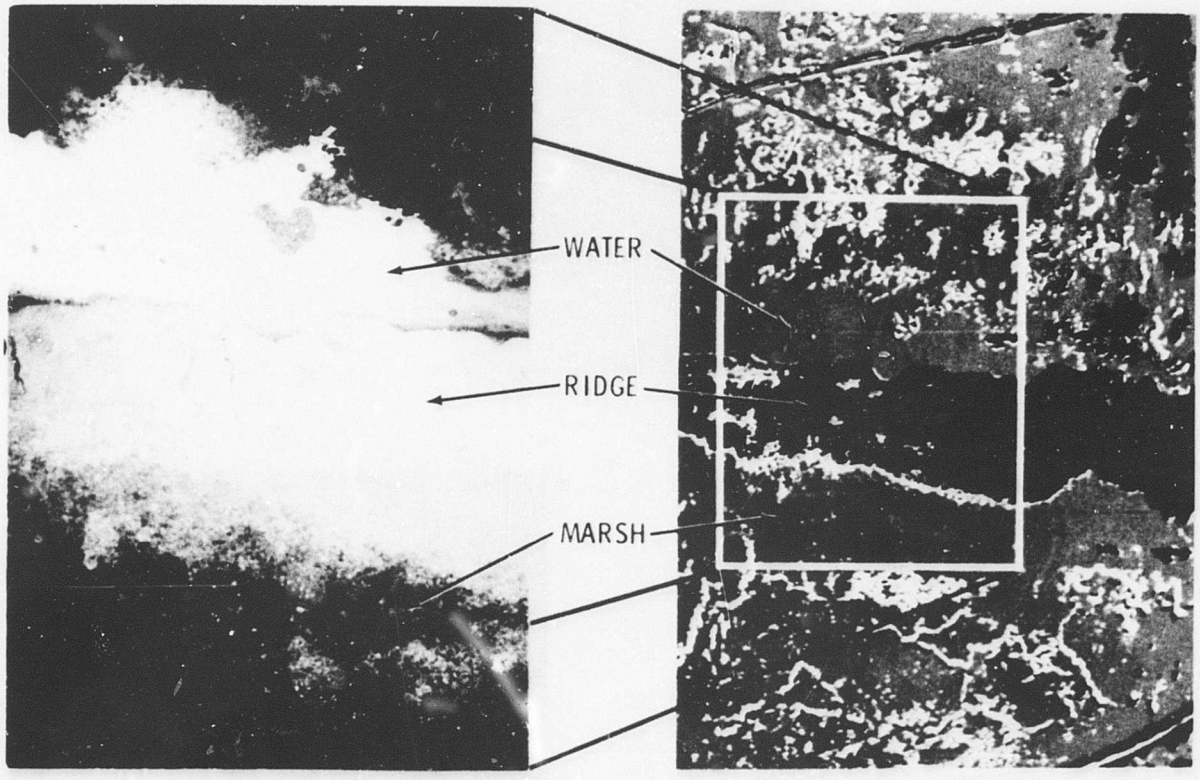




Separation of Pumpkin Ridge into Three Materials

FIG. 13

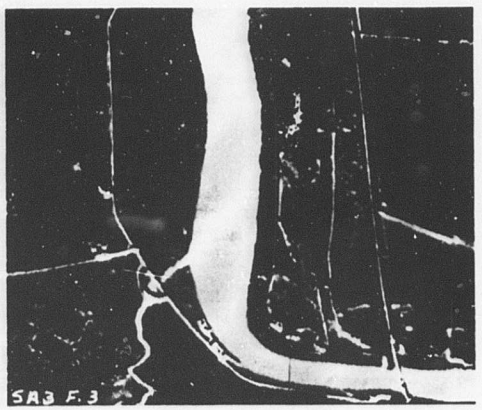
False Color Classification of Multispectral Data From Computer Analysis



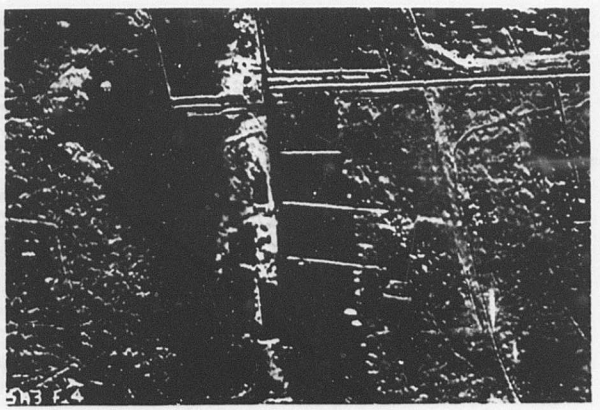
THIS BEACH RIDGE IS A POTENTIAL LOCATION OF SAND AND GRAVEL DEPOSITS. MAPPING OF POTENTIAL DEPOSITS SIMPLIFIES THE SEARCH FOR ENGINEERING CONSTRUCTION MATERIALS. THE PHOTOGRAPH WAS MADE ON AERIAL EKTACHROME FILM NEAR LAKE CHARLES, LOUISIANA AT AN ALTITUDE OF 2000 FEET ON MARCH 24, 1969.

THE COLOR CODED IMAGE IS BASED ON COMPUTER ANALYSIS OF MULTISPECTRAL SCANNER VIDEO OF THE SAME LOCATION. THE CLASSIFICATION OF MARSH, WATER, AND BEACH RIDGE IS PERFORMED AUTOMATICALLY TO MAP TARGETS OF INTEREST OVER EXTENDED AREAS.

FIG. 14



Sand A3 Factor 3
Enhanced Imagery



Sand A3 Factor 4
Enhanced Imagery

FIG. 15

BLANK PAGE

MIXTURE PROBLEMS IN PATTERN RECOGNITION

William John Sacco
Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

ABSTRACT. This paper is concerned with the analysis of a class of problems involving mixture variates (functions of several random variables). The explicit inclusion of a mixture as a separate category in pattern recognition problems is studied, and plausible forms for the mixing densities in such problems are derived. Methods for the estimation of mixing parameters and mixing distributions, using mathematical programming techniques, are presented.

TABLE OF CONTENTS

The remainder of this article has been reproduced photographically from the author's manuscript.

INTRODUCTION

This paper is concerned with the analysis of a class of statistical problems involving "mixtures". The problems studied include the explicit consideration of mixtures in pattern recognition, and the estimation of mixing distributions. (Mixture problems have also been extensively studied in pattern recognition in connection with unsupervised estimation[1,2,3,4].)

Let $K(w)$ and $L(u|w)$ be distribution functions defined for w in a set W. The distribution function $H(u)$ is said to be the mixture of $L(u|w)$ and $K(w)$ if

$$H(u) = \int_W L(u|w)dK(w) \ . \tag{1}$$

In this expression, $K(w)$ is called the mixing distribution. Similarly, let $k(w)$ and $\ell(u|w)$ be density functions defined for $w \in W$; then the density function $h(u)$ is called the mixture of $\ell(u|w)$ and $k(w)$ if

$$h(u) = \int_W \ell(u|w)k(w)dw \ , \tag{2}$$

and $k(w)$ is called the mixing density.

Let the dependence of u on w be given by an expression of the

form

$$u = g(w,t_1,\ldots,t_m) \tag{3}$$

where $t_1,\ldots,t_m$ may themselves be functions of w. If $w,t_1,\ldots,t_m$ are random variables of known densities, then the density of u can be expressed in the form (2), where $\ell(u|w)$ can be derived from (3) and the known densities; and analogously for distributions.

In pattern recognition, if u is an observed measurement which we wish to classify, we can apply statistical decision-theoretic methods if we know the density of u under the hypotheses that u comes from each of the classes in question. A special case of pattern recognition is signal (or target) detection; here there are usually considered to be two classes, "noise" and "signal". Under many circumstances, it is more realistic to consider the possibility of a mixture class, part noise and part signal. Here the measurement u will depend on a mixing variate w, as well as on other variates which may be functions of w. Thus the problem of determining the density of u becomes one of computing a mixture density, given the densities of w and the other variates.

In Chapter 1 of this paper we review basics of pattern recognition which are needed in the remainder of the paper. Chapter 2 discusses recognition and detection situations in which various simple types of mixing occur, as expressed by simple forms of the dependence (3); methods of determining h(u) in these cases are presented. Chapter 3 derives plausible forms for the mixing density

k(w) in such situations.

The pattern classification problem treated in Chapter 2 involves assigning a measurement u to one of a set of classes, where the classes are defined in terms of values of a mixing variate w. This problem can be reformulated as that of estimating w, given u. Chapter 4 discusses the computation of maximum-likelihood estimators for w for several of the cases treated in Chapter 2.

Chapter 5 introduces the problem of estimation of mixing distributions. Let Ω be a set of distributions K(w), each defined on W, and let H(u) be defined by (1). Given a set $u_1,\ldots,u_n$ of independent measurements of u, we wish to determine the mixing distribution $K(w) \in \Omega$ which produced H(u). Conditions under which H(u) determines a unique K(w) ("identifiability") are reviewed. In general, the empirical distribution $H_n(u)$ defined by $u_1,\ldots,u_n$ may not itself be of the form (1); however, we prove, under suitable assumptions on Ω, that there always exists a unique H(u) of the form (1) which is "closest" to $H_n(u)$ in a certain sense.

Chapter 6 reviews the methods which have been used to estimate mixing distributions, and formulates mathematical programming approaches to the estimation problem.

CHAPTER 1

PRELIMINARIES ON PATTERN RECOGNITION

In this chapter we review some basic facts about pattern recognition. The pattern classification problem is formulated as a problem of statistical hypothesis testing, and the optimum Bayes discriminant functions are defined. For the two-class case, e.g. the case of signal detection, the concept of an Operating Characteristic (OC) curve is introduced; this curve describes the tradeoff between the two possible types of error. The sequential approach to pattern classification is formulated, and a method for its solution using dynamic programming is given.

The problem of pattern recognition is often said to consist of two parts: (1) the extraction of pattern features (the characterization problem) and (2) the optimum classification of pattern classes (the classification problem). The former is concerned with finding a comprehensive set of distinguishing characteristics and the latter deals with the problem of making optimal decisions in classifications.

We adopt the usual approach of characterization which determines a mapping of an event into a set of real variables $x_1, x_2, \ldots, x_d$, called attributes, which hopefully characterize the event. It is convenient to represent the pattern by a point X in d-dimensional space. Thus $X = (x_1, x_2, \ldots, x_d)$. Most successful solutions in the field of pattern recognition hinge on the judicious selection of attributes. Unfortunately there are few general results that can serve as guides when selecting attributes.

A pattern classifier[5] is a device which assigns a vector $X = (x_1, x_2, \ldots, x_d)$ to one of several categories $R_1, R_2, \ldots, R_n$. An objective of the pattern classification scheme is to select a set of scalar functions $d_1(X), d_2(X), \ldots, d_n(X)$, called discriminant functions, so that for all $X \in R_i$

$$d_i(X) > d_j(X) \quad i,j=1,2,\ldots,n, \quad j \neq i$$

This is equivalent to partitioning X space into n sets. For each X in R_i, $d_i(X) > d_j(X)$ for all $j \neq i$.

Two approaches for determining discriminant functions are frequently used; the non-parametric (non-statistical) approach, and the

statistical approach. In this paper we shall restrict ourselves to the statistical approach.

In the statistical approach, the classification problem reduces to that of statistical hypothesis testing. Each pattern class, R_i, is assumed to have two probability functions, $p(X|i)$ and $p(i)$, associated with it. The function $p(X|i)$ is the probability of occurrence of pattern X given that it is sampled from class i. The function $p(i)$ is the probability of occurrence of class i. A pattern classifier is said to be optimum in a Bayes sense if decisions are made according to a rule which minimizes the average "loss" of decision making. The discriminant functions which are used to implement Bayes decision making are

$$d_i(X) = - \sum_{j=1}^{R} \lambda(i|j)p(X|j)p(j)$$

where $\lambda(i|j)$, the so-called "loss" function, represents the loss incurred when the classifier places a pattern actually belonging to the class R_j into category R_i.

The classifications are made in the following manner:

i. The vector X is presented to the classifier.

ii. The classifier computes $\text{Max}[d_i(X)] = d_k(X)$, say, and assigns the vector X to class R_k. For the special important case $n = 2$ and $\lambda(1|1) = \lambda(2|2) = 0$, the decision rule reduces to the following:

$$\text{If } \frac{p(X|1)}{p(X|2)} > \frac{p(2)\lambda(1|2)}{p(1)\lambda(2|1)} \text{ , assign X to class } R_1$$

Otherwise, assign X to R_2.

In many applied problems the $\lambda(i|j)$ and $p(i)$ are not available. For this reason it is important to consider decision rules which do not explicitly include this information. In the Appendix we shall have occasion to use Operating Characteristic (OC) curves[6] which require prior knowledge only of $p(X|1)$ and $p(X|2)$ for the two-class problem. For the two-class problem we wish to partition all of X space into two sets Ω_1 and Ω_2, where

$$\Omega_1 = \{X|X \text{ is assigned to } R_1\} ,$$

and

$$\Omega_2 = \{X|X \text{ is assigned to } R_2\}.$$

Associated with the partition are two error probabilities,

$$\alpha = \int_{\Omega_1} p(X|2)dX ,$$

and

$$\beta = \int_{\Omega_2} p(X|1)dX .$$

The quantity α is the probability of assigning a sample X to class R_1 when it truly belongs to class R_2, and vice versa for β.

Thus, for each partition there is associated the errors α and β. An OC curve is a plot of α versus β for a class of partitions.

As an example, let X be a scalar quantity; let $p(X|1)$ be a normal distribution with mean μ_1 and standard deviation σ_1, and let $p(X|2)$ be a normal distribution with mean μ_2 and standard deviation σ_2 (see Figure 1.1).

For each value X^* on the line we have a partition, $\Omega_1 = \{X| \ -\infty < X \leq X^*\}$ and $\Omega_2 = \{X|X^* < X < \infty\}$. Moreover,

$$\alpha = \int_{-\infty}^{X^*} \frac{1}{\sqrt{2\pi}\,\sigma_2} \exp - \left[(X - \mu_2)^2/2\sigma_2^2\right] dX \ .$$

$$\beta = \int_{X^*}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_1} \exp - \left[(X - \mu_1)^2/2\sigma_1^2\right] dX \ .$$

From Figure 1.1 it can be seen that as X^* increases, α increases and β decreases. This relationship between α and β can be plotted as a curve. This curve is called the OC curve and is illustrated in Figure 1.2.

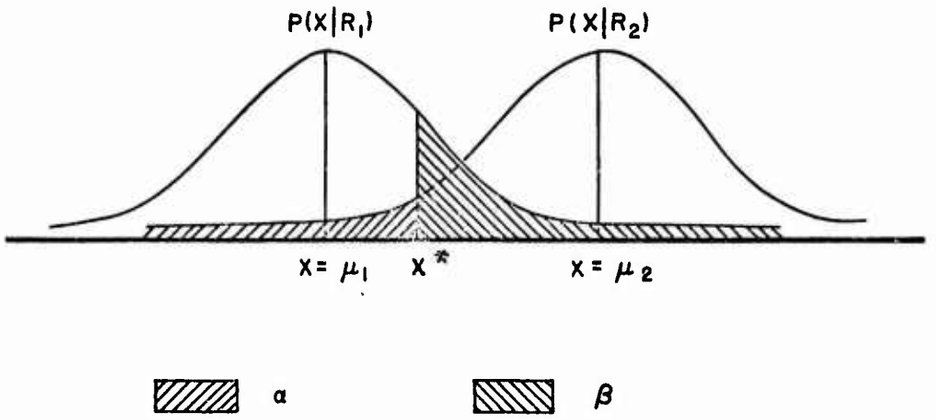


Figure I.I Probability Density Functions

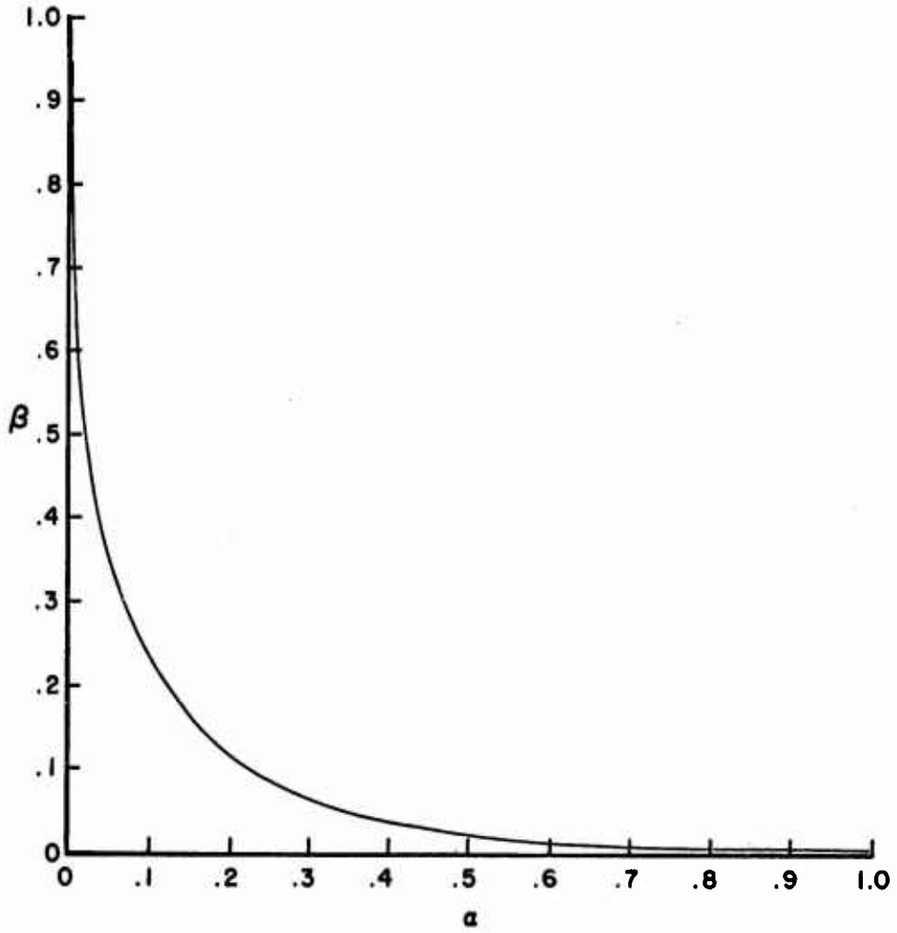


Figure 1.2 Operating Characteristic (OC) Curve

In many situations the classication decision need not be made on the basis of a single measurement. Thus the observer need not always make one of the decisions D_i (that class R_i is present) but instead may make the decision W ("wait for another measurement"). The loss incurred by this delay will in general be dependent on which class R_i we are measuring. Let W_i equal the delay loss incurred if R_i is truly the class being measured. The objective of the observer is to minimize the expected cost. The decision policy that achieves the minimum expected cost is called the "optimal" policy.

A more realistic problem arises when one considers that there exist at most (say) n available observations remaining before a terminal decision D_i must be made. If the decision W is made, at the next decision there will be only n-1 possible observations left. If n = 0, then one of the terminal decisions must be made. Let $f_n(p_1,p_2)$ be the cost of search using an optimal policy, given that there are n available observations remaining before a terminal decision must be made. We can write a functional equation which will yield the optimal policy. The equation is an application of Bellman's Principle of Optimality.[7] In the case of three classes, the equation is

$$f_n(p_1,p_2) = \min \begin{cases} p_2\lambda(1|2) + p_3\lambda(1|3); & \text{decision } D_1 \\ p_1\lambda(2|1) + p_3\lambda(2|3); & \text{decision } D_2 \\ p_1\lambda(3|1) + p_2\lambda(3|2); & \text{decision } D_3 \\ p_1W_1 + p_2W_2 + p_3W_3 + \int_0^{\infty} \sum_{i=1}^{3} p_i p(x|R_i) f_{n-1}(p_1',p_2')\, dx\,; & \text{decision } W. \end{cases}$$

For n = 0,

$$f_0(p_1,p_2) = \min \begin{cases} p_2\lambda(1|2) + p_3\lambda(1|3); & D_1 \\ p_1\lambda(2|1) + p_3\lambda(2|3); & D_2 \\ p_1\lambda(3|1) + p_2\lambda(3|2); & D_3 \end{cases}$$

If the decision is $\begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix}$ respectively, the expected losses incurred due to terminal incorrect decisions are $\begin{Bmatrix} p_2\lambda(1|2) + p_3\lambda(1|3) \\ p_1\lambda(2|1) + p_3\lambda(2|3) \\ p_1\lambda(3|1) + p_2\lambda(3|2) \end{Bmatrix}$ respectively.

If the decision is W, the expected cost is $p_1W_1 + p_2W_2 + p_3W_3$ plus the cost of continuing from that point on, having observed some value of x. The probability of observing a value between x and x + dx is $[p_1p(x|R_1) + p_2p(x|R_2) + p_3p(x|R_3)]\, dx$. Having observed the value x, the probabilities

$$p_1' = p(R_1|x) = \frac{p(x|R_1)p_1}{p_1p(x|R_1) + p_2p(x|R_2) + p_3p(x|R_3)},$$

and

$$p_2' = p(R_2|x) = \frac{p(x|R_2)p_2}{p_1p(x|R_1) + p_2p(x|R_2) + p_3p(x|R_3)}$$

are obtained from Bayes rule.

CHAPTER 2

COMPUTATION OF MIXTURE DENSITIES

In classical detection problems, it is usual to consider two alternatives or states: (S_1) the target (or other event of interest) alone is present; (S_2) background (or noise) alone is present. In many situations, alternative (S_1) is unrealistic; rather, it is appropriate to consider an alternative (S_3) in which a <u>mixture</u> of target and background is present.

Suppose, for simplicity, that we are making only a single measurement, so that the vector X of Chapter 1 becomes a scalar, call it u. (This restriction will be relaxed at the end of the chapter). We wish to decide whether an observed measurement u arises from alternative (S_2) or alternative (S_3). In order to use the methods of Chapter 1, we need to know the probability density function of u under hypotheses (S_2) and (S_3).

The density function of u under hypotheses (S_3) will in general depend on the proportion w in which target and background are mixed. In this chapter, we consider several specific forms for this dependence, namely

$$(1) \quad u = wr + (1-w)s$$

$$(2) \quad u = wr(w) + (1-w)s(w)$$

$$(3) \quad u = r(w) + s(w)$$

where r and s, as well as w, are random variables. In the following sections we shall provide motivation for the forms (1) - (3), and shall derive probability density functions for u assuming various density functions for w, r, and s.

2.1 $u = wr + (1-w)s$

Suppose one wishes to scan a photograph or a natural scene with a sensor having a field of view of fixed size and shape in order to identify an event of interest. For each positioning of the sensor one obtains a measurement u. The problems that one can consider vary in accordance with the properties of the sensor. The first class of problems that we shall discuss involves a sensor which is assumed to measure an average "intensity" over its field of view. Each element in the field of view is assumed to receive equal weight in the averaging process. An example of this situation occurs when one wishes to detect a target embedded in foliage. One can visualize the field of view V of the sensor as partitioned into two regions, R_1 and R_2, (see Figure 2.1) where R_1 contains target elements only and R_2 contains foliage only. (Note that R_1 and R_2 need not necessarily be connected regions.) Let w be the fraction of V containing target elements, and let $w = \frac{\text{area of } R_1}{\text{area of V}}$. Then the measurement u taken over the region V can be represented as $u = wr + (1-w)s$, where

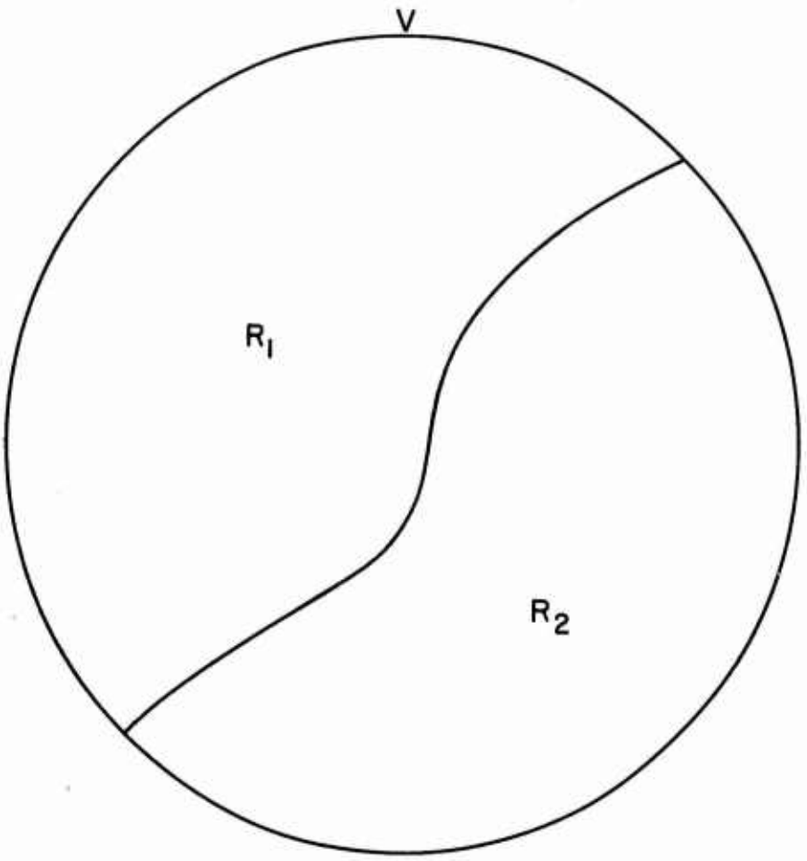


Figure 2.1 Field of View of Sensor

r = the average intensity over R_1 ,

s = the average intensity over R_2 ,

and

$$0 \leq w \leq 1 .$$

The quantities r, s, and w are assumed to be random variables.

Given the measurement u, we wish to decide whether state S_2 or state S_3 obtains. In order to use the statistical decision theory approach of Chapter 1 it will be necessary to compute the probability density function of u under hypothesis S_3. We denote this function by $p(u|S_3)$. Let $p(u|S_1)$ be the density function for u under hypothesis S_1 (corresponding to $w = 1$) and let $p(u|S_2)$ be the density function for u under hypothesis S_2 (corresponding to $w = 0$). To facilitate further discussion let $f_1(r) \equiv p(u|S_1)$, $f_2(s) \equiv p(u|S_2)$, and $h(u) \equiv p(u|S_3)$. We wish to determine h(u), where

$$u = wr + (1-w)s,$$

and w, r, and s are random variables with density functions k(w), $f_1(r)$, and $f_2(s)$, respectively. Let R be the range of r, S the range of s, and W the range of w.

In deriving h(u) we choose to fix the variables r and s at constant values initially. Then u behaves as a monotonic function of w when r and s are held constant. Thus the conditional density $\ell(u|r,s)$ can be obtained from the probability density, k(w), or w by a univariate method of derived distributions[8] yielding

$$\ell(u|r,s) = k\left(\frac{u-s}{r-s}\right) \frac{1}{|r-s|}$$

Moreover, one can derive the joint density function $g(u,r,s)$ from the expression

$$g(u,r,s) = f(r,s)\ell(u|r,s)$$

where $f(r,s)$ is the joint density function of r and s. Finally the density of u is obtained by integrating $g(u,r,s)$ with respect to r and s. Thus we have

$$h(u) = \int\int_{T(u)} f(r,s)k\left(\frac{u-s}{r-s}\right) \frac{1}{|r-s|}\, dr\, ds$$

where

$$T(u) = \{(r,s)|r \in R,\ s \in S, \text{ and } \frac{u-s}{r-s} \in W\} .$$

Letting $R = [c,d]$, $S = [a,b]$, $W = [0,1]$ and assuming that r, s, and w are independent, we obtain the expression

$$\begin{aligned} h(u) = &\int_{\gamma(u)}^{d}\int_{a}^{\beta(u)} f_1(r)f_2(s)k\left(\frac{u-s}{r-s}\right)\left(\frac{1}{r-s}\right) ds\, dr \\ &+ \int_{\gamma'(u)}^{b}\int_{c}^{\beta'(u)} f_1(r)f_2(s)k\left(\frac{u-s}{r-s}\right)\left(\frac{1}{s-r}\right) dr\, ds, \end{aligned} \tag{2.1.1}$$

where

$$\beta(u) = \max\{a,\min(u,b)\},\ \gamma(u) = \min\{d,\max(c,u)\} ,$$

$$\beta'(u) = \max\{c,\min(u,d)\},\ \gamma'(u) = \min\{b,\max(u,a)\} .$$

The two integrals evolve as a result of partitioning T(u) into two sets $T_1(u)$ and $T_2(u)$, where

$$T_1(u) = \{(r,s) | r \in R,\ s \in S,\ r > s, \text{ and } 0 \leq \frac{u-s}{r-s} \leq 1\} ,$$

and

$$T_2(u) = \{(r,s) | r \in R,\ s \in S,\ r < s, \text{ and } 0 \leq \frac{u-s}{r-s} \leq 1\} ,$$

or equivalently,

$$T_1(u) = \{(r,s) | r \in R,\ s \in S, \text{ and } s \leq u \leq r\} ,$$

and

$$T_2(u) = \{(r,s) | r \in R,\ s \in S, \text{ and } r \leq u \leq s\} .$$

Example 1. Suppose

(i) r and s are constants.

(ii) $s > r$.

(iii) w is uniformly distributed on $[\alpha,\beta]$ where $0 \leq \alpha < \beta \leq 1$ (see Chapter 3 on the plausibility of this).

Then u is uniformly distributed on $[s - \beta(s-r),\ s - \alpha(s-r)]$.

Example 2. Suppose

(i) r is a constant.

(ii) s and w are statistically independent.

(iii) s is uniformly distributed on [a,b].

(iv) w is uniformly distributed on [0,1].

Then

$$h(u) = \int_{T(u)} \frac{1}{b-a} \left(\frac{1}{1-w}\right) dw ,$$

or

$$(b-a)h(u) = \int_{T(u)} \frac{1}{1-w}\, dw ,$$

where

$$T(u) = \{w \mid 0 \le w \le 1 \text{ and } a \le \frac{u-wr}{1-w} \le b\} .$$

We consider three subcases:

(i) $r < a$. Then

$$T(u) = \begin{cases} \left[\frac{a-u}{a-r}, \frac{b-u}{b-r}\right] & \text{for} \quad r \le u \le a \\ \left[0, \frac{b-u}{b-r}\right] & \text{for} \quad a \le u < b \end{cases}$$

and

$$(b-a)h(u) = \begin{cases} \ell n\left(\frac{b-r}{a-r}\right) & \text{for} \quad r \le u \le a \\ \ell n\left(\frac{b-r}{u-r}\right) & \text{for} \quad a \le u \le b \end{cases}$$

The graph of this function has the form

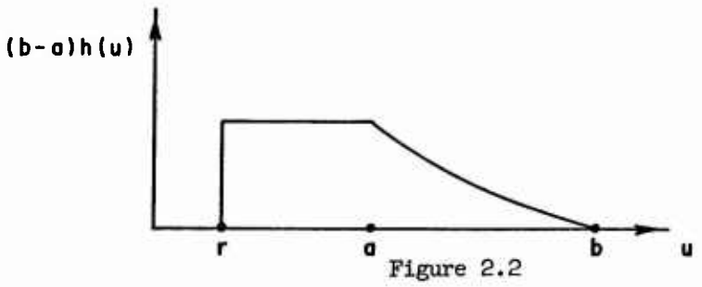


Figure 2.2

(ii) $a < r < b$. Then

$$T(u) = \begin{cases} (0, \frac{u-a}{r-a}) & \text{for} \quad a < u < r \\ (0, \frac{b-u}{b-r}) & \text{for} \quad r < u \leq b \end{cases}$$

and

$$(b-a)h(u) = \begin{cases} \ell n(\frac{r-a}{r-u}) & \text{for} \quad a < u < r \\ \infty & \text{for} \quad u = r \\ \ell n(\frac{b-r}{u-r}) & \text{for} \quad r < u < b \end{cases}$$

whose graph has the form

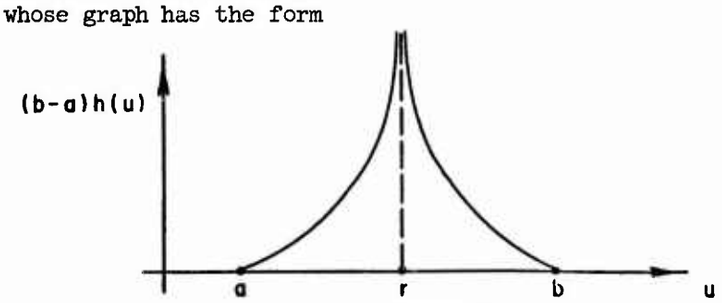


Figure 2.3

(iii) $r > b$. Then

$$T(u) = [a, \min(u,b)],$$

and

$$(b-a)h(u) = \begin{cases} \ell n(\frac{r-a}{r-u}) & \text{for} \quad a \leq u \leq b \\ \ell n(\frac{r-a}{r-b}) & \text{for} \quad b \leq u \leq r \end{cases}$$

The graph of $(b-a)h(u)$ has the form

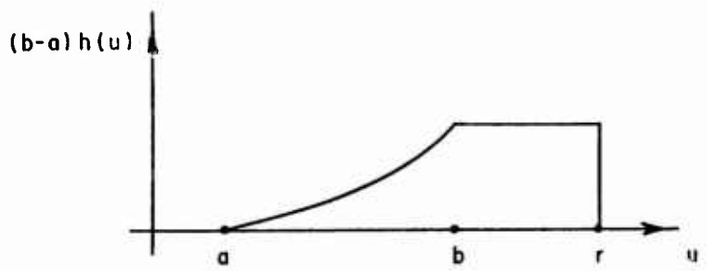


Figure 2.4

Example 3. Suppose

(i) r, s, and w are statistically independent, uniformly distributed random variables.

(ii) R = [c,d], S = [a,b], and W = [0,1].

(iii) For definiteness, let $0 < c < a < d < b$.

Then

$$(d-c)(b-a)h(u) = I_1(u) + I_2(u) ,$$

where,

$$I_1(u) = \begin{cases} 0 & \text{for } u \le a \\ (d-a)\ell n(d-a) + [\gamma(u) - \beta(u)]\ell n[\gamma(u) - \beta(u)] & \\ -[d-\beta(u)]\ell n[d-\beta(u)] - [\gamma(u)-a]\ell n[\gamma(u)-a] & \text{for } u > a \end{cases}$$

and

$$I_2(u) = (b-c)\ell n(b-c) + [\gamma'(u)]\ell n[\gamma'(u) - \beta'(u)]$$
$$-[b-\beta'(u)]\ell n[b-\beta'(u)] - [\gamma'(u)-c]\ell n[\gamma'(u)-c]$$

Furthermore,

$$(d-c)(b-a)h(u) = \begin{cases} k_1 + h_1(u) & \text{for} \quad c \le u < a \\ k_2 - h_2(u) & \text{for} \quad a \le u \le d \\ k_3 + h_3(u) & \text{for} \quad d < u \le b \end{cases}$$

where

$$k_1 = (b-c)\ell n(b-c) - (a-c)\ell n(a-c) ,$$

$$k_2 = (d-a)\ell n(d-a) + (b-c)\ell n(b-c) ,$$

$$k_3 = (b-c)\ell n(b-c) - (b-d)\ell n(b-d) ,$$

$$h_1(u) = (a-u)\ell n(a-u) - (b-u)\ell n(b-u) ,$$

$$h_2(u) = (d-u)\ell n(d-u) + (u-a)\ell n(u-a) + (b-u)\ell n(b-u) + (u-c)\ell n(u-c) ,$$

and

$$h_3(u) = (u-d)\ell n(u-d) - (u-c)\ell n(u-c) .$$

Let c=1, a=2, d=3, and b=4. Then $h(u) = \dfrac{I_1(u) + I_2(u)}{4}$, and $k_1 = k_2 = k_3 = 3\ \ell n\ 3$. The graph of h(u) is given in the following figure.

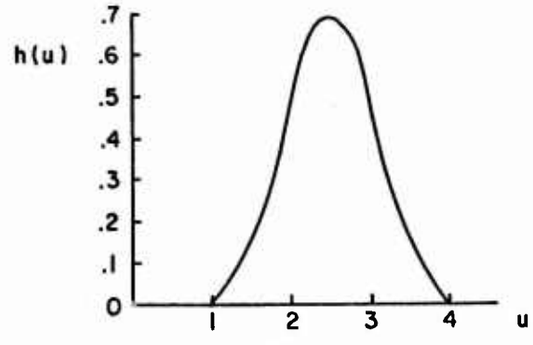


Figure 2.5

Example 4. Suppose

(i) r, s, and w are independent random variables.

(ii) r and s are normally distributed.

(iii) $f_1(r) \equiv N(\mu_1, \sigma_1^2)$; $f_2(s) \equiv N(\mu_2, \sigma_2^2)$.

Then

$$h(u) = \int_W k(w)\ell(u|w)dw$$

where

$$\ell(u|w) \equiv N(w\mu_1 + (1-w)\mu_2 \; ; \; w^2\sigma_1^2 + (1-w)^2\sigma_2^2) \; .$$

Example 5. Suppose

(i) r, s, and w are independent random variables.

(ii) r and s are distributed exponentially.

(iii) $f_1(r) \equiv \lambda_1 e^{-\lambda_1 r}$; $f_2(s) \equiv \lambda_2 e^{-\lambda_2 s}$ $\quad r,s \geq 0$

Then

$$h(u) = \int_W k(w)\ell(u|w)dw \; ,$$

where

$$\ell(u|w) = \frac{1}{w} \int_{T(u,w)} f_2(s) f_1\left(\frac{u-(1-w)s}{w}\right) ds \; ,$$

and

$$T(u,w) = \{s: 0 \le s \le \frac{u}{1-w}\} .$$

Substituting for f_1 and f_2, integrating and simplifying, we obtain

$$\ell(u|w) = \frac{\lambda_1\lambda_2}{(1-w)\lambda_1 - w\lambda_2}\left[e^{-\frac{\lambda_2 u}{1-w}} - e^{-\frac{\lambda_1 u}{w}}\right] .$$

Example 6. The Step Function Case.

In many pattern recognition problems, probability density functions for r and s are approximated by histograms obtained from sample measurements. With this in mind we define

$$f_1(r) = \sum_{i=1}^{m} a_i H(r-r_i)$$

and

$$f_2(s) = \sum_{j=1}^{n} b_j H(s-s_j)$$

where

$$H(x) = \begin{cases} 1 & ; \ x > 0 \\ 0 & ; \ x \le 0 \end{cases} .$$

We shall determine the function h(u) with $f_1(r)$ and $f_2(s)$ defined as above, and with a uniform density function for w. It follows from equation (2.1.1) that

$$h(u) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j [I_1(i,j,u) + I_2(i,j,u)]$$

where

$$I_1(i,j,u) = \int_{\gamma_i(u)}^{a} \int_{s_j}^{\beta_i(u)} k\left(\frac{u-s}{r-s}\right) \frac{1}{r-s}\, ds\, dr ,$$

and

$$I_2(i,j,u) = \int_{\gamma_j(u)}^{b} \int_{c}^{\beta_j(u)} k\left(\frac{u-s}{r-s}\right) \frac{1}{s-r}\, dr\, ds ,$$

and

$$\beta_j(u) = \max [s_j, \min(u,b)] ,$$

$$\gamma_i(u) = \min [d, \max(r_i,u)] ,$$

$$\beta_i(u) = \max [r_i, \min(u,d)] ,$$

$$\gamma_j(u) = \min [b, \max(u,s_j)] .$$

For w uniformly distributed on [0,1], we obtain

$$I_1(i,j,u) = (d-s_j)\ell n(d-s_j) + [\gamma_i(u) - \beta_j(u)]\ell n[\gamma_i(u) - \beta_j(u)]$$
$$- [d - \beta_j(u)]\ell n[d - \beta_j(u)] - [\gamma_i(u) - s_j]\ell n[\gamma_i(u) - s_j] ,$$

and

$$I_2(i,j,u) = (b-r_i)\ell n(b-r_i) + [\gamma_j(u) - \beta_i(u)]\ell n[\gamma_j(u) - \beta_i(u)]$$
$$- [b - \beta_i(u)]\ell n[b - \beta_i(u)] - [\gamma_j(u) - r_i]\ell n[\gamma_j(u) - r_i] .$$

2.2 u = wr(w) + (1-w)s(w)

In Section 2.1 we considered the "mix" variate u, where

$$u = wr + (1-w)s.$$

This model represented a sensor which obtains an average of "intensities" over its field of view and the average intensities r and s were not dependent on w.

One can conceive of situations when r and s would be dependent on w. Indeed, for an integrating sensor, the value of w corresponds to a sample size in the sense that a value of w close to 0 leads to greater variability in the measurement of r, while a value close to 1 makes the measurement of s more variable. This would give rise to an expression of the form

$$u = wr(w) + (1-w)s(w) .$$

Our objective is again to derive an expression for the density function h(u). The quantities r, s, and w are assumed to be random variables with densities $f_1(r)$, $f_2(s)$, and k(w), respectively. The variate u is a monotonic function of s when r(w) and w are held constant. To obtain the distribution of u we imagine that the two variates r(w) and w are held fixed at some arbitrary value while s varies over its possible range according to its distribution. The probability distribution of u which then results is the conditional distribution of u given r(w) and w. Moreover, u becomes a monotonic function of the variable s although the equation relating u and s will contain the parameters w and r(w). Hence the conditional density $\varphi[u|w, r(w)]$ can be derived from $f_2(s|w)$ by a univariate method[8] giving $\varphi[u|w, r(w)] = \frac{1}{1-w} f_2\left[\frac{u-wr(w)}{1-w} \,\middle|\, w\right]$. Furthermore one can then produce the joint density function, $g[u,r(w)|w]$, of u and r(w) conditioned upon w by the expression

$$g[u,r(w)|w] = \frac{1}{1-w} f_1[r(w)|w] f_2\left[\frac{u-wr(w)}{1-w} \,\middle|\, w\right] .$$

One obtains the density of u (conditioned upon w) by integrating the function $g[u,r(w)|w]$ with respect to r(w), taking pains to keep the argument of the function f_2 in the range of the variable s. This yields

$$\ell(u|w) = \frac{1}{1-w} \int_{T(u,w)} f_1[r(w)|w] f_2\left[\frac{u-wr(w)}{1-w} \,\middle|\, w\right] dr .$$

In the above, the set $T(u,w) = \{r(w)|r(w) \in R(w)$ and $\frac{u-wr(w)}{1-w} \in S(w)\}$, where R(w) is the range of r(w) and S(w) is the range of s(w). If we let $R(w) = [c(w), d(w)]$, and $S(w) = [a(w), b(w)]$, then

$$T(u,w) = \{r|c(w) \le r \le d(w) \text{ and } a(w) \le \frac{u-wr}{1-w} \le b(w)\} .$$

From the inequalities $a(w) \le \frac{u-wr}{1-w} \le b(w)$, we obtain $(1-w)a(w) + wr(w) \le u \le (1-w)b(w) + wr(w)$. From Figure 2.6, we visualize the dependence of the limits of integration upon u and w.

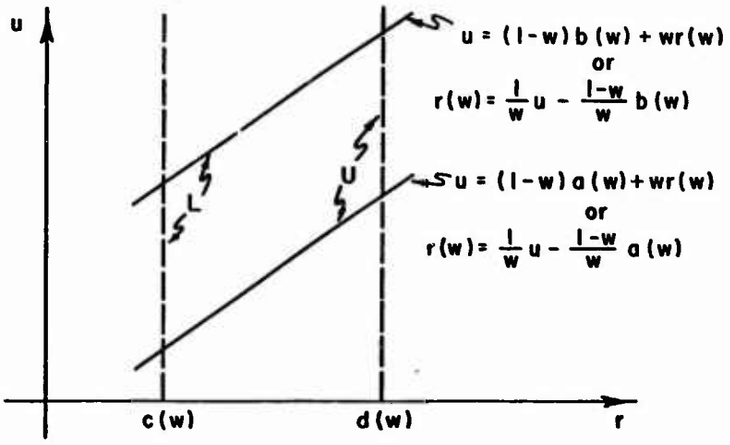


Figure 2.6

Hence

$$\ell(u|w) = \frac{1}{1-w} \int_{L(u,w)}^{U(u,w)} f[r(w)|w]\, f_2\left[\frac{u-wr(w)}{1-w} \;\middle|\; w\right] dr \ ,$$

with

$$L = \max\left[c(w);\ \frac{1}{w}\, u - \frac{(1-w)}{w}\, b(w)\right] \ ,$$

and

$$U = \min\left[\frac{1}{w}\, u - \frac{1-w}{w}\, a(w);\ d(w)\right] \ .$$

There are two cases to distinguish. If

$(1-w)a(w) + wd(w) \leq (1-w)b(w) + wc(w)$, then $T(u,w)$ is

$$\begin{cases} [c(w),\ \frac{1}{w}u - (\frac{1-w}{w})a(w)] \text{ for } (1-w)a(w) + wc(w) \leq u \leq (1-w)a(w) + wd(w) \\ [c(w),\ d(w)] \text{ for } (1-w)a(w) + wd(w) \leq u \leq (1-w)b(w) + wc(w) \\ [\frac{1}{w}\, u - \frac{1-w}{w}\, b(w),\ d(w)] \text{ for } (1-w)b(w) + wc(w) \leq u \leq (1-w)b(w) + wd(w) \ . \end{cases}$$

If $(1-w)a(w) + wd(w) \geq (1-w)b(w) + wc(w)$, then $T(u,w)$ is

$$\begin{cases} [c(w),\ \frac{1}{w}u - (\frac{1-w}{w})\, a(w)] \text{ for } (1-w)a(w) + wc(w) \leq u \leq (1-w)b(w) + wc(w) \\ [\frac{1}{w}u - (\frac{1-w}{w})b(w), \frac{1}{w}u - (\frac{1-w}{w})a(w)] \text{ for } (1-w)b(w) + wc(w) \leq u \leq (1-w)a(w) + wd(w) \\ [\frac{1}{w}u - (\frac{1-w}{w})b(w),\ d(w)] \text{ for } (1-w)a(w) + wd(w) \leq u \leq (1-w)b(w) + wd(w) \ . \end{cases}$$

Finally, to determine $h(u)$ we must integrate with respect to w; the result is

$$h(u) = \int_W k(w)\ell(u|w)\, dw \ .$$

2.3 $u = r(w) + s(w)$

Suppose the variable u is a count of the number of occurrences of a certain event in the field of view of a sensor or perhaps in a sub-region of a picture. Then a model for u which arises quite naturally is

$$u = r(w) + s(w) \ .$$

Examples of this model might arise as follows: Suppose that the sensor does not simply integrate the detected intensity over its field of view, but instead forms an image of it and analyzes the image. Here the analysis may involve features or subimages of a given size, e.g. edges of a given length or "blobs" of a given area, such that a single feature either lies entirely in R_1 or entirely in R_2. If what we are measuring is the <u>number</u> of such features, then the measured count over the field of view will be the sum of the counts over R_1 and over R_2. In general, these counts will depend on w, since if R_1 (say) is small, there may be no room for the features to occur in it.

Proceeding as in the earlier sections, we obtain

$$h(u) = \int_W k(w) \int_{T'(u,w)} f_1(r|w) f_2(u-r(w)) dr \, dw \ .$$

$$T'(u,w) = \{r \mid c(w) \le r \le d(w) \ ; \ a(w) \le w - r \le b(w)\}$$

As an example suppose r(w) and s(w) are independent random variables possessing Poisson distributions dependent on w. Then

$$f_1(r(w)) \equiv e^{-\lambda(w)} \frac{[\lambda(w)]^r}{r!} \qquad r=0,1,2,\ldots ,$$

$$f_2(s(w)) \equiv e^{-\beta(w)} \frac{[\beta(w)]^s}{s!} \qquad s=0,1,2,\ldots ,$$

and since the sum of two independent Poisson variates is a Poisson variate,

$$\ell(u|w) \equiv e^{-\gamma(w)} \frac{[\gamma(w)]^u}{u!} \qquad u=0,1,2,\ldots ,$$

where

$$\gamma(w) = \lambda(w) + \beta(w) .$$

For $\lambda(w) = w\lambda$, $\beta(w) = (1-w)\beta$, where λ and β are constants,

$$\ell(u|w) = e^{-[w\lambda+(1-w)\beta]} \frac{[w\lambda+(1-w)\beta]^u}{u!} ,$$

$$= e^{-\beta} e^{-[\lambda-\beta]w} \frac{[(\lambda-\beta)w+\beta]^u}{u!} .$$

a. Suppose w is distributed uniformly on [a,b] where $0 \le a < b \le 1$. Then

$$h(u) = \int_a^b k(w)\ell(u|w) = \frac{e^{-\beta}}{(b-a)u!} \int_a^b e^{-(\lambda-\beta)w}[(\lambda-\beta)w + \beta]^u \, dw.$$

Repeated integration by parts yields

$$h(u) = [(\lambda-\beta)(b-a)u!]^{-1}[e^{-p}\{p^u + up^{u-1} + u(u-1)p^{u-2} +\ldots+ u!\}$$

$$- e^{-q}\{q^u + uq^{u-1} + u(u-1)q^{u-2} +\ldots+ u!\}] ,$$

where

$$p = a\lambda + (1-a)\beta ,$$

and

$$q = b\lambda + (1-b)\beta .$$

b. Suppose w is distributed as follows:

$$k(w) = a_i + b_i w ; \quad w_i \le w \le w_{i+1} ; \quad i=1,2,\ldots,n-1$$

where $k(w)$ is continuous and

$$0 = w_1 < w_2 < \ldots < w_n = 1 .$$

Then

$$h(u) = \sum_{i=1}^{n} \int_{w_i}^{w_{i+1}} (a_i+b_i w)\ell(u|w)dw$$

$$= \sum_{i=1}^{n} \left[a_i \int_{w_i}^{w_{i+1}} \ell(u|w)dw + b_i \int_{w_i}^{w_{i+1}} w\ell(u|w)dw \right] .$$

Substituting for $\ell(u|w)$ and integrating both integrals by parts, repeatedly, we obtain

$$h(u) = [(\lambda-\beta)u!]^{-1} \sum_{i=1}^{n} a_i \left[e^{-p_i} \left\{ p_i^u + up_i^{u-1} + u(u-1)p_i^{u-2} + \ldots + u!p_i + u! \right\} \right.$$

$$\left. - e^{-q_i} \left\{ q_i^u + uq_i^{u-1} + u(u-1)q_i^{u-2} + \ldots + u!q_i + u! \right\} \right]$$

$$+ [(\lambda-\beta)^2 u!]^{-1} \sum_{i=1}^{n} b_i \left[e^{-p_i} \left\{ p_i^{u+1} + (u+1)p_i^u + (u+1)up_i^{u-1} + \ldots + (u+1)!p_i \right.\right.$$

$$\left. + (u+1)! \right\} - e^{-q_i} \left\{ q_i^{u+1} + (u+1)q_i^u + (u+1)uq_i^{j-1} + \ldots + (u+1)!q_i + (u+1)! \right\}$$

$$+ \beta e^{-q_i} \left\{ q_i^u + uq_i^{u-1} + u(u-1)q_i^{u-2} + \ldots + u!q_i + u! \right\}$$

$$\left. - \beta e^{-p_i} \left\{ p_i^u + up_i^{u-1} + u(u-1)p_i^{u-2} + \ldots + u!p_i + u! \right\} \right] ,$$

where

$$p_i = (\lambda-\beta)w_i + \beta ,$$

and

$$q_i = (\lambda-\beta)w_{i+1} + \beta .$$

Extension to a Set of Measurements

Let us suppose that for each positioning of the sensor one obtains several measurements, $x_1, x_2, \ldots, x_d$. Let $X = (x_1, \ldots, x_d)$. Let us assume that for each of the classes S_1, S_2, and S_3 the components of X are statistically independent. This allows us to write

$$p(X|S_i) = p(x_1|S_i)p(x_2|S_i)\ldots p(x_d|S_i); \quad i=1,2,3.$$

Each x_i is then a mixture variate and the theory discussed in Sections 2.1 to 2.3 is applicable.

CHAPTER 3

SOME PLAUSIBLE MIXING DENSITIES

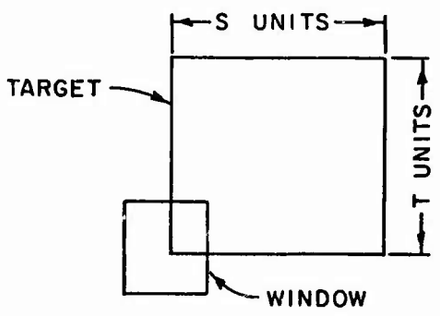
In this chapter we derive several plausible density functions, $k(w)$, for the random variable w.

Consider the problem of scanning, with an aperture or window, a two-dimensional photograph (or natural scene). Suppose

(i) The target and photograph are rectangular-shaped.

(ii) The window is square-shaped.

(iii) The edges of the window and target are parallel to the edges of the photograph as illustrated:



(iiii) Each possible position of the window is equally likely.

Let the window be a unit square and let the target be $S(> 1)$ units

long and T(> 1) units wide. Let w be the ratio

$$\frac{\text{area of overlap of target and window}}{\text{area of window}} .$$

We wish to find the probability density function, k(w), given there exists an overlap of window and target. Assuming that each admissible position of the window with respect to the target is equally likely, we can derive the expression

$$k(w) = \frac{S + T - 2 - 2 \ell n\, w}{S + T} \; ; \; 0 < w \leq 1 .$$

For S = T, we obtain

$$k(w) = \frac{S - 1 - \ell n\, w}{S} \; ; \; 0 < w \leq 1 .$$

As either S or T approaches infinity, k(w) approaches a uniform distribution on [0,1].

Suppose next that

(i) The target and window are circular-shaped with radii of R(> 1) units and one unit, respectively.

(ii) Each possible position of the window is equally likely.

(iii) w(s) is the ratio

$$\frac{\text{area of overlap of the target and window}}{\text{area of window}} .$$

Then

$$w(s) = \frac{S_1(s) + S_2(s)}{\pi}$$

$$S_1(s) = \begin{cases} \frac{\pi}{2} - \left[y(s) \sqrt{1-y^2(s)} + \sin^{-1} y(s) \right] & \text{for} \quad 0 \leq y(s) \leq 1 \\ -\frac{\pi}{2} + \left[|y(s)| \sqrt{1-y^2(s)} + \sin^{-1} |y(s)| \right] & \text{for} \quad -1 \leq y(s) \leq 0 \end{cases} ,$$

$$S_2(s) = \frac{\pi R^2}{2} - [t(s) - y(s)]\sqrt{R^2 - [t(s) - y(s)]^2} - R\sin^{-1}\left[\frac{t(s)-y(s)}{R}\right],$$

$$t(s) = R - 1 + s ,$$

and

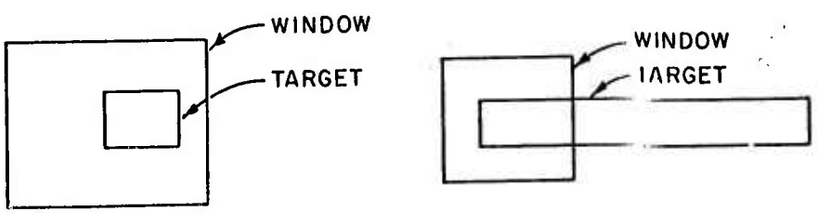
$$y(s) = \frac{s^2 + 2(R-1)(s-1)}{2(s + R-1)} .$$

The variable s is a random variable having a density function

$$\ell(s) = \frac{R + 1 - s}{2R} \quad 0 \le s \le 2 .$$

Estimates of the probability density function for w may be obtained by computer simulation. One could sample s, many times, from a discrete version of its cumulative distribution, evaluate w(s), and construct a discrete approximation to k(w). The procedure could be repeated for a family of R values.

For the discussion above, we selected two examples where the variate w could range over the entire interval [0,1]. Another possibility is suggested by the two cases illustrated below:



In these cases the target is smaller than the window in either one or both dimensions, so that the target never completely fills the window. These examples show that the values of w might be restricted to an interval, say, [0,m], where m < 1.

Even in an application where target shape is not regular, the above analyses may provide useful approximations. Moreover, knowledge of the properties of density functions for w in various cases can be used as a guide in selecting window shapes.

CHAPTER 4

ESTIMATION OF MIXING PARAMETERS

In Chapter 2 our aim was to assign the given measurement u to one of two classes, S_2 or S_3, where S_2 corresponded to background only and S_3 to a mixture of target and background. The dependence of u on the mixing variate w was assumed to be of the form, e.g.,

$$u = wr(w) + (1-w)s(w)$$

where $w = 0$ corresponded to case S_2 and $w \neq 0$ to case S_3. Thus our problem in Chapter 2 was, in effect, to decide whether $w = 0$ or $w \neq 0$ for the given u; the problem could thus have been regarded as one of estimating w, given u.

If we know the density function $\ell(u|w)$ of u conditioned on w, it is reasonable to choose as our estimate the value of w that most likely caused the given value of u to occur, i.e. the value $\max_w \ell(u|w)$. In Chapter 2 we determined $\ell(u|w)$ in various cases; some of these are summarized below.

Example 1'. This example corresponds to Example 1 of Section 2.1. Solving $u = wr + (1-w)s$ for w we obtain

$$w = \frac{s-u}{s-r} .$$

Since r and s are constants a measurement u uniquely determines w.

Example 2'. This example corresponds to Example 2 of Section 2.1. For definiteness suppose $r < a < b$. It is easy to show that $\ell(u|w)$ is uniformly distributed on $[wr + (1-w)a,\ wr + (1-w)b]$, that is,

$$\ell(u|w) = \frac{1}{(b-a)(1-w)} \quad ; \quad wr + (1-w)a \le u \le wr + (1-w)b \ .$$

Also it is easy to show that

$$\max_{0 \le w \le 1} \ell(u|w) = \ell(u|w^*)$$

where

$$w^* = \frac{b-u}{b-r} \ .$$

Example 4'. For Example 4 we obtained

$$\ell(u|w) \equiv N(w\mu_1 + (1-w)\mu_2 \ ; \ w^2\sigma_1^2 + (1-w)^2\sigma_2^2) \ . \qquad (1)$$

Example 5'. For Example 5 we obtained

$$\ell(u|w) \equiv \frac{\lambda_1\lambda_2}{(1-w)\lambda_1 - w\lambda_2}\left[e^{-\frac{\lambda_1 u}{1-w}} - e^{-\frac{\lambda_1 u}{w}}\right] . \qquad (2)$$

Analytical determination of $\max_{0 \le w \le 1} \ell(u|w)$ appears to be difficult for equations (1) and (2); the best way to determine the maximum is probably by direct search. For the next example, however, we can give an analytical treatment.

Example 6'. For the Poisson example in Section 2.3 we obtained the likelihood function

$$\ell(u|w) = e^{-\beta} e^{-[\lambda-\beta]w} \frac{[w\lambda+(1-w)\beta]^u}{u!} . \qquad (3)$$

Suppose, for definiteness, that $\lambda < \beta$. The maximum likelihood estimate $\hat{w}(u)$ is the value of w in $[0,1]$ at which the likelihood function is a maximum. In this case we choose to work with the logarithm, $\ell n\ \ell(u|w)$. If the maximum belongs to the open interval $(0,1)$, then a necessary condition on $\hat{w}$ is obtained by setting to zero the derivative of $\ell n\ \ell(u|w)$:

$$\frac{\partial\ \ell n\ \ell(u|w)}{\partial w} = 0 . \qquad (4)$$

From Equation (3), we have

$$\frac{\partial\ \ell n\ \ell(u|w)}{\partial w} = (\beta-\lambda)\left(1 - \frac{u}{w\lambda + (1-w)\beta}\right)$$

Substituting in Equation 4 we get

$$\hat{w} = \frac{\beta-u}{\beta-\lambda} \quad \text{for} \quad \lambda \le u \le \beta .$$

Since the sign of the derivative is positive for all values of w when $u < \lambda$ and negative for all values of w when $u > \beta$ we have the complete solution

$$\hat{w}(u) = \begin{cases} 1 & ; \quad u < \lambda \\ \frac{\beta-u}{\beta-\lambda} & ; \quad \lambda \le u \le \beta \\ 0 & ; \quad u > \beta \end{cases} .$$

CHAPTER 5

MIXTURE DISTRIBUTIONS

Chapter 4 considered the problem of estimating the mixing parameter w, given a single observation u. In this and the following two chapters we treat the more difficult problem of determining the mixing distribution K(w), given a set of independent observations $u_1,\ldots,u_n$ which define an empirical distribution $H_n(u)$, and similarly for densities.

Let Ω be a set of distributions K(w), each defined on a set W. (For example, W might be the interval [0,1], and Ω might be the set of uniform distributions on subintervals of W.) Let

$$H(u) = \int_W L(u|w)dK(w) \tag{1}$$

where for each w, L(u|w) is a distribution. Given $u_1,\ldots,u_n$, we wish to estimate the particular distribution K(w) in Ω which produced $u_1,\ldots,u_n$. This is essentially the problem of unsupervised estimation,[1,2,3,4] which arises in connection with communications, control, and pattern recognition.

5.1 Identifiability

The first consideration in attempting to find K(w), given $u_1,\ldots,u_n$, is whether specifying H(u) does in fact uniquely determine K(w). Let

$$\Omega' = \left\{H(u):\ H(u) = \int_W L(u|w)dK(w)\ ;\ K(w)\ \epsilon\ \Omega\right\} .$$

Let A denote the mapping from Ω to Ω', i.e. H=AK. We say that Ω' is identifiable if A is one-to-one. In this case, A^{-1} exists and we have $K=A^{-1}H$. Teicher[9] has shown (generalizing a result of Feller[10]) that mixtures of Poisson distributions L(u|w) are always identifiable, but that mixtures of binomial or uniform distributions are not in general identifiable.

H(u) is called a finite mixture if it has the form

$$H(u) = p_1L(u;\theta_1) +\ldots+ p_M(u;\theta_M) \quad ;\quad p_i \geq 0,\ \sum_{i=1}^{M} p_i = 1 .$$

Here M is finite, but its value may not be known; in addition, the p's and θ's may not be known. In the finite case, the estimation problem reduces to that of estimating M, $\theta = (\theta_1,\ldots,\theta_M)$, and $P = (p_1,\ldots,p_M)$, given $u_1,\ldots,u_n$. Teicher[11] has established a sufficient condition for the identifiability of finite mixtures, and used it to prove the identifiability of all finite mixtures of gamma or one-dimensional Gaussian distributions. Yakowitz and Spragins[12] have generalized these results by proving that the class of finite

mixtures of a family $\mathfrak{F}$ of distributions is identifiable if and only if $\mathfrak{F}$ is linearly independent over the field of real numbers. In particular, they have proved identifiability for such families as multivariate normal, products of univariate exponentials, two-parameter Cauchy or negative binomial, and one-parameter translations.

5.2 Compactness and Convexity

In general, the empirical distribution $H_n(u)$ defined by the observations $u_1,\ldots,u_n$ does not belong to Ω', so that even if identifiability holds, $A^{-1}H_n$ may not belong to Ω. In this section we prove, for a wide class of Ω's, that there always exists a unique $H_{K^*}(u)$ in Ω' which is "closest" to $H_n(u)$ in a certain sense.

Let $\mathfrak{F}$ be the class of all distribution functions on the real line. The Lévy metric on $\mathfrak{F}$ is defined[13] by

$$L(F,G) = \inf\{h: F(x-h)-h \le G(x) \le F(x+h) + h\}$$

for all F, G in $\mathfrak{F}$. Let Ω be a subset of $\mathfrak{F}$ which is convex and compact in the Lévy metric. The function A maps Ω onto Ω', which can be regarded as a subset of $\mathfrak{L}^2$. Using the metric defined by the $\mathfrak{L}^2$ norm, we have

Proposition 1. Suppose the function $L(u|w)$ is, for each $w \in W$, a distribution on a closed subset of the real line. Suppose $L(u|w)$ is continuous in w for each fixed value of u. Then A is continuous.

Proof: The weak convergence of K_n to K is assured by the condition $L(K_n,K) \to 0$ [13]. This means $H_{K_n}(u)$ converges pointwise to $H_K(u)$. The desired result then follows from the dominated convergence theorem for $\mathcal{L}^2$.

Corollary. Ω' is closed.

Proof: Ω is compact and a continuous image of a compact set is compact. Thus Ω' is compact. Since $\mathcal{L}^2$ is a metric space it follows that Ω' is closed.

Proposition 2. Ω' is convex.

Proof: Let $H_1(u) = \int L(u|w)dK_1(w)$ and $H_2 = \int L(u|w)dK_2(w)$ belong to Ω'. Then $K_1(w)$ and $K_2(w) \in \Omega$. Let $0 \le \alpha \le 1$. Then

$$H(u) = \alpha H_1(u) + (1-\alpha)H_2(u)$$

$$= \alpha \int L(u|w)dK_1(w) + (1-\alpha) \int L(u|w)dK_2(w)$$

$$= \int L(u|w)(\alpha dK_1(w) + (1-\alpha)dK_2(w))$$

$$= \int L(u|w)d[\alpha K_1(w) + (1-\alpha)K_2(w)] .$$

Since Ω is convex $\alpha K_1(w) + (1-\alpha)K_2(w) \in \Omega$. Thus $H(u) \in \Omega'$.

A well-known theorem in approximation theory is the following: Let S be a closed and convex subset of a Hilbert space $\mathcal{H}$. Given any $X \in \mathcal{H}$ there exists a unique vector $Y_o \in S$ such that $\|X - Y_o\| \le \|X - Y\|$ for all $Y \in S$. In our case we identify Ω' with S, $\mathcal{L}^2$ with $\mathcal{H}$, $H_r(u)$ with X, and H_{K^*} with Y_o to obtain the desired result.

To see that compactness and convexity of Ω are reasonable assumptions, let [a,b] be any closed interval on the real line, and let $\Omega_{a,b}$ be the class of all distribution functions K such that $K(w) = 0$ for $w < a$ and $K(w) = 1$ for $w > b$.

Proposition 3. $\Omega_{a,b}$ is convex.

Proof: Let $K_1(w)$, $K_2(w) \in \Omega$ and let $0 \leq \alpha \leq 1$. Then $K(w) = \alpha K_1(w) + (1-\alpha)K_2(w)$ is continuous from the left, and non-decreasing. Furthermore $K(a) = 0$, $K(b) = 1$. Therefore $K(w) \in \Omega$.

Proposition 4. $\Omega_{a,b}$ is compact.

Proof: Since Ω is a metric space, it is compact if and only if it is complete and totally bounded. (A set Q of a metric space is totally bounded if for every $\epsilon > 0$ it is possible to cover Q by a finite number of spheres $B_\epsilon(q_i)$ $(i=1,2,\ldots,n)$ with centers in Q.) The proof for completeness is essentially the proof given in [13]: Let $F_1, F_2, \ldots$ be a sequence of functions in Ω that satisfy the Cauchy condition, $L(F_n, F_m) \to 0$ as $n,m \to \infty$. Consider the set of rationals

$M = \{x_1, x_2, \ldots, x_s, \ldots\}$, ordered as a sequence, on the interval $[a,b]$. Since the values of $F_n(x_s)$ are bounded, we know that there exists a subsequence of functions $F_{n_1}(x), F_{n_2}(x), \ldots, F_{n_k}(x), \ldots$ which converges at every point of M. The limit $v(x_s) = \lim_{k \to \infty} F_{n_k}(x_s)$ is defined on M and is a non-decreasing function on M. Set

$$F(x) = \sup_{x_s < x} v(x_s).$$

The function $F(x)$ is defined everywhere on $[a,b]$ is non-decreasing and continuous from left. Also $F(a) = 0$, $F(b) = 1$. Indeed for any $\epsilon > 0$ there exists an n such that $L(F_n, F_m) < \epsilon$ for $m \geq n$. We can find a z such that $F_n(z) < \epsilon$. Then for $x_s < z - \epsilon$

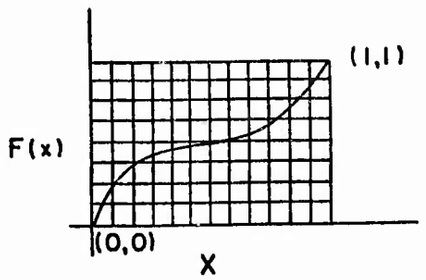
$$F_{n_k}(x_s) \leq F_n(z) + \epsilon \leq 2\epsilon; \qquad n_k \geq n$$

and therefore

$$v(x_s) \leq 2\epsilon.$$

Since $\epsilon > 0$ is arbitrary, $F(a) = 0$. A similar argument proves $F(b) = 1$. It is easy to see that $F_{n_k}(x)$ converges to $F(x)$ at every continuity point of $F(x)$. Thus $\lim_{k \to \infty} L(F_{n_k}, F) \to 0$. From this last result and $L(F_n, F_m) \to 0$ it follows that $\lim_{n \to \infty} L(F_n, F) \to 0$.

We next show that Ω is totally bounded. There is no loss of generality in letting $a = 0$, $b = 1$. Given ϵ, we consider a grid of n^2 squares on the unit square, $0 \leq x \leq 1$, $0 \leq F(x) \leq 1$, where $\frac{1}{n} > \epsilon$. This grid is pictured below.



The ϵ-net for Ω is composed of all "East-North" paths from (0,0) to (1,1). Each distribution function is within a distance ϵ (in Lévy metric sense) from one of the East-North paths. Indeed, given a distribution function $G(x)$ we may select $F(x)$ to be the closest East-North path subject to the condition $F(x) \leq G(x)$. Then $F(x + \frac{1}{n}) + \frac{1}{n} \geq G(x)$ which implies that $L(F,G) \leq \epsilon$.

CHAPTER 6

ESTIMATION OF MIXING DISTRIBUTIONS

The estimation problems associated with mixtures have received considerable attention during the last five years. Many authors, both mathematicians and specialists, have dealt with these problems, not only because of their interesting mathematical content but because of their great importance in many theoretical and applied sciences. The large variety of problems that have been considered is due to different choices for objective functions and different assumptions regarding the a priori knowledge available. We shall now review some of the solutions to estimation problems.

We shall first treat finite mixture problems. We will review the results of Doetsch[14], Medgyessy[15], Stanat[16], Sammon[17], McCormick[18] and Choi[19]. Then we present 1) a non-linear programming formulation of the Choi problem using an objective function based on the uniform norm and 2) a Lagrange multiplier solution to yet another mixture problem.

Turning to arbitrary mixtures we review the Keely-Kruse results and present a direct linear programming formulation of their problem.

6.1 Finite Mixtures

We stated in Section 5.1 that $H(u)$ is a finite mixture if it is of the form $H(u) = \sum_{i=1}^{M} p_i L_i(u;\theta_i)$ (or in terms of probability density functions, $h(u) = \sum_{i=1}^{M} p_i \ell_i(u;\theta_i)$) where $p_i \geq 0$ and $\sum_{i=1}^{M} p_i = 1$. M is finite, but its value may be unknown. Each $L_i(u;\theta_i)$ ($\ell_i(u;\theta_i)$) is a distribution function (density function) with a parameter θ_i which may be unknown. We assume that we are given a sequence $u_1, u_2, \ldots, u_n$ of independent observations selected according to the mixture $H(u)$. Our objective is to estimate the unknown parameters among M, P, and θ.

There are several cases to consider.

Case 1. M, P, and θ are all unknown

Only a few special problems have been treated when M, θ, and P are all unknown.

Doetsch[14] has assumed $H(u)$ to be a mixture of normal functions and exhibited a linear operator which reduced the variances of the summand functions without changing their weights or means. Medgyessy[15] extended Doetsch's work to a large class of finite mixtures with the restriction that the summand functions be univariate and have no more than two unknown parameters.

Stanat[16] has exploited the techniques of Doetsch and Medgyessy for the purpose of treating empirical functions and extended the techniques to include summand functions which are multivariate. Two cases were studied in detail, the multivariate normal and the multivariate Bernoulli.

To illustrate the technique we describe the univariate, normal summand case due to Doetsch.

Assume $h(u)$ is equal to a weighted sum of univariate normal functions,

$$h(u) = \sum_{i=1}^{M} p_i N(\mu_i, \sigma_i^2);\ p_i > 0,\ \sum_{i=1}^{M} p_i = 1\ ,$$

where the parameters p_i, μ_i, σ_i, and M are all unknown. Also, for definiteness assume $\sigma_i \leq \sigma_{i+1}$ for all i. Let $\Phi(\omega)$ denote the Fourier transform of $h(u)$ and $\Phi_i(\omega)$ that of the summand function $N(\mu_i, \sigma_i^2)$. Then

$$\Phi(\omega) = \sum_{i=1}^{M} p_i \Phi_i(\omega)$$

$$= \sum p_i \exp(j\mu_i\omega - \frac{1}{2}\sigma_i^2\omega^2).$$

A function $Z(\lambda,\omega) = \exp(\frac{1}{2}\lambda^2\omega^2)$ is defined which will determine a linear operator on the summand functions of $\Phi(\omega)$, that is,

$$\Phi^*(\omega) = \Phi(\omega)\cdot Z(\lambda,\omega) = \sum_1^M p_i \Phi_i(\omega)\cdot Z(\lambda,\omega)$$

$$= \sum_1^M p_i \exp(j\mu_i\omega - \tfrac{1}{2}(\sigma_i^2 - \lambda^2)\omega^2).$$

The properties of $\Phi^*(\omega)$ are dependent on λ as follows:

(1) For $\lambda = 0$, $\Phi^*(\omega) = \Phi(\omega)$.

(2) For $0 \le \lambda \le \sigma_1$, $\Phi^*(\omega)$ is the Fourier transform of a weighted sum of normal probability functions.

(3) For $\lambda > \sigma_1$, $\Phi^*(\lambda)$ is not a characteristic function.

(4) For $\lambda = \sigma_1$, $\Phi^*(\omega)$ is the Fourier transform of a weighted sum of normal probability functions, at least one of which is degenerate. In particular if $h^*(u)$ is defined to be the inverse transform of $\Phi^*(\omega)$, then for $\lambda = \sigma_1$,

$$h^*(u) = p_1 N_1(\mu_1,0) + \sum_{i=2}^M p_i N(\mu_i,\sigma_i^2 - \sigma_1^2).$$

If $H^*(u)$ is defined to be the cumulative function of $h^*(u)$, then $H^*(u)$ has a discontinuity of height p_1 at the value μ_1. This fact and the knowledge of the critical value of λ are sufficient to specify completely the summand function with minimum variance. This function can be subtracted from the original function $h(u)$ and the process repeated.

Sammon[17] has considered a mixture

$$h(u) = \sum_{i=1}^{M} p_i \ell(u-\mu_i) \quad ; \quad p_i \geq 0, \sum p_i = 1$$

where $\ell(u)$ is of a particular form. The aim is to determine the unknown parameters M, p_i, and μ_i.

The method of solution involves the application of a linear operator T to h(u). Then $T[h(u)] = \sum_{i=1}^{M} p_i \alpha(u-\mu_i)$ where $T[\ell] = \alpha$. He shows that a linear operator can, in general, be found so that each element $\alpha(u-\mu_i)$ has a large narrow positive peak in the neighborhood of its location parameter μ_i and relatively small amplitude elsewhere. Therefore, the location of the individual elements μ_i and the weights p_i can be found by inspection. The parameter n is found by counting the number of individual peaks.

Case 2. M is Known; θ and P are to be Determined

Maximum likelihood procedures together with mathematical programming provide an approach[18] to the problem of estimating P and θ. Let $Q = (p_1,p_2,\ldots,p_M;\theta_1,\theta_2,\ldots,\theta_M)$ and $h(u;Q) = \sum_{i=1}^{M} p_i \ell_i(u;\theta_i)$. Let $u_1,u_2,\ldots,u_n$ be observations of the random variable u. The likelihood function associated with these observations is

$$\lambda(u_1,\ldots,u_n;Q) = h(u_1;Q)h(u_2;Q)\ldots h(u_n;Q)$$

The $\bar{P}$ and $\bar{\theta}$ which maximize λ are called maximum likelihood estimates and have many desirable statistical properties. The problem of maximizing λ subject to the constraints on P is a mathematical programming problem. Sometimes it is also desirable to constrain P so that certain physical requirements are not violated.

To describe another approach (involving a different objective function), due to Choi[19], we need to introduce additional notation. Let $H_n(u)$ be the empirical distribution based on the samples $u_1, u_2, \ldots, u_n$, i.e., $H_n(u) = (\text{number of } u_i \leq u)/n$. Let the subscript n on Q_n denote estimators based on n samples. Choi has treated the problem of finding $\bar{Q}_n$ which minimizes

$$S(Q_n) = \int \left[H(u;Q_n) - H_n(u)\right]^2 dH_n(u).$$

He proves that $\bar{Q}_n$ is a strongly consistent estimator of Q and that for all Q in Ω and for all n sufficiently large, with probability one, the unique $\bar{Q}_n$ exists and it is the only solution of the normal equations of $S(Q_n)$.

We shall now present a non-linear programming approach to this case, using an objective function different from that used by Choi — namely, the uniform norm

$$\|H(u) - G(u)\| = \sup_u |H(u) - G(u)| .$$

Thus $$\|H_n(u) - H_Q(u)\| \approx \max_{i=1,2,\ldots,n} |H_n(u_i) - H_Q(u_i)|.$$

$$= \max\left|H_n(u_i) - \sum_{j=1}^{M} p_j L(u_i;\theta_j)\right|$$

$$= S_1(Q) = S_1(P,\theta) = t$$

Problem: $$\underset{P,Q}{\text{Minimize}}\ S_1(P,\theta) = t$$

$$p_i \geq 0\ ,$$

$$a_i \leq \theta_i \leq b_i\ ,$$

$$\sum p_i = 1.$$

The expression $$t \geq \left|H_n(u_i) - \sum_{j=1}^{M} p_j L(u_i;\theta_j)\right|$$

implies

$$t \geq H_n(u_i) - \sum p_j L_j(u_i;\theta_j)$$

and

$$t \geq \sum p_j L_j(u_i;\theta_j) - H_n(u_i)$$

so the problem becomes a non-linear programming problem for which numerical solutions may be obtained using any of a number of techniques[18]. The problem reads

$$\min t$$

where

$$t + \sum_{j=1}^{M} p_j L_j(u_i;\theta_j) \geq H_n(u_i); \quad i=1,2,\ldots,n$$

$$t - \sum_{j=1}^{M} p_j L_j(u_i;t_j) \geq - H_n(u_i); \quad i=1,2,\ldots,n$$

$$p_i \geq 0$$

$$a_i \leq \theta_i \leq b_i$$

$$p_1 + p_2 + \ldots + p_M = 1.$$

Case 3. M and θ are Known; P to be Determined

Letting $L_i(u) = L_i(u;\theta_i)$, we have $H(u) = \sum_{i=1}^{M} p_i L_i(u)$ where $\Sigma\, p_i = 1$ and $p_i \geq 0$. Given $u_1,u_2,\ldots,u_n$ we wish to find the vector P which minimizes

$$S(P) = \int (H_n(u) - \sum_{i=1}^{M} p_i L_i(u))^2 du.$$

The Lagrange multiplier approach is to minimize

$$J(P) = S(P) + 2\lambda(\sum_{i=1}^{M} p_i - 1)$$

with respect to P where λ is a Lagrange multiplier. The solution is obtained by solving simultaneously the system of linear equations

obtained by setting $\frac{\partial J(P)}{\partial p_i} = 0$ for i=1,2,...,M, namely,

$$AP = b - \lambda U ,$$

where $A = (a_{ij})$ and $a_{ij} = \int L_i(u)L_j(u)du$, i,j=1,2,...,M; b is a column vector with $b_i = \int L_i(u)H_n(u)du$, i=1,2,...,M, and U is a column vector of M components all equal to 1.

6.2 Arbitrary Mixtures

In 1964 Robbins[20] studied estimating sequences for the general mixture problem and suggested the problem of obtaining explicit construction of the sequences. A method for constructing such sequences has been presented by Deely and Kruse[21]. Their results are obtained under the following assumptions:

1. The class $\Omega = \{K(w)\}$ of mixing distributions is defined on a compact subset W of the real line.

2. For each $w \in W$, $L(u|w)$ is a distribution on a closed subset U of the real line.

For each natural number n, Deely and Kruse define Ω_n to be the class of discrete distributions on W with weights at $w_{1n}, w_{2n}, \ldots, w_{nn}$, where the w_{in} are chosen so that for any $K \in \Omega$ there is a sequence $\{K_n\}$ with $K_n \in \Omega_n$ which converges weakly to K. Their approach is to find $K_n^* \in \Omega_n$ which minimizes $\|H_K(u) - H_n(u)\| = \sup_u |H_K(u) - H_n(u)|$ for $K \in \Omega_n$, where $H_n(u)$ is the empirical distribution based on the first n observations. They prove the following theorem which guarantees the desired convergence.

Theorem: Suppose

(i) $L(u|w)$ is continuous on $U \times W$.

(ii) $H_{K_1} = H_{K_2}$ for $K_1, K_2 \in \Omega$ implies $K_1 = K_2$

(identifiability condition).

(iii) $\{K_n^*(w)\}$ is a sequence for which

$K_n^*(w) \in \Omega$ and $\|H_{K_n^*} - H_n\| = \inf_{K \in \Omega_n} \|H_K(u) - H_n(u)\|$.

Then Prob $\{ \lim_{n \to \infty} K_n^*(w) = K(w)$; w any continuity point of $K\} = 1$.

Finding the desired distribution $K_n^*(w)$ is shown to be equivalent to finding an optimal strategy in a certain game which in turn is equivalent to solving a linear programming problem.

A more direct linear programming formulation will now be presented. A discrete approximation of $H(u)$ is given by

$$H(u) \approx k_1 L(u|w_1) + k_2 L(u|w_2) + \ldots + k_m L(u|w_m),$$

where $k_i \geq 0$ and $\Sigma k_i = 1$. Then

$$\|H_n(u) - H(u)\| \approx \max_{i=1,2,\ldots,n} |H_n(u_i) - H(u_i)|$$

$$= \max \left| H_n(u_i) - \sum_{j=1}^{m} k_j L(u_i|w_j) \right|$$

$$= M(k_1, k_2, \ldots, k_m)$$

The problem is to determine

$$\underset{(k_1,k_2,\ldots,k_m)}{\text{minimum}} M(k_1, k_2, \ldots, k_m)$$

subject to the conditions

$$k_i \geq 0$$

and

$$k_1 + k_2 + \ldots + k_m = 1.$$

Let $L_j(u_i) = L(u_i|w_j)$, $t = \min M(k_1, k_2, \ldots, k_m)$.

Then $t \geq \left|H_n(u_i) - \sum_{j=1}^{m} k_j L_j(u_i)\right|$ implies that

$$t \geq H_n(u_i) - \sum_{j=1}^{m} k_j L_j(u_i) ,$$

and

$$t \geq \sum_{j=1}^{m} k_j L_j(u_i) - H_n(u_i).$$

The problem reduces to finding

$$\min t$$

where

$$t + \sum_{j=1}^{m} k_j L_j(u_i) \geq H_n(u_i); \quad i=1,2,\ldots,n$$

$$t - \sum_{j=1}^{m} k_j L_j(u_i) \geq - H_n(u_i); \quad i=1,2,\ldots,n$$

$$k_1 + k_2 + \ldots + k_m = 1$$

$$t, k_i \geq 0 .$$

This is a linear programming problem with $2n + 1$ constraints and $m + 1$ unknown quantities $k_1, k_2, \ldots, k_m, t$. We may write the problem in the standard matrix form

$$\min_X Z = CX$$

subject to the constraints

$$AX = b$$

$$X \geq 0$$

where $X = (x_1, x_2, \ldots, x_{m+2})$ is a column vector; $x_1 = t$, $x_2 = k_1, \ldots, x_{m+1} = k_m$, x_{m+2} is a surplus variable; $b = (b_1, b_2, \ldots, b_{2m+1})$ is a column vector with

$$b_i = H_n(u_i); \quad i=1,2,\ldots,n$$

$$b_{n+j} = -H_n(u_i); \quad j=1,2,\ldots,n$$

$$b_{2n+1} = 1;$$

and $C = (c_1, c_2, \ldots, c_{n+2})$ is a row vector with $c_1 = 1$ and $c_2 = c_3 = \ldots = c_{m+2} = 0$; $A = (a_{ij})$ is a $(2n+1) \times (m+2)$ order matrix;

$$A = \begin{pmatrix} 1 & L_1(u_1) & \cdots & L_m(u_1) & -1 \\ 1 & L_1(u_2) & \cdots & L_m(u_2) & -1 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & L_1(u_n) & \cdots & L_m(u_n) & -1 \\ 1 & -L_1(u_1) & \cdots & -L_m(u_1) & -1 \\ 1 & -L_1(u_2) & \cdots & -L_m(u_2) & -1 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & -L_1(u_n) & \cdots & -L_m(u_n) & -1 \\ 0 & 1 & 1 \cdots & 1 & 0 \end{pmatrix}$$

REFERENCES

1. Hancock, J. and Patrick, E.: "Learning Probability Spaces for Classification and Recognition of Patterns With or Without Supervision." Purdue University Report TR-EE65-21. 1965.

2. Robbins, H.: "The Empirical Bayes Approach to Statistical Decision Problems." Ann. Math. Statist. 1964. 35; 1-20.

3. Choi, K.: "Empirical Bayes Procedure for Classification With Stochastic Learning." University of Missouri Technical Report No. 22. 1967.

4. Spragins, J.: "Learning Without a Teacher." IEEE Trans. on Info. Theory. 1966. IT-12. No. 2; 223-230.

5. Nilsson, N.: Learning Machines. McGraw-Hill, New York, New York, 1965.

6. Van Trees, H.: Detection, Estimation, and Modulation Theory. Part I. John Wiley, New York, New York, 1968.

7. Bellman, R. and Dreyfus, S.: Applied Dynamic Programming. Princeton University Press, Princeton, New Jersey, 1962.

8. Wadsworth, G. and Bryan, J.: Introduction to Probability and Random Variables. McGraw-Hill, New York, New York, 1960.

9. Teicher, H.: "Identifiability of Mixtures." Ann. Math. Statist. 1961. 32; 244-248.

10. Feller, W.: "On a General Class of Contagious Distributions." Ann. Math. Statist. 1943. 14; 389-399.

11. Teicher, H.: "Identifiability of Finite Mixtures." Ann. Math. Statist. 1963. 34; 1265-1269.

12. Yakowitz, S. and Spragins, J : "On the Identifiability of Finite Mixtures." Ann. Math. Statist. 1968. 39; 209-214.

13. Gnedenko, B. and Kolmogorov, A. Limit Distributions for Sums of Independent Random Variables. Addison-Wesley, Cambridge, Mass., 1954.

14. Doetsch, G.: "Zerlegung einer Funktion in Gausche Fehlerkurven und zeitliche Zurulkverfolgung eines Temperaturzustandes." Mathematische Zeitschrift. 1936. 41; 283-318.

15. Medgyessy, P.: Decomposition of Superpositions of Distribution Functions. Hungarian Academy of Sciences, Budapest, Hungary, 1961.

16. Stanat, D.: "Unsupervised Learning of Mixtures of Probability Functions" appears in Pattern Recognition edited by L. Kanal. Thompson Book Co. Washington, D.C. 1968.

17. Sammon, J.: "Adaptive Decomposition of Superpositions of Probability Density Functions." 1966.

18. Bracken, J. and McCormick, G.: "Selected Applications of Nonlinear Programming." Research Analysic Corporation Report RAC-R-68. 1968.

19. Choi, K.: "Estimators for the Parameters of a Finite Mixture of Distributions." University of Missouri Technical Report No. 18. 1966.

20. Robbins, H.: "The Empirical Bayes Approach to Statistical Decision Problems." Ann. Math. Statist. 1964. 35; 1-20.

21. Deely, J. and Kruse, R. "Construction of Sequences Estimating the Mixing Distribution." Ann. Math. Statist. 1968. 39; 286-288.

ANALYSIS OF GEOMETRICAL MOMENT FEATURES EXTRACTED FROM DIGITIZED TANK PHOTOGRAPHS USING AN ON-LINE PATTERN ANALYSIS AND RECOGNITION SYSTEM*

James R. Rapp

Ballistic Research Laboratories
U. S. Army Aberdeen Research and Development Center
Aberdeen Proving Ground, Maryland

Success in classifying patterns using geometrical moment features has been reported by a number of investigators, including M-K Hu,[1] F. L. Alt,[2] and J. W. Butler.[3] These encouraging results, obtained with patterns of fairly simple nature, indicate that moment features may have significant potential for discrimination of the more complex spatial patterns representing signatures of army targets and their diverse environments. We use the term signature here to refer to the output of a sensor of some kind which is used to derive information about the presence or location of targets. In the present investigation, photographic signatures of a tank in a wooded environment are used, as they provide a convenient source of input data for an initial study such as this. The differences in the spatial composition of contrast details within photographic signatures of tank and background areas indicate that perhaps geometrical moment features might provide a basis for distinguishing between them. The main object of the experiment, then, is to determine how useful certain moment features are in characterizing the photographic signatures of tanks and typical backgrounds.

*This paper has been reproduced photographically from the author's notes. Messrs. Roberts and Rapp submitted their paper in two parts. Part II by Mr. Roberts can be found directly following the present article.

A digital computer was used to generate moment feature data for specific pattern classes. In order to provide input data in a discrete format compatible with digital processing methods, digitized picture samples were derived in the manner shown schematically in Figure 1. Raw signature data consisted of a set of 11 photographic transparencies containing different aspect views of a medium tank located just in front of a thick grove of trees. All photographs were taken at one site and different tank views were obtained by in-place rotations of the tank in 30-degree increments through 300 degrees. The resulting gray scale transparencies were subsequently digitized by a flying spot scanner system and recorded on magnetic tapes. These digital pictures contain over one million picture elements with density information resolved to 64 levels of gray. Portions of these digital pictures were used as samples of target and background for this investigation. The dimensions of the sample areas were chosen so that samples in the target class contained a practical minimum of nontank picture elements. Square sample areas, 100 elements on a side, were chosen on this basis. In order to reduce the amount of computer time required to perform the moment feature computations on these sample picture areas, the effective resolution within the sample areas was reduced fourfold by averaging over nonoverlapping 16-element arrays to produce new samples with only 25 elements on a side. The corresponding size of these elements on the actual tank is roughly 6 centimeters. A test was made to determine the effects of this averaging process on the values of moment features

and only minor changes in moment values for corresponding areas were observed. These differences should not affect the results of this investigation significantly.

The reduced-resolution gray scale picture samples were also preprocessed for detail contrast enhancement using two different algorithms. Basically, these procedures are intended to increase the level of contrast, i.e., the numerical difference between picture elements, in regions of the picture samples where significant "detail" exists. The resulting preprocessed picture samples are thus referred to as detail enhanced pictures. One preprocessing routine transforms each gray level picture array into a new array consisting of the sum of squared first differences at each point in the original picture. This form of preprocessing is a close digital approximation to the process of taking the squared magnitude of the gradient function for a picture defined by a continuous picture function. In view of this similarity, we refer to these new samples as gradient preprocessed samples or simply gradient samples. The second preprocessing routine is a digital approximation to the process of forming the Laplacian of a picture function in the analogous case of a continuous picture sample. A 17-point Laplacian, equivalent to that used by Bogusz et al. in their work,[3] is implemented in a manner which can be conveniently thought of as a convolution process wherein the original picture sample is convolved with an appropriate Laplacian mask. This mask would consist of a 5 x 5 array of weighting elements. The 16 elements in the periphery of the array are -1's, the central elements are 0's. The absolute value of the Laplacian is used

and the resultant samples are referred to simply as Laplacian samples. Both sets of preprocessed picture samples were used in addition to the basic gray scale samples as inputs for the moment feature extraction program to determine if there might be an advantage to using detail enhancement prior to moment feature computation.

The input samples were chosen from two basic classes, target and nontarget or background. Illustrations of the kind of information contained in these samples are given in Figure 2 in which ten representative background samples are outlined together with five samples from the target class. Since all eleven pictures in the data set contain identical backgrounds, all of the background samples, comprising 64 nonoverlapping areas such as those shown outlined here, were selected from this picture. The diversity of picture content represented by the samples within each class is evident in this picture, which is only one of eleven different aspect views of the tank. A finer breakdown of these sample areas was established to create subclasses within which the variation in subjective picture content is considerably reduced. Accordingly, two subclasses of background were established, trees and foreground; both are illustrated in the pictorial example. The target class was similarly divided into three subclasses: fore and aft, turret, and wheels. The outlined portions covering the tank include two turret samples and three samples of fore and aft. Figure 3 illustrates some additional sample areas in the target class from a different tank picture. The two upper blocks represent samples from the turret subclass, the two lower left blocks are samples from the fore and aft subclass, and the

two blocks in the lower right are sample areas representing the wheels subclass. A total of 70 samples make up the target class population which is comprised of 22 turret samples, 32 wheels samples, and 16 fore and aft samples. The input sample distribution is illustrated in Figure 4. The three different digital versions of a particular sample area are shown in Figure 5. Blanks appear here in places where zeros normally occur in these printouts.

Figure 6 illustrates the form of the basic geometrical moments. The pictorial data are represented by a pattern function P(x,y) designating the encoded density at discrete values of the coordinates (x,y). The order of the moments is specified by i and j which are integer values in the range [0,4]. Moments of this form are independent of overall pattern "intensity" since they are expressed in terms of the normalized pattern value D(x,y). This normalization effectively removes the sensitivity of these moments to such factors as scene brightness and exposure times which typically affect photographic renditions of real-world scenes. Capital letters and the symbol '1' are used to designate the 27 moment features. Ten of these moment features are basic moments and the remainder are algebraic combinations of these moments, some resembling statistical parameters commonly used to characterize distributions. Six of the moment features are equivalent to those constructed by Hu in his investigation of the recognition of well-defined shapes such as those represented by the printed characters in the English alphabet. These features are independent of the position, size, and orientation of the pattern and are referred to as moment invariants. It

is noted, however, that substantial differences exist between the types of visual patterns considered by Hu and those with which the present investigations are concerned. Specifically, the patterns used here were derived from portions of photographs of a tank scene, and therefore the patterns do not contain a single object of interest surrounded by a featureless border. For this reason the invariance properties of these moment features are not as significant for this investigation as for Hu's work.

In the present investigation, we have attempted only to determine the degree of separation provided by each moment feature individually. In this endeavor, we establish the threshold value which best separates the input samples into their prescribed classes. This is done by assuming that the errors in classification which result from assigning either class of input to the wrong class are of equal importance. In this situation then, the criterion for best separation is that the threshold be set to maximize the separation factor, which is the average of the fractions of correct classifications for both classes. This form of separation factor does not unduly weight the classification rate for the class having the larger number of samples. Results of the BRL investigation are based upon class separation data of this kind.

The feature value distributions within each class and subclass will be illustrated by the histograms in Figures 7, 8, 9, and 10 which are representative samples of the data for the set of 27 moment features. In Figure 7, the histograms for Feature A measured on gray scale samples are shown. and the distributions for different classes may be readily

compared. The abscissa axes for these histograms represent the computed values of Feature A normalized with respect to the mean value for all picture samples. It is evident that very little separation between target and background classes could be achieved using this feature. Furthermore, the distributions for the subclasses also overlap to such an extent that no pair of subclasses can be reasonably separated. The distributions for all subclasses are quite similar except that the samples from the foreground class are somewhat more clustered about the mean. Although the data are not shown here, the histograms for Feature A measured for gradient and Laplacian samples reveal that the same sort of situation exists. The preprocessing methods provided no significant improvement in the class separation achievable using this feature. Of all the features examined in the study, Feature A provided the least discrimination between the basic target and background classes. Feature A is a basic second order moment about a vertical axis through the pattern centroid. Feature B is the corresponding second order moment about a horizontal axis through the pattern centroid; histograms showing the results using this feature are presented in Figure 8. Although a considerable amount of overlap in class distributions is evident here as well, the means for the target and background classes are different. There is an observable trend toward higher values among the target class samples than among the background samples. Quantitatively, the difference in potential for target and background discrimination for these two features, A and B, is indicated by their respective separation factors, 0.58 vs 0.74. The difference in separability is probably due

to the differences in the horizontally oriented pattern detail compositions being more pronounced than the vertically oriented detail structure. Other features also tend to reveal this disparity between class separabilities for moment features which differ only in the coordinate axis involved in their definitions.

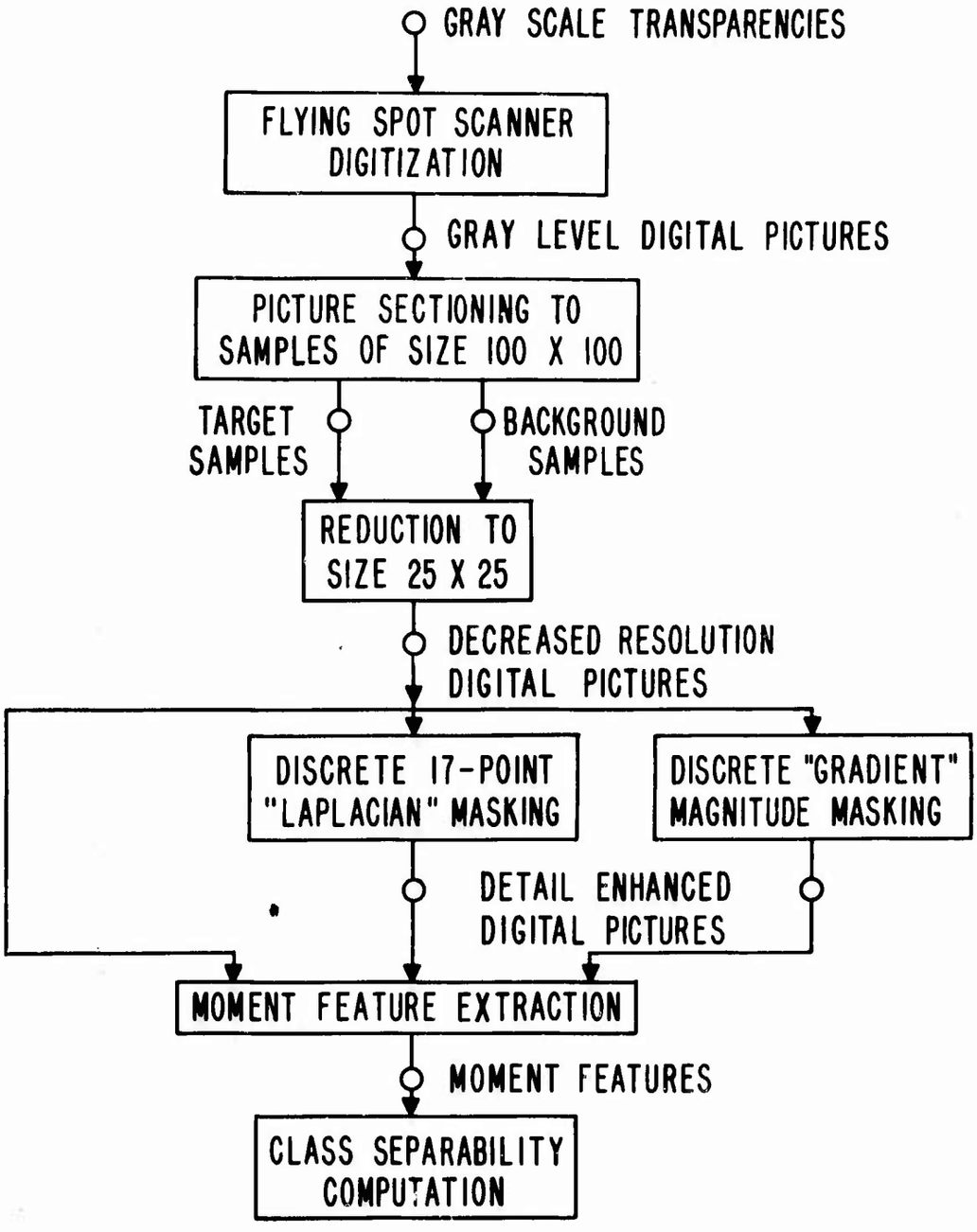
Figure 9 shows the histograms for Feature G measured on gray scale samples. Feature G is a basic fourth order moment measured with respect to the horizontal coordinate axis, and it was found to be effective in a number of specific situations. An example is the separation of wheels and trees subclasses, for which the separation factor was found to be 0.96. The effects of preprocessing in conjunction with this feature are interesting to note. The use of gradient preprocessing, for example, reduced the separability of the trees and wheels samples, and likewise the Laplacian preprocessed samples were less well separated for these subclasses. However, the separability of the basic target and background classes was improved by preprocessing. The histograms for Feature G measured on gradient samples are presented in Figure 10. The separation of target and background samples afforded by this feature is the best that could be attained in the course of this investigation. The separation factor in this instance was 0.84. A summary of the best features and their separation factors for several classes and subclasses is presented in Figure 11. This table shows that no single feature was outstanding for all the cases tested. Also, the use of preprocessing methods for detail enhancement is not consistently beneficial, even though in some cases the preprocessed data could be separated to a greater extent

than the gray scale data. Even for the same feature, the effects of preprocessing were not uniform in the sense of producing similar differences in the separation factors for each pair of classes and subclasses. As a result, it is not possible to determine whether preprocessing should be used in conjunction with a particular feature without specifying which classes are involved. Similarly, the best feature for separation of a given pair of classes indeed depends upon whether or not preprocessing is performed and, if so, which technique is used.

A number of factors may be cited which influence the results of this experiment directly and serve to limit their applicability. The selection of sample dimensions, for example, was made rather arbitrarily for convenience in processing, and it is well to note that the results of this investigation might have been significantly different had sample areas of a different size been chosen. Certainly the choice of position for the sampling window with respect to the picture details is somewhat arbitrary also. Both of these factors are related to such operational parameters as target range, sensor configuration, signature processing mode, and requirements for target identification. An important limitation insofar as interpretation of the results is concerned is that the number of samples in each class is relatively small. On the whole, however, the extensive intraclass spread of values for many of the features tested clearly indicates that they would not be suitable as discriminants for this task. In general, those features which were defined exclusively in terms of moments with respect to the vertical axis through the pattern centroid seemed to demonstrate the least potential. The more complex,

"invariant" features were relatively unsuccessful also, which suggests that their additional complexity and attendant processing difficulty is probably not warranted for this application.

It is difficult to fully assess the comparative potential of geometric moment features on the basis of this experiment alone. Little information exists concerning the efficacy of other types of features for pattern recognition involving patterns as complex and diverse as those of the signatures of army targets in typical backgrounds. However, this experiment demonstrates that the information contained in certain relatively low order geometric moments does serve to characterize spatial patterns of this sort to some extent. Additional testing of the more promising features would appear to be worth consideration. It might also be profitable to examine the feasibility of using a number of geometric features jointly or in conjunction with other kinds of features to test their effectiveness for target and background discrimination.



PICTURE PROCESSING FLOW DIAGRAM

FIGURE 1

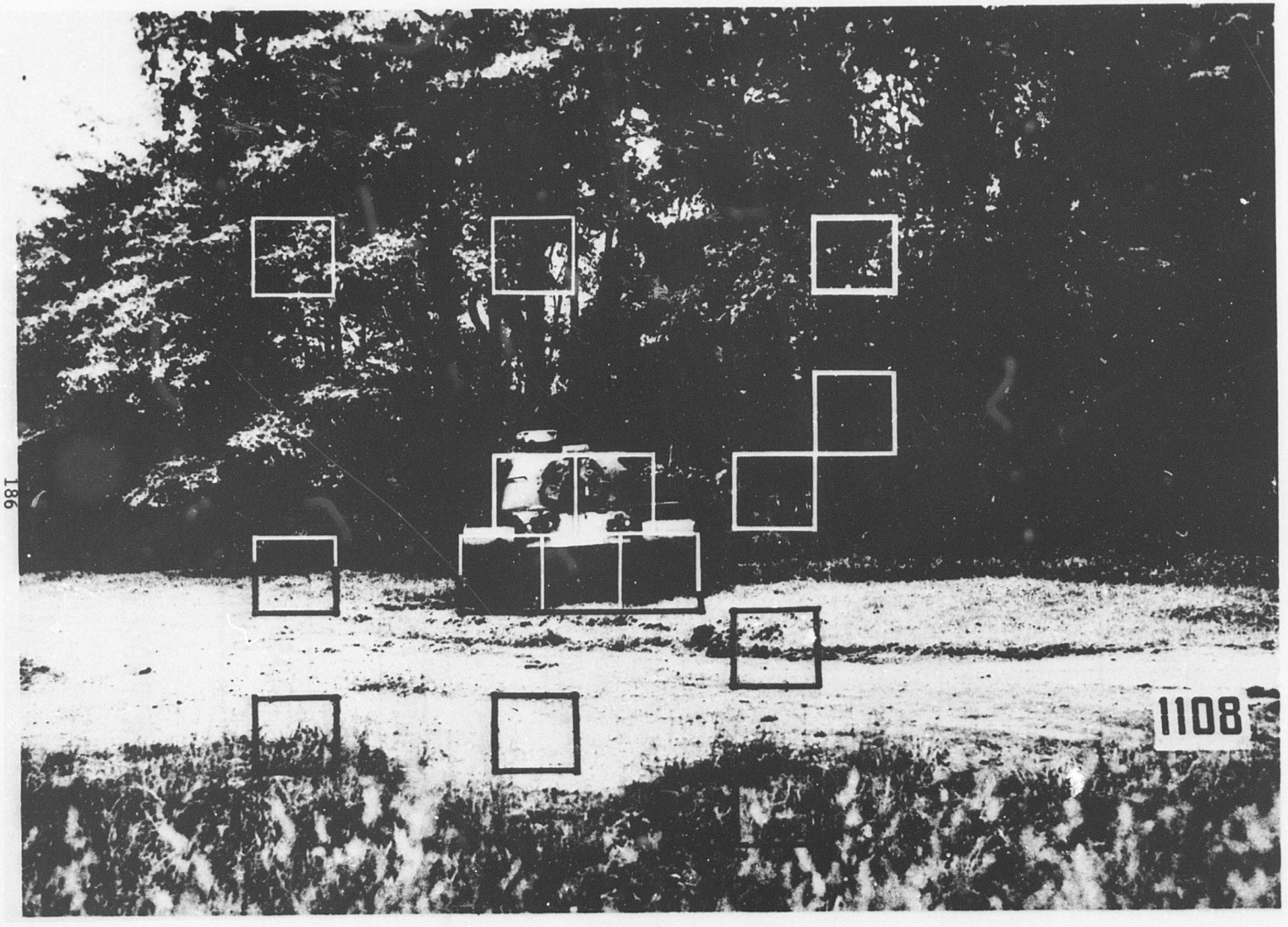


FIGURE 2

TARGET AND BACKGROUND SAMPLE AREAS

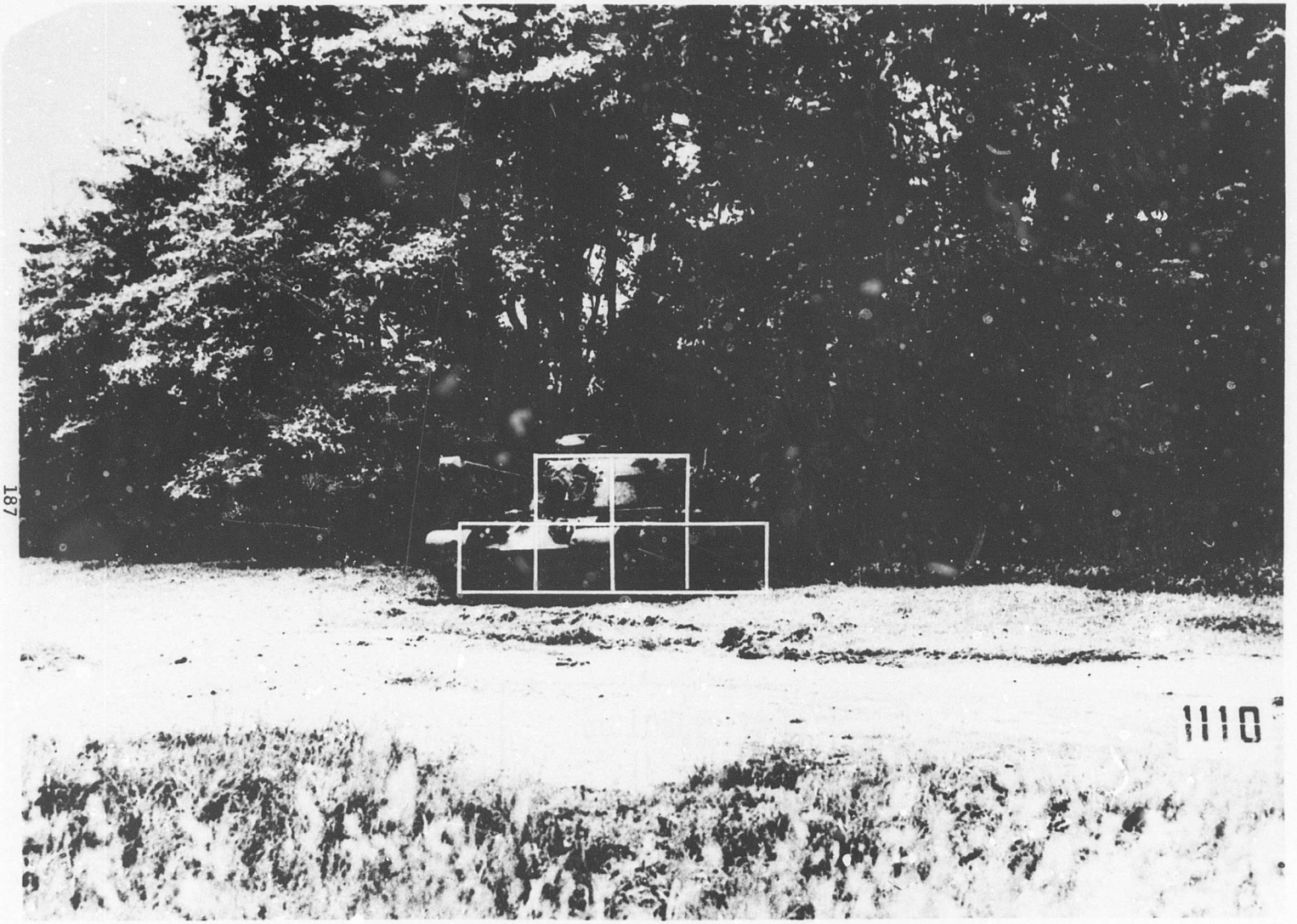


FIGURE 3

TARGET SAMPLES, OBLIQUE TANK ASPECT

SAMPLE CLASSES AND SUBCLASSES

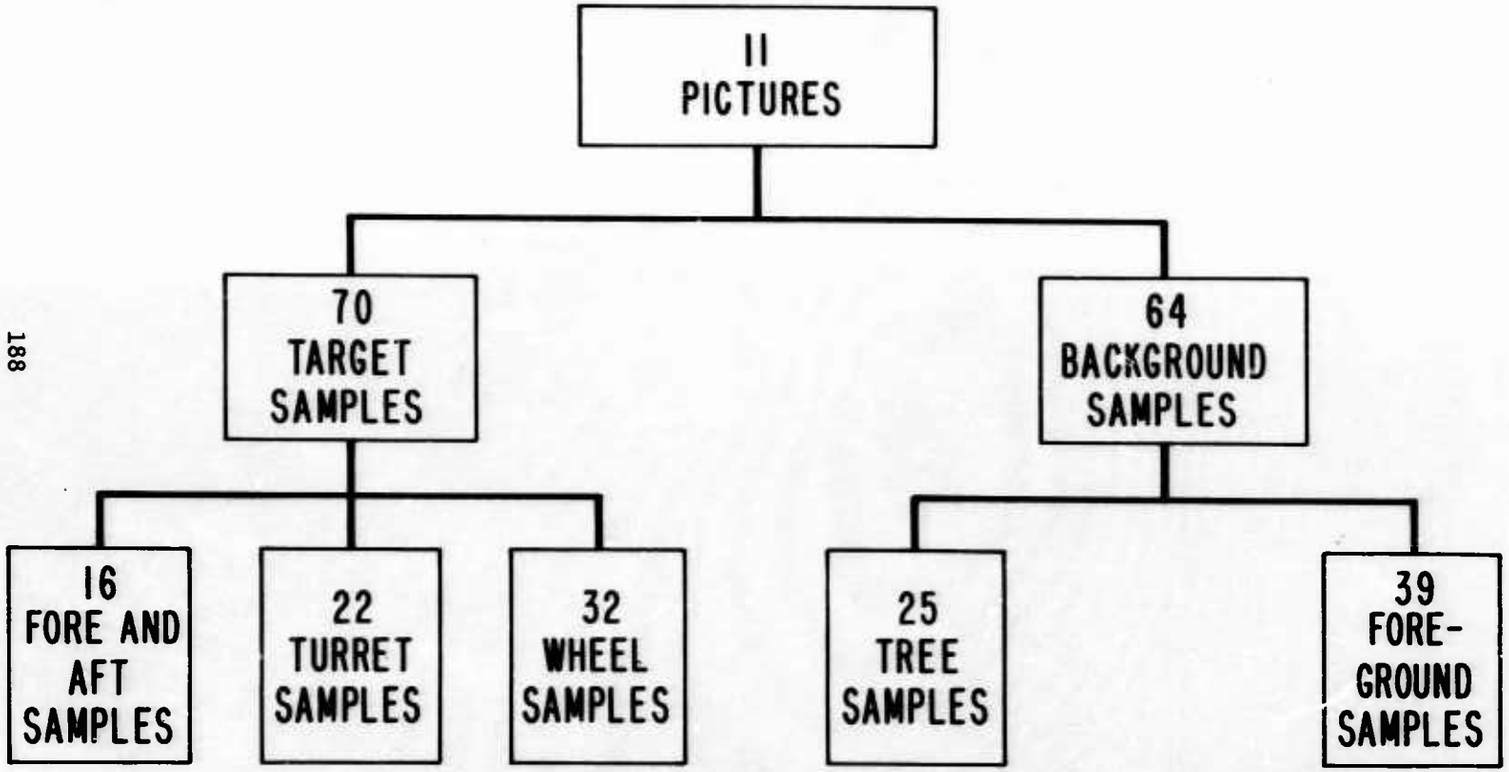


FIGURE 4

WHEELS SAMPLE AND DIGITAL VERSIONS

GRAY SCALE

GRADIENT

LAPLACIAN

FIGURE 5

GEOMETRICAL MOMENT DEFINITION

$$\mu_{i,j} = \sum_{x=1}^{n} \sum_{y=1}^{m} (x-\bar{x})^i (y-\bar{y})^j D(x,y)$$

$$i = o, 1, 2 \cdots \qquad j = o, 1, 2 \cdots$$

$$\bar{x} = \sum_{x=1}^{n} \sum_{y=1}^{m} x D(x,y)$$

$$\bar{y} = \sum_{x=1}^{n} \sum_{y=1}^{m} y D(x,y)$$

$$D(x,y) = \frac{P(x,y)}{\sum_{x=1}^{n} \sum_{y=1}^{m} P(x,y)}$$

FIGURE 6

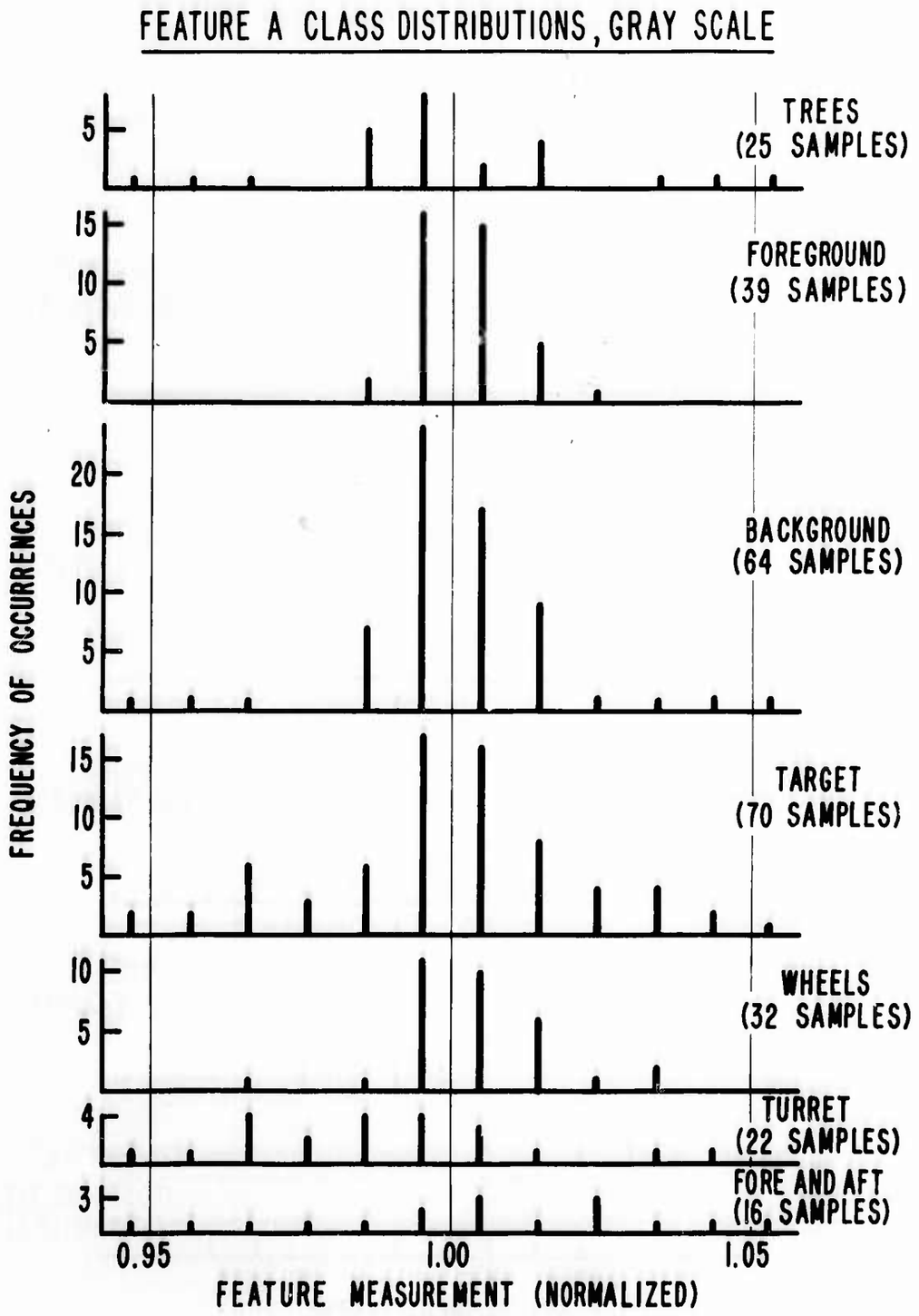

FEATURE A CLASS DISTRIBUTIONS, GRAY SCALE
TREES
(25 SAMPLES)
FOREGROUND
(39 SAMPLES)
BACKGROUND
(64 SAMPLES)
TARGET
(70 SAMPLES)
WHEELS
(32 SAMPLES)
TURRET
(22 SAMPLES)
FORE AND AFT
(16 SAMPLES)
FREQUENCY OF OCCURRENCES
5
15
10
5
20
15
10
5
15
10
5
10
5
4
3
0.95
1.00
1.05
FEATURE MEASUREMENT (NORMALIZED)

FIGURE 7

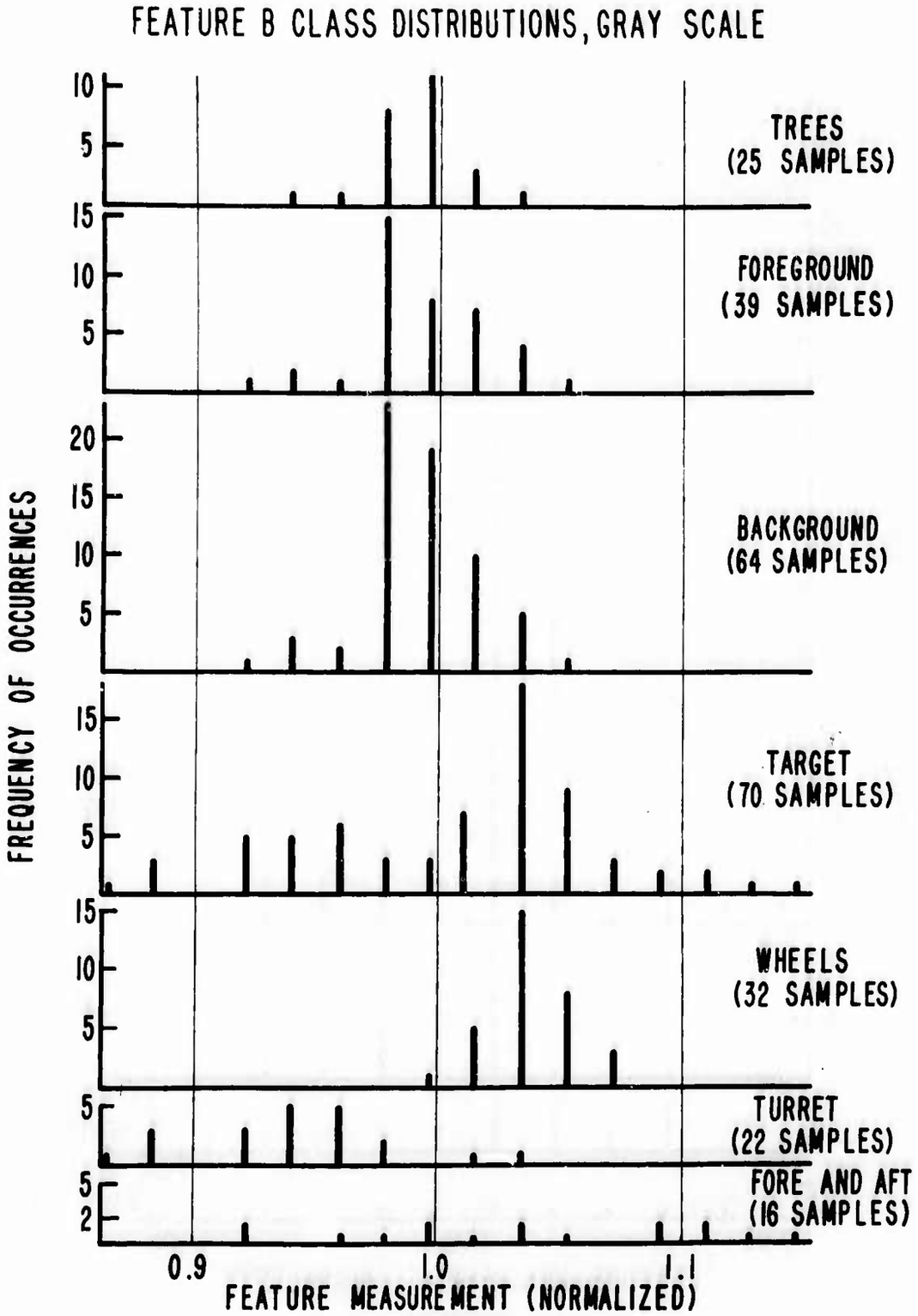

FEATURE B CLASS DISTRIBUTIONS, GRAY SCALE
FREQUENCY OF OCCURRENCES
10
5
TREES
(25 SAMPLES)
15
10
5
FOREGROUND
(39 SAMPLES)
20
15
10
5
BACKGROUND
(64 SAMPLES)
15
10
5
TARGET
(70 SAMPLES)
15
10
5
WHEELS
(32 SAMPLES)
5
TURRET
(22 SAMPLES)
5
2
FORE AND AFT
(16 SAMPLES)
0.9
1.0
1.1
FEATURE MEASUREMENT (NORMALIZED)

FIGURE 8

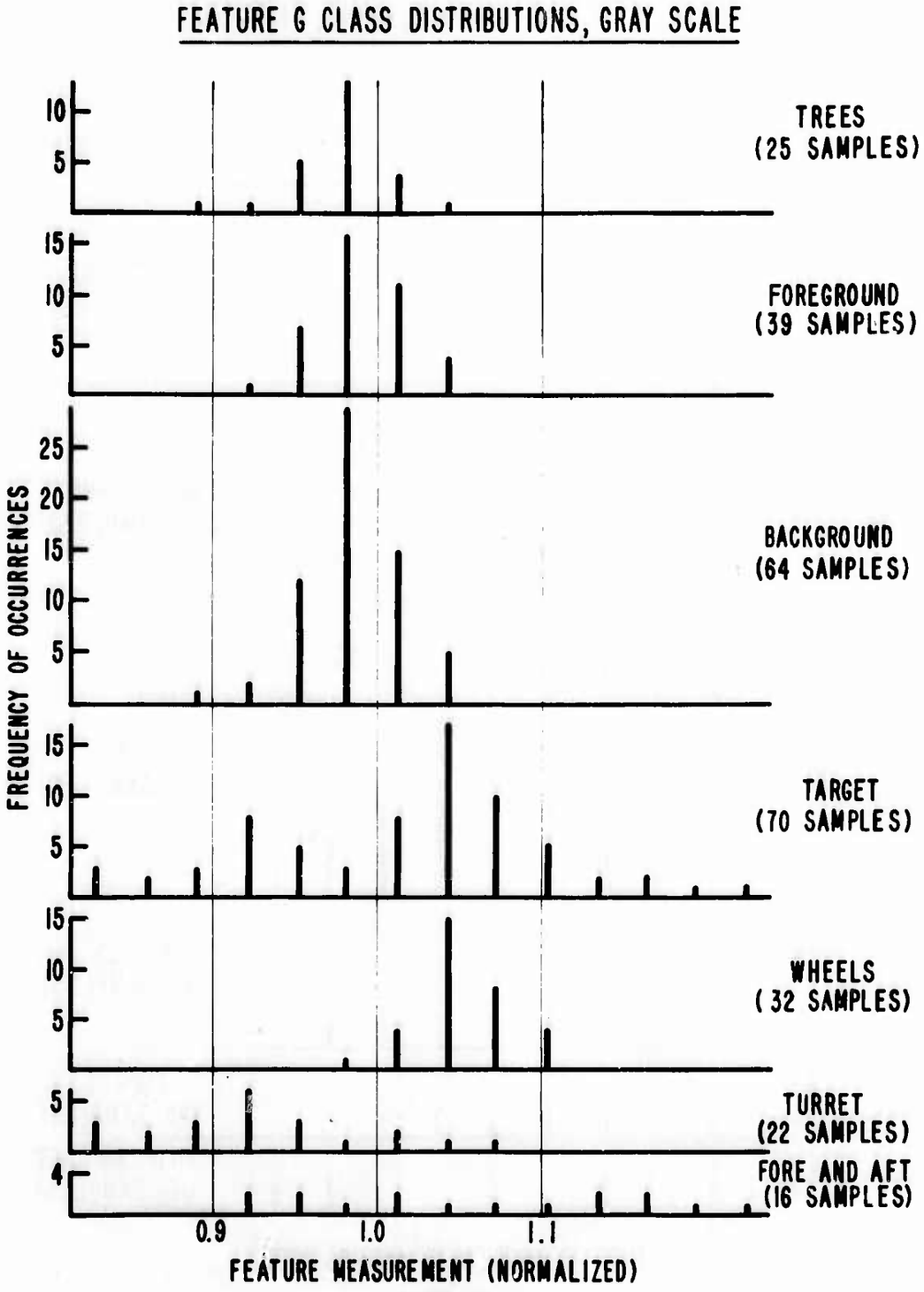

FEATURE G CLASS DISTRIBUTIONS, GRAY SCALE
TREES
(25 SAMPLES)
FOREGROUND
(39 SAMPLES)
BACKGROUND
(64 SAMPLES)
TARGET
(70 SAMPLES)
WHEELS
(32 SAMPLES)
TURRET
(22 SAMPLES)
FORE AND AFT
(16 SAMPLES)
FREQUENCY OF OCCURRENCES
0.9
1.0
1.1
FEATURE MEASUREMENT (NORMALIZED)

FIGURE 9

FEATURE G CLASS DISTRIBUTIONS, GRADIENT

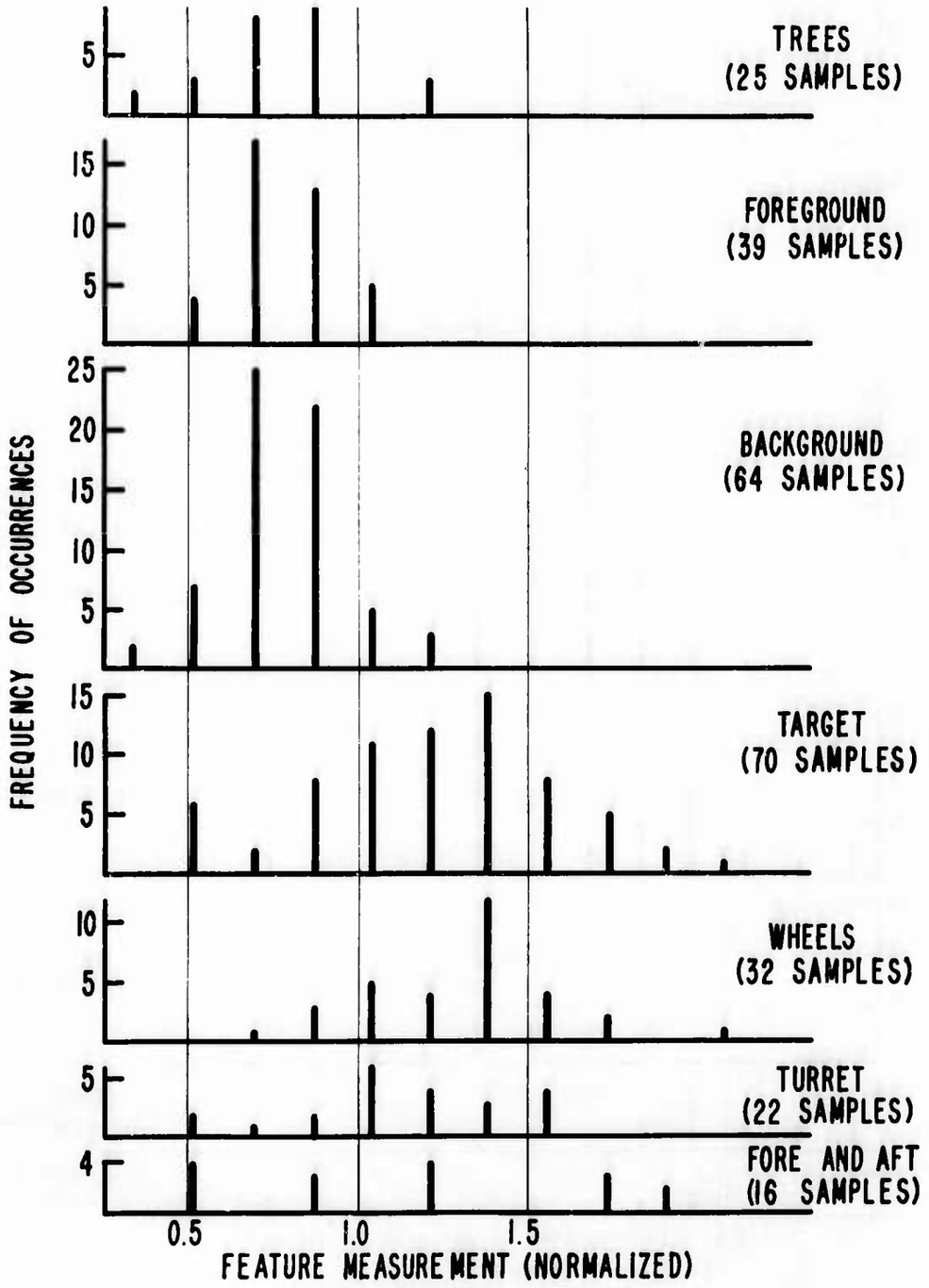


FIGURE 10

BEST MOMENT FEATURES AND SEPARATION FACTORS

PICTURE DATA / CLASSES	GRAY SCALE	GRADIENT	LAPLACIAN
TARGET VS BACKGROUND	U 0.78	G 0.84	K 0.83
WHEELS VS BACKGROUND	B, G 0.91	G 0.89	K 0.86
TURRET VS BACKGROUND	I, L 0.85	G 0.83	K 0.80
FORE AND AFT VS BACKGROUND	U 0.84	X 0.89	Z 0.86

FIGURE 11

BLANK PAGE

ANALYSIS OF GEOMETRICAL MOMENT FEATURES EXTRACTED FROM DIGITIZED TANK PHOTOGRAPHS USING AN ON-LINE PATTERN ANALYSIS AND RECOGNITION SYSTEM

PART II

Donald F. Roberts
Information Sciences Division
Rome Air Development Center
Griffiss Air Force Base, New York

PROBLEM DEFINITION. The feature evaluation study conducted at Aberdeen concentrated on analyzing each feature independently. In order to determine the discriminating power of the 27 moment features jointly, the gray scale, gradient and Laplacian pattern features were used in an independent study conducted at Rome Air Development Center. The study was carried out using the On-Line Pattern Analysis and Recognition System (OLPARS), developed at RADC and involved:

(1) Analysis of the geometric structure of the data.
(2) Design of classification logic.
(3) Evaluation of the given features.

Before reporting the results of this study, I would like to briefly describe the OLPARS system. A detailed description of the system is contained in Reference [1].

PATTERN RECOGNITION: AN OVERVIEW. First let us examine the relationship of OLPARS to pattern recognition in general. A typical pattern recognition system can be described as having two components.

A Feature Extractor and a Classification Device (Figure 1). OLPARS represents the second component. It is an on-line graphics oriented computer system used in solving two related pattern recognition problems: pattern analysis and pattern classification.

The pattern classification problem can be stated as [1]: "Given a set of L-dimensional statistically generated sample vectors (i.e. L-measurements) from each of K class (i.e. K states of the environment) design K decision regions (possibly K + 1 if a reject region is included) in the L-space, according to some goodness criterion. An L-dimensional vector from an unknown class is then classified by determing the decision region which contains the vector."

The pattern analysis problem involves analyzing the sample vectors in the L-space and selected subspaces in order to determine "structure" in the data. The term structure is used to denote inherent geometrical relationships between and within the K classes of data. Some examples are: regions of dense vector population, modality of the classes, and presence of wild shots. OLPARS allows the analyst to graphically identify this structure.

Sammon [1] makes the following observation concerning the utility of pattern analysis: "Although the pattern classification problem is defined independently of the pattern analysis problem, it is our contention that the pattern classification problem can be solved far more effectively by first performing the analysis function. The reason we believe this to be true is that, with the possible exception of some non-parametric techniques, the existing pattern classification algorithms either explicitly or implicitly make definite assumptions regarding the class data structure.* Hence, if the data structure assumptions implied by the use of a particular pattern classification algorithm are not valid, the resulting decision logic in general cannot be expected to be good. If, however, the data were first analyzed and then the appropriate pattern classification algorithm were selected to fit the data structure, one could expect a superior solution."

OLPARS IMPLEMENTATION. In designing the OLPARS system we have taken into consideration the following system objectives [1]:

(1) The system must be capable of processing large quantities of high-dimensional vector data.

(2) The system must provide efficient means for the user to interact with the data.

(3) The system must be capable of being expanded in a simple and straightforward manner to allow the addition of new mathematical and graphical routines.

(4) The system must provide the user with a convenient means for performing pattern analysis and pattern classification.

(5) The system should provide a great deal of flexibility in selecting mathematical techniques for solving the analysis and classification functions.

---

*Some examples of the kinds of implied assumptions made by various pattern classification algorithms are: the assumption that the class conditional probability density functions are multivariate gaussian, the L-measurements are statistically independent, the class conditional probability density functions are unimodal, or the distribution of the classes in the L-dimensional measurement space are pairwise linearly separable.

This last objective is very important. Often the solving of a pattern recognition problem is a learning process. That is, at any point in the process, the decision as to which mathematical algorithm or technique should be applied to the data, must be based on previous results. For example, some of the analysis algorithms are iterative techniques requiring adjustment of parameters before each iteration. Often automatically constructed decision boundaries require modification; therefore, we have put the man in the middle, giving him control of the system and providing him with information through a graphics console.

The processing of large quantities of vector data is supervised by the OLPARS Vector Filing System. Data sets that are too large to be stored in the main computer memory are stored on a peripheral high speed drum. Segments of the data are then rapidly swapped in and out of memory as required.

The Vector Filing System also provides data interaction capabilities. New data sets may be created by applying transformations and logical operations to existing data sets. A variety of transformations are available. The logical operations of union, not union, intersection and not intersection may be applied to subclasses of the data.

To fulfill the analysis requirements, OLPARS provides the analyst with clustering algorithms, transformations and mappings. Clustering algorithms, such as ISODATA [2] and Similarity Matrix [1], permit the analyst to identify data structure in the original vector space. A non-linear mapping technique developed at RADC [3], and eigenvector and discriminant plane projections allow the analyst to view data structure in one, two, and three dimensional displays on the C.R.T.

In pattern analysis, the problem typically involves identification of multimodal classes and the partitioning of these classes into unimodal subclasses. The motivation for such a procedure is that classification logic consisting of simple linear or piecewise classes often require complicated decision logic for separation and usually inhibit the classification process in general. For example, consider the data set consisting of K classes represented by the tree structure in Figure 2A. If class C, of the set were found to be bimodal, the class would be subdivided and the resulting data structure would be as in Figure 2B. Classification logic must then be designed for K + 1 classes.

The classification module of OLPARS provides the analyst with three techniques for the design of logic. The first technique involves on-line creation by the analyst of decision regions in eigenvector planes. Before the analyst can graphically create decision regions, he must first have the data vectors in a perceivable frame of reference. The L-dimensional vectors must therefore be mapped into a fitting subspace. It has been shown that the

eigenvectors of the lumped covariance matrix define such a subspace. In particular, the two eigenvectors corresponding to the maximum two eigenvalues define the best fitting two dimensional plane in the least squared sense.

The second method is an automatic thresholding technique using Fisher's Linear Discriminant for all pairs of classes. Fisher's Discriminant is the direction in the L-space which maximizes the difference between the class means relative to the sum of their variances. In the event that there is statistical overlap of target classes (non-linearly separable data), a likelihood ratio technique is available to the analyst to give the best separation under cost control. In addition, the data may be projected on to the discriminant plans defined by the Fisher Discriminant and a similarly defined vector orthogonal to it. This option is available to the analyst for on-line modification of automatically created decision logic.

The third classification technique available consists of simple nearest mean vector logic. Once the logic design is completed, the decision boundaries are stored in the computer and an independent set of data may be tested.

HARDWARD. The hardware for OLPARS is a CDC 160 4B computer driving a BR-85 graphics display. The analyst communicates with the computer and manipulates data via lightgun action, alphanumeric keyboard, and function keys.

OLPARS CONTROL TREE. The OLPARS system can be described functionally by a tree structure (Figure 3). Here the nodes of the tree represent options available to the analyst. These options are displayed in the form of a menu on the extreme right and left sides of the C.R.T.(Figure 4). By light gunning these options, the analyst automatically moves to another level in the control tree and calls on the mathematical algorithms and graphic displays needed to analyze his data.

ANALYSIS AND CLASSIFICATION OF TANK DATA. The normal procedure for a pattern recognition problem is to arbitrarily divide the original data vectors into two sets. One to be used for the design of logic and the other for the testing of that logic. However, since there were relatively few sample vectors in each class, it was not feasible to split this particular data set. Consequently, the results reported are only for a design set of sample vectors and are therefore considered preliminary. We hope in the future to obtain additional sample vectors for testing purposes.

The 27 dimensional Gray scale, Gradient and Laplacian data vectors are represented on the OLPARS system by the tree structure as shown in Figure 5. Each of the sets contained the following:

CLASS SYMBOL	CLASS DESIGNATION	NUMBER OF SAMPLES
W	Wheels	32
E	End views (fore & aft)	16
T	Turrets	22
F	Foreground	39
B	Background (Trees)	25

Preprocessing of the data consisted of normalizing along each of the 27 coordinate directions (i.e. $\mu = 0$, $\sigma = 1$).

Pattern analysis was carried out using the Non-Linear Mapping algorithm and Eigenvector projections.

For logic design Fisher's Linear Discriminant, and the Discriminant Plane Techniques were utilized. Classification logic was designed for the five class problem and results for the two class problems were obtained by combining the appropriate decision regions. (i.e. tank vs. non-tank)

LAPLACIAN. Pattern analysis of the Laplacian data revealed no particular iosyncrasies in the data, and all five subclasses appeared to be uni-modal.

Results of the logic design are given by the confusion matrix in Figure 6. For the five class problem there were 23 errors or 83.4% correct classification. As intuitively expected, most of the confusion occurred between the Foreground and Background classes. However, since the real problem is to separate tank from non-tank, these errors are not significant.

GRADIENT. Pattern analysis of the gradient data produced significant results. On both, the Non-Linear Mapping Plot and the projection of the data onto the two eigenvectors corresponding to the largest two eigenvalues, we were able to identify a "wildshot."

Any sample vector that is geometrically located, a relatively large distance from the mass of the sample vectors may be designated a "wild shot" by the analyst. The underlying reasons for such an occurrence range from card punching er[illegible] o faulty sensing equipment in the field. In any case, since it is desi[illegible] have operationally clean data for the design of classification logic, such vectors are often eliminated by the analyst.

Analysis also showed the E (fore and aft) class to be bimodal. Consequently, the class was divided into two subclasses. The resulting structure is shown in Figure 7.

Classification logic was designed for the six classes shown in Figure 7. For the six class problem, there were 13 errors or 90.9% correct classification. As a result of combining decision regions, there were 3 errors or 98.4% correct classification for the two class problems (Figure 8).

GRAY SCALE. The structure of the Gray Scale Data was very similar to that of the Gradient Data. Analysis showed the E (fore and aft) class to be bi-modal and the existence again of a "wild shot" in this class. However, the "wild shot" was not the same vector identified in the Gradient Data Set.

The six class problem yielded 7 errors or 94.7% correct classification. For the two class problem there was 1 error or 99.2% correct classification (Figure 9).

CONCLUSIONS. A few conclusions may be drawn, based on the above results. The original gray scale features appear to contain suffieient information to detect the tank in the given environment.*

Some discriminatory information has been lost by implementing the gradient transformation and a significant amount of discriminatory information lost in the Laplacian transformation.

In addition, by projecting the data on to the various Discriminant Planes, [1] we were able to hypothesize that the gradient classification logic is probably not as good as the percentages indicated. The pairwise two dimensional plots showed that in most instances, the two classes were relative close with points from one or both classes located very close to the separting threshold or boundary. Similar gray scale plots revealed distinct class separation in most instances.

Future studies might explore the possibility of designing classification logic with a subset of the original 27 gray scale features. An on-line feature evaluation module will be added to the OLPARS system in the near future. This module will provide the capability to conduct such a study. The first objective, however, is to obtain additional sample vectors with which to test the logic already designed.

---

*Although the scope of this problem was narrow (i.e. tank in one environment), the gray scale extraction technique may prove useful in a wide variety photo reconnaissance problem and warrants further study.

REFERENCES

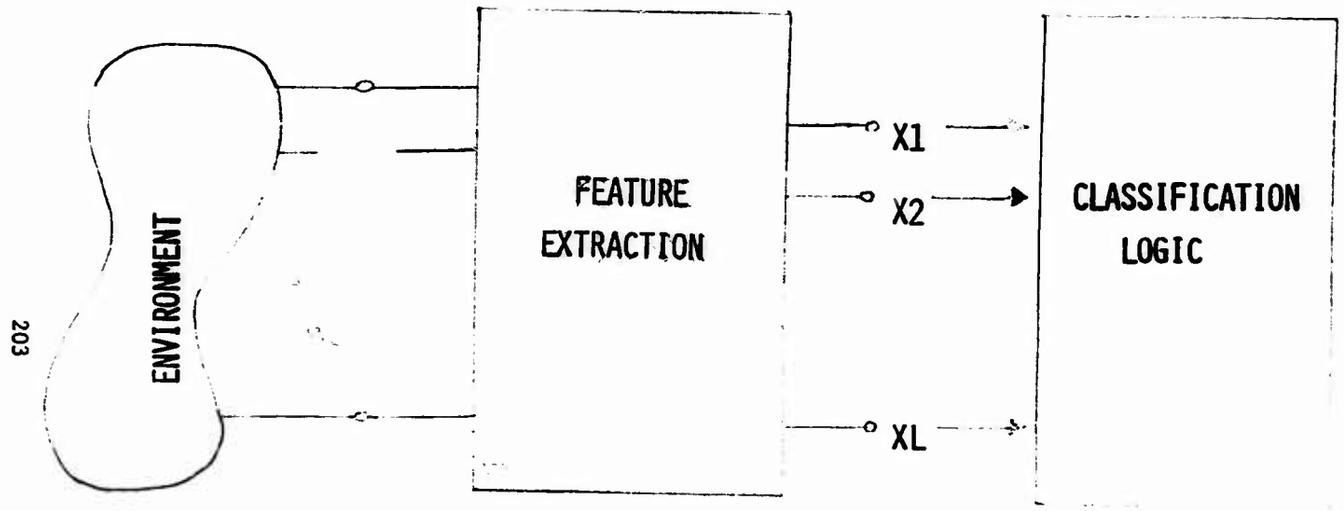
[1] J.W. Sammon, "On-Line Pattern Analysis and Recognition System (OLPARS)," Rome Air Development Center, Rome, N.Y., TR RADC - TR-68-263, August 1968.

[2] G. H. Ball and D. J. Hall, "ISODATA, An Interactive Method of Clustering Multivariate Data," Stanford Research Institute, Menlo Park, California, Technical Report SRI Project 5533, 1968.

[3] J.W. Sammon, "A Non-Linear Mapping for Data Structure Analysis," IEEE Trans. Computers, Vol. C-18, pp. 401-409, May, 1969.

[4] B. M. Landson, Captain, "On-Line Pattern Analysis and Recognition," S&E Symposium, Brooks AFB, Texas, 7-9 October 1969.

PATTERN RECOGNITION



203

PATTERN RECOGNITION = FEATURE DEFINITION + PATTERN CLASSIFICATION

FIGURE 1

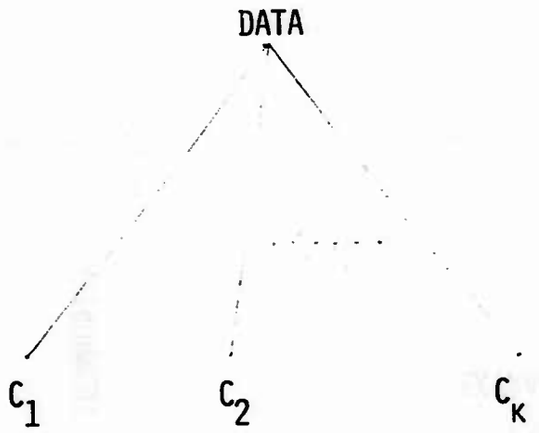


FIGURE 2a

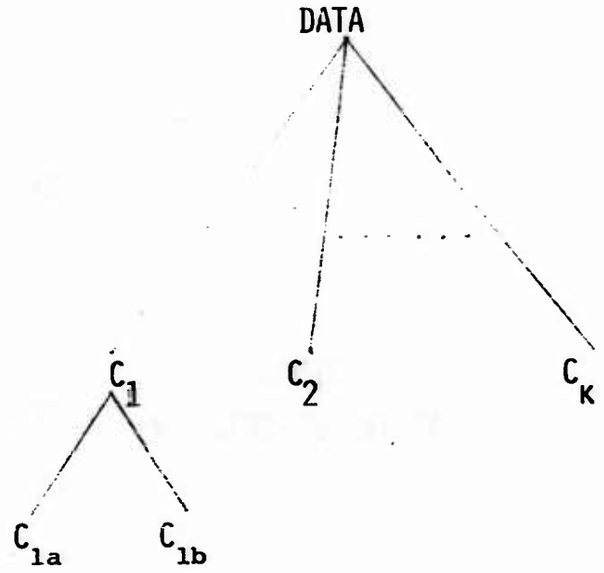


FIGURE 2b

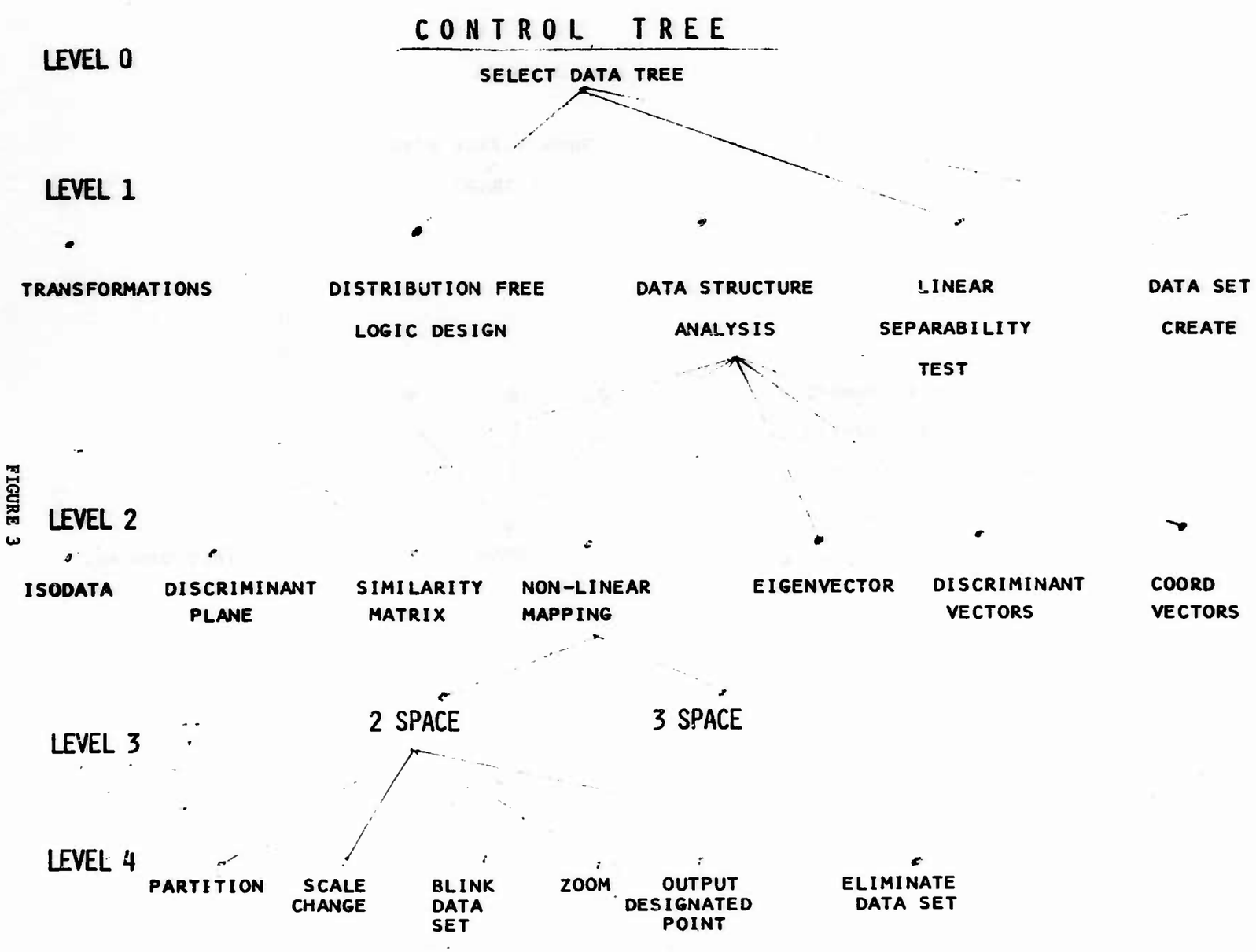
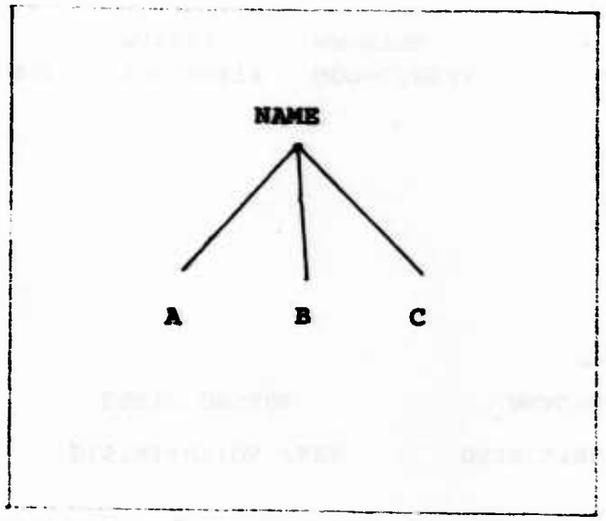

CONTROL TREE
LEVEL 0
SELECT DATA TREE
LEVEL 1
TRANSFORMATIONS
DISTRIBUTION FREE LOGIC DESIGN
DATA STRUCTURE ANALYSIS
LINEAR SEPARABILITY TEST
DATA SET CREATE
LEVEL 2
ISODATA
DISCRIMINANT PLANE
SIMILARITY MATRIX
NON-LINEAR MAPPING
EIGENVECTOR
DISCRIMINANT VECTORS
COORD VECTORS
LEVEL 3
2 SPACE
3 SPACE
LEVEL 4
PARTITION
SCALE CHANGE
BLINK DATA SET
ZOOM
OUTPUT DESIGNATED POINT
ELIMINATE DATA SET

FIGURE 3

CRT DISPLAY OF OPTIONS

LEVEL 1

*Return L(0)



* Transformations
* Distribution Free
* Data Structure Analysis
* Linear Separability Test
* Create Data Set

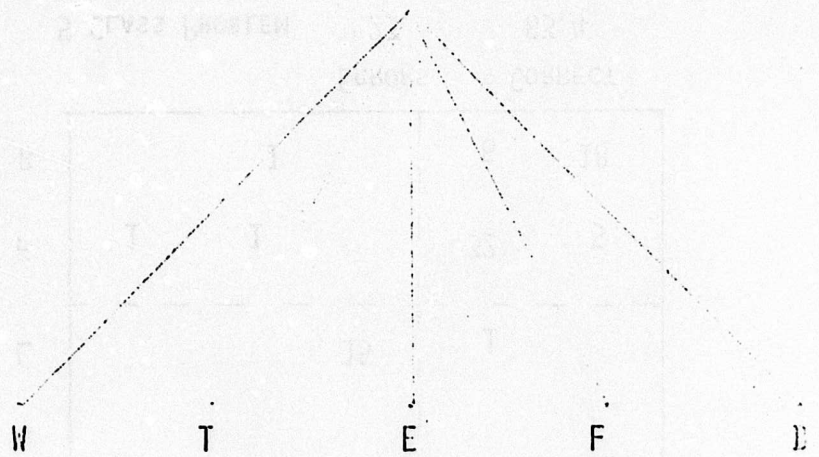
LEVEL 1

DATA TREE = NAME

FIGURE 4

FIGURE 5

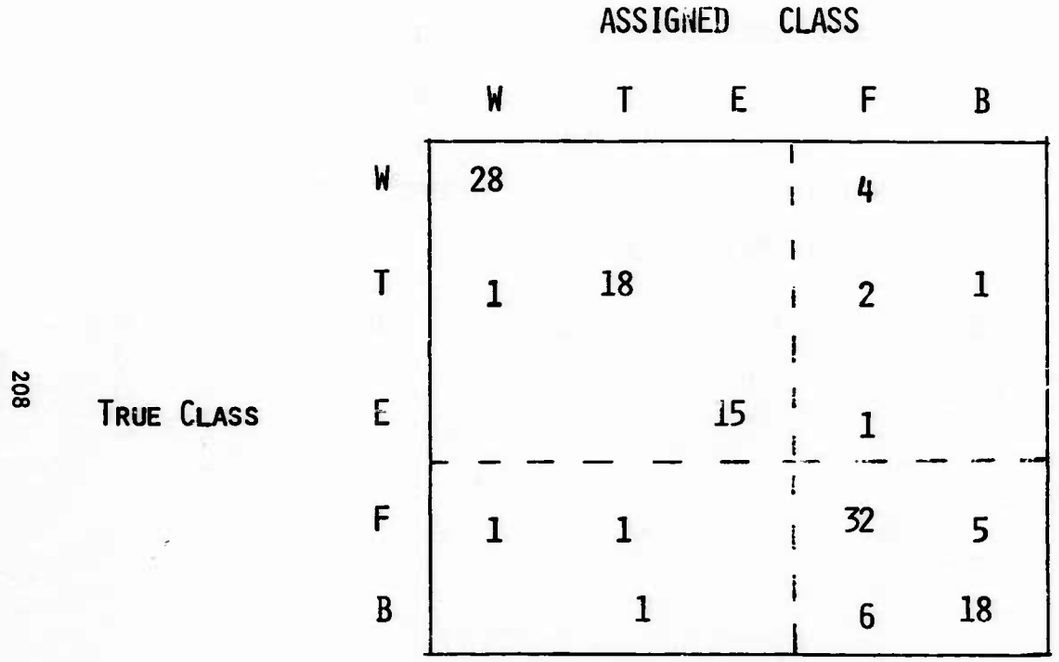
ABERDEEN TANK DATA



FIVE CLASS PROBLEM

CLASS SYMBOL	CLASS DESIGNATION	NUMBER OF SAMPLES
W	WHEELS	32
T	TURRETS	22
E	END VIEWS (FORE AND AFT)	16
F	FOREGROUND	30
B	BACKGROUND (TREES)	25

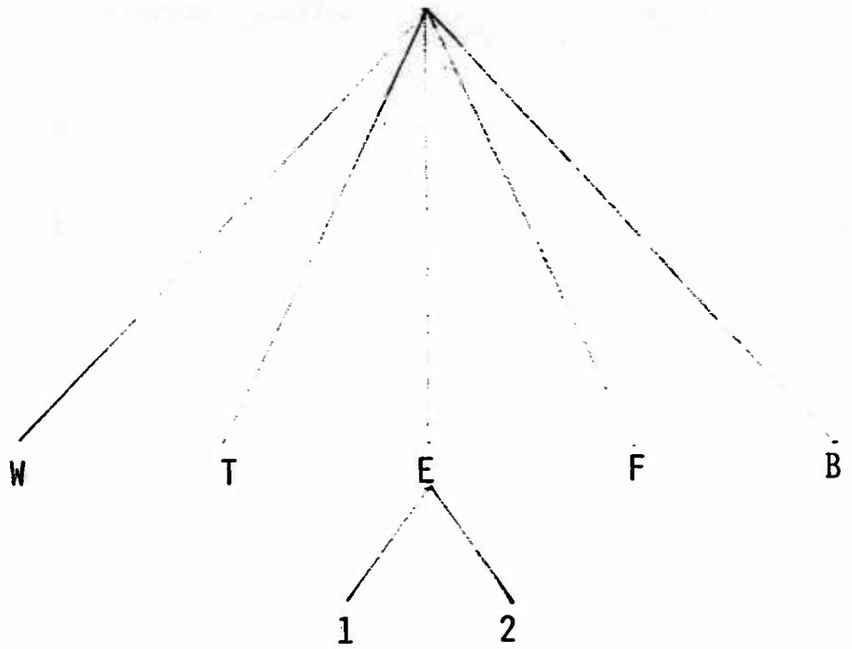
CONFUSION MATRIX LAPLACIAN DATA



	Errors	% Correct
5 Class Problem	23	83.4
2 Class Problem	11	91.7

FIGURE 6

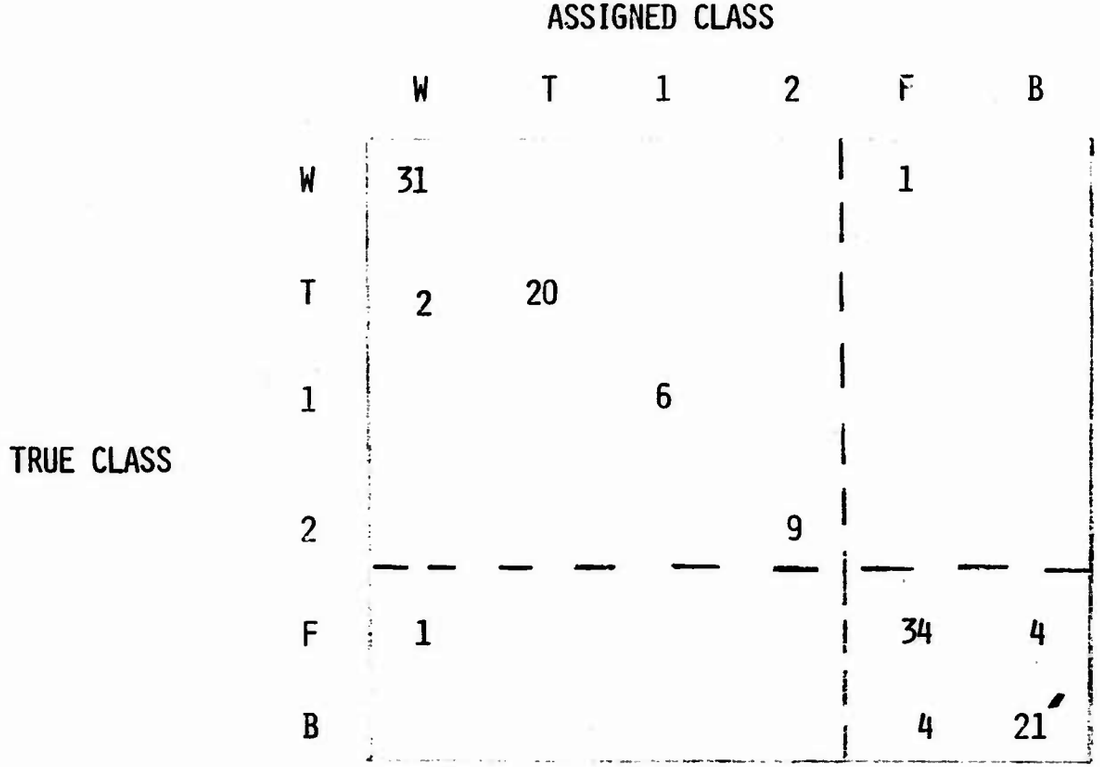
RESTRUCTURED GRADIENT DATA



SIX CLASS PROBLEM

FIGURE 7

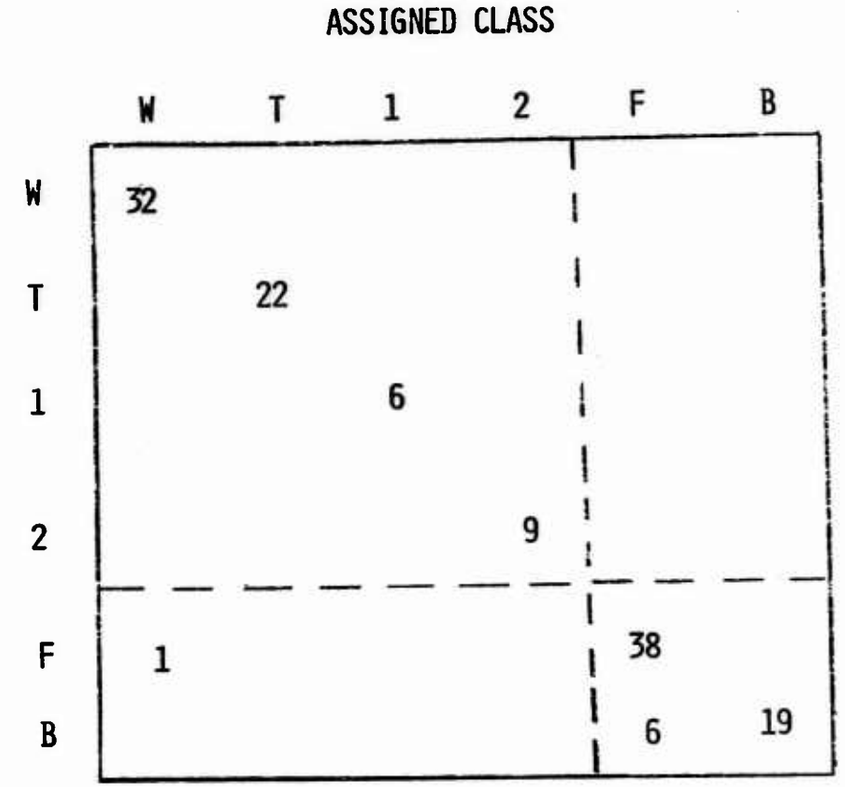
CONFUSION MATRIX GRADIENT DATA



	Errors	% Correct
6 Class Problem	12	90.9
2 Class Problem	2	98.4

FIGURE 8

CONFUSION MATRIX GRAY SCALE DATA



	Errors	% Correct
6 Class Problem	7	94.7
2 Class Problem	1	99.2

FIGURE 9

BLANK PAGE

A FAMILY OF CUBIC SPLINE SUBROUTINES

R. D. Scanlon
Benet R&D Laboratories
Watervliet Arsenal
Watervliet, New York

ABSTRACT. A subroutine package designed to evaluate the coefficients of interpolating and approximating cubic splines under a large variety of circumstances is described. The advantages of the package are argued. An application involving tangential strain is examined.

INTRODUCTION. A computist is frequently asked to manipulate experimental data in accordance with a mathematical analysis of the physical situation...a typical example being a request for the first derivative. Since there is almost always a shortage of time and money, there is a great interest in something fast and cheap.

The first difficulty is noise, the random element which overlies the desired signal. Traditionally one improves the signal by concentrating on the randomicity of the noise and the coherence of the underlying process but in doing this, we may run counter to strong opinions the experimentalist has about characteristics of the physical process. He may insist that it is monotonic over some interval or that its slope is monotonic. He may have certain notions of smoothness; he might prefer that the fitted function would take the form of a curve which a graphic artist would draw through the data.

In practice, this means that the fitted curve should be about as smooth as the artist would draw it but must not depart too far from any one group of the data points. If we attempt to fit the data with a function which is analytic over the whole interval, we will almost inevitably do violence to one or the other of these constraints.

A second problem is that of scale. The fit must be precise, but not too precise; smooth, but not too smooth. Over-precision takes time and in computation, time is money.

A piecewise analytic function preserving some degree of continuity across the joins has proved an excellent working tool in both of these areas.

A separate but related question is overall efficiency in computation. This is not at all the same thing as efficient coding. The polishing of inner loops has a place in large scale computations but is a positive evil in jobshop scientific computation. The computist should be enabled to keep his eye on the desired course of the calculations without getting lost in the intricacies of threading the problem through the machine.

One solution is the problem oriented language where the statements are directed solely to the problem and all machine oriented questions are laid aside. It is quite likely this is the road which scientific computation will ultimately take. Lacking such a language, it is our contention that in the meantime it is possible to write families of subroutines which embed a problem oriented structure in Fortran.

This paper is a description of such a family of subroutines designed to handle many of the difficulties associated with the manipulation of discrete data as though it were continuous. They are also designed to provide results quickly and easily with a minimum of specialized coding and a maximum of help in avoiding pitfalls. These routines are programmed for reasaonble efficiency but it is hoped that flexibility and ease of use have been preferred over mindless polishing. One notes that in most scientific computation, program run time is a small proportion of program preparation time.

The Spline and Its Mathematical Analog

A spline is a thin strip of metal, wood, or plastic. When such a strip is bent, it takes on a shape of smoothly distributed tension which is pleasing to the eye and at the same time mechanically elegant. The use of the spline to fair ships' lines probably started shortly after man began building planked vessels and the resulting hulls are noted for their blending of beauty and utility. Of equal antiquity is the use of a spline, known as a tawami jaku, by the Japanese carpenter to produce the powerful, sweeping curves which distinguish Japanese rooflines.

One form of the spline, as used in lofts when fairing lines on full-scale plans, is a grooved strip constrained to pass through chosen points by means of weights which have projections that ride in the groove. The bending movement of such a spline, considered as a thin beam, is defined by

$$M(x) = \frac{E}{R(x)} \int z^2 dA$$

where E is Young's modulus and R the radius of curvature.

We are allowed by custom to replace R(x) by 1/y" and since the spline is of constant cross section, we write

$$M(x) = ky".$$

Because of the way that the weights are free to move along the groove, this is a simply supported beam: the variation of M(x) between the weights is linear and is continuous at the constraining points.

This approximation can be represented by a piecewise cubic polynomial with a continuous second derivative and possible jump discontinuities in the third derivative at the knots, nodes, or joins which replace the constraining weights.

Definition of a Cubic Spline

Given an interval

$$a \leq x \leq b,$$

a mesh on the interval

$$\Delta: \quad a = x_1 < x_2 < \ldots < x_N = b,$$

and an associated set of ordinates

$$Y: \quad y_1, y_2, \ldots, y_N \quad ,$$

then a cubic spline satisfies

$$S_\Delta (Y;x) \in C^2 \quad \text{on} \quad [a,b]$$

$$S_\Delta (Y;x_j) = y_j \quad (j = 1,2,\ldots,N)$$

and is coincident with a cubic on each subinterval

$$x_{j-1} \leq x \leq x_j \quad (j = 2,3,\ldots,N)$$

If in addition

$$S_\Delta^{(p)}(a+) = S_\Delta^{(p)}(b-) \qquad (p=0,1,2)$$

the spline is said to be periodic with period (b - a).

If, alternatively,

$$S_\Delta(Y;x_j) = y_j + \varepsilon_j$$

with the ε_j subject to some minimizing constraint, it is said to be approximating rather than an interpolating spline.

Foundations of the Algorithms

Without going into detail, we desire to indicate the general paths which can be followed in constructing cubic spline algorithms.

Since M(x) is linear between joins, we have on the kth subinterval

$$S'' = M_{k-1}\frac{x_k - x}{\ell_k} + M_k\frac{x - x_{k-1}}{\ell_k} \quad , \quad x_{k-1} \le x \le x_k$$

where

$$M_k = S''_\Delta(Y;x_k)$$

$$\ell_k = x_k - x_{k-1}$$

By integrating twice and making use of

$$S \in C^2$$

we are lead to

$$\frac{\ell_k}{6}M_{k-1} + \frac{\ell_k + \ell_{k+1}}{3}M_k + \frac{\ell_{k+1}}{6}M_{k+1} = \frac{y_{k+1} - y_k}{\ell_{k+1}} - \frac{y_k - y_{k-1}}{\ell_k}$$

This set of N-2 conditions with two additional constraints on the end conditions will allow us to compute a solution for an interpolating cubic spline.

For an approximating cubic spline, we can minimize

$$\int_a^b \{S''(x)\}^2\,dx + \lambda\sum_{j=1}^{N}\{S(x_j) - y_j\}^2$$

where the first term is a popular measure of smoothness and the second term is an even more popular measure of the departure from the data. The parameter λ is introduced to strike a balance between the two.

For the case where there is a join corresponding to each data point, a variational demonstration leads to

$$S'''(x_j-) - S'''(x_j+) = \lambda \{S(x_j) - y_j\}$$

and this, together with the relationship derived for the interpolating spline gives us

$$\frac{1}{\lambda \ell_k \ell_{k-1}} M_{k-2}$$

$$+ \{\frac{\ell_k}{6} - \frac{1}{\lambda} [\frac{1}{\ell_{k+1}\ell_k} + \frac{2}{\ell_k^2} + \frac{1}{\ell_k \ell_{k-1}}]\} M_{k-1}$$

$$+ \{\frac{\ell_k+\ell_{k+1}}{3} + \frac{2}{2} [\frac{1}{\ell_{k+1}^2} + \frac{1}{\ell_{k+1}\ell_k} + \frac{1}{\ell_k^2}]\} M_k$$

$$+ \{\frac{\ell_{k+1}}{6} - \frac{1}{\lambda} [\frac{1}{\ell_{k+2}\ell_{k+1}} + \frac{2}{\ell_{k+1}^2} + \frac{1}{\ell_{k+1}\ell_k}]\} M_{k+1}$$

$$+ \frac{1}{\lambda \ell_{k+2} \ell_{k+1}} M_{k+2}$$

$$= \frac{y_{k+1}-y_k}{\ell_{k+1}} - \frac{y_k-y_{k-1}}{\ell_k}$$

With suitable constraints on the end points, we will be able to compute a solution for an approximating cubic spline with N data points and N joins.

For the case of N data points and M joins where N > M we define a set of N + 2 cardinal splines

$$\left.\begin{array}{ll} S_k(x_j) = \delta_{k,j} & (j=1,2,\ldots,N) \\ S_k'(x_i) = 0 & (i=1 \text{ and } N) \end{array}\right\} \quad (k = 1,2,\ldots,N)$$

$$\left.\begin{aligned} \bar{S}_k(x_j) &= 0 \qquad (j = 1,2,\ldots,N) \\ \bar{S}_k'(x_i) &= \delta_{k,i} \qquad (i = 1 \text{ and } N) \end{aligned}\right\} \quad (k = 1 \text{ and } N)$$

It is readily seen that these form a basis set for all cubic splines over the given mesh and the minimization problem can be solved directly.

Classification of Spline Routines

The names of the routines (consisting of 5 or 6 letter groups) are codified here so that their interrelationships may be seen at a glance. Expanded information about calling sequences and returned information is given for those routines which are starred.

Interpolating, $S_\Delta(Y;x_j) = y_j$:

Periodic

Normal Precision - SPLNP*
Extended Precision - DSPLNP

Non-Periodic

Normal Precision - SPLND
Extended Precision - DSPLND

Approximating, $S_\Delta(Y;x_j) = y_j + \varepsilon_j$, N points, M joins :

N = M

Periodic

Normal Precision - SPLPW
Extended Precision - DSPLPW

Non-Periodic

Normal Precision - SPLSW*
Extended Precision - DSPLSW

N > M

Periodic

Normal Precision - SPLFP
Extended Precision - DSPLFP

Non-Periodic

Normal Precision - SPLFT
Extended Precision - DSPLFT

Evaluating:

(These routines evaluate a previously constructed spline function or its derivatives at a given point.)

Periodic

Normal Precision - SNTRP
Extended Precision - DSNTRP

Non-Periodic

Normal Precision - SNTER*
Extended Precision - DSNTER

SPLNP, an example of an interpolating, periodic, normal precision, cubic spline.

SUBROUTINE SPLNP (N, X, Y, DY, A, DUM)

N	The number of data points.
X	Vector of x-coordinates of the data points. Dimensioned $\geq$ N.
Y	Vector of corresponding y-coordinates. Dimensioned $\geq$ N.
DY	Upon return, this vector will contain (as a by-product) the fitted slopes corresponding to X and Y. Dimensioned $\geq$ N.
A(J,K)	Upon return, this array will contain the coefficients of the interpolating, cubic spline. Dimensioned at J = 4, K $\geq$ N.

$$Y = \sum_{i=0}^{3} a_i z^i$$

where $z = x - X(K-1)$

for $X(K-1) \leq x < X(K)$

$$J = \begin{cases} 1, & a_0 \\ 2, & a_1 \\ 3, & a_2 \\ 4, & a_3 \end{cases}$$

$$K = \begin{cases} 1, & x < X(2) \\ 2, & X(2) \leq x < X(3) \\ \ldots & \\ N-1 & X(N-1) \leq x < X(N) \\ n & X(N) \leq x \end{cases}$$

DUM(J,K) A working space. Dimensioned at J = 2, K $\geq$ N.

SPLSW, an example of an approximating, non-periodic, normal precision cubic spline.

SUBROUTINES SPLSW (N,X,Y,W,YDP,A,IA,VA,IB,VB,ALAM)

N The number of data points.

X Vector of x-coordinates of the data points. Dimensioned $\geq$ N.

Y Vector of corresponding y-coordinates. Dimensioned $\geq$ N.

W Vector of non-negative weights assigned to corresponding data points. Dimensioned $\geq$ N.

YDP Upon return, this vector will contain (as a by-product) the fitted second derivatives at the corresponding x-coordinates. Dimensioned $\geq$ N.

A(J,K) Upon return, this array will contain the coefficients of the approximating cubic spline. Dimensioned at J = 5, K $\geq$ N.

$$Y = \sum_{i=0}^{3} a_i z^i$$

where $z = x - x_{k-1}$

for $x_{k-1} \leq x < x_k$

$$J = \begin{cases} 1, & a_0 \\ 2, & a_1 \\ 3, & a_2 \\ 4, & a_3 \\ 5, & 0 \end{cases}$$

$$K = \begin{cases} 1, & x < X(2) \\ 2, & X(2) \leq x < X(3) \\ \ldots & \\ N-1, & X(N-1) \leq x < X(N) \\ N, & X(N) \leq x \end{cases}$$

IA,VA Desired constraint at leftmost point.

$$IA = \begin{cases} 1, & y'_L = VA \\ 2, & y''_L = VA \\ 3, & y''_L = VA * y''_{L+1} \end{cases}$$

IB,VB Desired constraint at rightmost point. Same structure as IA,VA.

ALAM Value of λ in

$$\int_a^b \{S''(x)\}^2 \, dx + \lambda \sum_{j=1}^{N} w_j^2 \{S(x_j) - y_j\}^2$$

SNTER, an example of a function with multiple entry points for evaluating a non-periodic, normal precision, cubic spline. Two of the entry points are, in effect, subroutines but the entire routine is defined as a function so that the proper return coding will be generated for those entries which are truly functions.

After the spline has been constructed and before any evaluation calls are made, a cell must be made to SINIT to initialize this routine.

FUNCTION SNTER (XA,Y, YP, YDP, YTP)

XA Desired argument.

Returned valves:

Y y
YP y'
YDP y"
YTP y'"

ENTRY SINIT (N,L,X,A)

An initializing entry.

N Number of joins.

L= { 4, interpolating spline
   { 5, approximating spline

X Vector of x-coordinates of joins. Dimensioned $\geq$ N.

A(J,K) Array of coefficients as returned by a spline routine such as SPLNP or SPLSW. Dimensioned at J = L, K $\geq$ N.

ENTRY G (XA)

Returns y as a function value.

ENTRY GP (XA)

Returns y' as a function value.

ENTRY GDP (XA)

Returns y'' as a function value.

ENTRY GTP (XA)

Returns y''' as a function value.

An Application

Several of these routines were used recently to perform a typical series of analytic manipulations on some empirical data. The results were satisfactory and form an interesting example.

A hollow cylinder with a crack in the inner surface was subjected to internal pressure. Circumferential strain readings were taken at 16 positions on the outer surface for 10 crack depths and 4 pressures. The strain gages lay in a plane perpendicular to the axis. Of the 640 possible readings, 16 were missing because the tube ruptured before the last scheduled measurement, 35 were missing because of gage failure, 3 were apparent gage failures, and one was judged a transcription error. These 39 missing points were replaced by bi-quadratic interpolation.

Considering these readings as representing the tangential strain at selected points on the circumference of a circle, it was desired to determine numerically the shape of the perimeter of the distorted figure and from this the increment, ΔA_o, in this area.

Then it was desired to evaluate

$$\frac{K_I}{P} = \left[\frac{E}{2(1-\nu)^2} \frac{d(\Delta V_o/P)}{}\right]^{1/2}$$

and

$$\frac{K_I^*}{P} = \frac{E}{2(1-\nu)^2} \frac{d(\Delta V_o/P)}{da^*}\Big]^{1/2}$$

where

$$a^* = a + \frac{1}{2\pi}\left[\frac{K_I}{\sigma_y}\right]^2$$

- a - crack depth
- E - Young's modulus
- ν - Poisson's ratio
- K_I - Opening mode stress intensity factor
- P - Pressure
- r_2 - Outer radius
- σy - 0.2% tensile yield stress
- ΔV_o - ΔA_o^* Unit length

The problem of determining the increment in area is somewhat ambiguous since we lack any direct knowledge of the radial displacment. However, if we can interpolate the tangential strain measurements with an angular function and assume that the departure of the distorted perimeter from the original circle is small, then two simplified approaches suggest themselves.

We can integrate the tangential strain to obtain the change in length of the perimeter

$$\Delta P = \int_o^{2\pi} r\epsilon_\theta(\theta)d\theta \ ,$$

and, assuming the distorted figure remains circular, the change in area inclosed by the perimeter is

$$\Delta A_{o_L} = \frac{2P\Delta P + (\Delta P)^2}{4\pi} \ .$$

As an alternative approach we can assume that any displacemnet of the perimeter is solely radial, that is

$$u_\theta\ (r_2, \theta) \equiv 0$$

and then from

$$\varepsilon_\theta\ (r_2, \theta) = \frac{1}{r_2} \frac{\partial u_\theta (r_2, \theta)}{\partial \theta} + \frac{u_r (r_2, \theta)}{r_2}$$

we have

$$u_r\ (r_2, \theta) = \varepsilon_\theta\ (r_2, \theta) \cdot r_2$$

and

$$\Delta A_o = \int_0^{2\pi} \int_{r_2}^{r_2 + u(r_2, \theta)} r\ dr\ d\theta$$

Since these two approaches give reasonable bounds on the distortion of a circular cross section, we can assume that the degree to which they approximate each other is an indication of the number of significant figures in our approximation of the increment in area.

Since ε_θ is a periodic function and the data was very smooth (Figure 1) we chose the interpolating, periodic spline routine, SPLNP, to perform these operations and were able to obtain agreement to about four figures.

To indicate the smooth nature of the strain data as interpolated by a periodic spline and the obvious nature of the transcription error, we show the data and the fitted spline functions for four pressures at a crack depth of 0.117 inches in Figure 2. The datum at $+86^\circ$ and 30,000 psi looks like an outlier. It was recorded as 945 μ in/in and quite likely was really 845 μ in/in as indicated by the dashed curve.

As an interesting sidelight, it is apparent upon close inspection of Figure 2 that the minimum strain is not at the indexed zero. Examination of the sectioned tube confirmed that the gages were, indeed, displaced a small positive angle from the true location of the crack. Of course, this has no effect on the computation.

To approximate ΔV_0 we used a non-periodic, approximating cubic spline routine, SPLSW. As a guide to the choice of λ we used the knowledge from physical considerations that

$$\frac{d^2K_I}{da^2}$$

is strictly monotonic increasing and so we wanted λ to be small enough to make this true of our approximating spline. On the other hand, we must evaluate a^* and K_I^* by an iterative process and λ must not be so small that this process is unstable. A value of $\lambda = 100$ proved to be a reasonable compromise leading to rapid and stable convergence of K_I^*. See Figures 3, 4, and 5.

Conclusions

The spline routines presented here have been used in a large variety of circumstances. They have proven easy to use and satisfactory in their results. Their use is unhesitatingly recommended in the manipulation of experimental data.

SELECTED BIBLIOGRAPHY

1. Ahlberg, J. H., Nilson, E. N., and Walsh, J, L., "The Theory of Splines and Their Application." Academic Press, New York, 1967.

2. Birkhoff, G. and DeBoor, C. R., "Piecewise Polynomial Interpolation and Approximation," "Approximation of Functions," H. L. Garabedian, ed., Elsevier, Amsterdam, 1965, pp. 164-190.

3. Greville, T. N. E., "Spline Functions, Interpolation, and Numerical Quadrature," "Mathematical Methods for Digital Computers," Vol. 2, A Ralston and H. S. Wilf, eds., Wiley, New York, Ch. 8, 1967, pp. 156-168.

4. Greville, T. N. E., "Introduction to Spline Functions," "Theory and Applications of Spline Functions," D. C. Handscomb, ed., Academic Press, New York, 1969, pp. 1-35.

5. Handscomb, D. C., "Spline Functions," "Methods of Numerical Approximation," D. C. Handscomb, ed., Pergamon Press, London, 1966, pp. 163-167.

6. Scanlon, R. D., "An Interpolating Cubic Spline Fortran Subroutine," Technical Report WVT-7010, Watervliet Arsenal, Watervliet, N. Y., 1970.

7. Underwood, J. H., Laselle, R. R., Scanlon, R. D., and Hussain, M. A,, "A Compliance K Calibration for a Pressurized Thick-Wall Cylinder with a Radial Crack," Technical Report WVT-7026, Watervliet Arsenal, Watervliet, New York, 1970.

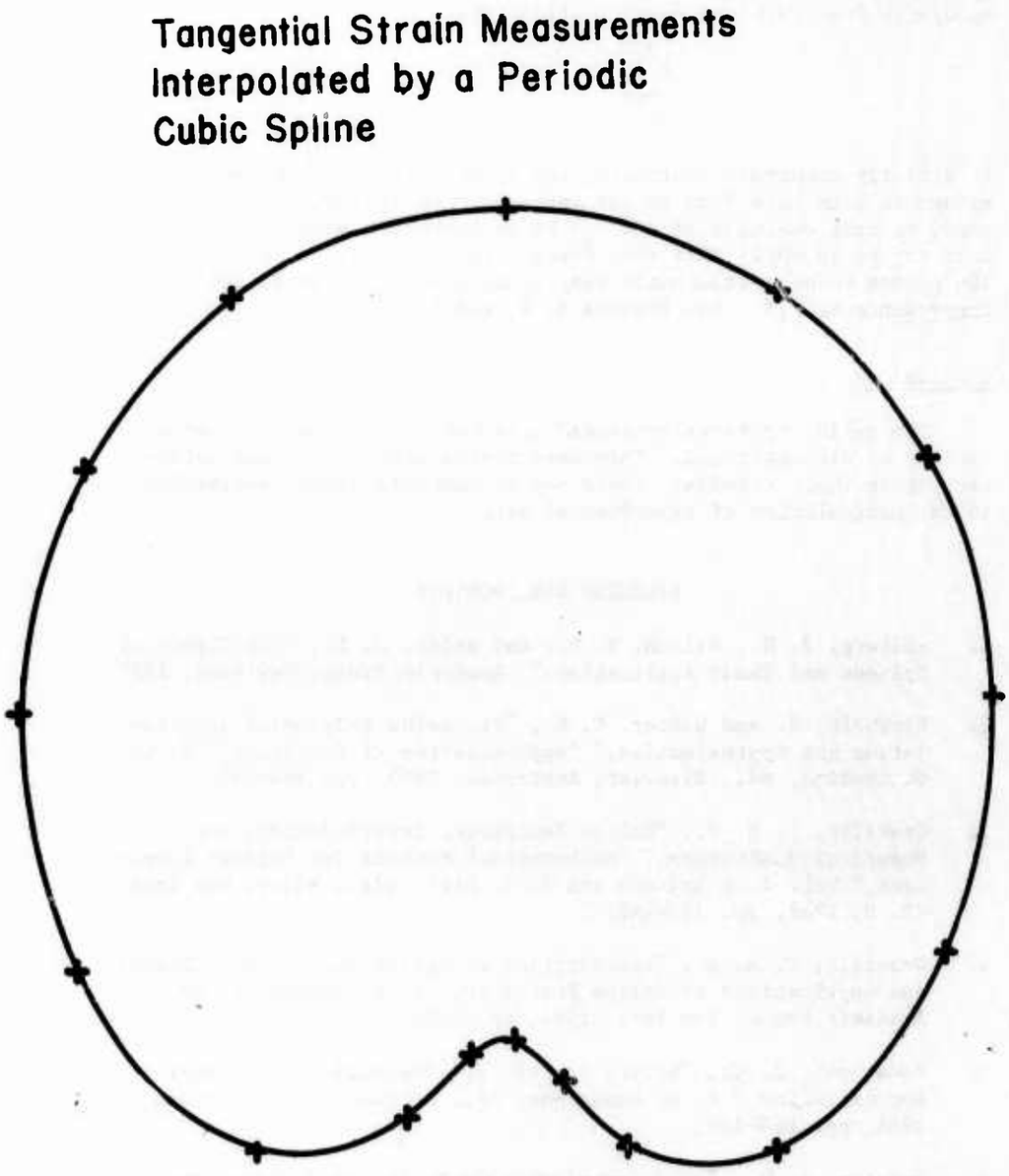


Figure 1.

Plot of data and interpolating, periodic spline in polar coordinates.

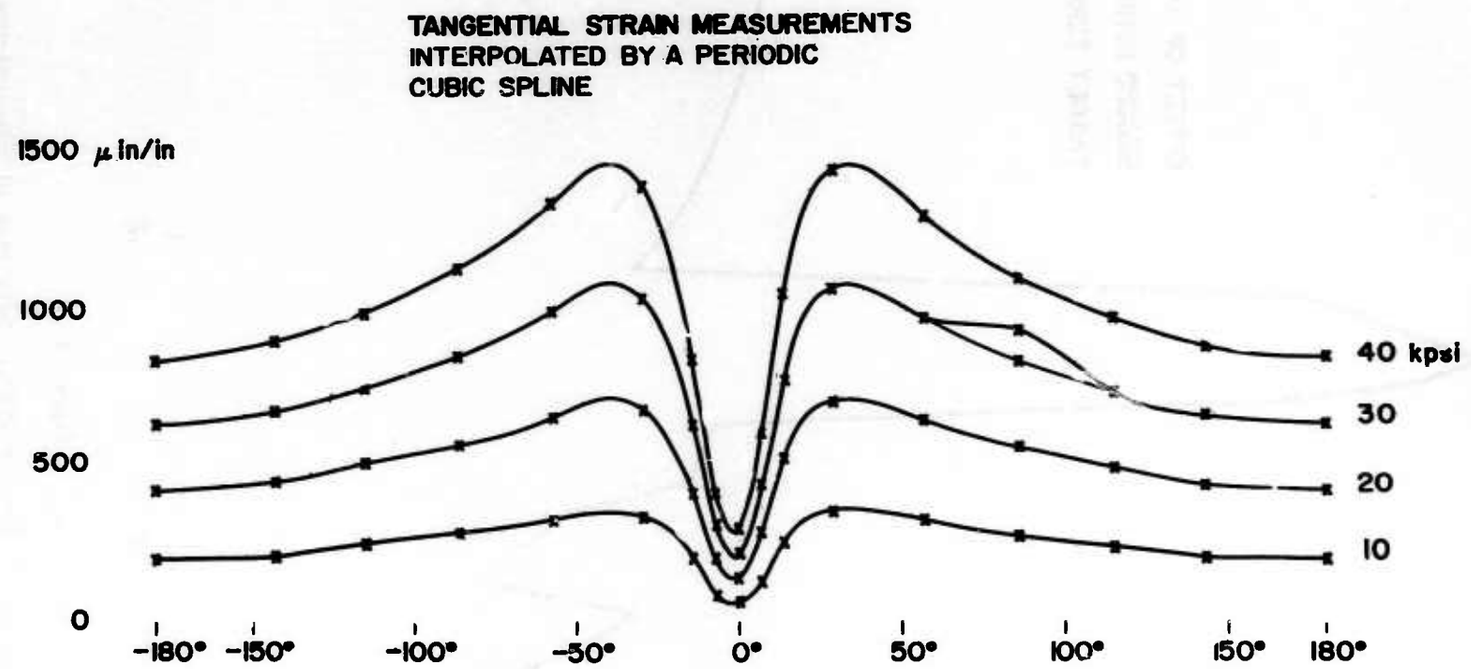


Figure 2.

Plot of data at 0.117 in. crack depth. One outlier is shown.

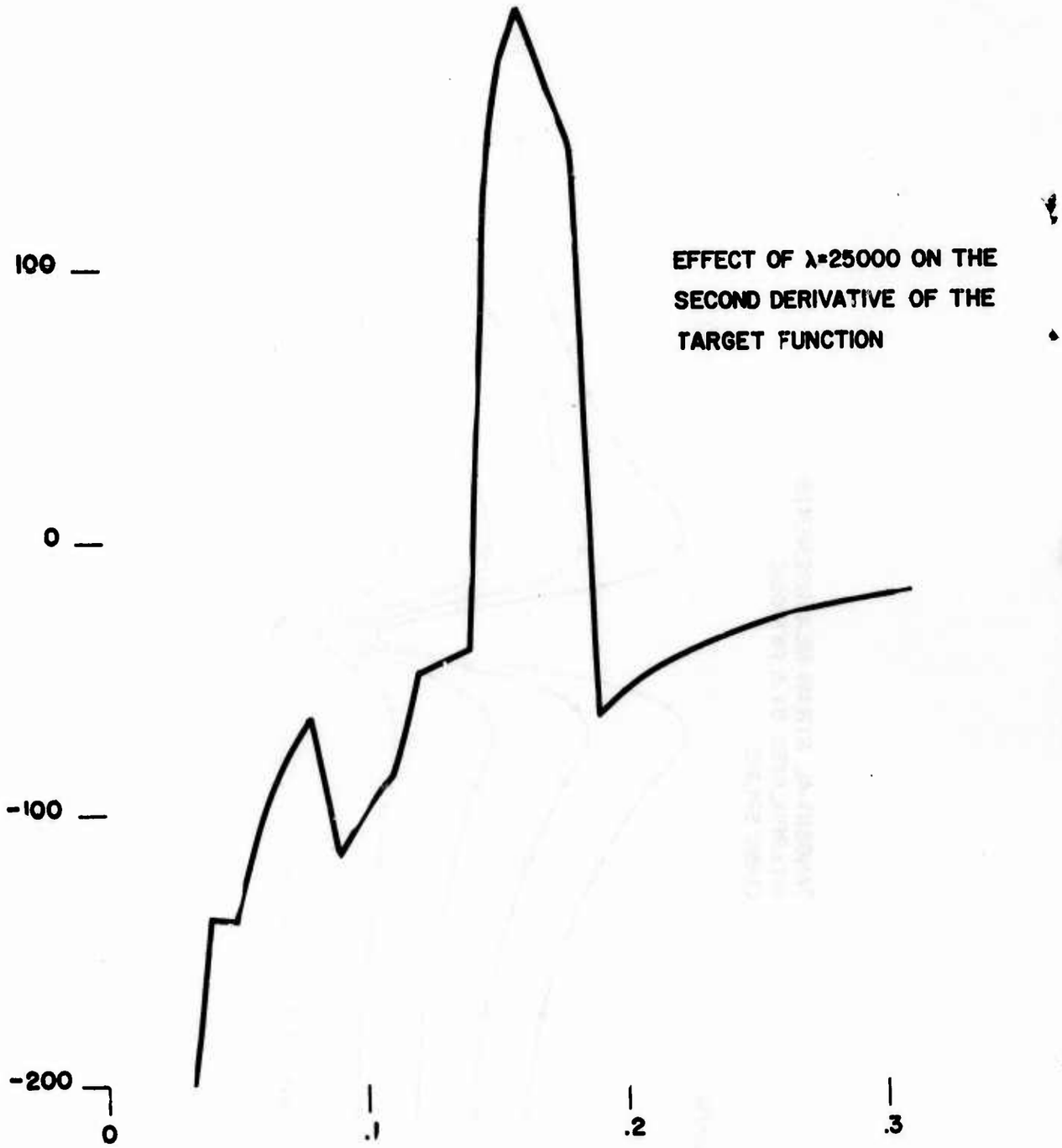
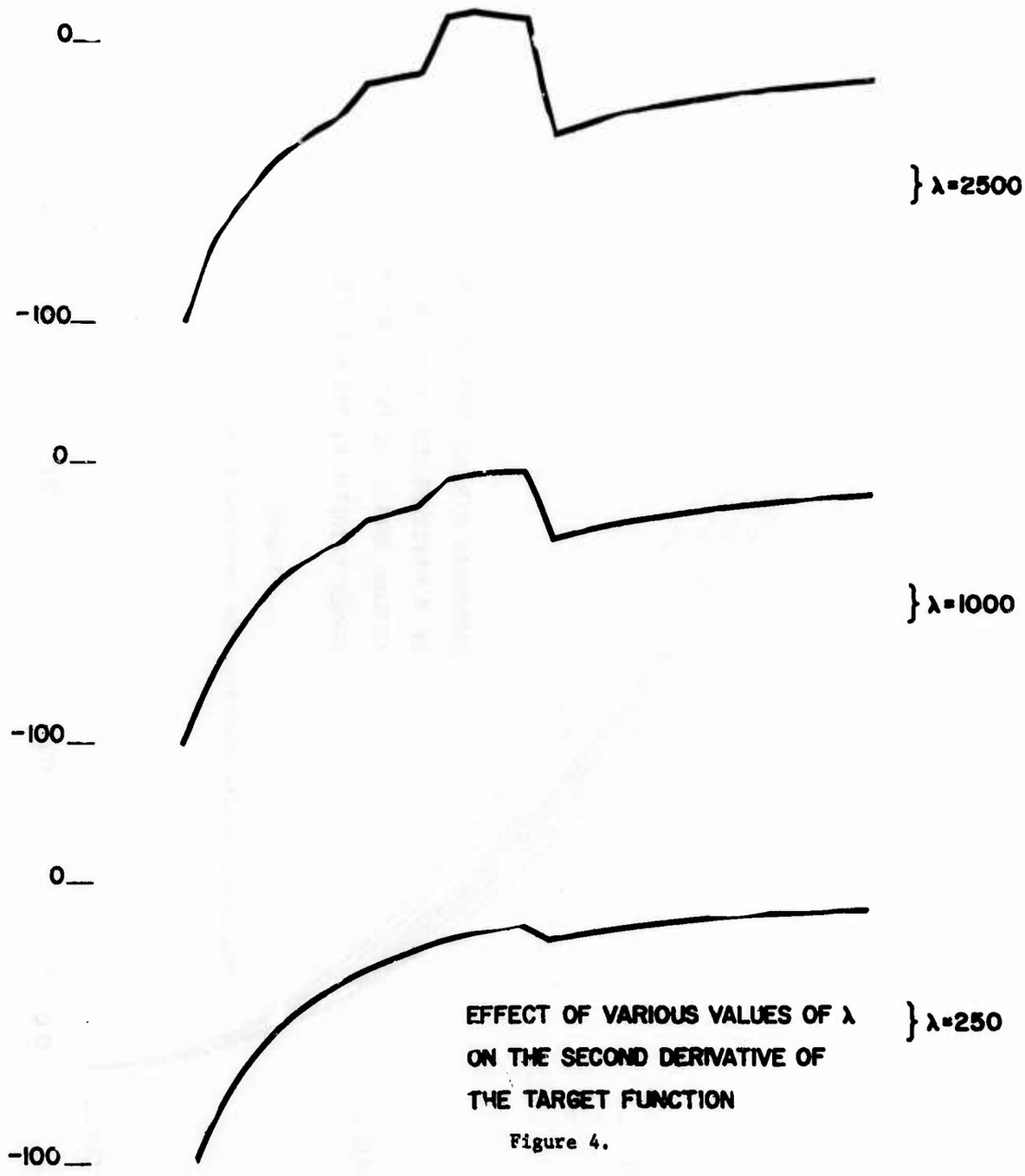


Figure 3.

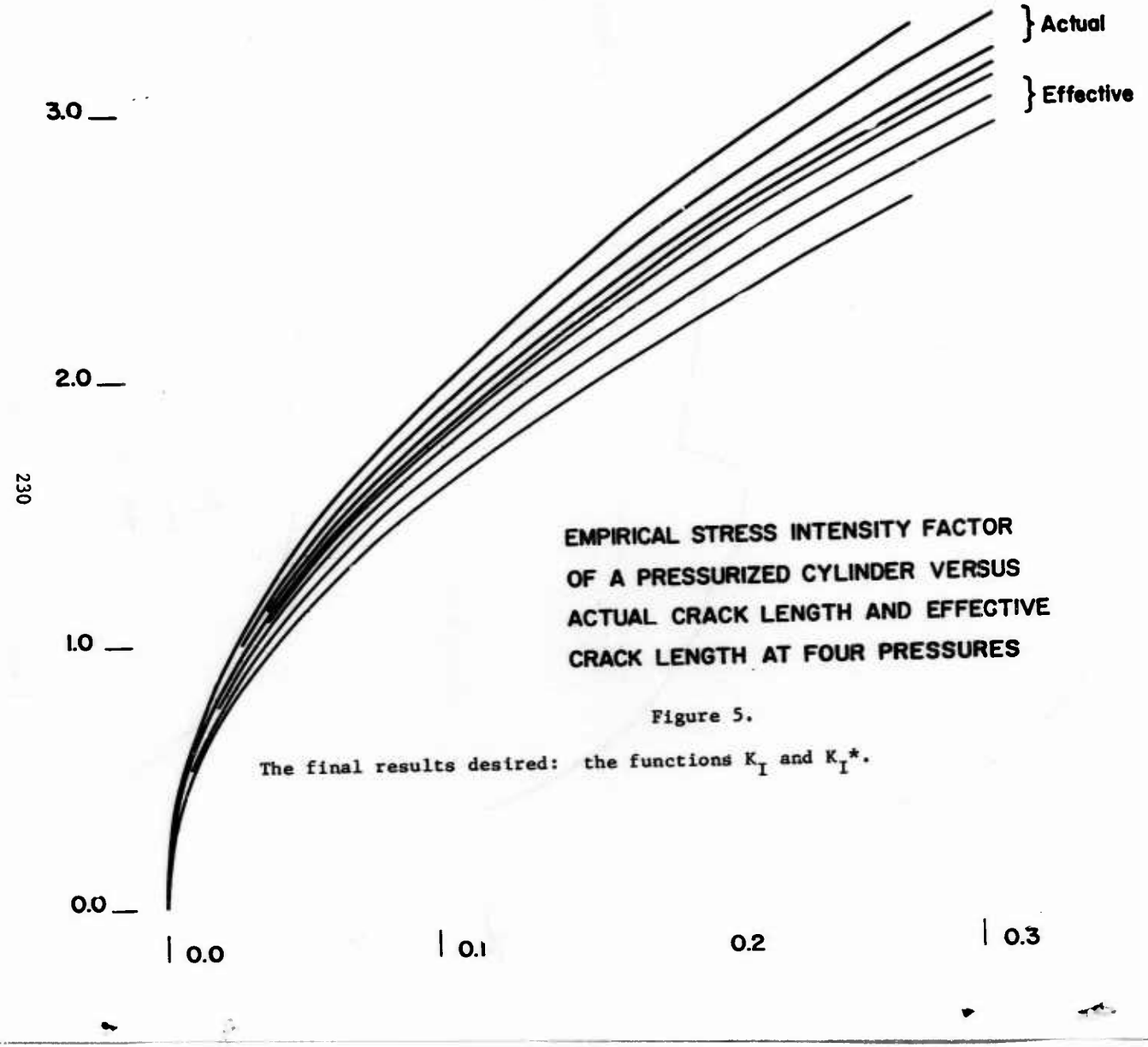
Second derivative of K_I^* for $\lambda = 25000$. Note lack of monotonicity.

Second derivative of K_I^*. As λ is decreased, the graphed function becomes nearly monotonic increasing.



EFFECT OF VARIOUS VALUES OF λ ON THE SECOND DERIVATIVE OF THE TARGET FUNCTION

Figure 4.



EMPIRICAL STRESS INTENSITY FACTOR OF A PRESSURIZED CYLINDER VERSUS ACTUAL CRACK LENGTH AND EFFECTIVE CRACK LENGTH AT FOUR PRESSURES

Figure 5.

The final results desired: the functions K_I and K_I^*.

PATTERN IDENTIFICATION - A REVIEW OF PERCEPTION RESEARCH*

Selby Evans
Texas Christian University

In the title of my paper I have proposed to review the psychological research on human pattern perception. This is a task at once easy and difficult. It is easy because, if one imposes a definition of pattern corresponding to the definition commonly used in the study of machine pattern recognition, the psychological research is disappointingly sparse. The task is difficult because those aspects of psychological research which might bear on machine pattern recognition appear in several specialized contexts with divergent orientations.

For the most part, psychology has not confronted the problem of pattern recognition as it is encountered in efforts to develop automatic pattern recognition systems. In fact, psychologists often use the term pattern to refer to any complex visual array, without regard to the question of whether the instances of a class could be recognized by a template method or not. As workers in machine pattern recognition know very well, this distinction is crucial; they have come to use the term pattern to refer to cases in which the variability within a class makes a template approach ineffective. That humans can recognize patterns of this sort has, of course, been known; but it has not been regarded as a problem meriting serious explanatory effort. Psychologists have dealt with the problem of assigning distinguishably different objects to the same class, but they have done so under the term "concept identification" and they have been principally concerned with objects differing along clearly defined and named dimensions, such as color, number or the like.

To understand why psychology has generally ignored the problem which has proved so formidable in machine pattern recognition, we have to consider a fundamental problem in the study of perception. In all sciences, perception is part of the system of observation, the system which carries to the scientist information about the object under study. Perception has worked very well in that role, as it has in the more general and essentially similar role of providing humans with information about the everyday environment. In this role, perception is almost never questioned. Indeed, it cannot be broadly called into question without conjuring up the ghost of unproductive solipsism, which asks "How do you know the world is really as you perceive it to be?"

The study of perception cannot dispense with perception as a part of the observing system. But perception is also, in this particular case, the object of study. This dual role can, I think, be handled if the two roles are carefully distinguished and kept conceptually separate. But the two roles have offered various opportunities for subtle confusions.

---

*I thank Dr. R. M. Fenker and Dr. Leona Aiken for comments on the manuscript.

Perhaps as a result, there has been the persistant tendency of perception psychologists to study special cases--cases which have what I call a high G factor. That is, they get people to say 'Gee, look at that!" Visual illusions, ambiguous figures, and so forth, have received a great deal of attention, I think, because they demonstrate that perception can properly be called into the question. Perhaps such demonstrations are necessary, but they may not get us far toward understanding real life perception, which is remarkably accurate.

Remarkably? Did I say remarkably? What, after all, is remarkable about recognizing a truck as a truck? After all, it is a truck. If someone thought it was a goldfish, that would be remarkable.

But you who have worked toward the development of machines to duplicate human pattern recognition know that the human perceptual process is indeed remarkable. We psychologists are indebted to the field of machine pattern recognition for defining the problem - a problem that has been invisible to most psychologists because they studied perception in terms of an environment known to them through their own perception systems. It was only when the environment was seen through a different system--as a raw data matrix of light intensities across a scene, for example--only then did the problem become apparent. In this review, I am going to focus primarily on work which I think contributes materially to solving this problem.

BACKGROUND. There is one topic, Gestalt psychology, which deserves mention in the context even though it did tend to confuse the roles of perception. Despite this confusion, Gestalt psychology embodies a number of important observations about human pattern perception. Perhaps the most important is the subjective phenomenon of figure-ground. The phenomenon is simply this: when we look at a scene, a drawing, or a photograph and attend to a particular object, we see the object as figure, an integral unit distinct from the rest of the scene (which is subjectively in the background). This perceptual organization, which delineates a region of visual scene as a unit to be recognized, is accomplished without any awareness on our part and we normally take it as given. The Gestalt psychologists made a number of observations about this phenomenon but I think the most important, from the standpoint of automatic pattern recognition, is the phenomenon itself.

Because our visual systems are so successful at selecting units to be recognized, it apparently did not occur to anyone that this selection might be difficult. It is quite difficult, of course, from the standpoint of machine pattern recognition, as anyone who has contemplated the problem of identifying particular objects in a scene can testify. In fact, many efforts at automatic pattern recognition have tended to avoid this problem, either by selecting such things as hand printed letters - in which the units are clearly delineated - or by working with problems such as terrain classification, in which the entire photograph can be used as a unit.

I would suggest that the ability to separate a visual scene into suitable units may play a key role in human pattern identification processes. Of course, in one sense it must, since it selects the parts of the scene which will be treated as a unit for the purpose of recognition. But quite aside from that essential and obvious role, if we acknowledge that a human can break up a scene into suitable units for recognition, we must acknowledge that he can also break up a single object into suitable and simpler units of recognition. The figure-ground phenomenon may thus give humans the capability of separating any complex figure into a set of simpler component units. Such separation could be recursively applied through several stages, if necessary, to reduce a complex figure to a number of simple units and their spatial relationships. Perhaps such an arrangement would simplify the pattern recognition process.

I am not going to review further the work of Gestalt psychology because its observations are too thoroughly contaminated by confusion between the two roles of perception. The observations have considerable intuitive appeal, but they were cast in terms of subjective descriptions. What was required to go beyond this level was a new methodology and a new frame of reference.

THE NEW METHODOLOGY - OBJECTIVE DESCRIPTION OF THE STIMULUS. Gestalt psychology had resisted the notion that a pattern - in psychological terms, a stimulus - could be analyzed strictly in terms of its physical characteristics. The Gestalt view, if correct, would leave automatic pattern recognition in a rather difficult situation. Automatic pattern recognition has nothing to work with except the physical characteristics of the stimulus. Moreover, one must wonder what it is that the human visual system could respond to in the stimulus other than its physical characteristics. We would have to grant, of course, that the human system has a great deal available to it in the form of memory and processing capability. But to the extent that the stimulus presented to a subject influences his response, one must assume, I think, that the response is a function of the physical characteristics of the stimulus. This point of view was set forth and championed vigorously by James J. Gibson (1959) and for these efforts he must be credited with challenging the domination of Gestalt psychology and reorienting at least some research in perception.

Another contribution to this line of development was made by Attneave and Arnoult (1956). They noted that most research on pattern or form perception was done with familiar or, at best, arbitrarily constructed figures. Such research, they pointed out, was limited in its conclusions to generalizations about the particular figures used in the study. In contrast, psychologists had long been using techniques of statistical inference which allowed them to generalize their results to populations of people. Attneave and Arnoult suggested that the same

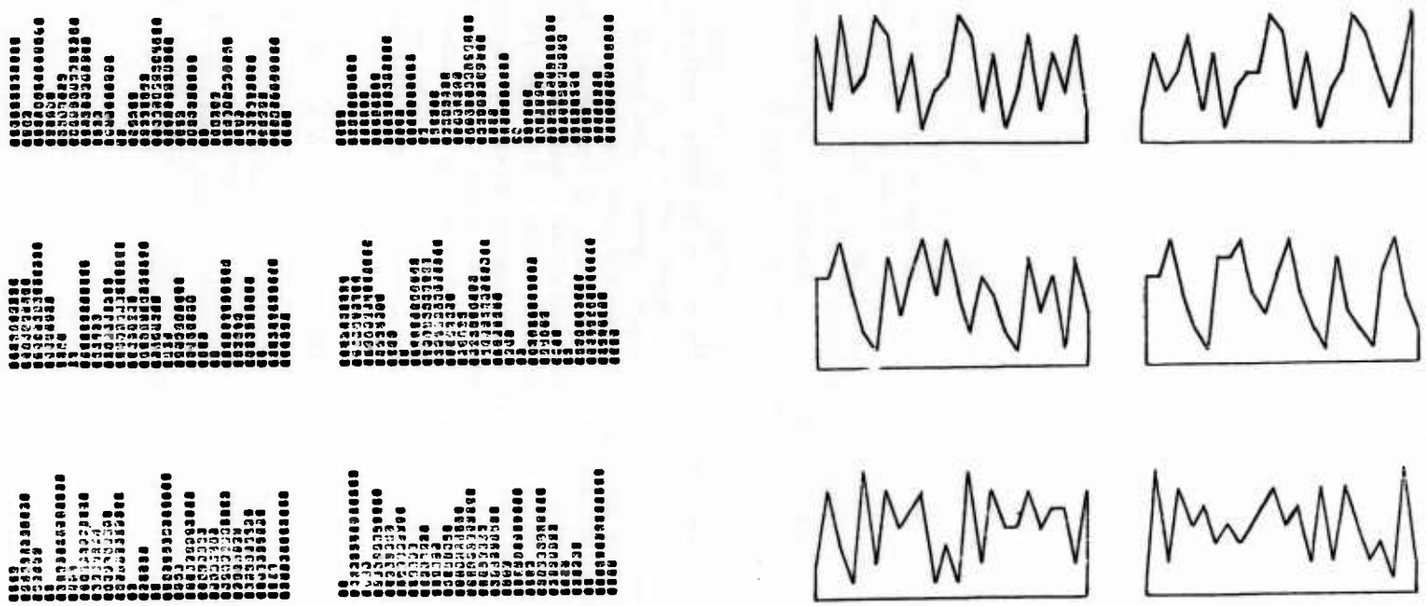
kind of generalization could be used for making inferences about a population of stimuli if there were methods for drawing random samples from such a population. These authors then proposed a number of methods for creating unfamiliar forms by rules which incorporated random processes. Thus a large number of forms could be drawn, each being a random selection from the population defined by the generating rules.

Still another significant contribution was made by Fitts (Fitts, Weinstein, Rappaport, Anderson & Leonard, 1956) and a number of his associates. Their interest was initially stimulated by developments in information theory and by efforts to apply these developments to the study of pattern perception. Fitts and his associates also devised methods for generating patterns so as to sample them from a defined population. These patterns were much simpler than those generated by Attneave and Arnoult, but this simplicity permitted Fitts and his associates to determine an information measure associated with the sampling procedure. Thus he could manipulate the average amount of information per stimulus. This measure has been used rather extensively since then in psychological research, although it has proved to be too ambiguous and too abstract to be related in any simple fashion to pattern identification performance. Its introduction, however, did constitute a major step toward the objective quantification of stimulus characteristics.

A fourth contributor was Brunswick (1958), who introduced the notion of representative design and what he termed ecological validity. With respect to pattern identification, the thrust of Brunswick's argument was that the stimuli of an experiment should be representative, in an appropriate sense, of real world stimuli. Brunswick's efforts in this direction were not very successful, but his proposal was appealing and efforts toward that objective have continued.

Out of these efforts, I think it has become clear that what is needed as a basis for research in human pattern identification is a specification of the relevant characteristics of real world stimuli. We know intuitively that our environment is very orderly, but we do not have any delineation of the rules of that orderliness. Perception surely makes extensive use of the orderliness, and we can hardly expect the characteristics of human perception to show up in our experiments until we can identify, manipulate and measure that orderliness. What we need is a kind of theory of environmental patterns, as a complement to any theory of pattern identification.

Let me now turn to some recent efforts in that direction. What constitutes a pattern? The best source of objective information at present seems to lie in the work on machine pattern recognition. From that work, it is clear that members of a pattern class should not be template-recognizable. On the other hand, they presumably must cluster in some feature space if they are to be recognizable at all. These principles provide some guidelines, and several pattern generating methods have been developed in terms of these guides. Figure 1, for example,



Examples of Vargus 7 patterns representing three different classes. The two patterns on each row belong to the same class. The patterns were produced by a Markov process which generated a sequence of numbers; each pattern class had a different most favored set of transitional probabilities for the numbers. The number sequences were plotted as heights to form the above figures.

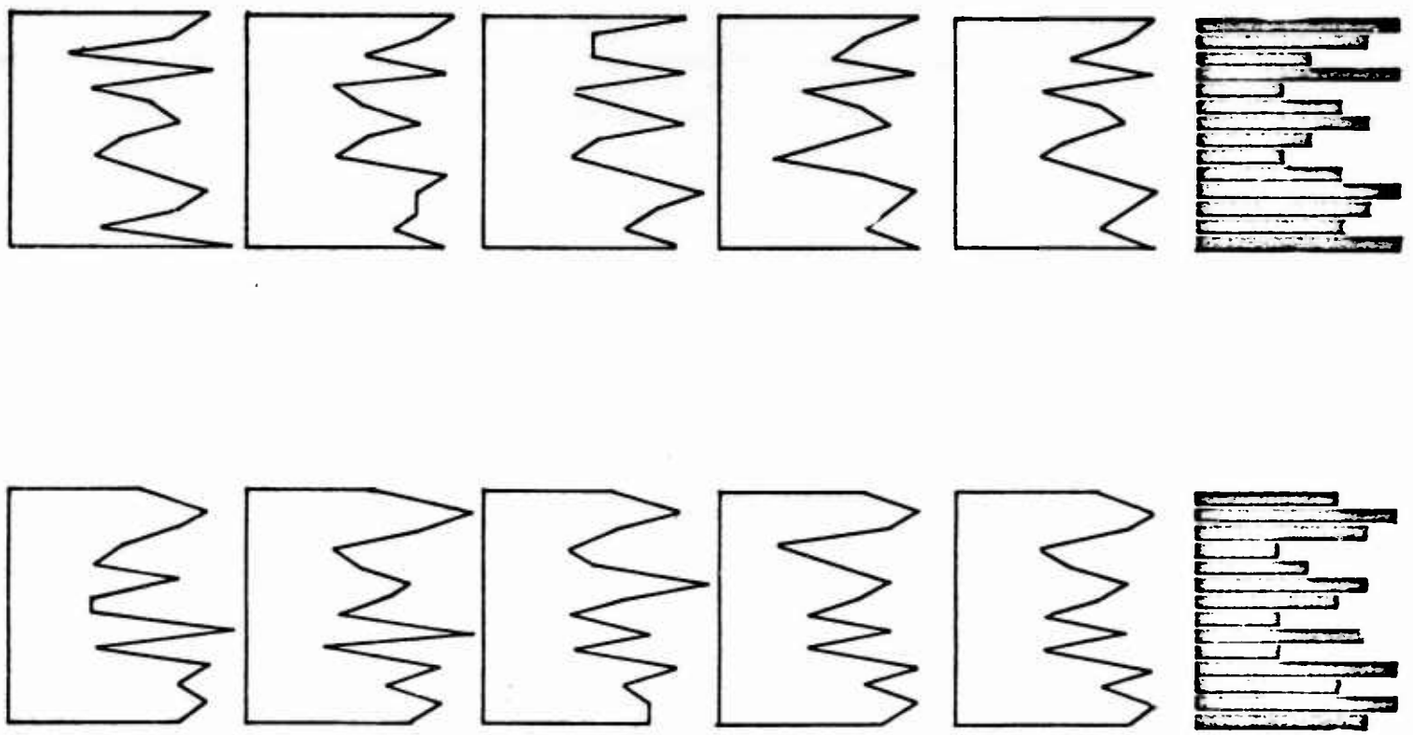
Figure 1

shows the results of a technique (Vargus 7) which introduces recurring regularities into a sequence of numbers or graph points (Evans, 1967). The regularities are introduced by favored transitional probabilities in a Markov process, a method which permits substantial control over the generating characteristics. In particular, it permits control of the information content of the stimuli and of their adherance to a pattern class. The stimuli generated in this fashion are relatively unfamiliar but even so they exhibit a discernible pattern. Their unfamiliarity permits the investigation of pattern learning as well as pattern identification.

In many cases, a pattern may be thought of as a prototype, a configuration typical of a population of patterns. In Figure 2, another pattern generating method is illustrated. This one (Evans and Mueller, 1966) was designed specifically to incorporate a prototype and to produce members of a population in the form of deviations from the prototype. The prototype may be thought of as a point in a multi-dimensional space. The deviations are produced by adding the equivalent of an error component, a random variable, unimodally and symetrically distributed about zero. This error term is independently obtained for each dimension of the prototype and produces a dispersion of points clustering around the prototype. In its original form, of course, the pattern is simply a sequence of numbers. These are then plotted as column heights or points on a graph. Alternative forms are possible, as illustrated by the schematic airplane shown in Figure 3.

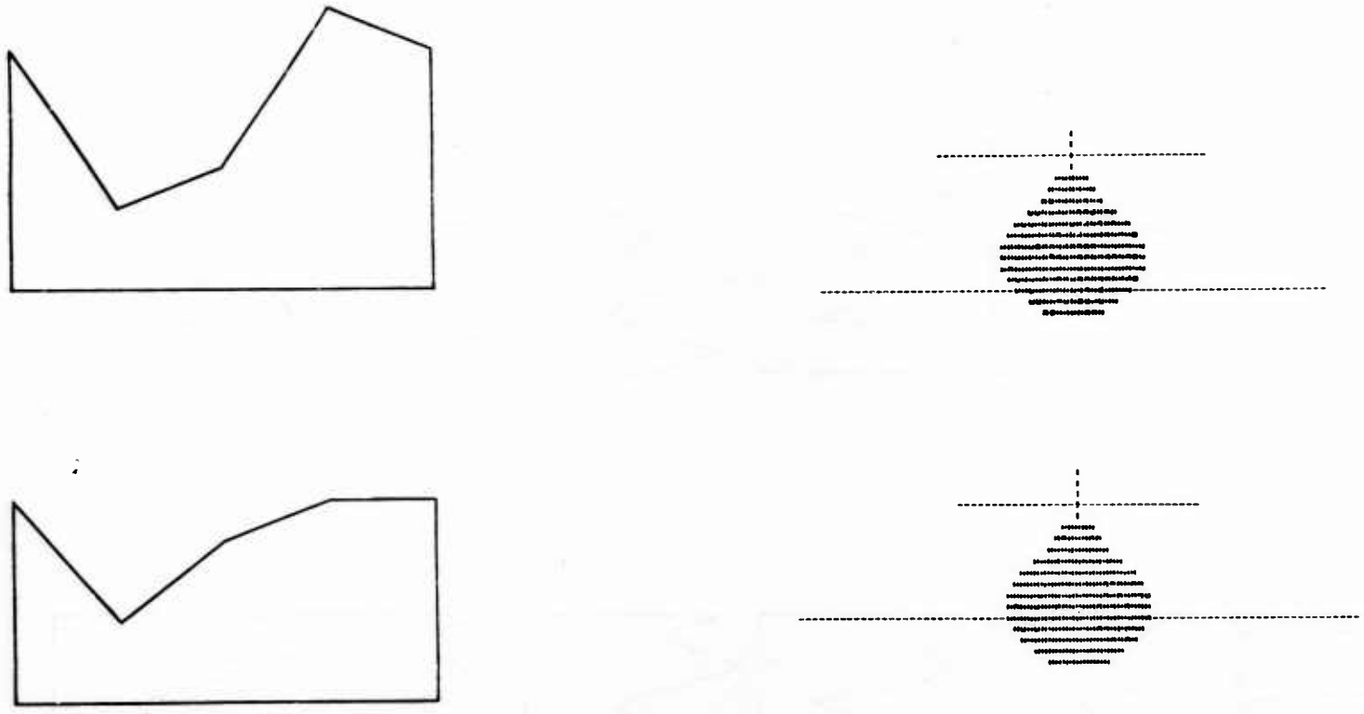
The preceding examples of pattern generation have been used principally with unfamiliar figures. These figures were designed to incorporate general properties which we assume characterize the orderliness of the natural environment. There are however, more specific kinds of orderliness - in particular the orderliness in familiar patterns - which presumably should be represented in research. In order to incorporate familiar configurations into stimuli appropriate for research, we borrowed a notion from Shannon (1949). Long ago in studying the transitional probabilities of English grammar, Shannon used humans as repositories of the statistical characteristics of grammar. In the same way, one might reasonably assume that humans are repositories of the statistical characteristics of the visual environment and that they make use of these characteristics when they draw figures.

Accordingly, we engaged a commercial artist to draw a number of figures in such a way as to allow us to decompose his figures into parts which could be reconnected in a number of different ways. Thus, by drawing at random from these components, we can construct patterns which are samples from the population of possible patterns made from these components. Examples of this generating procedure, Vargus 10 (Evans, Hoffman, Arnoult & Zinser, 1968) are shown in Figure 4.



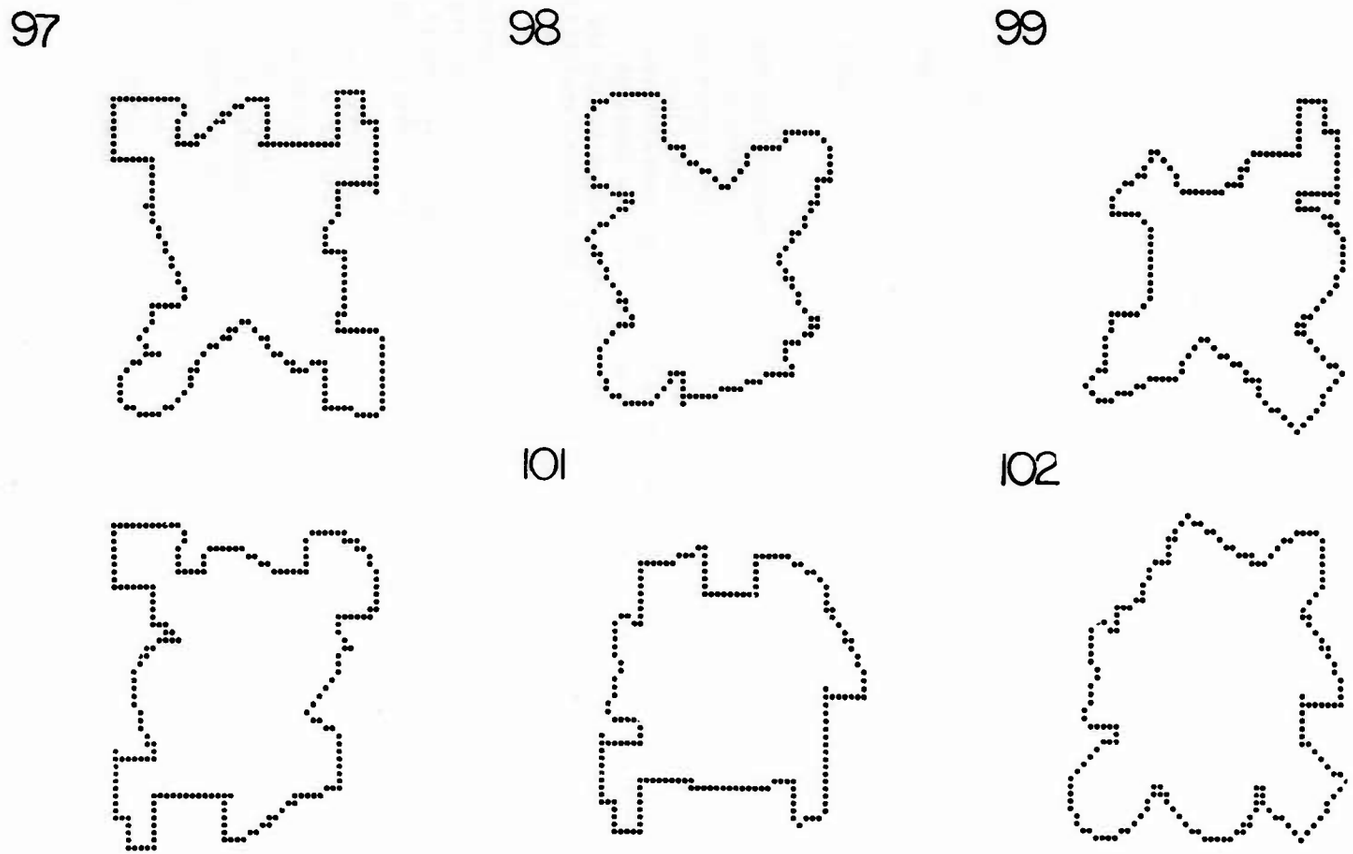
Examples of Vargus 9 patterns representing two different classes. At the top of each column is the class prototype; increasingly deviant examples appear in descending order.

Figure 2



Additional examples of Vargus 9 patterns, showing transformations from graph format to "schematic airplane" format. The graphs at left were mapped into the planes by associating points on the graph with attributes of the planes, such as wing heights. The altitudes of the points were used to determine the magnitudes of the attributes.

Figure 3



Examples of Vargus 10 patterns.

Figure 4

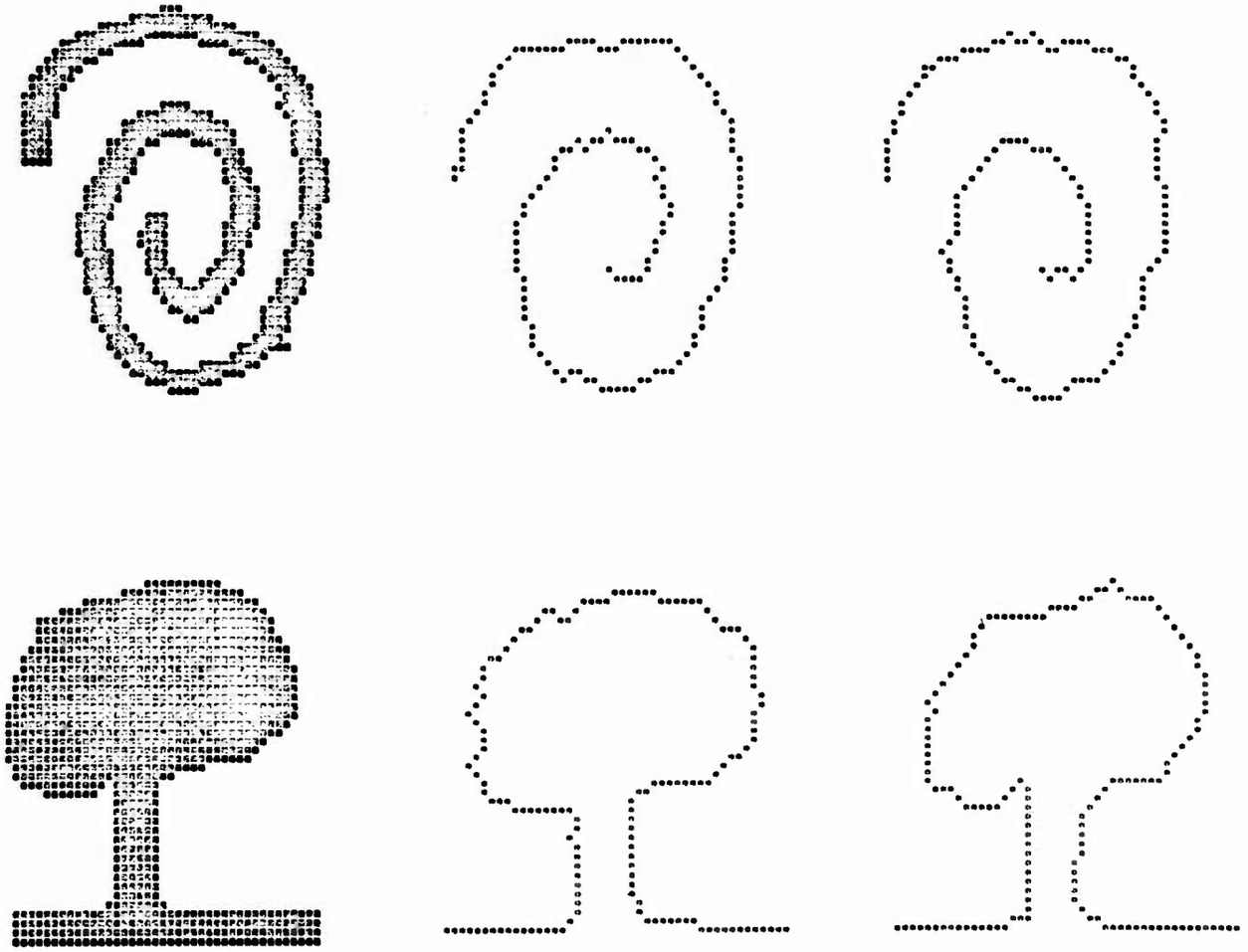
In producing the Vargus 10 patterns we have been able to introduce familiar components but we have sacrificed the characteristics achieved earlier, that of producing patterns which are deviations from a prototype. In Figure 5 is illustrated a method* for producing deviations on a prototype when these are drawings of fimiliar figures of the Vargus 10 class. These figures are achieved by taking a single point, which might be thought of as a pen point, on a strongly constrained "random" walk. The walk is guided by the path which the drawing point must follow if it is to construct the required figure. The constraint allows the drawing point to wander a bit, and that wandering introduces the required variability from instance to instance. The amount of variability is, of course, subject to program control.

Other researchers have used approaches similar to this. For example, Posner (1964) has used figures composed of an array of dots and produced deviations by allowing the dots to take on different locations in different instances. Ellis (Ellis & Feuge, 1966) used Attneave and Arnoult polygons and introduced deviations by small changes in the locations of the corners. A number of texture generating methods have also been developed, but I think this presentation is better limited to the generation of forms.

In this section I have reviewed some approaches to a methodology for research on pattern identification. Taken together they constitute a methodology which allows psychologists to sample from a defined population of relatively unfamiliar stimuli, to introduce some relevant variables associated with pattern characteristics, and to measure or control these variables. No doubt new pattern generation techniques will be required, but the present methodology has permitted several important steps toward the understanding of human pattern identification. Let me now turn to empirical and theoretical contributions toward that objective.

EMPIRICAL AND THEORETICAL CONTRIBUTIONS. The great majority of machine pattern recognition techniques can be regarded as measuring an input pattern in terms of a number of features and then locating the pattern as a point in a multidimensional feature space. The classification of the pattern is obtained by associating its location in some way with a particular pattern class. If we take this process as a preliminary model for human pattern identification, we can view one major line of empirical research as being a selection of a feature space for human pattern perception. This research has been done principally by D. R. Brown and his associates. Brown worked with patterns generated by one of the techniques proposed by Attneave and Arnoult, a technique which produces randomly shaped polygons of a predetermined number of sides. Pattterns of this kind have been used by various researchers, and some of these devised reasonable ways to measure the patterns so as to obtain quantities representing subjective attributes. This work can now be viewed as an exploratory effort toward defining a feature space composed of dimensions relevant to human performance.

*This effort was supported by the Department of Defense Project THEMIS Contract (DAA05-68-C-0176), under the Department of Army to the Institute for the Study of Cognitive Systems through the TCU Research Foundation.



Examples of variable instances produced from prototypes (left) by a constrained random walk process.

Figure 5

Brown (Brown & Owen, 1967) and his associates set out to extend this effort in an orderly fashion. They assembled all the measures used by previous researchers and added new measures where these seemed likely to subsume several previous measures or to provide better representation of a perceptually relevant variable. This step resulted in an initial collection of 130 measures. The set was reduced to 80 by eliminating measures which were essentially duplications of others.

The 80 measures, of course, would not be expected to be independent of one another. In fact, one ought to expect substantial intercorrelations among 80 measures taken on the same pattern. Brown's next step therefore, was to reduce the 80 measures to a much smaller set of linearly independent dimensions. He did so by measuring 1000 random polygons in terms of these dimensions, obtaining the intercorrelations among the measures and applying factor analytic techniques to obtain a small number of factors which would represent most of the variance of the 80 measures.

Factor analysis has been described by a previous paper in this meeting, so I will not elaborate on that. The technique might be regarded as a brute force approach, since there would be mathematical relations among various measures taken on a figure. But in this case brute force promised to be faster and considerably less expensive. The factor analysis, augmented by some judgment, produced 12 factors which accounted for most of the variance in the 80 measures.

These independent dimensions, of course, do not correspond completely to any of the original measures. In general, they represent clusters of measures. Nevertheless, by considering measures which relate strongly to these dimensions, one can assign a name to each dimension. I will here mention only a few of the most important, by way of suggesting the kind of features that humans may attend to when they identify patterns.

Compactness: This measure was most strongly related to the area of the enclosing rectangle. (Since the shapes were standardized for actual area before the measures were computed, the area of the enclosing rectangle would indicate the extent to which the patterns spread out to occupy a larger area.)

Jaggedness: This measure principally represented the first four moments of the interior angles of the contour.

Skewness: Two measures are represented in this case; one relative to the X axis and one relative to the Y axis. Skewness relative to the X axis was simply the skewness of the distribution of X coordinates of the contour. Skewness relative to the Y axis was similarly obtained with the Y coordinate.

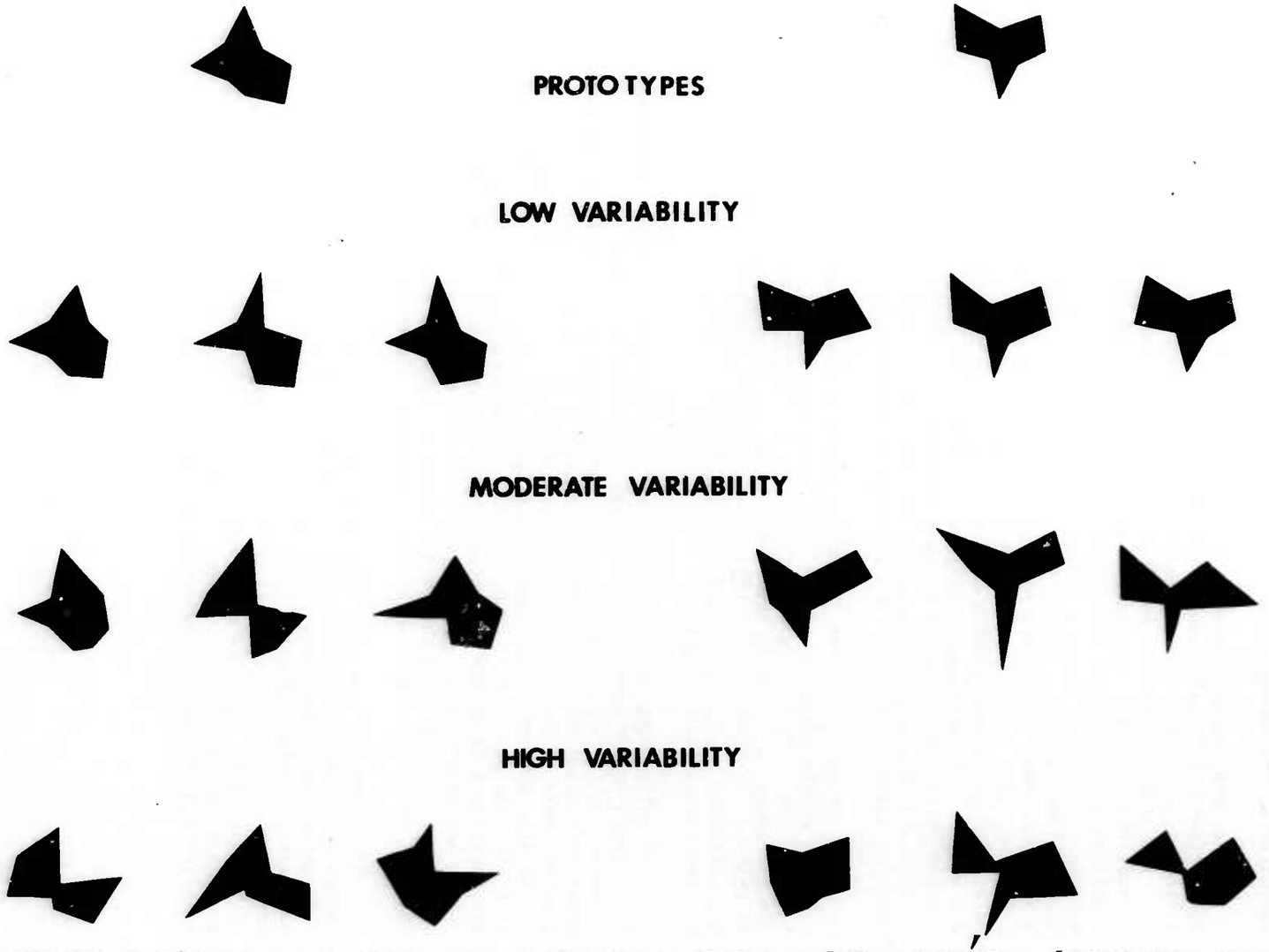
Major axis: This dimension reflected the extent to which the pattern exhibited an enlongation on a major axis.

This brief summary cannot, of course, do justice to the scope and thoroughness of Brown's work in determining these measures. They provide a tremendous opportunity for studying human performance as a function of objective and quantitative measures of forms. I expect that they will provide the basis for substantial advance in the study of form perception. Brown and his associates have already conducted a number of studies with these measures (Aiken & Brown, 1969a; Aiken & Brown, 1969b; Aiken & Brown, 1969c; Behrman & Brown, 1968; Brown & Andrews, 1968; Brown & Brumaghim, 1968; Brumaghim & Brown, 1968; Mavrides & Brown, 1969; Brown & LoSassa, 1967). The general trend of the results is to indicate strong relationships between Brown's measures, especially those mentioned above, and such performance variables as similarity judgments and discrimination difficulty. Most of these studies have used Attneave and Arnoult figures and have not introduced the within-class variability that I consider essential to represent a proper pattern recognition task.

More recently, however, Brown and two of his associates have used patterns generated by techniques like the Vargus 9 system described earlier. In order to permit measurements of the patterns, however, it was necessary to transform them into closed polygons. This was accomplished by graphing the patterns in polar coordinates rather than in the cartesian coordinate system used in Figure 2. Examples of such patterns are shown in Figure 6.

Let me describe one study based on these patterns so as to illustrate the techniques and some of the findings. Mavrides and Brown (1970) presented subjects with patterns or forms representing a single prototype; these were mixed with other forms generated by a random process. The task she gave to the subjects was to select those forms which adhered to the same pattern. This task was analogous to the unsupervised learning task that has been described by researchers such as Cooper & Cooper (1964) and Haralick & Kelly (1969). Research in our laboratory, to be discussed later, had demonstrated that subjects could do this task with Vargus 7 and Vargus 9 patterns in the format shown in Figures 1 and 2; Mavrides' interest was whether this performance could be related to Brown's measures.

To answer this question, Mavrides employed a discriminant function technique, in accord with an earlier suggestion by Rodwan & Hake (1964). This technique is doubtless familiar to most members of this audience, but perhaps I should mention that it includes finding a set of weights for a linear combination of measures so as to maximize the combination's correlation with a prescribed dichotomy. It is usually employed to provide a classification function, but in the present application the function was treated as a descriptive model of human performance.



Examples of polygons constructed in polar coordinates as deviations from a prototype. (Illustration supplied by Leona Aiken.)

Figure 6

The experiment contained a number of groups differing in their previous experience with the patterns, a variable which naturally produced different levels of performance. A discriminant function was obtained for each group, relating its modal classifications of patterns to the pattern measures. An optimal function was also obtained relating the "true" classifications to the measures. For the group which classified in close accord with the true classes, the discriminant function yielded weights similar to the optimal function. For groups which were less successful in this sense, the discriminant function indicated heavier weightings on measures which were less closely related to the true class. In both cases the function proved to be quite successful as a model for the human classification performance.

This study indicates that human pattern identification can be usefully described in terms of a discriminant function and that, in conjunction with Brown's measures, the weights obtained for the function can be used as indications of the features to which the subjects are responding. In this particular study, for example, some of the groups which did not classify according to the true categories apparently attended more strongly to jaggedness and skewness than they should have for optimal discrimination. More important than the particular variables, however, is the demonstration of a technique which would permit a systematic investigation of the features to which people respond in a pattern identification task.

Another possibility offered by this technique is the comparison of human pattern identification with an optimal model. Mavrides used two sets of patterns, with differing levels of deviation from the prototype. With the set which had a low level of deviation, humans could do about as well as the model. But with the more deviant set of patterns, the optimal model did better than did the humans.

The comparison is not entirely proper, because the model was based on knowledge of the true classes and the humans were not given this knowledge. Nevertheless, the comparison is worth noting in view of the general superiority of humans in pattern identification. Mavrides' results suggest that a relatively simple linear model can surpass human performance when the relevant feature space is known. On the basis of this and of observations in our laboratory, I would be willing to speculate that presently available statistical decision techniques are substantially superior to human capabilities when the two are on equal footing. The secret of human success at pattern recognition must lie somewhere else.

I have reviewed this piece of research in some detail to show what can be done with the measures developed in Brown's research program, especially when they are combined with sophistical statistical techniques.

Rather than review other studies in this area, may I just summarize a few major conclusions: Subjects change the features they use in accordance with the task imposed upon them, the viewing conditions (such as visual noise), and the statistical distributions of features within the set of patterns. In the case of a pattern identification task, for example, the subjects find without assistance a set of features which discriminate between the classes of patterns. In fact, subjects in this task will discard features they normally prefer in favor of other features if the other features provide strong clustering (Aiken, 1970). These results therefore suggest that one basis for the highly successful pattern identification by humans lies in their ability to select an appropirate feature space for a particular set of circumstances and to do so spontaneously.

In view of the suggestion above, one might reasonably ask, do humans achieve their pattern identification success by using a feature space of high dimensionality? This question has been addressed by Fenker (Fenker and Brown, 1969). Fenker had a subject make similarity judgments about patterns in terms of a number of different conceptual dimensions. The subject judged how similar patterns were in complexity, for example, in esthetic appeal, and in other such properties. He then employed a multi-dimensional scaling technique (Kruskal, 1964) which allows an inference as to the number of underlying dimensions the subjects were using when they formed their similarity judgments. This technique does not provide a completely explicit method for inferring the number of dimensions, and inferences are not entirely indisputable. Nevertheless, Fenker's results point rather convincingly to the conclusion that subjects typically used three dimensions as a basis for their similarity judgments. A number of other studies (Behrman & Brown, 1968; Brown & Andrews, 1968; Aiken & Brown, 1969a; Aiken & Brown, 1969b) corroborate this conclusion. Thus it appears that the key to human success in pattern identification does not lie in the dimensionality of the decision space but in a highly effective ability to select an appropriate feature space for the given circumstances.

One should not conclude, of course, that humans only use three dimensions when they classify patterns. A more reasonable inference would be that they only use three dimensions at a time. If I may again speculate, I would suggest that humans probably execute pattern identification as a hierarchical classification process, assigning a pattern to some very broad category, then assiging it to one of a small set of divisions within that category, and proceeding recursively to greater precision of classification until some desired level is reached. Such a process would keep the number of categories at any decision point small. In turn, three well chosen dimensions might suffice. The process would require facility in choosing appropriate dimensions, and that seems to be available to humans.

One implication of this classification procedure is that it would almost never result in a pattern which was completely unclassifiable. An unfamiliar object would simply not allow pursuit as far down the classification tree. Common observation suggests that humans are indeed able to classify any pattern presented to them. Those of us who have tried to conjure up unfamiliar figures. such as the ones I presented earlier, have long been aware that there is no such thing as an unfamiliar figure to a human subject. It may simply be a jagged silhouette or some sort of a graph, but it certainly belongs to some class. Similarly, in the natural environment the uninitiate may encounter a large number of crawling things which he calls bugs or insects. To the biologist, of course, this gross classification is distressingly inaccurate. But the tolerance limits of this category make it possible for the observer to include within it many new species that he has never seen before.

Thus the hierarchical process seems to be compatible with some human characteristics. It has a serious weakness in that any inaccuracy at the upper level classification dooms the remaining steps to failure and jeopardizes the whole process, unless there is some capability for detecting this circumstance and retracing or restarting. Even this weakness, however, may be compatible with human characteristics, because parallel processing--which must be abundant in human perception--could compensate for the weakness by pursuing all reasonable paths until one is definitely confirmed.

There is a great deal more that could be said about the work of Brown and his associates, but time requires me to move on to another, rather different topic.

PRIMITIVE FEATURES. Research in physiological psychology has developed evidence which may indicate the nature of the elementary or primitive features used in the visual system of higher mamals. This work has been carried out primarily by Hubel and Wiesel (Hubel, 1963; Hubel & Wiesel, 1965; Hubel & Wiesel, 1962; Hubel & Wiesel, 1963; Hubel & Wiesel, 1959; Hubel & Wiesel, 1969). Their strategy has been to establish relationships between characteristics of stimulation on the retina and the activity of individual neurons in the higher visual centers.

A subject, generally a cat or a monkey, is first anesthetized and placed in an apparatus which holds the head still and ensures that its eyes are pointed in a particular direction. An electrode is implanted into a fairly well specified position in the target region, such as the primary visual cortex. The electrode senses the activity of a single cell; readings taken from it are monitored while a stimulus (usually a spot of light) is moved around in the visual field. Different placements of the light may result in changes in the activity sensed by the electrode.

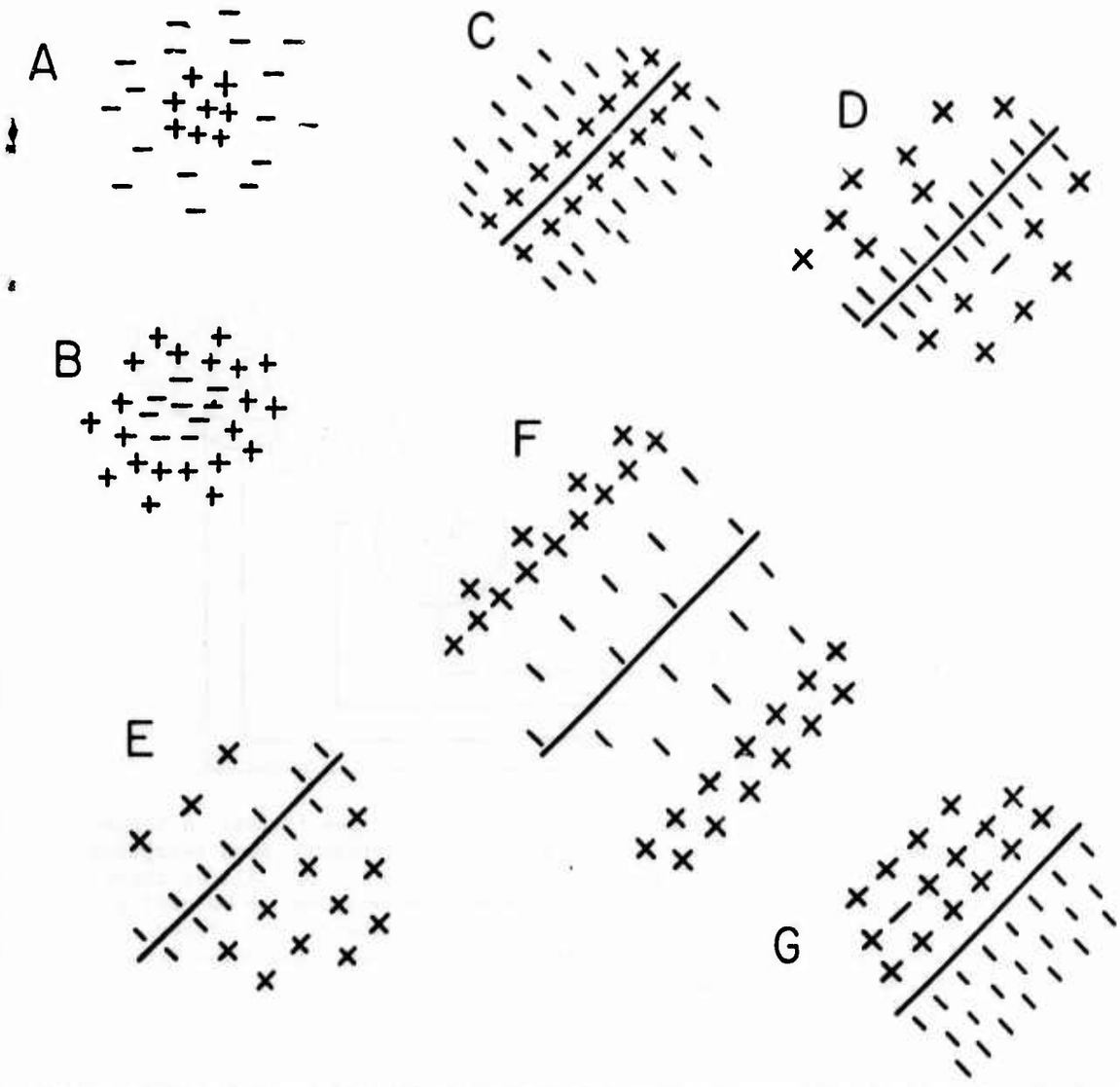
If such changes are found, there will normally be a region, more or less round or rectangular, in which the location or movement of the light makes a difference in the cell's activity. Beyond this region, the light will be found to have no appreciable effect. This region in the visual field projects through the lens of the eye to a region on the retina, and this latter region is referred to as the receptive field. It is assumed that the receptor cells in this region project, perhaps through several stages, to the cortical cell being sensed by the electrode. But that is merely an assumption; the receptive field has only a functional definition. It is that part of the retina in which stimulation influences the functioning of the cortical cell.

The receptive fields reported by Hubel and Wiesel have functions which seem to be analogous to the two dimensional linear operators, or filters, which have been used in some work on pattern recognition and pattern enhancement. I will therefore describe the receptive fields as if they were two dimensional filters, eventhough that description is probably an oversimplification.

The simplest fields may be regarded as spot detectors (see Figure 7). Some of these fields have positive effects in the center and negative effects on the periphery; thus they would maximally respond to a spot of light striking only the center of the field. Other receptive fields of similar shape and size have an inverse arrangement of positive and negative effects. They would thus be maximally activated by a bright field with a single dark spot lying over the center of the receptive field. In general, complementary receptive fields seem to be available for detecting similar dark configurations on a bright background. Thus I will not make any further distinction between the two kind of fields and will simply discuss the kinds of configuration the fields respond to.

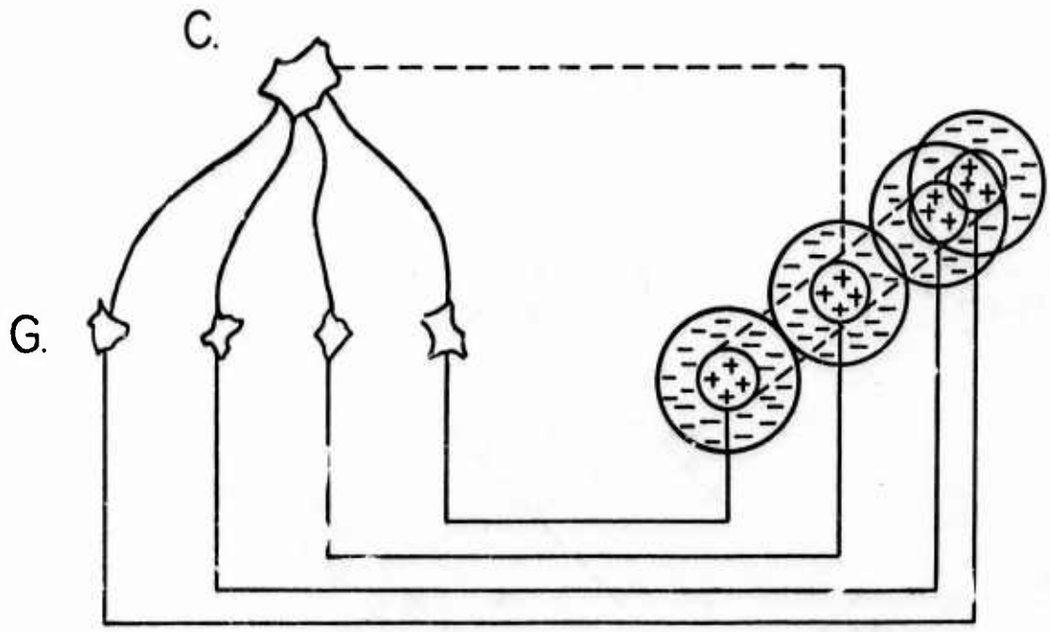
The spot detectors clearly could function as differentiating filters which might reduce an image to contour lines. Indeed the spot detectors at the cortical level apparently do not respond substantially either to brightness alone or to darkness alone but only to differential levels of brightness lying properly across them.

At the next level, again from the functional standpoint, are the line and edge detectors also shown in Figure 7. These fields are maximally activated by lines or edges properly oriented and located with respect to the fields. Hubel and Wiesel have suggested that these detectors might be produced by a hierarchical organization on the spot detectors. In other words, a single line detector might be formed by an appropriately chosen set of spot detectors. This arrangement is illustrated in Figure 8.



Common arrangements of lateral geniculate and cortical receptive fields. A. "on"-center geniculate receptive field. B. "off"-center geniculate receptive field. C - G. Various arrangements of simple cortical receptive fields. +, areas giving excitatory responses ("on" responses); -, areas giving inhibitory responses ("off" responses). Receptive-field axes are shown by continuous lines through field centers; in the figure these are all oblique, but each arrangement occurs in all orientations. (From Hubel and Wiesel, 1962, reproduced with permission.)

Figure 7



Possible scheme for explaining the organization of simple receptive fields. A large number of lateral geniculate cells (G), of which four are illustrated, have receptive fields with "on" centers arranged along a straight line on the retina. All of these project upon a single cortical cell (C), and the synapses are supposed to be excitatory. The receptive field of the cortical cell will then have an elongated "on" center indicated by the interrupted lines in the receptive-field diagram. (Modified from Hubel and Wiesel, 1962, reproduced with permission.)

Figure 8

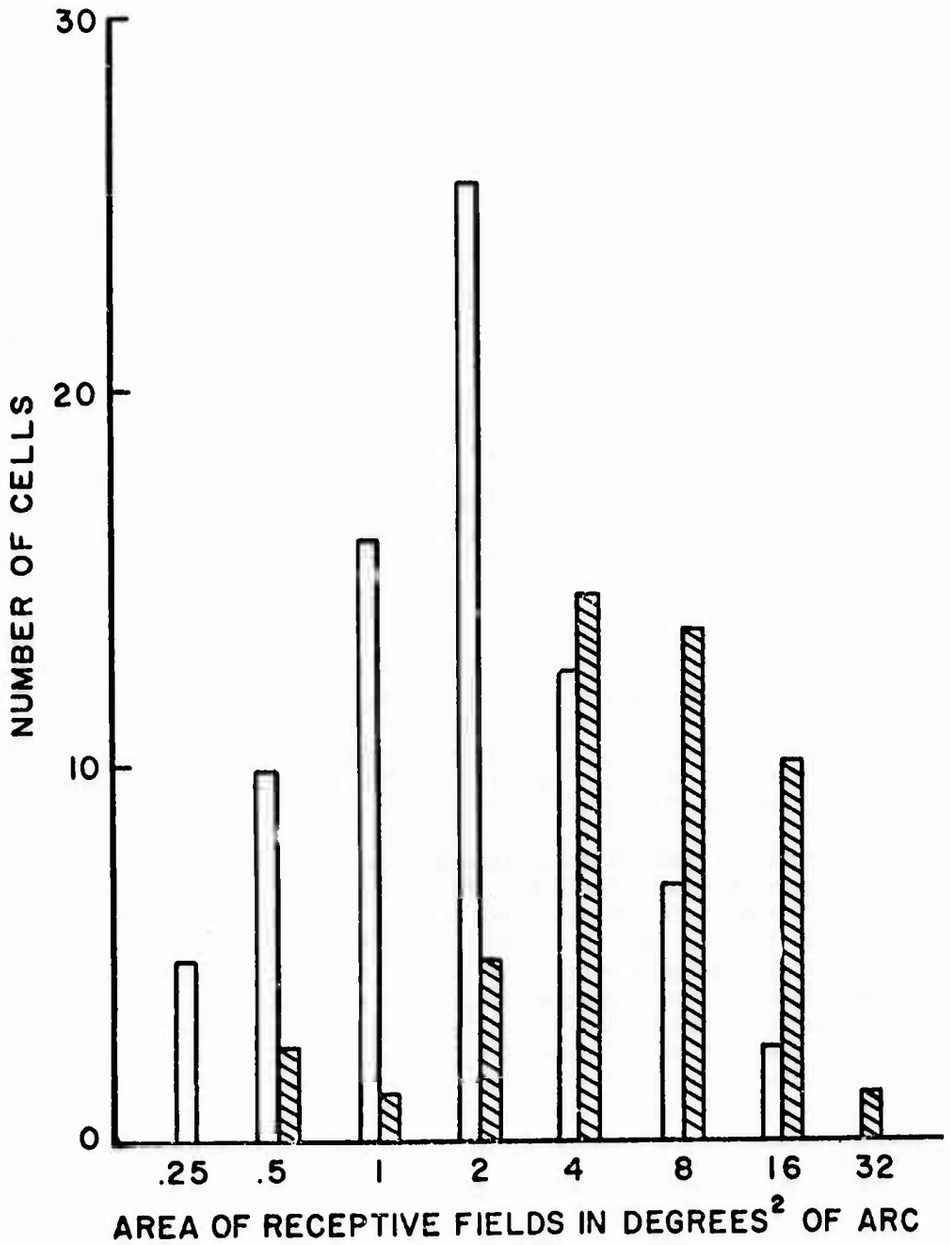
Still higher order receptive fields have also been described. These respond when an appropriate stimulus, for example, a properly oriented line, appears anywhere within the receptive field. These fields could be thought of as feature detectors which are insensitive to the exact location of the feature.

At the next higher level of organization, Hubel and Wiesel have identified receptive fields which respond to stopped lines or edges and to corners or projections. It is worth noting that this last set of receptive fields seems to function so as to detect abrupt changes in contour and these abrupt changes may represent the most important information in the figure.

There appears to be a wide range of sizes for these receptive fields. Figure 9 presents the distribution of sizes observed by Hubel and Wiesel for both the simple spot detectors and the slightly more complex edge and line detectors. (These results were obtained in cats.) It should be noted in interpreting this distribution that one degree of visual angle subtends about half an inch at 30 inches. Thus these receptive fields are not small compared to objects ordinarily recognized by cats. One should not assume, then, that the function of a receptive field is merely to enhance the edges or contours. On the basis of research presented by Dr. Zinser in another session of this meeting, I would suggest that these fields are better interpreted as contributing to an analysis of the figure into elementary components. Nor should we assume that they are simply operating on contours. The larger receptive fields are large enough to contain the entire image of small objects at normal viewing distance. They could be used to obtain information about the gross characteristics of the image, such as the direction of a major axis.

It is obviously impossible to conduct this kind of research on humans. But supplementary evidence (Antelman, Olson, & Orbach, 1969) has been developed as a result of the Hubel and Wiesel research. The technique for obtaining this evidence was to present subjects with a visual scene composed entirely of lines having a particular slant. After the subjects had stared at this scene for a substantial period of time, one might expect that any line detectors suitable for that slant would be somewhat fatigued. Thus, if the subjects were tested with a new set of lines of the same slant, the subjects might be less successful at detecting them than at detecting lines of different slants. This expectation was confirmed.

On the basis of these data, one might speculate that the mammalian visual system analyses patterns into contours composed of lines or edges of various slopes. It then registers the presence and approximate location of corners, projections, and invaginations. These features may serve as the basic description in terms of which pattern identification operates.



Distribution of 119 cells in the visual cortex with respect to approximate area of their receptive fields. White columns indicate cells with simple receptive fields; shaded columns, cells with complex fields. (From Hubel and Wiesel, 1962, reproduced with permission.)

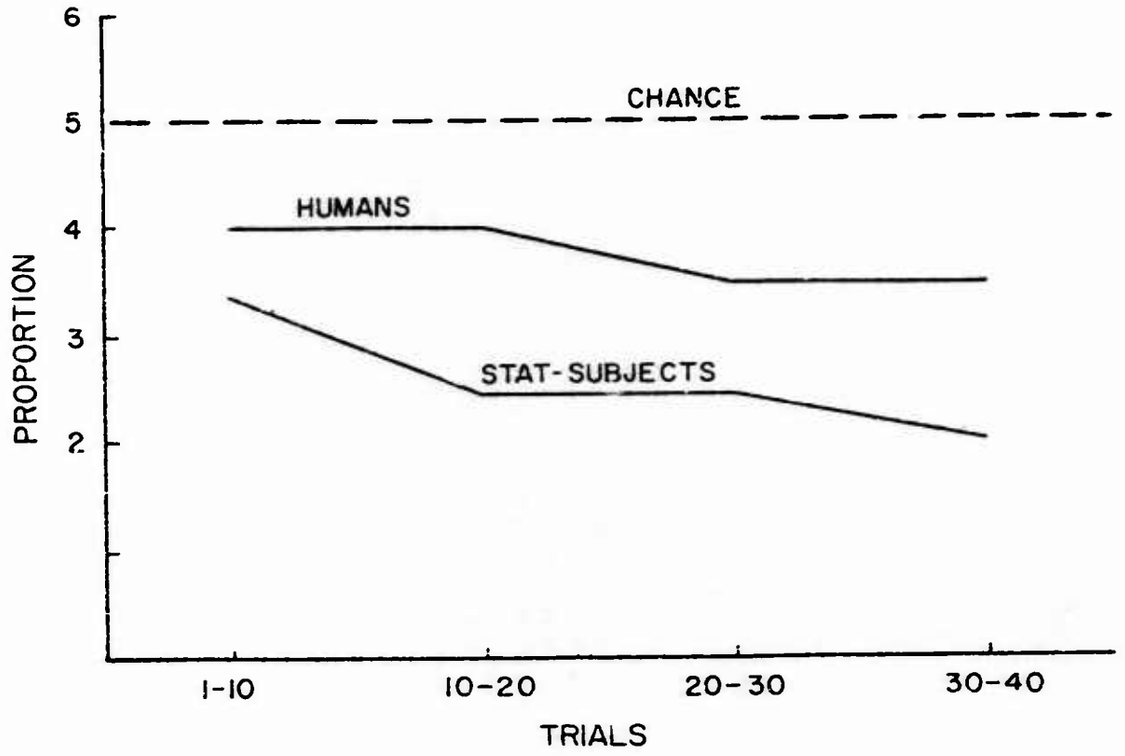
Figure 9

Speculation of this sort on the basis of physiological data is always hazardous because alternative interpretations are always possible and one can seldom establish an unassailable case for a particular interpretation. Nevertheless, this evidence can be used, I think, as a basis for hypotheses about preprocessing mechanisms in human pattern perception, as discussed by Dr. Zinser in another paper in this meeting.

SCHEMATIC CONCEPT FORMATION. A final topic I should like to discuss briefly is one which we have been studying intensively in our laboratory. Pattern identification is generally thought of as assigning patterns to previously established classes. There has, however, been some work in unsupervised learning (Cooper & Cooper, 1964; Cooper, 1967; Nagy, 1968; Haralick & Kelly, 1969) in which classes are found without previous identification or external advice from a trainer. As noted earlier in connection with the Mavrides' study, humans can also form pattern classes without assistance. A number of studies (Brown & Evans, 1969; Edmonds, Mueller & Evans, 1966; Edmonds & Mueller, 1967; Evans, 1964; Rankin & Evans, 1968; Rosser, 1967; Bersted, Brown, & Evans, 1969) have shown that unassisted humans can sort or distinguish patterns, such as those shown in Figures 1 and 2, according to prototype or generating source. On the basis of this research we conclude that humans can select a feature space which yields well defined clusters.

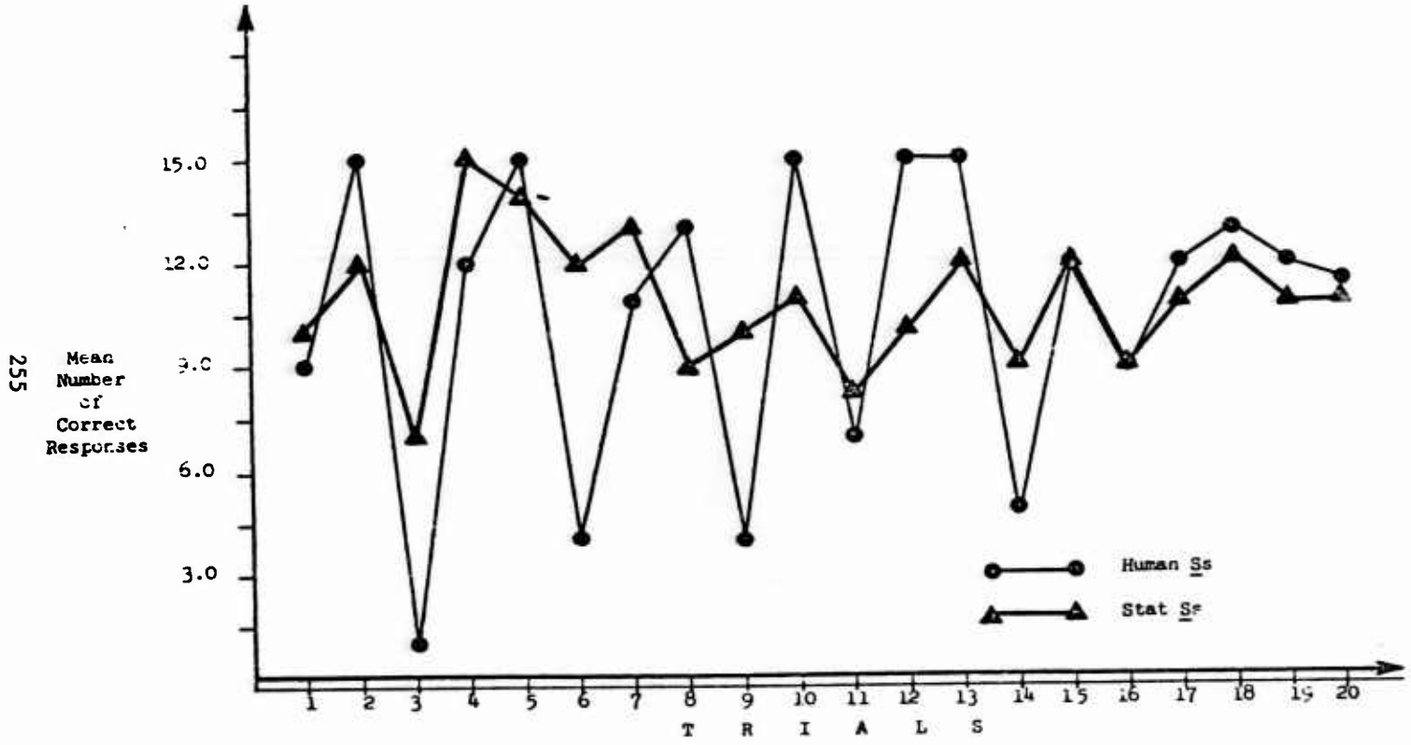
This process seems to have the potential for an important role in human pattern identification, and we have been attempting to model it. Time is too short for me to go into detail about these efforts, but I would like to illustrate some of our results. Figure 10 shows the earliest effort (Evans, 1969). In this and subsequent models, we have sought to use elementary processes which were compatible with psychological theories of learning. From the viewpoint of machine pattern recognition, however, our models may be regarded as hierarchically organized probabilistic adaptive automata which learn to select functions so as to minimize an error signal derived by trying to predict one part of a pattern on the basis of another part. This first model performed at levels somewhat above that of the human subjects. Later studies (Brown & Evans, 1970), with somewhat improved instructions and stimulus formats, have found better human performance. The model has proved encouraging with respect to prediction of performance in this limited task; comparisons with human performance, as found in a recent study, are shown in Figures 11 and 12.

This model is part of our effort to develop a model for human pattern identification. Many of the other points I mentioned earlier will also be incorporated. We are hopeful that developments on this model will allow psychology in the future to give more construcitve answers to the question of how humans recognize patterns.



Proportion of errors in unsupervised discrimination between two pattern classes, for human subjects and for "stat-subjects" resulting from the operation of a probabilistic adaptive model. (From Evans, 1969.)

Figure 10

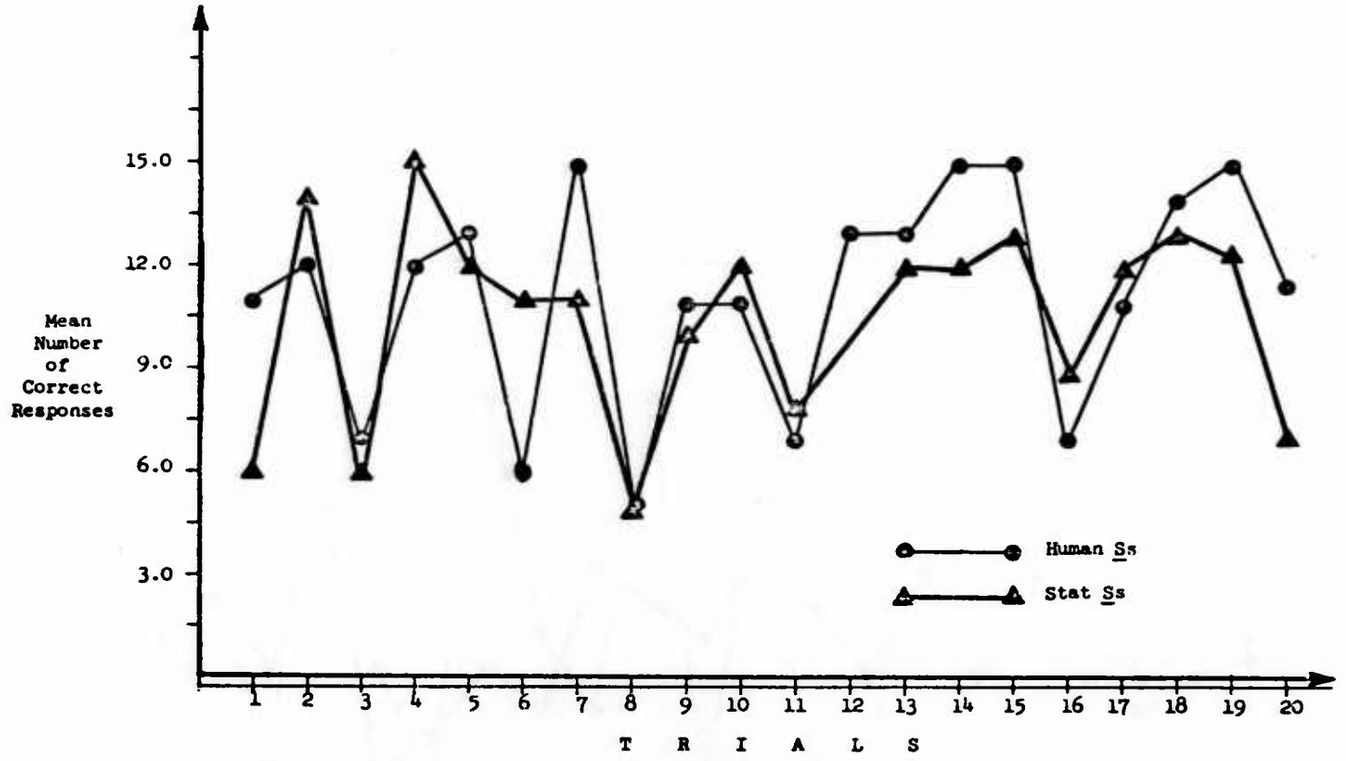


255

Trial by trial plot of the mean number of correct responses given to the A-C Histoform patterns.

Mean number of correct responses in an unsupervised pattern discrimination task for human subject and for "stat-subjects." The stat-subjects represent data produced by a probabilistic adaptive model. The stimuli in this task were Vargus 9 patterns plotted as histograms. (From Wright, 1969).

Figure 11



Trial by trial plot of the mean number of correct responses given to the A-C Aircraft patterns.

Mean number of correct responses in an unsupervised pattern discrimination task for human subjects and for "stat-subjects." The stat-subjects represent data produced by a probabilistic adaptive model. The stimuli in this task were in the schematic airplane format shown in Figure 3. (From Wright, 1969).

Figure 12

REFERENCES

1. Aiken, Leona. A feature utilization analysis of pattern class structure. Perception and Psychophysics, 1970, in press.

2. Aiken, Leona and Brown, D. R. Visual form perception: Congruence among spatial configurations. Perception and Psychophysics, 1969a, 5 (3), 155-160.

3. Aiken, Leona and Brown, D. R. A spatial analysis of the discriminability of forms in noise. Perception and Psychophysics, 1969b, 5 (3), 171-175.

4. Aiken, Leona, and Brown, D. R. Visual form perception: Task demands and congruence among spatial solutions. Psychonomic Science, 1969c, 15 (4), 219-220.

5. Antelman, S. M., Olson, A. and Orbach, J. Orientation-specific mechanisms in the human visual system. Perception and Psychophysics, 1961, 6, 266-268.

6. Attneave, F. and Arnoult, M. D. The quantitative study of shape and pattern perception. Psychological Bulletin, 1956, 53, 452-271.

7. Behrman, B. W. and Brown, D. R. Multidimensional scaling of form: A psychophysical analysis. Perception and Psychophysics, 1968, 4, 19-25.

8. Bersted, C. T., Brown, B. R. and Evans, S. H. Free sorting with stimuli clustered in a multidimensional attribute space. Perception and Psychophysics, 1969, 6 (6B), 409-413.

9. Brown, D. R. and Andrews, M. H. Visual form discrimination: Multidimensional analyses. Perception and Psychophysics, 1968, 3, 401-406.

10. Brown, D. R. and Brumaghim, S. H. Perceptual equivalence, pattern perception, and multidimensional methods. Perception and Psychophysics, 1968, 4 (4), 253-256.

11. Brown, B. R. and Evans, S. H. Further applications of the random adaptive module (RAM) system to schema theory. Human Engineering Laboratories Technical Memorandum 4-70, March 1970.

12. Brown, B. R. and Evans, S. H. Perceptual learning in pattern discrimination tasks with two and three schema categories. Psychonomic Science, 1969, 15 (2), 101-103.

13. Brown, D. R. and LoSasso, J. S. Pattern degradation, discrimination difficulty, and quantified stimulus attributed. Psychomonic Science, 1967, 9 (6), 351-352.

14. Brown, D. R. and Owen, D. H. The metrices of visual form: Methodological dyspepsia. Psychological Bulletin, 1967, 68, (4), 243-259.

15. Brumaghim, S. H. and Brown, D. R. Perceptual equivalence between visual and tactual pattern perception: An anchoring study. Perception and Psychophysics, 1968, 4 (3), 175-179.

16. Brunswick, E. Perception and the representative design of experiments. Berkeley: University of California Press, 1956.

17. Cooper, D. B. and Cooper, P. W. Nonsupervised adaptive signal detection and pattern recognition. Information and Control, 1964, 7 (3), 416-444.

18. Cooper, P. W. Some topics on nonsupervised adaptive detection for multivariate normal distributions. Applied Optics and Optical Engineering, 1967, 4, 123-246.

19. Edmonds, E. M. and Mueller, M. R. The role of schemata in perceptual learning. Psychonomic Science, 1967, 8 (6), 239-240.

20. Edmonds, E. M., Mueller, M. R., and Evans, S. H. Effects of knowledge of results on mixed schema discrimination. Psychonomic Science, 1966, 6 (8), 377-378.

21. Ellis, H. C. and Feuge, R. L. Transfer of predifferentiation training to gradients of generalization in shape recognition. Journal of Experimental Psychology, 1966, 71 (4), 539-542.

22. Evans, S. H., Hoffman, A. A. J., Arnoult, M. D., and Zinser, O. Pattern enhancement with schematic operators, Behavioral Science, 1968, 13 (5), 402-404.

23. Evans, S. H. and Mueller, M. R. VARGUS 9: Computed stimuli for schema research. Psychonomic Science, 1966, 6 (12), 511-512.

24. Evans, S. H. A model for perceptual category formation. Unpublished doctoral dissertation, Texas Christian University, 1964.

25. Evans, S. H. The random adaptive module (RAM): A formal system for S-R modeling, with applications to schema theory, Human Engineering Laboratories Technical Memorandum 24-69, December, 1969.

26. Evans, S. H. VARGUS 7: Computed patterns from Markov processes. Behavioral Science, 1967, 12 (4), 323-328.

27. Fenker, R. M., Jr., and Brown, D. R. Pattern perception, conceptual spaces, and dimensional limitations on information processing. Multivariate Behavioral Research, 1969, 4,257-272.

28. Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, Nancy, and Leonard, J. A. Stimulus correlates of visual pattern recognition: A probability approach. Journal of Experimental Psychology, 1956, 51 (1), 1-11.

29. Gibson, J. J. Perception as a function of stimulation. In S.Koch (Ed.), Psychology: A study of a science. Vol. 1. New York: McGraw-Hill, 1959.

30. Haralick, R. M. and Kelly, G. L. Pattern recognition with measurement space and spatial clustering for multiple images. Proceedings of the IEEE, 1969, 57 (4), 654-665.

31. Hubel, D. H. The visual cortex of the brain. Scientific American, 1963, 209, 54-78.

32. Hubel, D. H. and Wiesel, T. H. Receptive fields of single neutrons in the cat's striate cortex. Journal of Physiology, 1959, 148, 574-591.

33. Hubel, D. H. and Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, Journal of Physiology, 1962, 160, 106-154.

34. Hubel, D. H. and Wiesel, T. N. Shape and arrangement of columns in cat's striate cortex. Journal of Physiology, 1963, 165, 559-568.

35. Hubel, D. H. and Wiesel, T. N. Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. Journal of Neurophysiology, 1965, 28, 229-289.

36. Hubel, D. H. and Wiesel, T. N. Anatomical demonstration of columns in the striate cortex. Nature, 1969, 221, 747-750.

37. Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. Psychometrika, 1964, 29, 1-27.

38. Mavrides, Cynthia M. and Brown, D. R. Discrimination and reproduction of patterns: Feature measures and constraint redundancy as predictors. Perception and Psychophysics, 1969, 6, 276-280.

39. Nagy, G. A. A transposed method of feature extraction for binary patterns. IEEE Systems Science and Conference Record, 1968, 41-47.

40. Posner, M. I. Uncertainty as a predictor of similarity in the study of generalization. Journal of Experimental Psychology, 1964, 68 (2), 113-118.

41. Rankin, W. C. and Evans, S. H. Facilitation of schematic concept formation as a function of two within-schema pretraining modes. Psychonomic Science, 1968, 13 (6), 325-326.

42. Rodwan, A. S. and Hake, H. W. The discriminant-function as a model for perception. American Journal of Psychology, 1964, 77, 380-392.

43. Rosser, E. M. Categorization and discrimination of tone sequences. Unpublished doctoral dissertation. Harvard University, April 1967.

44. Shannon, C. E. and Weaver, W. The mathematical theory of communication. Urbana, Illinois: The University of Illinois Press, 1949.

45. Wright, A. D. Schematic concept formation: An extension of previous models, Unpublished doctoral dissertation, Texas Christian University, 1969.

NOTES ON THE KALMAN FILTERING TECHNIQUE

John Matsushino, SSAD, SAFEGUARD
Communications Agency Analysis Division
Fort Huachuca, Arizona

I. INTRODUCTION. The Kalman filter is a digital filter which is widely used on digital computers as an effective data processing technique. Its successful applications are due to Rudolph E. Kalman.

According to classical data processing methods, the observation vector y_i is obtained sequentially at discrete time t_1, t_2, t_3, ..., t_k, that is, y_1 at t_1, y_2 at t_2,...sequentially. At each instant of time, an estimate of the x vector is formed based on all previous data plus the data at that time instant. Thus, if we have a set of observation vectors, then

$$\begin{aligned} y_1 &= H_1 x + v_1 \\ y_2 &= H_2 x + v_2 \\ y_3 &= H_3 x + v_3 \\ &\vdots \\ y_k &= H_k x + v_k \ . \end{aligned}$$

where H_1, H_2, ..., H_k are fixed and known matrices, x is the parameter to be estimated, and v_1, v_2, v_3, ..., v_k are the measurement noises, this type of data processing results in an enormous amount of data to be stored. In particular, matrix inversion becomes a difficult task since the array of vectors is quite large.

Based on the estimation technique we are interested in, the procedures are a bit different, that is, at time t_{i-1}, we assume that we have an estimate x_i based on the observation y_{i-1}. Then at time t_i, a new set of data is obtained and a new estimate x_{i+1}, is formed based on x_i and the observation y_i. Thus, at $t = 0$, we assume we have an estimate x_1 based on the observation y_o; at $t = t_1$, we form an estimate x_2 based on the previous estimate x_1 and the observation y_1; at $t = t_2$, a set of observation data y_2 becomes available and we form a new estimate x_3 based on x_2 and y_2. This procedure requires storage

of the current estimate x_i and the current data y_i only, and the size of matrix to be inverted becomes much smaller compared to the classical method previously discussed. In particular, if the matrix elements are scalar, matrix inversion is a simple division.

The lastly discussed data processing method is stated by a mathematical model defined by the prediction equation and the observation as

$$x(k+1) = \phi(k)\, x(k) + u(k) \quad \ldots \text{ prediction}$$

$$z(k) = Hx(k) + v(k) \quad \ldots \text{ observation}$$

where: $x(k+1)$ the state to be estimated at time t_k

t_k time at k th sampling instant

$x(k)$ the state of the linear dynamical system at t_k

$u(k)$ vector control function in linear, dynamical system at t_k

$z(k)$ observation data at t_k linearly related to $x(k)$

H observation matrix at t_k

$v(k)$ measurement noise at t_k

$\phi(k)$ state transition matrix relating to $x(k)$

Given a state $x(k)$ at t_k, the problem is -- what are the methods we can use in order to obtain a new estimate $x(k+1)$? The Kalman filter is that method which will give the desired estimate.

II. THE RECURSIVE KALMAN FILTER EQUATIONS. The Kalman filter equations are used in the order they are given below; that is, steps 1 through 5. Having completed step 5 return to step one and go on through 5 recursively until the $P(k)$ matrix indicates that $\hat{x}(k)$ is sufficiently accurate. In this respect, the Kalman filter is recursive.

1. Predicted linear estimate of the state $x(k)$ at t_k before $z(k)$ is processed

$$x_p(k+1) = \phi(k)\hat{x}(k)$$

2. Covariance of error in the predicted estimate of the state t_k before $z(k)$ is processed

$$P_p(k+1) = \phi(k)\, P(k)\phi^T(k)$$

3. Optimal gain matrix $K(k)$ for the Kalman filter at t_k such that error existing between the state and the predicted state will be a minimum.

$$K(k) = P_p(k)\, H^T[H\, P_p(k)\, H^T + R(k)]^{-1};\ R(k) = E[v(k)\, v^T(k)]$$

Let $x(k+1) = x(k) + \dot{x}(k)\Delta t$.

By definition, $x(k+1) = \phi(k)x(k)$, so that

$$\phi(k)x(k) = x(k) + \dot{x}(k)\Delta t \rightarrow \phi(k) = 1 + \frac{\dot{x}(k)}{x(k)}\Delta t$$

Assuming linearity, $\phi(k)$ can be written in vector form as

$$\phi(k) = 1 + \frac{\partial \bar{x}(k)}{\partial \bar{x}(k)} \frac{\dot{\bar{x}}(k)}{\bar{x}(k)} = 1 + \frac{\partial \dot{\bar{x}}(k)}{\partial \bar{x}(k)} \frac{\bar{x}(k)}{\bar{x}(k)} \Delta t$$

$$\phi(k) = 1 + \frac{\partial \dot{\bar{x}}(k)}{\partial \bar{x}(k)} \Delta t = 1 + \begin{bmatrix} \frac{\partial \dot{x}}{\partial x} & \frac{\partial \dot{x}}{\partial \dot{x}} \\ \frac{\partial \ddot{x}}{\partial x} & \frac{\partial \ddot{x}}{\partial \dot{x}} \end{bmatrix} \Delta t$$

Let Δt = unity. Then

$$\phi(k) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad \phi^T(k) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

IV. GAIN INITIALIZATION. During a computation, the covariance matrix P(k) may intermittenly give the values of its elements practically equal to zero. When this happens the computer will gyrate with the zero elements of the P(k) elements, without giving values for the $P_p(k+1)$ computation of the next stage. In order to cure this, the gain K(k) must be initialized so that the computer can go on. The gain by definition is

$$K(k) = P_p(k)H^T \left[HP_p(k)H^T\right]^{-1}$$

$$P_p(k) = E \begin{bmatrix} [x_p(k) - x(k)]^2 & [x_p(k) - x(k)]\,[\dot{x}_p(k) - \dot{x}(k)] \\ [\dot{x}_p(k) - \dot{x}(k)]\,[x_p(k) - x(k)] & [\dot{x}_p(k) - \dot{x}(k)]^2 \end{bmatrix}$$

$$= E \begin{bmatrix} \Delta x^2 & \Delta x \Delta \dot{x} \\ \Delta \dot{x} \Delta x & \Delta \dot{x}^2 \end{bmatrix}$$

Therefore,

4. Linear estimate $\hat{x}(k)$ of the state $x(k)$ at t_k using the data $z(k)$

$$\hat{x}(k) = x_p(k) + K(k)[z(k) - H(k)x_p(k)]$$

5. Covariance of error matrix in the estimate of state at t_k;

$$P(k) \equiv E[\hat{x}(k) - x(k)]^T[\hat{x}(k) - x(k)]$$

$$P(k) = P_p(k) - K(k)\, HP_p(k);$$

Note that k is zero in steps 1 and 2 initially and that k is 1 in steps 3, 4, 5. Then in the next stage k becomes 1 in steps 1 and 2 and k is 2 in steps 3, 4, and 5.

III. THE STATE TRANSITION MATRIX $\phi(k)$.

Case 1. $\phi(k)$ is a constant matrix associated with a first-order differential equation with constant coefficients. In this event, $\phi(k)$ causes no trouble.

Case 2. $\phi(k)$ is time varying due to time-varying coefficients. In this case, let

$$x(k+1) = \phi(k)x(k) + B(k)u(k)$$

$$k = 0,\ x(1) = \phi(0)\,x(0) + B(0)u(0)$$

$$k = 1,\ x(2) = \phi(1)x(1) + B(1)u(1)$$

$$= \phi(1)[\phi(0)x(0) + B(0)u(0)] + B(1)u(1)$$

Therefore

$$x(2) = \phi(1)\phi(0) + \phi(1)B(0)u(0) + B(1)u(1).$$

In general, for $k > j$, $\phi(k,j) = \prod_{i=j}^{k-1} \phi(i) = \phi(j)\phi(j+1) \ldots \phi(k-2)\phi(k-1)$.

Since $\phi(k,j)$ is the product of all the previous transition matrices, to compute it becomes very difficult even on computers.

Case 3. An easy method for computing $\phi(k)$.

$$K(k) = \begin{bmatrix} \Delta x^2 & \Delta x\Delta\dot{x} \\ \Delta x\Delta\dot{x} & \Delta\dot{x}^2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \left\{ [1 \quad 0] \begin{bmatrix} \Delta x^2 & \Delta x\Delta\dot{x} \\ \Delta x\Delta\dot{x} & \Delta\dot{x}^2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}^{-1}$$

$$= \begin{bmatrix} \Delta x^2 \\ \Delta x\Delta\dot{x} \end{bmatrix} \left\{ [\Delta x^2 \quad \Delta x\Delta\dot{x}] \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}^{-1} = \begin{bmatrix} 1 \\ \frac{\Delta\dot{x}}{\Delta x} \end{bmatrix}_{\Delta t = 1} = K(k) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The element $\Delta x\Delta\dot{x}$ for $P_p(k)$ is zero by definition in this case since $E[\Delta x\Delta\dot{x}] = 0$.

V. A SIMPLE NUMERICAL EXAMPLE. To illustrate the use of the Kalman filter, assume that observation data is given in the table below.

k	x	$\dot{x}$
0	0	966
1	966	↑
2	1930	
3	2900	
4	3860	↓
5	4840	966

Before the first stage computation can start, some preliminary calculations must be made as follows:

a) $P_p(1) = E \begin{bmatrix} \Delta x^2 & \Delta x\Delta\dot{x} \\ \Delta x\Delta\dot{x} & \Delta\dot{x}^2 \end{bmatrix}$ from previous page

Let Δx be a convenient value such as 1 meter
Δt be 1 second. Since $E[\Delta x\Delta x] = 0$,

$$P_p(1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

b) From case 3, in section III,

$$\phi(k) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \phi^T(k) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

c) $H = [1 \quad 0)$, $H^T = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

d) Assume that $R(k) = E[v(k)\, v^T(k)] = 0.10$.

1. $x_p(1) = \begin{bmatrix} 0 \\ 960 \end{bmatrix}$

2. $P_p(1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

3. $K(1) = P_p(1)H^T[HP_p(1)H^T + 0.10]^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \left\{ [1 \quad 0] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.1 \right\}^{-1}$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \left\{ [1 \quad 0] \begin{bmatrix} 1 \\ 0 \end{bmatrix} + .1 \right\}^{-1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} [1 + .1]^{-1} = \begin{bmatrix} \frac{1}{1.1} \\ 0 \end{bmatrix} = \begin{bmatrix} .91 \\ 0 \end{bmatrix}.$$

Note that matrix inversion is a simple diversion since the elements are scalar.

4. $\hat{x}(1) = x_p(1) + K(1)[z(1) - Hx_p(1)]$

$$= \begin{bmatrix} 0 \\ 960 \end{bmatrix} + \begin{bmatrix} .91 \\ 0 \end{bmatrix} \left\{ 960 - [1 \quad 0] \begin{bmatrix} 0 \\ 960 \end{bmatrix} \right\} = \begin{bmatrix} 880 \\ 960 \end{bmatrix}$$

5. $P(1) = P_p(1) - K(1)HP_p(1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 91 \\ 0 \end{bmatrix} [1 \quad 0] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} .09 & 0 \\ 0 & .09 \end{bmatrix}$

Return to step 1 for the next stage.

VI. DERIVATION OF THE KALMAN FILTERING EQUATIONS.

A. Predicted state $x_p(k)$ and Best Estimate $\hat{x}(k)$

Suppose the dynamics are described by the homogeneous, linear difference equation

(1) $$x(k) = \Phi(k,k-1)\, x(k-1)$$

and the measurements are given by

(2) $$z(k) = H\, x(k) + v(k)$$

where: $x(k)$ is the state to be estimated at t_{k-1}

$\Phi(k,k-1)$ is the transition matrix

$x(k-1)$ is the state at t_{k-1}

The initial state x_o is a vector random variable with the known statistics:

$$E\, x_o = 0$$

$$E\, x_o x_o^T \text{ is a known quantity}$$

$z(k)$ is the measurement at t_k

H is the observation matrix

$v(k)$ is the measurement noise

The statistical characteristics for the observation are:

$$E\, v(k) = 0, \quad E\, v(k)v^T(k) = R(k), \quad E\, v(k)x^T(k) = 0 \text{ for all } k.$$

$$E[\Delta x \Delta \dot{x}] = 0 \text{ for all } k.$$

Given the model defined by equations (1) and (2), the recursive estimation problem is to determine best estimate $\hat{x}(k)$ which is a linear combination of $x(k-1)$ and the measurement $z(k)$. The estimate is best in the sense that the expected value of the sum of the squares of the error in the estimate is a minimum, that is,

$$E[(\hat{x}(k) - x(k)]^T\, [\hat{x}(k) - x(k)] = \text{a minimum}$$

$x_p(k)$ = predicted state before the data is processed

$\hat{x}(k)$ = best estimate of $x(k)$ using data.

$x(k)$ = the state of a given linear dynamic system at t_k

$\tilde{x}(k) = \hat{x}(k) - x(k)$

The state is given according to (1) so that, given an estimate $\hat{x}(k-1)$ at time t_{k-1}, it is reasonable to predict the estimate at time t_k as

$$x_p(k) = \phi(k,k-1)\hat{x}(k-1),$$

or

$$\boxed{x_p(k+1) = \phi(k)\,\hat{x}(k)} \tag{3}$$

$x_p(k+1)$ is the predicted linear estimate at t_k before the data is processed.

Based on (2) and (3), one would expect the measurement $z(k)$ at time t_k to be

$$z_p(k) = Hx_p(k). \tag{4}$$

But an error in the estimate is reflected by a corresponding error in the expected measurement so that

$$\begin{aligned} \text{Error}(k) &= z(k) - z_p(k) = z(k) - Hx_p(k) \\ &= z(k) - H(k)\phi(k,k-1)\hat{x}(k-1). \end{aligned} \tag{5}$$

Now define the gain matrix $K(k)$ such that the estimate $\hat{x}(k)$ is given by

$$\begin{aligned} \hat{x}(k) &= \phi(k,k-1)\hat{x}(k-1) + K(k)[z(k) - z_p(k)] \\ \hat{x}(k) &= \underbrace{\phi(k,k-1)\hat{x}(k-1)}_{x_p(k)} + K(k)[z(k) - \underbrace{H\phi(k,k-1)\hat{x}(k-1)}_{x_p(k)}] \end{aligned} \tag{6}$$

Therefore

$$\boxed{\hat{x}(k) = x_p(k) + K(k)[z(k) - Hx_p(k)]} \tag{6-1}$$

B. The gain matrix $K(k)$.

Let $\tilde{x}(k)$ be the error in the estimate at time t_k, that is,

$$\tilde{x}(k) = \hat{x}(k) - x(k) \tag{7}$$

Then, $E[\tilde{x}(k)\,x(k)] = E[\hat{x}(k)-x(k)]^T[\hat{x}(k) - x(k)]$, and define $P(k)$ as

$$P(k) = E[\tilde{x}(k)\,\tilde{x}^T(k)].$$

By 6, 6-1, and 7,

$$\tilde{x}(k) = \hat{x}(k) - x(k) = \phi(k,k-1)\hat{x}(k-1) + K(k)[z(k) - H\phi(k,k-1)\hat{x}(k-1)] - \underbrace{\phi(k,k-1)x(k-1)}_{x(k)}$$

$$= \phi(k,k-1)\hat{x}(k) + K(k)z(k) - K(k)H\phi(k,k-1)\hat{x}(k-1) - \phi(k,k-1)x(k-1)$$

$$= \phi(k,k-1)[\hat{x}(k-1) - x(k-1)] - K(k)H\phi(k,k-1)\hat{x}(k-1) + K(k)z(k).$$

Note that $\hat{x}(k-1) - x(k-1) = \tilde{x}(k-1)$, $z(k) = Hx(k) + v(k)$.

Therefore

(8) $$\tilde{x}(k) = \phi(k,k-1)\tilde{x}(k-1) - K(k)H\phi(k,k-1)\hat{x}(k-1) + K(k)[Hx(k) + v(k)]$$

Performing the multiplication in (8), we get

(9) $$\tilde{x}(k) = \phi(k,k-1)\tilde{x}(k-1) - K(k)H\phi(k,k-1)\hat{x}(k-1) + K(k)Hx(k) + K(k)v(k)$$

Since $x(k) = \phi(k,k-1)x(k-1)$ by (1)

$$\tilde{x}(k) = \phi(k,k-1)\tilde{x}(k-1) - K(k)H\phi(k,k-1)\hat{x}(k-1) + K(k)H\phi(k,k-1)x(k-1) + K(k)v(k)$$

$$\tilde{x}(k) = \phi(k,k-1)\tilde{x}(k-1) - [\hat{x}(k-1) - x(k-1)]\phi(k,k-1)K(k)H + K(k)v(k).$$

$\tilde{x}(k-1) - x(k-1) = \hat{x}(k-1)$ by definition; so that,

$$\tilde{x}(k) = \phi(k,k-1)\tilde{x}(k-1) - \phi(k,k-1)K(k)H\tilde{x}(k-1) + K(k)v(k).$$

Therefore

(10) $$\tilde{x}(k) = [1-K(k)H]\phi(k,k-1)\tilde{x}(k-1) + K(k)v(k).$$

Since $P(k) = E\,\tilde{x}(k)\tilde{x}^T(k)$,

$$P(k) = E[\{[1-K(k)H]\phi(k,k-1)\tilde{x}(k-1) + K(k)v(k)\}$$
$$\{[1-K(k)H]\phi(k,k-1)\tilde{x}(k-1) + K(k)v(k)\}^T]$$

$$\overbrace{\qquad\qquad\qquad\qquad\qquad\qquad}^{P_p(k)}$$

(11) $$P(k) = E\{[1-K(k)H]\phi(k,k-1)[\tilde{x}(k-1)\tilde{x}(k-1)^T\phi^T(k,k-1)][1-K(k)H]^T$$
$$+ K(k)v(k)\tilde{x}(k-1)[1-K(k)H]^T\phi^T(k,k-1)$$
$$+ [1-K(k)H]\phi(k,k-1)\tilde{x}(k-1)v^T(k)K(k) + K(k)K^T(k)\underbrace{v(k)v(k)^T}_{R(k)}\}.$$

$Ev(k)\tilde{x}(k-1) = 0$, $E\tilde{x}(k-1)v^T(k) = 0$.

Since $P(k) = E\,\tilde{x}(k)\tilde{x}^T(k)$, $E\,\tilde{x}(k-1)\tilde{x}^T(k-1) = P(k-1)$,

$E\,v(k)v(k)^T = R(k)$, $E\,v(k)v^T(k) = 0$

$$P_p(k) = \phi(k,k-1)P(k-1)\phi^T(k,k-1) \tag{11-1}$$

The expression in (11) can be rewritten as

$$P(k) = [1-K(k)H]P_p(k)[1-K(k)H]^T + K(k)R(k)K^T(k)$$

$$P(k) = P_p(k)[1-K(k)H - H^TK^T(k) + K(k)HH^TK^T(k)] + K(k)R(k)K^T(k).$$

$$P(k) = P_p(k) - P_p(k)K(k)H - P_p(k)H^TK^T(k) + P_p(k)K(k)HH^TK^T(k) + K(k)R(k)K^T(k),$$

or

$$\boxed{P(k) = P_p(k) - P_p(k)K(k)H - P_p(k)H^TK^T(k) + K(k)[HP_p(k)H^T + R(k)]K^T(k)}. \tag{12}$$

$P(k)$ is the covariance of the error. $E\tilde{x}(k)\tilde{x}^T(k) = E\,[\hat{x}(k) - x(k)][\hat{x}(k) - x(k)]^T$.

The term $[HP_p(k)H^T + R(k)]$ in (12) is symmetric and nonnegative since the product of any matrix and its tranpose is symmetric.

Define the symmetric matrix $S(k)$ and $S(k)^T$ such that

$$\boxed{S(k)S(k)^T = H(k)P_p(k)H^T + R(k).} \tag{13}$$

The last three terms in (12) are a quadratic matrix polynomial in $K(k)$. Substitute (13) into (12) and hypothesize the existence of a matrix $A(k)$ such that

$$P(k) = P_p(k) + [K(k)S(k) - A(k)][K(k)S(k) - A(k)]^T - A(k)A(k)^T \tag{14}$$

We proceed to check the validity of this assumption as follows:

$$P(k) = P_p(k) + K(k)K^T(k)S(k)S^T(k) - A(k)K^T(k)S^T(k) - K(k)S(k)A^T(k) + A(k)A(k)^T - A(k)A^T(k) \tag{15}$$

$$A(k) = P_p(k)H^T[S(k)^{-1}]^T \tag{15-1}$$

$$A(k)K^T(k)S(k)^T = \{P_p(k)H^T[S(k)^{-1}]^T\}K^T(k)S(k)^T$$

(15-2) $$K(k)S(k)A^T(k) = K(k)S(k)\{P_p(k)H^T[(S(k))^{-1}]^T\}^T.$$

Therefore

$$P(k) = P_p(k) + K(k)K^T(k)S(k)S^T(k) - [P_p(k)H^T(S(k)^{-1})^T]K^T(k)S^T(k)$$

$$- K(k)S(k)[P_p(k)H^T(S(k)^{-1})^T]^T$$

Now making use of formula (13) we obtain

$$P(k) = P_p(k) + K(k)[HP_p(k)H^T + R(k)]K^T(k) - P_p(k)H^TK^T(k) - K(k)S(k)[S(k)^{-1}HP_p(k)^T]$$

$$P(k) = P_p(k) + K(k)[HP_p(k)H^T + R(k)]K^T(k) - P_p(k)K^T(k)H^T - K(k)HP_p{}^T(k),$$

From formula (12) we have

$$P(k) = P_p(k) + K(k)[H(k)P_p(k)H^T + R(k)]K^T(k) - P_p(k)K^T(k)H^T - K(k)HP_p(k).$$

We shall now show that $P_p(k) = P_p{}^T(k)$, so that the above representations of $P(k)$ are equal. By definition,

$$P_p(k) = \phi(k,k-1)P(k-1)\phi^T(k,k-1)$$

$$P_p{}^T(k) = [\phi(k,k-1)P(k-1)\phi^T(k-1)]^T = \phi(k-1)P^T(k-1)\phi^T(k,k-1)$$

But $P_p{}^T(k-1) = E[\tilde{x}(k-1)\tilde{x}(k-1)^T]^T = [\tilde{x}(k-1)\tilde{x}(k-1)^T] = P_p(k-1)$.

Therefore

$$P_p{}^T(k) = \phi(k,k-1)P^T(k-1)\phi^T(k,k-1) = P_p(k)$$

Hence the expression in (15) is identical with the expression in (12). We now minimize the trace of $P(k) = E[\hat{x}(k) - x(k)]^T[\hat{x}(k) - x(k)]$ by choosing $K(k)$ such that

(15-2) $$K(k)S(k) = P_p(k)H^T[S(k)^{-1}]^T.$$

We solve for $K(k)$,

$$K(k) = P_p(k)H^T[S(k)^{-1}]^TS(k)^{-1}; \text{ but } [S(k)^{-1}]^TS(k)^{-1} = [S(k)^TS(k)]^{-1}.$$

Since $S(k)^TS(k) = HP_p(k)H^T + R(k)$ by (13),

$$[S(k)^TS(k)]^{-1} = [HP_p(k)H^T + R(k)]^{-1}.$$

Therefore

(16) $$\boxed{K(k) = P_p(k)H^T[HP_p(k)H^T + R(k)]^{-1}}$$

The gain K(k) can be derived by Wiener's notations:

1. x and v are random vectors
2. The arbitrary estimate $\hat{x}$ of the state x is a linear combination of the measurement of y
3. Best estimate $\hat{x}$ minimized the quantity $E(x-\hat{x})^T(x-\hat{x})$
4. The linear estimate $\hat{x}$ can be expressed as $\hat{x} = \hat{K}y$
5. The statistical characteristics are:
   a) $Ex = 0$, $Ev = 0$, $Ex^Tv = 0$
   b) $Exx^T = P$, $Evv^T = R$.

PROBLEM: TO DETERMINE THE OPTIMAL GAIN K(k)

$$E(x-\hat{x})^T(x-\hat{x}) = E(x-\hat{K}y)^T(x-\hat{K}y) = E\{x^Tx-x\hat{K}^Ty^T-x^T\hat{K}y + \hat{K}^T\hat{K}y^Ty\}$$

$$\frac{\partial}{\partial \hat{K}} E(x-\hat{x})^T(x-\hat{x}) = 0 \rightarrow E\{-x^Ty + \hat{K}^Ty^Ty\} = 0 \rightarrow E\hat{K}^Ty^Ty = Ex^Ty$$

$$\hat{K}Ey^Ty = Ex^Ty \rightarrow \hat{K} = \{Ex^Ty(Ey^Ty)^{-1}\}^{-1} = Exy^T[Ey^{-1}(y^T)^{-1}] = Exy^TE(yy^T)^{-1}$$

$$\hat{K} = Exy^T(E(yy^T))^{-1}$$

$$y = Hx + v.$$

Therefore

$$\hat{K} = [x(Hx + v)^T]\ \{E(Hx + v)(Hx + v)^T\}^{-1}$$

$$\hat{K} = E[xv^T + x(Hx)^T]\ E[Hxx^TH^T + (Hx)^Tv + Hxv^T + vv^T]^{-1}$$

$$E(Hx)^Tx = H^TP,\ Ex^Tv = 0,\ HH^Txx^T = HH^TP$$

$$EHxv^T = 0,\ Evv^T = R,$$

Therefore

$$\boxed{\hat{K} = H^TP_p[HH^TP_p + R]^{-1}}$$

Note that $\hat{K}$ has the same form as the expression in (16).

C. The Error Covariance P(k)

With $K(k) = P_p(k)H^T[HP_p(k)H^T + R(k)]^{-1}$,

(15-1) $\quad A(k) = P_p(k)H^T[S(k)^{-1}]^T, \quad A^T(k) = S(k)^{-1}HP_p(k)^T$

Hypothesizing the existence of the matrix A(k) such that P(k) is expressed as in (14),

(14) $\quad P(k) = P_p(k) + [K(k)S(k) - A(k)]\,[K(k)S(k) - A(k)]^T - A(k)A(k)^T$

$$P(k) = P_p(k) + \{[P_p(k)H^T(HP_p(k)H^T + R(k)^{-1}]S(k) - P_p(k)H^T[S(k)^{-1}]^T\}$$

$$\{[P_p(k)H^T(HP_p(k)H^T + R(k)^{-1}]S(k) - P_p(k)H^T[S(k)^{-1}]^T\}^T$$

$$P(k) = P_p(k) + [K(k)S(k) - P_p(k)H^T(S(k)^{-1})^T]\,[K(k)S(k) - P_p(k)H^T(S(k)^{-1})^T]^T$$

$$- A(k)A(k)^T$$

$$= P_p(k) + K(k)S(k)S^T(k)K^T(k) - P_p(k)H^T(S(k)^{-1})^TS(k)K(k) - K(k)S(k)S(k)^{-1}HP_p{}^T(k)$$

$$+ \underbrace{P_p(k)H^T(S(k)^{-1})^T}_{A(k)} \underbrace{(S(k)^{-1}HP_p(k)^T)}_{A^T(k)} - A(k)A(k)^T$$

$$P(k) = P_p(k) + K(k)S(k)S(k)^TK(k)^T - P_p(k)H^TK^T(k)(S(k)^{-1})^TS(k)^T$$

$$- K(k)S(k)S(k)^{-1}HP_p(k)^T$$

(16-1) $\quad P(k) = P_p(k) + K(k)S(k)S(k)^TK(k)^T - P_p(k)H^TK^T(k) - K(k)HP_p(k).$

Recall from (15-2),

$E[\hat{x}(k) - x(k)]^T[\hat{x}(k) - x(k)]$ is minimized by choosing K(k) such that

$K(k)S(k) = P_p(k)H^T(S(k)^{-1})^T$

Making use of the above relation in the second term of (16-1), we obtain

$$K(k)S(k)S(k)^{T}K(k)^{T} = P_p(k)H^{T}(S(k)^{-1})^{T}S^{T}(k)K^{T}(k) = P_p(k)H^{T}(S(k)^{T})^{-1}(S(k)^{T})K^{T}(k)$$

$$= P_p(k)H^{T}K^{T}(k),$$

The expression in (16-1) becomes:

$$P(k) = P_p(k) + P_p(k)H^{T}K^{T}(k) - P_p(k)H^{T}K^{T}(k) - K(k)HP_p(k)$$

or

(17) $$\boxed{P(k) = P_p(k) - K(k)HP_p(k)}$$

We now have the Kalman Filtering equations as follows:

1. Predicted covariance before the data from (11-1)

   $$P_p(k+1) = \phi(k)P(k)\phi^{T}(k)$$

2. Predicted estimate of the state from (3)

   $$x_p(k+1) = \phi(k)\hat{x}(k)$$

3. The gain K(k) such that $E[\hat{x}(k) - x(k)]^{T}[\hat{x}(k) - x(k)]$ is a minimum from (16)

   $$K(k) = P_p(k)H^{T}[HP_p(k)H^{T} + R(k)]^{-1}, \quad R(k) = Ev(k)v^{T}(k)$$

4. Best estimate $\hat{x}(k)$ using data from (6-1)

   $$\hat{x}(k) = x_p(k) + K(k)[z(k) - H(k)x_p(k)]$$

5. The covariance error matrix to check accuracy of $\hat{x}(k)$ from (17)

   $$P(k) = P_p(k) - K(k)HP_p(k).$$

REFERENCES

1. R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, J. Basic Eng. 82D (1960).

2. R. E. Kalman, New Results in Linear Filtering and Prediction Theory, J. Basic Eng. 83D (1961).

3. H.W. Sorenson, Kalman Filtering Techniques, Advances in Control Systems, Vol. 3, 1966.

4. Ralph Deutsch, Estimation Theory, Prentice Hall, 1965.

CLASSIFICATION TECHNIQUES FOR STRIP CHART RECORDINGS

Barry Rodin, Theodore Hlibka and William Sacco
Ballistic Research Laboratories
U.S. Army Aberdeen Research and Development Center
Aberdeen Proving Ground, Maryland

ABSTRACT. Many chemical and physical devices such as chromatographs and spectrometers record their results on a strip chart. Such devices are often used to classify a sample as belonging to one of several possible classes. Quite often, however, it is not known in advance which parts of the record are significant for discriminating between the given classes. The paper presents classification techniques for such recordings. Several typical recordings from each class will be used as a training set. We assume that the equipment is properly calibrated; however, no a priori knowledge of the physical or chemical composition of the classes is assumed.

Current approaches to this problem divide the parts of the chart deemed to be significant into n non-overlapping segments. Some function of each segment such as average intensity, maximum intensity, or total area under the curve is obtained. The pattern is thereby characterized as an n-dimensional vector to which standard pattern recognition techniques are applied. Inaccuracies due to characterizing the recording as an n-dimensional vector are discussed. New approaches are given which eliminate these objections. For example, a record may be characterized as a set of peaks and areas under each peak. Distance measures for this characterization are proposed and clustering and classification techniques are presented. Preliminary computational experience is discussed.

INTRODUCTION. Strip-chart recorders are widely used in many scientific fields. For example, infra-red spectrometers record absorbence or transmittance vs wavelength, and gas chromatographs record elution rate vs retention time or the normalized retention time given in Korats Index units. This paper presents methods for classifying a strip chart recording as belonging to one of several given classes. Throughout this paper we shall speak in terms of gas chromatography for the sake of concreteness and since that instrument is the source of most of our data although our methods are quite general for all types of strip-chart recorders.

Important Characteristics of the Graph

Several different characteristics of the graph may be considered to be important: the value of the dependent variable, such as time, at which the peaks occur, the height of the peak, or the area of the peak. Relative heights or areas of two or more peaks may also be considered; however, this requires an a priori knowledge of what to look for since there are too many combinations to search them all to see which are significant.

Sources of Error and Variability

Two strip-chart recordings representing the same class of objects may differ from each other considerably. This is due to many sources of error and variability such as:

1. Peaks may be only partially resolved so that for one run of the experiment we might find one large peak whereas for another run we might resolve this peak into two smaller peaks.

2. There is a problem in determining the base-line which represents the absence of a peak and from which the height or area of the peak would be measured. This problem may sometimes only involve the proper calibration of the equipment. However, for temperature programmed chromatography determination of the base line is more complex since the base line rises as heating takes place. Identification of this base line is an art at which the chromatographer is quite capable and so we envision that this part of the analysis will be done manually before the data is prepared for the automatic pattern analysis.

3. The pen may go off scale making an estimate of the correct value necessary.

4. The attenuation of the instrument might be changed during the course of the experiment and this fact must be properly taken into account.

5. The instrument itself has only a certain amount of accuracy and so variability due to the equipment contributes to the variability observed in the strip-chart recording.

6. Different scientists have somewhat different experimental techniques. For example, some people will round off their numbers more than others or measure heights or areas differently. Even the same person may record slightly different numbers for the same measurements if done on two different days.

7. The most important contributing factor to the variability of the strip-chart recording for items of the same class is the variability of the item itself. For example, in the chromatographic analysis of humans different people give off different set of effluents and even the same person may give off widely different effluents on different days.

Normalization

Since different samples may have different concentrations, it is often necessary to normalize the recorded values so as to be able to compare them. Several different types of normalizations might be tried, such as the following:

1. If a peak is larger than a certain threshold value, we consider it to be present, otherwise we consider it to be absent. For this normalization approach, we do not consider differences in the size of the peaks.

2. We might normalize with respect to the largest peak; that is, set the largest peak equal to one unit and let each of the other peaks be the appropriate fraction.

3. Some specific peak or a set of peaks might be found to be present in most or all of the samples and thus could be chosen for normalization.

4. The average size of a peak may be used for normalization. That is, we add the areas of all the peaks of a given recording and divide by the number of peaks to find the average area of a peak. We then divide the area of each peak of that recording by the average value to find its normalized value.

5. If it is known which peaks are chemically significant, that is, which peaks are likely to be present in one of the classes but absent from the other class then these peaks may be chosen as the standard for normalization. However, in this paper we shall not assume any a priori knowledge of the chemical composition of the different classes.

6. We may normalize by the total area. That is, we divide the area of each peak by the total area and thus obtain, as the normalized value, the percentage of the total area contained in the given peak. This is the form of normalization which will be used throughout the rest of this paper.

7. We might normalize in several steps. That is, first we normalize, using one of the methods mentioned above, considering all values of the independent variables to be equally important. Using this normalization on the training sets we may find some values to be relatively unimportant for classification purposes because there is considerable overlap among the recordings for the different classes. We may then normalize only with respect to the values of the independent variable which are found to be significant. We may do this process over again; that is, we may use this new normalization to find additional values of the independent variable which are relatively unimportant and then normalize only with respect to the values which are still considered to be important.

8. Finally, we might use several different types of normalizations such as discussed above. We may use a classification scheme separately or each set of normalized data and then take some kind of average to determine the class to which we will assign the test recording. Alternatively, we might employ a classification scheme that simultaneously utilizes the data of the various normalizations.

Chromatography Data

For the last three years Dr. A. Dravnieks and associates of Illinois Institute of Technology Research Institute (IITRI), under contract to the Ballistic Research Laboratories, have been working on the detection of human effluents. During this time they have designed and built an apparatus to collect effluents from humans in a specially designed trap. They have also devised methods of collecting air samples from various environments and analyzing them by means of gas chromatography.

As the gas chromatographic data on chemical signatures of humans began to accumulate, it became evident that the composition of the organic vapors from humans varied significantly from individual to individual. Visual inspection of the chromatograms did not indicate obvious common patterns. At that stage of the research each chromatogram consisted of 32 to 48 peaks. The sizes of the peaks in the same gas chromatographic locations varied considerably from individual to individual.

Because of the large number of peaks involved it became necessary to develop criteria for the selection of those peaks that would be most significant for identification and classification of human chemical signatures. To this end a histogram study was conducted to find intervals of elution times in which peaks most frequently occurred. Using this procedure 34 intervals along the time axis were selected. Only peaks occurring in these intervals were considered. Therefore each chromatogram was made to correspond with a 34-dimensional vector whose ith component was the normalized peak area appearing in the ith interval. Hence each chromatogram is expressed in the same number of variables. For three statistical experiments a procedure, called Stepwise Discriminant Analysis, was then used to obtain the significant discriminating variables from among the 65 variables.

Then various discriminant analyses were conducted including human versus non-human signatures, Caucasian male versus Caucasian female versus Indian males.

Discriminant functions were obtained using the stepwise discriminant analysis program. This procedure selects important variables in a stepwise manner. It extracts variables (peaks) one by one in order of their decreasing statistical significance (down to a certain statistical noise level) for the purpose of separating the pattern classes. Then the procedure computes linear discriminant functions based on the variables selected.

In the first study, 36 human chemical signatures, randomly selected from the total of 65 available, and all 10 environment (non-human) signatures were used to derive the discriminant function. Then the same discriminant function was applied to classify the remaining 29 human signatures. Only 4 out of 29 were misclassified for a discriminant function based on 15 variables.

A number of other class separations were tried e.g. Caucasian males versus Caucasian females versus Indian males with some success.

New Studies

Because of the promising results obtained by IITRI, we exposed the 5 and 15 dimensional data to cluster analysis using a procedure called ISODATA.* The principal objective of cluster analysis is to gain more information about the structure of a data set. Isodata is an iterative process which attempts to find clusters in the data set. Indeed, in each iteration, we begin with a set of cluster centers and apply the following procedure one or more times: partition the data set by placing each pattern in the subset associated with that cluster to which it is closest as measured by Euclidean distance. We then calculate the average of this subset and use this point as a cluster center and reiterate the procedure.

We discovered that the good results obtained using stepwise discriminant analysis were not reflected in the cluster analysis. Contrary to expectation the clusters generally were neither tight nor homogeneous (i.e. the clusters consisted of samples from both of the classes we were trying to separate). Although the data was multimodal the stepwise discriminant procedure managed to find a separating hyperplane. This is not always possible and led us to believe that the findings were rather fortuitious.

---

*G. H. Ball and D. J. Hall, "ISODATA, A Novel Method of Data Analysis and Pattern Classification," Stanford Research Institute, Menlo Park, California, (Apr. 1965).

These isodata findings as well as our reservations regarding the "interval approach" for selection of peaks led us to consider a non-Euclidean measurement of distance between two chromatograms which we shall now describe.

Distance Between Chromatograms

Each chromatogram is characterized as a set of ordered pairs of numbers--the first number of the pair is the time (or Kovats Index) at which a peak occurred, while the second number represents the area (perhaps normalized) under the peak. Now suppose we have two chromatograms

$$C_1 = \{(x_i, y_i)\}; \quad i = 1.2,\ldots,n \quad \text{and} \quad C_2 = \{(\alpha_j, \beta_j)\}; \quad j = 1,2,\ldots,m.$$

Let $X = \{x_i\}$, $A = \{\alpha_j\}$, and T be a given number

Then we define the distance , $\rho(C_1,C_2)$, to be

$$\rho(C_1,C_2) = \frac{1}{2}\left[\rho_1(C_1,C_2) + \rho_2(C_1,C_2)\right]$$

where

$$\rho_2(C_1,C_2) = \rho_1(C_2,C_1)$$

and

$$\rho_1(C_1,C_2) = d(Y,\beta) = \sum_{i=1}^{q}\left(\frac{2(y_i - b_i)}{y_i + b_i}\right)^2$$

where

$$Y = (y_1,\ldots,y_i,\ldots,y_n,\overbrace{0,0,\ldots,0}^{s}),$$

$$\beta = (b_1,\ldots,b_j,\ldots,b_n,b_{n+1},\ldots,b_{n+s}) ,$$

$$q = n + s ,$$

$$b_i = \sum_{j\in A_i} \beta_j \quad ; \quad i = 1,2,\ldots,n ,$$

$$A_i = \{\alpha_j: \min_r |x_r - \alpha_j| = |x_i - \alpha_j| \text{ and } |x_i - \alpha_j| < T\}$$

and

$$b_{n+1} = \beta_{j_1}$$

$$b_{n+2} = \beta_{j_2}$$

$$\cdot \quad \cdot \quad \cdot$$

$$b_{n+s} = \beta_{j_s} .$$

β_{j_t} is the peak area associated with an α_{j_t} which is not within T units of any x_i.

This distance measure is used later together with a Fix-Hodges procedure for classification studies.

Fix-Hodges Technique

The Fix-Hodges technique is a non-parameteric classification procedure whose use implies only that probability density functions exist for samples from each pattern class and that they are continuous. The objective of the technique is to determine discriminant functions, say $g_1(x)$, $g_2(x),\ldots,g_k(x)$ where x is a pattern vector. Discriminant functions are defined as follows: Suppose we have k pattern classes $R_1,R_2,\ldots R_k$ and k single-valued, scalar functions $g_1(x),g_2(x),\ldots,g_k(x)$ having the property that for all x in R_i,

$$g_i(x) > g_j(x) \quad \text{for } i,j = 1,\ldots,R,\ j \neq i.$$

The $g_i(x)$ are called discriminant functions.

In the Fix-Hodges procedure, the discriminant functions are determined by selecting a positive integer k, which is small compared to N, the total number of patterns in the training sets. A training set, T_i, is a set of sample vectors from class R_i. In order to classify an unknown pattern X, the patterns in the K training sets $T_1,T_2,\ldots,T_k$ are pooled. We then find those k patterns from $T = \bigcup_1^k T_i$ which are "closest" to X. Suppose that of the k closest patterns, n_1 patterns belong to T_1, n_2 patterns belong to $T_2,\ldots,$ and n_k patterns belong to T_k where $n_1+n_2+\ldots+n_k = k$. We then define

$$g_1(x) = n_1$$

$$g_2(x) = n_2$$

$$\vdots$$

$$g_k(x) = n_k$$

and we assign X to class R_j if $n_j > n_k$ for $k \neq j$.

The nearest neighbor procedure is the special case (k=1) of the Fix-Hodge procedure. For example let $K = 2$, $T_1 = \{X_1,\ldots,X_m\}$, $T_2 = \{X_{m+1},\ldots,X_{2m}\}$, $R = R_1 \cup R_2$, $T = T_1 \cup T_2$. Then if we wish to assign $X \in Z \setminus T$ to either R_1 or R_2 we compute

$$\text{men}_{X_j \in T}\ \rho(X_j,X) = \rho(X_\alpha,X)$$

and assign X to $\begin{cases} R_1 \text{ if } X_\alpha \in T_1 \\ R_2 \text{ if } X_\alpha \in T_2 \end{cases}$.

where ρ is the metric of our choice.

The Fix-Hodges technique was used where the non-Euclidean distance (described earlier) for ρ to conduct various discrimination studies. The results are summarized in the following tables.

CONFUSION MATRICES USING FIX-HODGE DISCRIMINATION

New "Distance"

Euclidean Distance
34 Dimensional "Bin" Data

White Male Vs. White Female

called / 15	WM	WF
WM	19	6
WF	8	15

called / 15	WM	WF
WM	7	18
WF	7	16

White Male Vs. Indian Male

called / 15	WM	IM
WM	18	6
IM	1	11

1 No Decision

called / 15	WM	IM
WM	15	10
IM	6	6

White Male Vs. Non-Humans

New "Distance"

called / 15	WM	NH
WM	20	5
NH	7	3

34 Dimensional "Bin" Data

called / 15	WM	NH
WM	24	1
NH	4	6

1 No Decision

White Female Vs. Non-Humans

called / 15	WF	NH
WF	19	6
NH	5	5

called / 15	WF	NH
WF	19	6
NH	7	3

Indian Male Vs. White Female

New "Distance"

called / 5	IM	F
IM	11	1
F	22	21

2 No Decision

Euclidean Distance
34 Dimensional "Bin" Data

called / 15	IM	F
IM	6	6
F	3	22

Indian Male Vs. Non-Humans

called / 15	IM	NH
IM	12	0
NH	2	7

called / 15	IM	NH
IM	12	0
NH	2	7

The Isodata program, as now constituted, could not be used with the non-Euclidean distance measure. So we were unable to obtain cluster results.

Criteria For A Good Classification Technique

We will present a new classification scheme which may be applied to strip chart recordings. We first list some properties which we desire our classification scheme to have.

1. It should emphasize some regions of the chart more than others based on the probability distributions of the training sets. That is, regions of time for which the values of the different classes overlap very little should be considered more important for classification purposes than regions of time for which there is considerable overlap in the values on the strip chart for the different classes.

2. On the other hand, we do not want to make our entire decision from just one part of the graph. By making our decision from several parts of the graph we will be able to minimize efforts at camouflage.

3. Our scheme should give an indication of our confidence in the decision; that is, sometimes the values of the item to be classified will be very close to typical values of one of the classes so that we can be very confident in our decision, however, at other times we will make a decision to put the test recording into one of the classes but it will be quite dissimilar from the training set of either class and so our confidence in that decision will be rather small.

4. We want the scheme to be sufficiently simple so that it may be implemented on a computer without requiring too much computer time or storage and without any human intervention.

5. We desire that the scheme should not require the use of any "chemical logic;" in other words, that the scheme require no a priori knowledge of which peaks are likely to appear in which classes.

6. The method should not be too sensitive to assumptions of the form of underlying probability distributions; that is, from the training sets we may infer on underlying probability distribution associated with each class. Since we make this inference from a small finite set of data it will be inaccurate. Our classification scheme, therefore, must not be too sensitive to such inaccuracies.

7. The most important criterion for judging the effectiveness of a classification technique must, in the final analysis, be the question: Does it work? That is, how successful is it in correctly classifying the data.

Classification Technique

We now propose a classification scheme which uses the training sets of each class to infer a probability distribution for that class. Any item to be classified is assigned to the class receiving the largest note based on the probabilities.

For this classification technique we characterize each recording as a vector $(r_1,r_2,\ldots,r_n)$ whose components are the value of the recording at successive small increments of time $t_1,t_2,\ldots,t_n$. For any given time we will assume the recorded value for elements of each class to be normally distributed. We shall use the data from the training set to estimate the unknown parameters; in this case the mean and variance. The normal distribution is by no means the only distribution which could be chosen for this purpose. As experience is gained with a specific problem a different family of probability distributions might be chosen.

We use the values from the training set for a given time t_i to estimate the mean M_{ij} and standard deviation S_{ij} for the ith time and the jth class. We thus characterize each class j as an ordered set of ordered pairs:

$$((M_{1j},S_{1j}),\ (M_{2j},S_{2j}),\ldots,(M_{nj},S_{nj})).$$

In order to classify the recording characterized by the vector $(r_1,r_2,\ldots,r_n)$ we investigate how significantly it differs from the average value of each class. Let α_{ij} be the value of the probability density function associated with time t_i and class j (in this case the normal distribution with mean M_{ij} and standard deviation S_{ij}) evaluated at r_i. We may then classify the recording as belonging to the class j which maximizes:

$$\prod_i \alpha_{ij}$$

Another approach is to choose the class j for which

$$\sum_i \frac{\alpha_{ij}}{\sum_k \alpha_{ik}}$$

is a maximum. If the correlations are taken into account, other procedures may be devised.

BLANK PAGE

A Computer Model of the Initial Stages of Mammalian Pattern Processing[1]

Otto Zinser
Night Vision Laboratory, Ft. Belvoir, Va.

Shelby H. Evans
Texas Christian University, Ft. Worth, Texas

Two forms of behavioral research suggest that the human visual system is sensitive to features. It has been reported that when patterns are presented in such a way as to compensate for involuntary eye movements (producing a stabilized image) features of patterns such as short straight lines, fade and reappear (Pritchard, 1961; Evans, 1967). Visual masking procedures (Kahneman, 1968; Weisstein, 1969) have shown that a masking stimulus (e.g. a short line of a particular orientation) will raise the threshold of recognition of a test stimulus of similar orientation, when presented shortly after the masking stimulus. The masking research suggests that line sensitive detectors in the brain become adapted to the masking stimulus and hence are slower to react to the test stimulus.

Physiologists and physiological psychologists have for the most part not been able to contribute much information concerning human pattern

---

[1]This research was supported by Grant No. NB7460-01 from the National Institute of Neurological Diseases and Blindness and by the Department of Defense, Project THEMIS Contract (DAAD05-68-C-0176), under the Department of the Army to the Institute for the Study of Cognitive Systems through the TCU Research Foundation. Further reproduction is authorized to satisfy needs of the U.S. Government.

processing. They have, however, provided somewhat more information on how infrahuman organisms respond to patterns; most research of this order has been conducted with cats and monkeys. Hubel and Wiesel's (1963) research with these organisms is now well known.

Briefly, Hubel and Wiesel found that nerve cells at the retinal (ganglion) and geniculate levels are most sensitive to spots and that cells at the cortical level are most sensitive to lines and edges of particular orientations. Responses from these cells were obtained by implanting electrodes in those regions of the visual system that were of interest and scanning the retina with spots, lines or edges. Those receptor cells of the retina which, when stimulated, affected the rate of firing of a cell, comprised the receptor field of that cell.

To Hubel and Wiesel these findings, and those of other researchers, suggested that the initial stages of the mammalian visual system are organized in a hierarchical manner, i.e., the output of retinal cells serves as input to geniculate cells and the output of geniculate cells serves as input to cortical cells.

The purpose of the research to be described was to model some of the hypothesized functions of the mammalian visual system. Hubel and Wiesel's data and their theorizing served as primary guidelines. When theory or data were lacking, invention, supported by empirical findings obtained from the computer, served as a basis for further model development. The model to be described is not a model of pattern recognition, i.e., it does not possess the ability to discriminate and classify patterns. It may be designated as a model of the initial stages of mammalian pattern processing.

A Model of Pattern Processing

Several principles of mammalian pattern processing are indicated in Hubel and Wiesel's research: antagonism between excitatory and inhibitory areas of receptor fields, the sensitivity of the levels of processing to specific pattern features (spots, lines and edges) and the hierarchical processing of pattern information. It has not been clearly established that the above principles of processing have validity; however, despite the tentativeness of the present conception of the visual system, it was considered desirable to study the results obtained from executing the total system on the computer.

Implementation of the above principles will now be discussed. Numerical analogs (5 x 5 matrices of digits) were designed to function in a manner similar to the receptor fields described by Hubel and Wiesel. A 5 x 5 random walk matrix (Fig. 1) was used as an analog of receptor fields of retinal and geniculate cells, to be sensitive to spots; each digit defines the number of unique ways one can move , in a specific number of steps, from the center cell of the matrix to the cell containing the digit.

A set of eight 5 x 5 matrices of digits (Fig. 2) was prepared to serve as analogs of receptor fields of edge-sensitive cortical cells. Positive and negative digits correspond, respectively, to excitatory and inhibitory areas of the receptor fields; it will be noted that the values in each matrix sum to zero. A different matrix was designed for each of four orientations; vertical, horizontal, slanted 45^{o} positive, and slanted 45^{o} negative. Two versions of each orientation were required because the matrices are not symmetrical; one matrix is necessary for the detection of one edge and the other for the detection of the opposing edge.

1	2	3	2	1
2	2	4	2	2
3	4	8	4	3
2	2	4	2	2
1	2	3	2	1

Fig. 1 - A 5x5 random walk matrix employed as a spot detector.

Vertical

(1)

2	2	2	-3	-3
2	2	2	-3	-3
2	2	2	-3	-3
2	2	2	-3	-3
2	2	2	-3	-3

(2)

-3	-3	2	2	2
-3	-3	2	2	2
-3	-3	2	2	2
-3	-3	2	2	2
-3	-3	2	2	2

Horizontal

(1)

2	2	2	2	2
2	2	2	2	2
2	2	2	2	2
-3	-3	-3	-3	-3
-3	-3	-3	-3	-3

(2)

-3	-3	-3	-3	-3
-3	-3	-3	-3	-3
2	2	2	2	2
2	2	2	2	2
2	2	2	2	2

Slanted 45° Positive

(1)

0	0	3	3	2
0	2	3	2	-3
3	3	2	-3	-3
3	2	-3	-3	-3
2	-3	-3	-3	-3

(2)

-3	-3	-3	-3	2
-3	-3	-3	2	3
-3	-3	2	3	3
-3	2	3	2	0
2	3	3	0	0

Slanted 45° Negative

(1)

2	3	3	0	0
-3	2	3	2	0
-3	-3	2	3	3
-3	-3	-3	2	3
-3	-3	-3	-3	2

(2)

2	-3	-3	-3	-3
3	2	-3	-3	-3
3	3	2	-3	-3
0	2	3	2	-3
0	0	3	3	2

Fig. 12 Edge operators (vertical, horizontal, slanted 45° positive and slanted 45° negative).

The analogs were designed within 5 x 5 matrices because it was judged that at that size they would be maximally sensitive to both small and large features of the patterns that were to be detected. Also, they were sufficiently large to include analogs of both excitatory and inhibitory areas. It will be noted, however, that a 5 x 5 matrix was not sufficiently large to include analogs of inhibitory areas within the spot detector. To compensate for the absence of the effect of negative values, a threshold value (described later), employed at a later stage of processing, was raised. The effect of the increase in the threshold value amounted to assigning a constant value to the inhibitory surround for all positions of the spot analog within the pattern.

The model took two forms with respect to the order of application of the analogs to the pattern. For one form (hierarchical method), the edge analogs were applied to the output of the spot analogs. The final version was formed by combining the spot detected version and the edge detected version. For the second form (nonhierarchical method), the edge analogs were applied directly to the patterns. The final version was formed again by combining the spot detected version and the edge detected version.

Pattern Generation

A computer program (VARGUS 10) was written which generated closed nonmeaningful patterns by randomly selecting segments of lines and then joining them end to end. Three of these patterns (Fig. 3) were selected for processing in this investigation.

Two other programs were prepared which introduced levels of two forms of degradation to the patterns. One program introduced random noise by

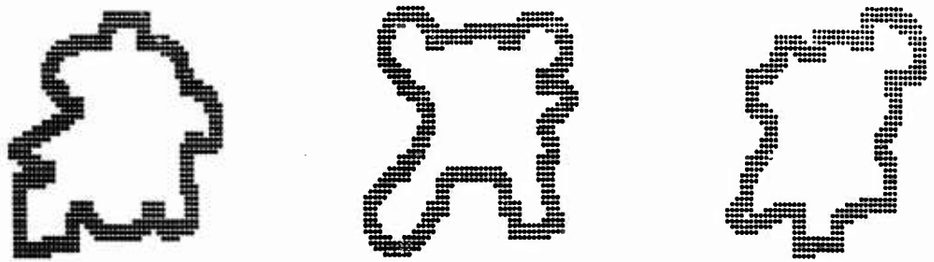


Fig. 3 - The three instances of VARGUS 10 patterns used in this study.

reversing the states of the cells (from black to white or white to black) in the pattern matrix at a specified probability level (Evans, Arnoult and Hoffman, 1966). At .50 the patterns would be completely lost and at 1.00 they would be left nondegraded; intermediate probability levels, appropriate to the degree of degradation desired, were selected by the programmer and given to the computer.

A second program degraded the patterns by blurring them. The blurring procedure involved the application of a 9 x 9 random walk matrix (Fig. 4). Progressively greater levels of blur were produced by successively reapplying the random walk matrix. Examples of random noise and blur may be found at the top of Fig. 7.

Pattern Processing of the Model

Both nondegraded and degraded versions were processed by applying the spot and edge detecting analogs. Each pattern was processed by two programs.

All patterns were coded in terms of ones (black) and zeros (white) to make them interpretable to the computer. The first program translated an appropriate analog over the pattern, beginning in the upper left hand corner and moving within rows, one cell at a time, until the analog matrix had occupied every possible position of the pattern matrix. At each position, the elements of the analog matrix were multiplied by the corresponding elements of the input pattern. All of the resulting products were then added and the sum of the products entered in the output matrix, in the position corresponding to the position over which the analog was centered. The spot analog and each of the edge analogs produced a separate output matrix. A schematic representation of the analog application procedure appears in Fig. 5.

1	4	10	16	19	16	10	4	1
4	12	28	40	48	40	28	12	4
10	28	70	100	124	100	70	28	10
16	40	100	132	168	132	100	40	16
19	48	124	168	216	168	124	48	19
16	40	100	132	168	132	100	40	16
10	28	70	100	124	100	70	28	10
4	12	28	40	48	40	28	12	4
1	4	10	16	19	16	10	4	1

Fig. 4 - A 9x9 random walk matrix employed to produce blurred patterns.

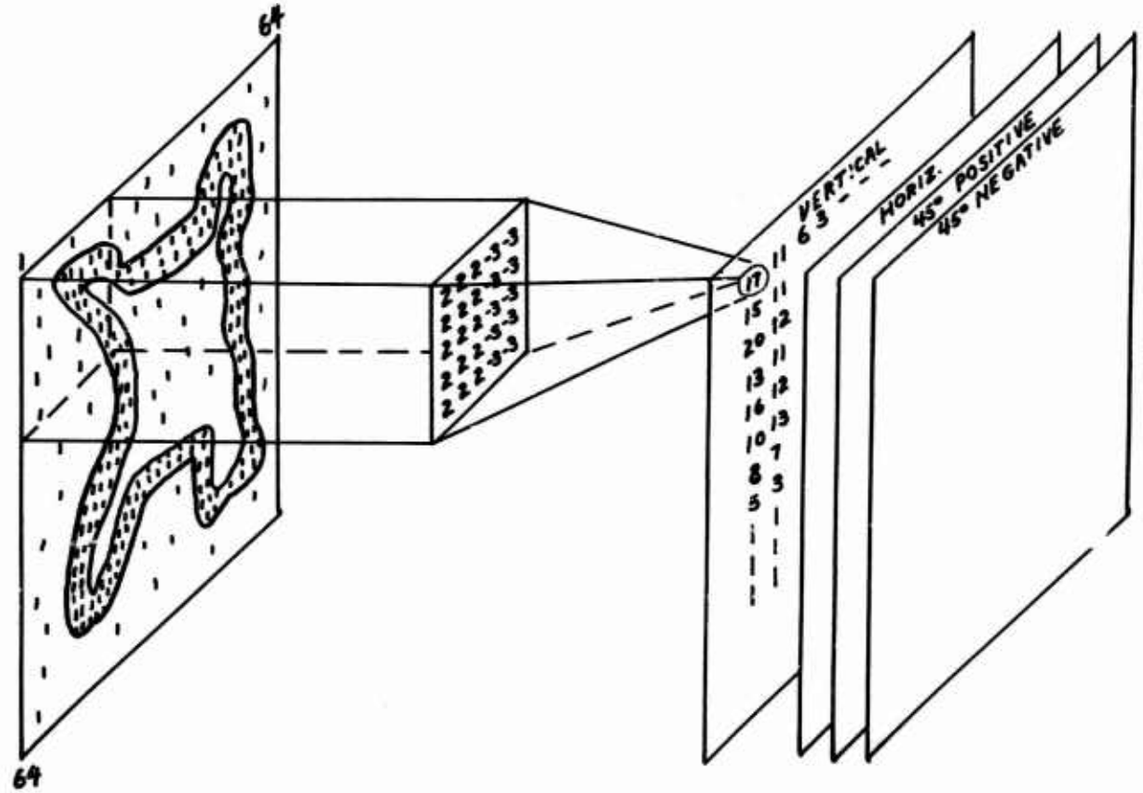


Fig. 5. A schematic representation of the operator application procedure: the vertical operator is shown being applied to the degraded version of the pattern SPIRAL. Output matrices for each of several operators are shown at the right.

Thereafter, each output matrix produced by either a spot or edge analog was processed by another program. The program employed a threshold value which was used to determine whether a cell would be designated black (with a one) or white (with a zero); those cells whose values exceeded the threshold value were designated black and those whose values were less than or equal to the threshold value were designated white. The threshold values were selected by the programmer; in each case that value was selected which according to his judgment maximized acceptance of detected pattern features and minimized acceptance of degradation.

For the case of the edge analog output matrices, after the thresholding procedure, the corresponding values of each of the eight matrices were added by a third program; those values of cells that were one or larger were replaced by a one, those which were zero remained zero. All spot analog and combined edge analog matrices were transformed by still another program which replaced cells with ones by an overstrike of an asterisk (*) and the letter O and left cells with zeros blank.

METHOD

Experimental Design

The nondegraded patterns and the processed nondegraded and degraded versions produced by the hierarchical processing method completed a three - factorial design (Fig. 6). Three instances of VARGUS 10 patterns varied factorially with seven levels of pattern degradation and three levels of pattern processing. The seven levels of pattern degradation consisted of three levels of random noise (at the .70, .75 and .80 probability levels of retaining a cell in its original state), three levels of blur (one and two or three successive applications of a 9 x 9 random walk matrix), and the nondegraded version. The above levels were selected on the basis of ratings provided by Ss whose task was to compare a wide range of degradation levels of each form of degradation with the nondegraded version; the levels of each form of degradation selected for the study, from lowest to highest, received respectively comparable ratings. Pattern processing included three conditions: one, a control, in which the patterns were left nondegraded, a second, for which a spot analog was applied, and a third, for which edge analogs were applied to the spot operator output, and the result added to the spot operator output.

Subjects were asked to rate the patterns according to how they compared to the original verson. Each of three groups (eight Ss per group) was given versions of one of the three patterns shown in Fig. 3. Each S rated all of the degraded versions of his assigned pattern as well as the versions prepared by the model; repeated measures were, therefore, taken over levels of pattern degradation and pattern processing. The Ss were introductory psychology students from Texas Christian University.

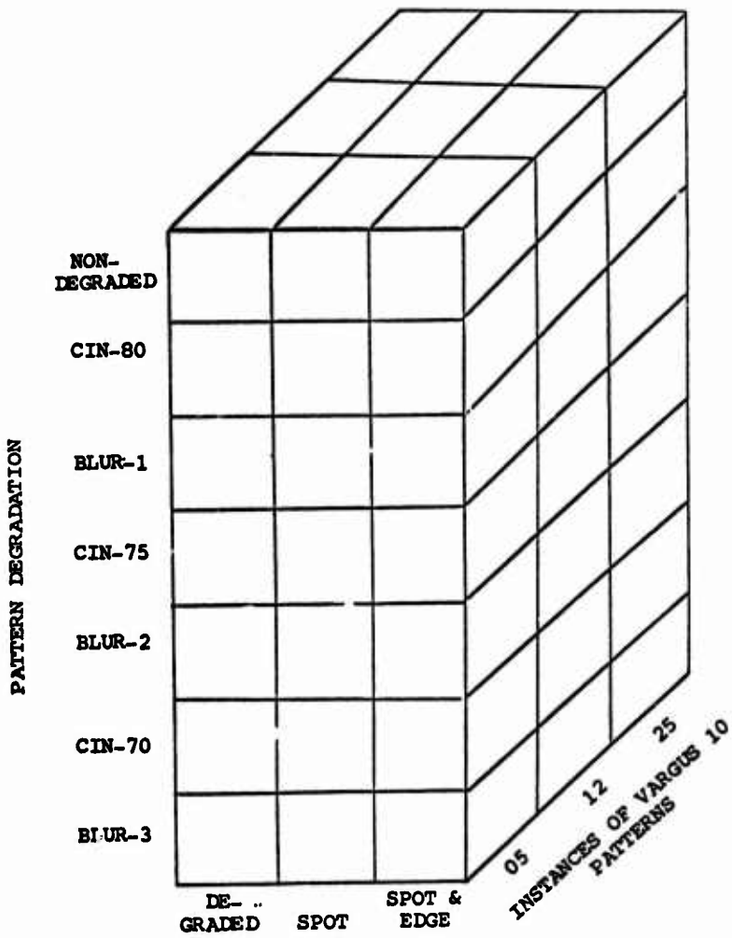


Fig. 17. A display of the three factorial design of this investigation (Pattern Degradation x Pattern Processing x Pattern Instances).

The design was not completed with the nonhierarchical procedure; only the nondegraded version and versions at the highest levels of each form of degradation were processed. Comparison made with the corresponding versions produced by the hierarchical method indicated that the nonhierarchical method led to poorer results. No ratings were obtained from versions produced by the nonhierarchical method.

Pattern Assessment Procedure

Subjects were instructed to rate the patterns according to how they compared to the nondegraded version; they were to assume that each version was a product of a hypothetical transmission system. Ss in each of the three groups rated 21 versions of the particular VARGUS 10 pattern that had been assigned to their group; all combinations of the levels of pattern degradation and pattern processing were represented in each set of 21 patterns. The patterns were reduced in size from the computer printout to the size of a matrix of about 3 in. long and 2 in. wide and were presented attached to 5 x 8 in. index cards.

All subjects were first seated at a table and given the following instruction: "First, I would like you to make yourself generally familiar with the set of patterns you see in front of you. They are probably not like any patterns you have seen before. I'll give you about one minute to study them." A set of eight VARGUS 10 patterns was displayed on the table, for the purpose of providing Ss with the opportunity to become acquainted with their properties. They were instances other than those selected for processing, and were in their computer printout form, of the size of about a 5 in. square.

Thereafter, all of the subjects were seated at another table and given a second set of instructions:

Let us assume that somebody has taken a pattern like those you've seen and attempted to transmit it, using a transmission system which may be described to be somewhat cruder than television. This is the pattern the person trasmitted. Now, I am going to give you a number of pictures, produced by the transmission system, which show the pattern being transmitted in varying degrees of clarity. I would like you to look at these pictures, paying particular attention to the range of quality of pattern transmission. I'll give you about 15 sec. for each picture and will remove each one at the end of that time period.

This phase of the procedure was intended to acquaint Ss with the type and range of pattern degradation inherent in the set of patterns they would subsequently be asked to rate. The patterns were presented in a stack, placed adjacent to the original nondegraded version. Each pattern was removed by the E at the end of its 15 sec. time period.

Finally, a third set of instructions was read to the Ss:

Next, I would like you to look at this set of pictures, but this time, after studying each, I would like you to indicate on a scale how well you think each pattern has been transmitted. Imagine that the clear pattern in front of you is located at one end of the scale and a picture where the pattern is not distinguishable at all is located at the other end. A scale for each of the patterns is drawn on this sheet. If you judge that the transmission is very good, place a checkmark on the scale toward the left; if you think that the transmission is of intermediate quality; place a checkmark somewhere within the moddle region; and if you judge that the transmission is very poor, place a checkmark toward the right end of the scale. You may place your checkmarks anywhere on the line. Ignore any differences in the darkness of the dots. I will give you about 15 secs. for each judgment. Do you have any questions?

This phase of the procedure differed from the previous pretraining phase only in that Ss were asked to provide ratings. The subjects rated the patterns eight times: a different set of the same patterns with a different random order was provided on each trial. The scales were graduated in increments of 10 from zero to 100. For the initial trials a time interval of 15 sec. was observed between ratings. On later trials Ss were allowed to work at their own speed.

RESULTS

The nondegraded and degraded versions and their processed versions are displayed in Fig. 7. The nondegraded version appears at the top left followed by (moving to the right) the three levels of random noise (increasing according to degree of degradation) and the three levels of blur. The processed versions are displayed directly underneath the nondegraded and degraded versions, appearing from top to bottom: the spot detected version, the edge detected version, and the combined version, resulting from combining the spot detected and edge detected versions. The versions produced by the nonhierarchical method of processing are displayed in Fig. 8; the results clearly indicate the inferiority of this method of processing over the hierarchical method. The other two patterns processed in this investigation are shown in Figs. 9 & 10. The results from the three patterns are generally similar.

The spot detected version of the nondegraded pattern exhibits the loss of some detail, such as the sharpness of corners. The edge detected version appears to have summarized the spot detected version quite adequately; however, nc improvement of detail over the spot detected output is evident. This conclusion is supported by the result obtained from combining the edge and spot detected versions; the combined result does not appear to be perceptably superior to the spot detected output.

The spot detected version of the pattern degraded at the .80 level shows that the pattern has been detected quite adequately; the background degradation has been removed and appropriate pattern cells have been filled in. The edge detected version shows the contour reproduced, but the combined version does not seem to show that the edge detected version contributed to the restoration

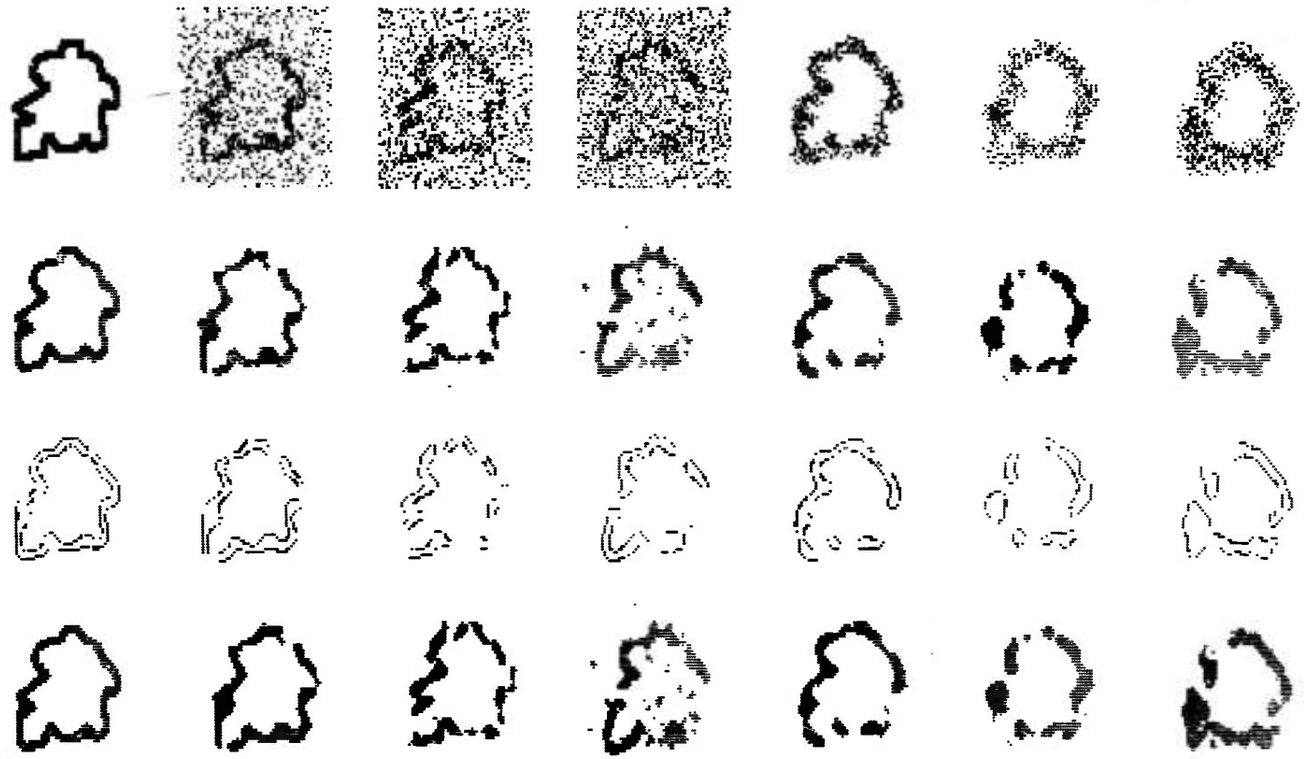


Fig. 7 - Pattern 5 shown nondegraded (top left) followed by versions (top row) with increasing levels of random noise and then increasing levels of blur. Directly beneath each of the top row versions (in descending order) are the spot detected version, the edge detected version and the version resulting from combining the edge and spot detected versions.

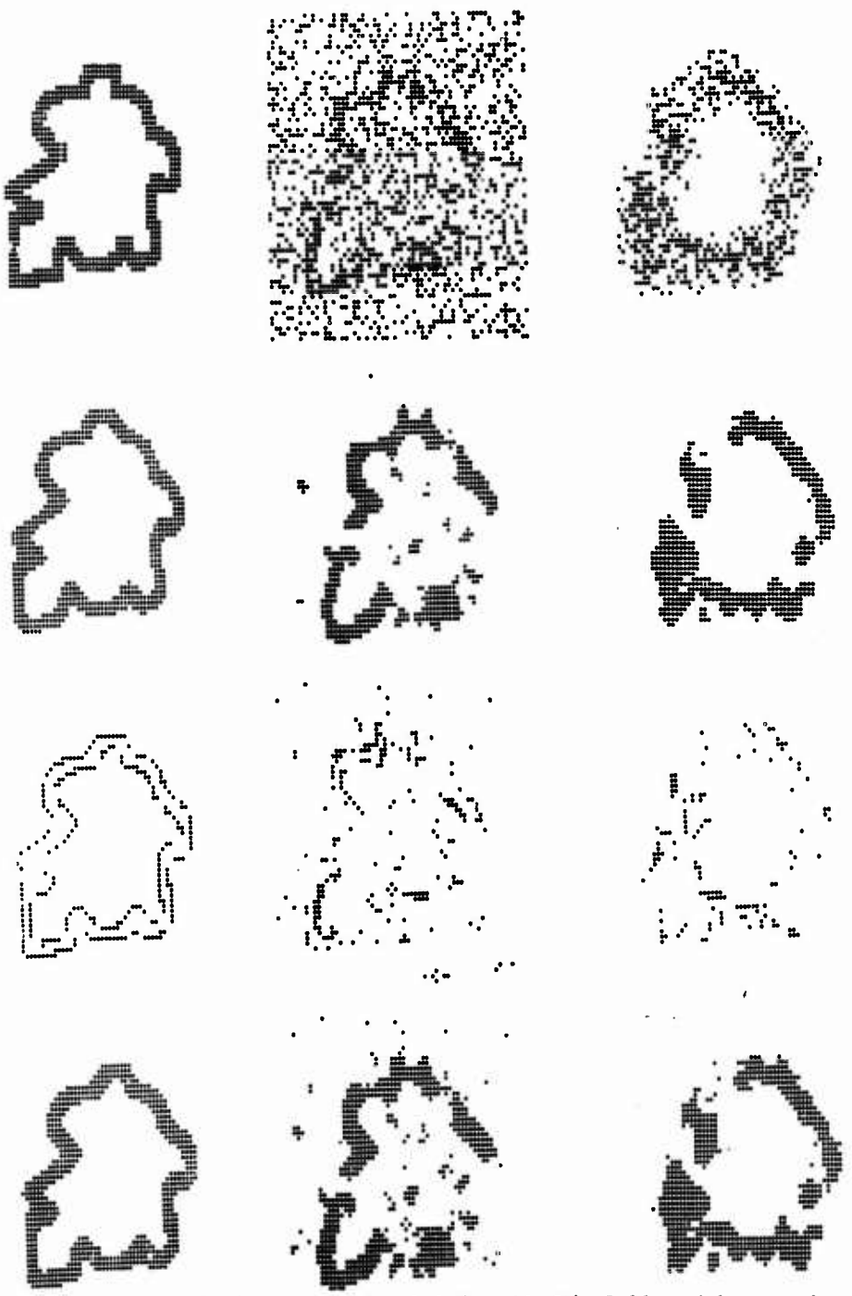


Fig. 8 - Pattern 5 shown nondegraded (top left) followed by versions with the highest level of random noise and the highest level of blur. Directly beneath each of the top row versions, in descending order, are the spot detected version, the edge detected version (nonhierarchical method) and the version resulting from combining the edge and spot detected versions.

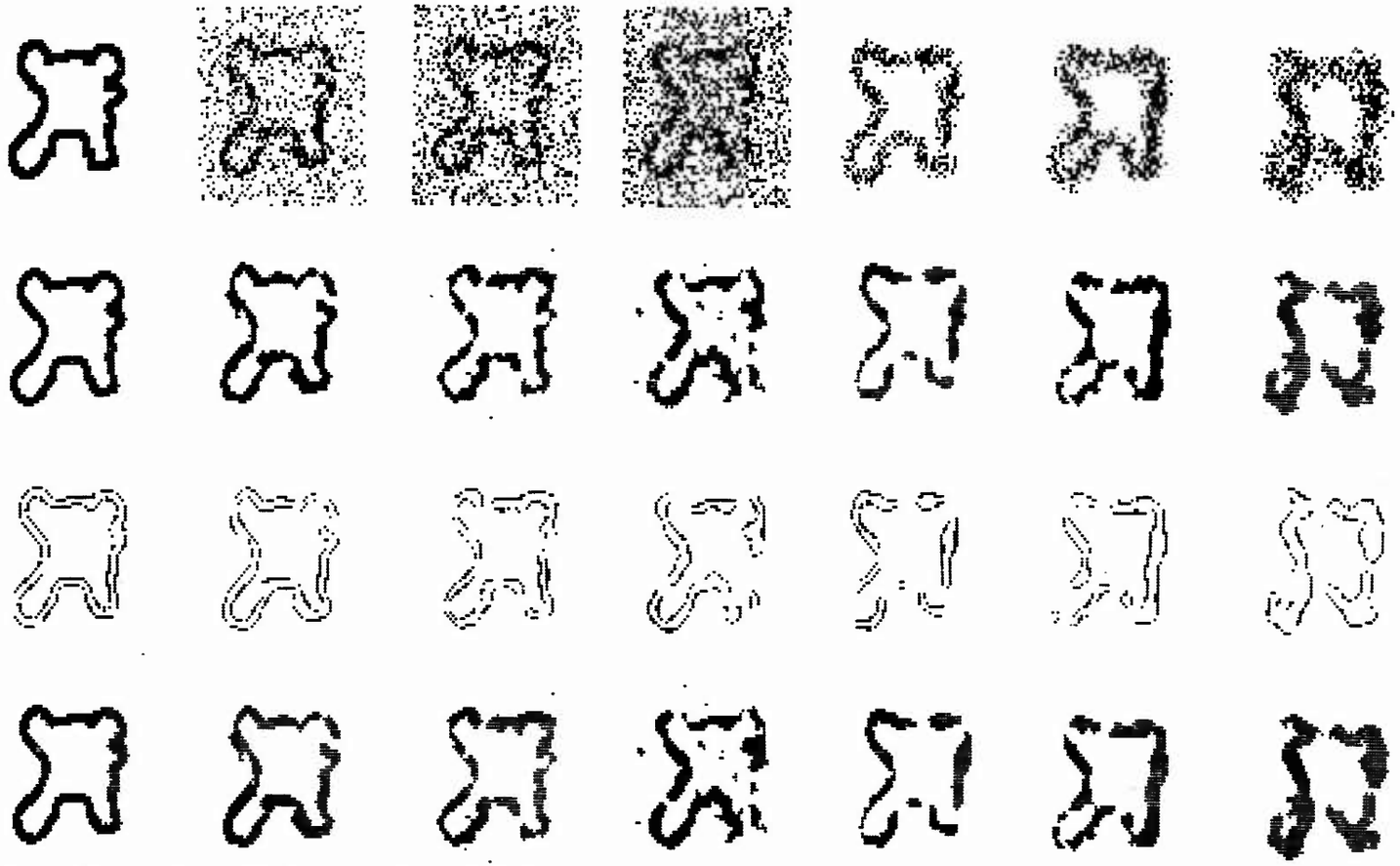


Fig. 9 - Pattern 12 shown nondegraded (top left) followed by versions with the highest level of random noise and the highest level of blur. Directly beneath each of the top row versions, in descending order, are the spot detected version, the edge detected version (nonhierarchical method) and the version resulting from combining the edge and spot detected versions.

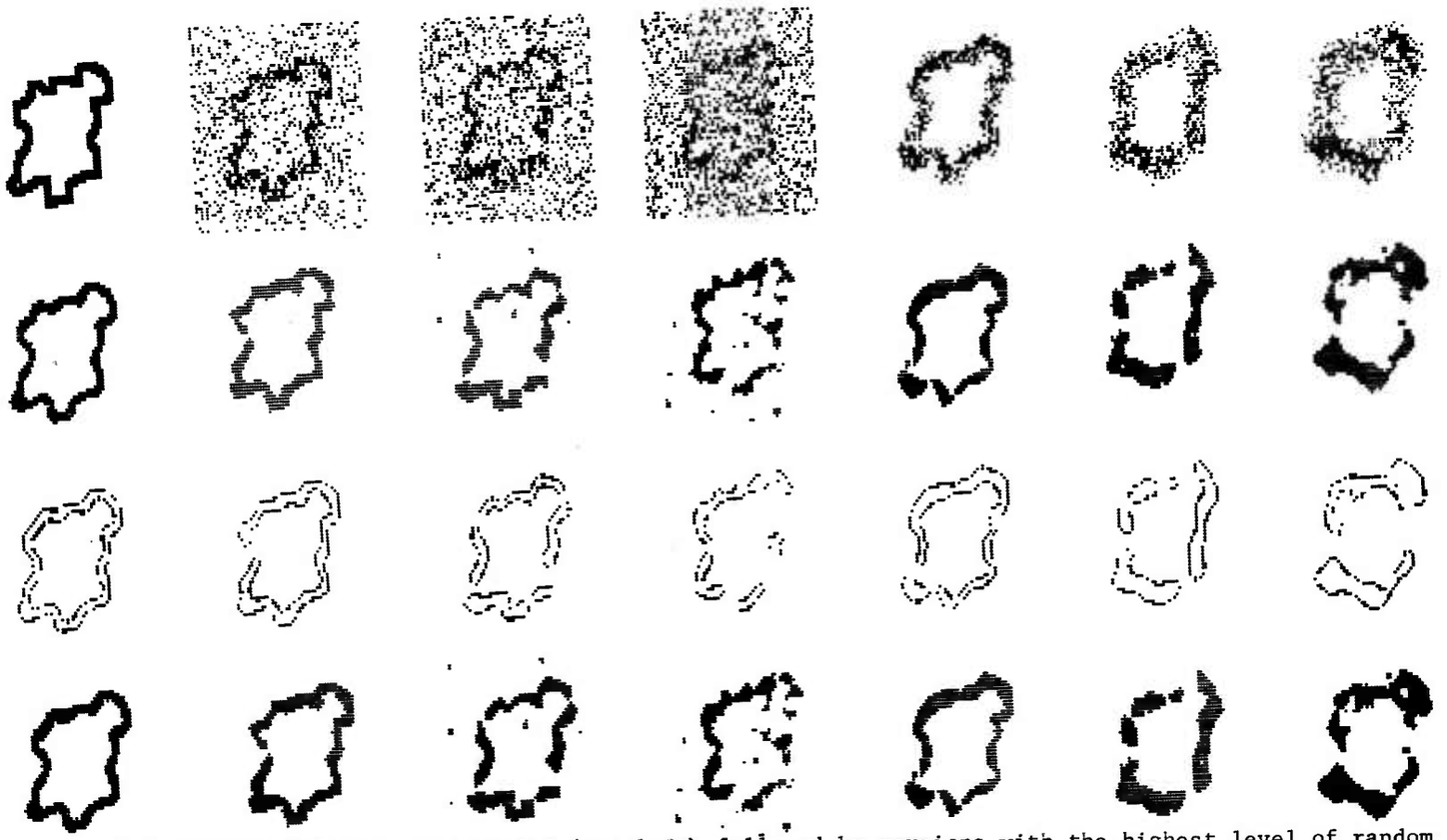


Fig. 10- Pattern 25 shown nondegraded (top left) followed by versions with the highest level of random noise and the highest level of blur. Directly beneath each of the top row versions, in descending order, are the spot detected version, the edge detected version (nonherarchical method) and the version resulting from combining the edge and spot detected versions.

of the original. The spot detected outputs of the pattern at .75 and .70 display progressively poorer pattern detection. Gaps in the patterns become more frequent and increasing amounts of background degradation are retained. The contributions of the edge detected versions continued to be negligible for both levels of degradation.

The spot detected version of the pattern degraded at the lowest level of blur shows that gaps have been filled in and some of the blur has been reduced. The edge detected version, again appears to be making a negligible contribution to the restoration of the pattern. The spot detected versions of the higher levels of blur show that poorer pattern detection has been achieved, and the edge detected versions continue to be making little contribution to the detection of the original.

The mean ratings over the three patterns processed, shown plotted in Fig. 11, provide support to the above observations. In addition, it will be noted that the spot detected versions of the random noise levels show, generally, considerable more improvement than those of the blur levels.

A 7x3x3 three factor analysis of variance with repeated measures on two factors (Pattern Processing and Patterns Degradation) was conducted. The pattern processing factor ($F(2,42) = 22.98$; $p < .01$) and the pattern degradation factor ($F(6,186) = 272.09$; $p < .01$) were found to be significant. The Newman - Keuls analysis for repeated measures was used to make comparisons among levels of pattern processing; the spot detected and the combined versions were each found to be significantly higher than the degraded versions, but the spot detected and combined versions were not significantly different. A significant interaction was obtained between the pattern degradation and pattern processing factors ($F(12,168) = 4,89$; $p < .01$). In addition, a significant interaction was obtained between the pattern degradation and pattern factors ($F(12,168) = 4.89$; $p < .01$).

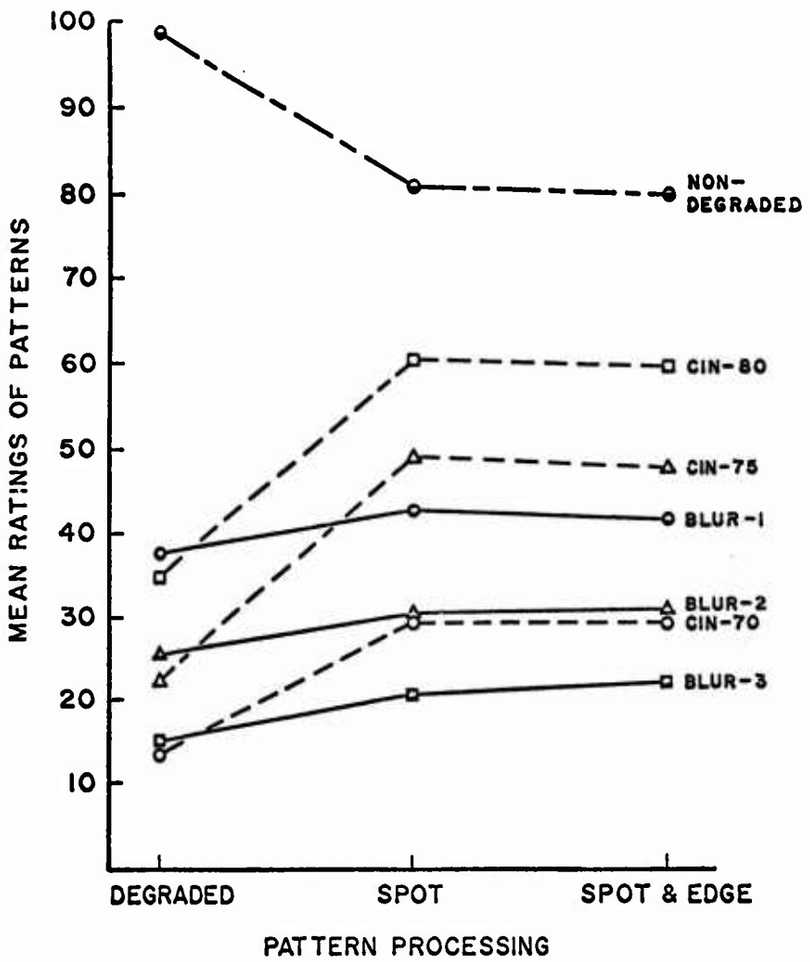


Fig. 11. Mean ratings over pattern instances: levels of Pattern Processing are plotted against levels of Pattern Degradation.

DISCUSSION

The purpose of this investigation was to develop and to assess the performance of a computer model of pattern processing. Hubel and Wiesel's data and theorizing served as the major basis for the model's development.

Numerical analogs of receptor fields were designed; excitatory and inhibitory areas were represented, respectively, by positive and negative values. Two types of analogs were prepared: an analog sensitive to spots and analogs sensitive to edges of four orientations. Finally, two procedures for applying the analogs were investigated: a hierarchical method and a nonhierarchical method.

The analogs were applied to nondegraded patterns and a variety of degraded versions. Three levels of each of two forms of degradation were introduced to the patterns: three levels of random noise and three levels of blur.

A pattern assessment methodology was developed to provide information concerning the effectiveness of the model in detecting degraded and nondegraded patterns. Subjects were asked to rate all pattern versions by comparing them to the original.

The results revealed that the hierarchical method was superior to the nonhierarchical method, in terms of the quality of the versions produced by the edge analogs. The edge detected versions produced by the hierarchical method did not contribute to the detection of the patterns nor did they degrade the spot detected version; they merely exhibit the contour of the edge detected version being summarized. The edge detected versions produced by the nonhierarchical method when added to the spot analog versions <u>did</u> degrade the spot detected versions.

It was expected that the edge analogs would show some effectiveness in detecting the contour of patterns degraded with blur, when applied directly to the blur versions (nonhierarchical method). This expectation was not realized. It remains possible that some measure of effectiveness may be demonstrable at lower levels of blur and with images which retain a gray scale.

The results further indicated that the spot analogs were considerably more effective in detecting patterns from random noise than from blur. The difference in the manner in which the random noise and blur procedures degraded patterns probably determined the disparity in the effectiveness of the spot analog to detect patterns from random noise and blur. Random noise destroyed both detail and the overall characteristics of patterns, whereas blur destroyed, at least at its lower levels, only detail (the contour). The spot analog appears to be particularly suited to detect the overall characteristics of patterns. Hence, there was greater opportunity for the spot analog to be effective in the random noise case.

The edge analogs were generally ineffective in contributing to the detection of the patterns from degraded conditions. This does not imply that the edge analogs would not be useful in other contexts. Their ability to summarize or detect detail would be particularly useful under circumstances in which detail was to be detected from nondegraded patterns. The availability of detail would be useful to a pattern recognition model, for example, which was given the task of discriminating between or among nondegraded patterns that were quite similar.

Human Ss were included in the procedure because of the belief that the human would in this context be the most appropriate pattern evaluation system.

Assuming the model is an adequate analog of the visual system, the model would be expected to be sensitive to the same features to which the human is sensitive. At the same time, the model would be expected to be insensitive to the same features to which the human is insensitive. Hence, it is conceivable that Ss would base their evaluations only upon those features to which they are sensitive and to which the model is sensitive. Another measurement system probably would not have operated on the same basis. It would probably not have placed similar weighting to those features the human regards as critical for pattern processing. It was argued that since human pattern detection was being modeled, it would be desirable to incorporate into the evaluation process the human abilities to process patterns.

In general, the results suggest that the spot analogs are effective in detecting degraded patterns whereas the edge analogs are not. The edge analogs are, however, effective in summarizing the contour. One may wish to speculate that the counterparts of the spot and edge analogs in the mammalian visual system function in a similar manner. The spot detectors, at the retinal and geniculate levels, may function as preprocessors, detecting patterns from whatever form of degradation that may exist. The edge detectors, at the cortical level, may function as feature analyzers for some higher level process.

Several matters may be gainfully explored in subsequent research. It would be of interest to determine the model's performance with photographs which include a full range of gray levels. Also, research which compared the performance of the model with that of humans would be of value. The development of a methodology would be required which would restrict Ss

to using only those capabilities being modeled. In addition, a means for comparing the performance of the model with that of humans would need development.

The present model may be viewed as contributing to several forms of research. First, the model may serve as the basis for the development of an automated pattern processing procedure; its function would be to enhance degraded photographs. Second, models of its type may potentially offer some clues to the functioning of the visual system, which physiologists could pursue in their research. An exchange between physiologists and researchers interested in computer modeling may contribute to the understanding of the functioning of the visual system. Third, the present model may serve as the preprocessing phase of a general model of pattern recognition. Higher phases of processing would perform property selection operations, with increasing orders of abstraction, and the highest phases would perform the requisite decision operations.

REFERENCES

Evans, C. R. Further studies of pattern perception and a stabilized retinal image: The use of prolonged after-images to achieve perfect stabilization. British J. Psychol. 1967, 58, 315-327.

Evans, S. H., Hoffman, A. J. & Arnoult, M. D. VARGUS 6D: A simple system for producing "noisy" patterns (for IBM 1620 and IBM 1401). Behavioral Science, 1967, 12, 268.

Evans, S. H., Hoffman, A. J., Arnoult, M. D. and Zinser, O. Pattern enhancement with schematic operators. Behavioral Science, 1968, 13 (5), 402-404.

Evans, S. H. An overview of machine and human pattern recognition. In Proc. of planning conference: Pattern identification by man and machine. Texas Christian University. Dec., 1968. Tech. Mem., U.S. Army Human Engineering Laboratories, Aberdeen Proving Ground, Maryland.

Gilinsky, A. S. & Doherty, R. S. Interocular transfer of orientational effects. Science, 1969, 164, 454-455.

Gose, E. E. Introduction to biological and mechanical pattern recognition To appear in M. S. Watanabe (Ed.) Methodologies of pattern recognition. Academic Press, 1969.

Grimsdale, R. L., Sumner, F. H., Tunis, C. J. and Kilburn, T. A system for the automatic recognition of patterns. Proc. IEEE, 1959, 106, 210-221.

Hubel, D. H. The visual cortex of the brain. Scientific American, 1963, 209, 54-78.

Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture in two nonstriate visual areas (18 & 19) of the cat. Journal of Neurophysiology, 1965, 28, 229-289.

Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. J. Physiol. 1968, 195, 215-243.

Kahneman, D. Method, findings and theory in studies of visual masking. Psychol. Bull. 1968, 70(6), 404-425.

Michael, C. R. Retinal processing of visual images. Sci. American, 1969, 105-114.

Nagy, G. State of the art in pattern recognition. Proc. IEEE, 1968, 56, 836-864.

Pritchard, R. M. Stabilized images on the retina. Scient. American, 1961, 204(6), 72-78.

Rosen, C. A. Pattern classification by adaptive machines. Science, 1967, 156, 38-44.

Rosenfield, A. Picture processing by computer. Computing Surveys, 1969, 1(3), 147-176.

Uhr, L. Pattern recognition, In L. Uhr (Ed.) Pattern recognition. New York: John Wiley & Sons, 1966.

Weisstein, Naomi. What the frog's eye tells the human brain: single cell analyzers in the human visual system. Psychol. Bull., 1969, 72(3), 157-176.

Zinser, O., Evans, S. H., and Hoffman, A. J. Pattern enhancement with schematic operators. Paper presented at Southwestern Psychological Assn. Convention, April, 1968.

Zinser, O., Evans, S. H. Pattern detection and enhancement with schematic operators. Paper presented at Southwestern Psychological Assn. Convention, April, 1969.

Zinser, O. A computer model of the initial stages of mammalian pattern processing. Unpublished Doctoral Dissertation, Texas Christian University, 1970.

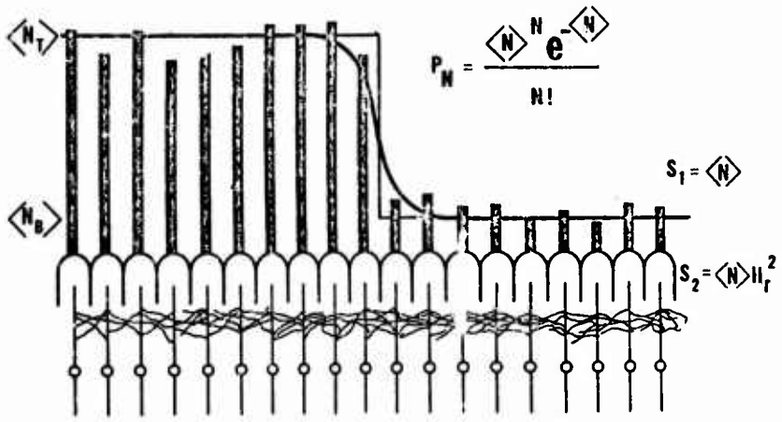
RESOLUTION AND NOISE LIMITATIONS OF NIGHT VISION DEVICES

Gerturde H. Kornfeld and Walter R. Lawson
Night Vision Laboratory
Visionics Technical Area
Fort Belvoir, Virginia

In preceding talks the detrimental effect of noise and degradation on pattern recognition was elaborated; we now show how reduction in picture quality is considered in the prediction of the performance of our night vision devices.

Theoretically it is possible to intensify the very weak signal from the night landscape to such an extent that it is projected with daylight intensity on the screen of our night vision devices. Is such strong intensification wise? This question is not simple to answer. If it is possible to keep noise and degradation on an acceptable level we have a better chance of detecting a target on the screen of daylight brightness than on a dimmer one. On the other hand, if the picture has inferior quality a compromise light level might be advantageous, even when the problem of weight and cost caused by additional intensification does not enter.

In this talk we outline a decision theoretical model used to predict the performance of our night vision equipment; with slight amendments it is based on a model we perfected for the prediction of the light level dependent performance of the human eye; we intend to illustrate the difference between high and low light level unaided vision first.



INTENSITY DISTRIBUTION ON THE RETINA
FIG. 1

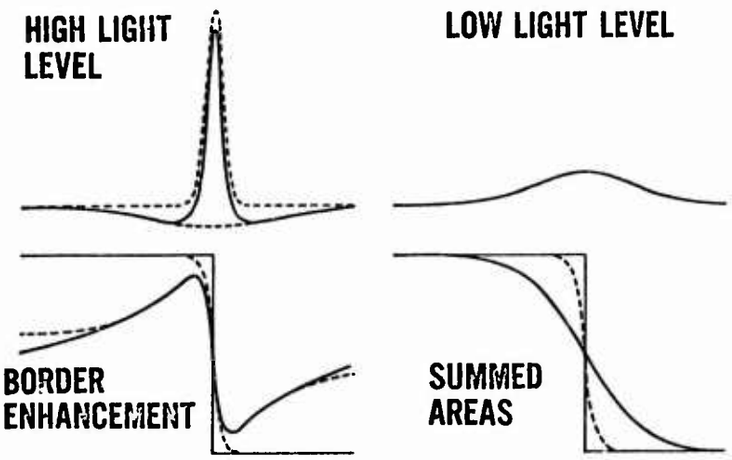
Figure 1 is a schematic drawing of the intensity distribution that reaches the retina. The thin line shows the ideal distribution having a step discontinuity between a background area of irradiance $\langle N_B \rangle$ and a target area of irradiance $\langle N_T \rangle$ photons per unit area. The actual distribution is, however, degraded by the eye's optics; the heavy line shows this distribution in the absence of any noise.

The irradiance reaching the retina is noisy, i.e. the irradiance at a point fluctuates randomly about some mean value. Noise is caused by the particle nature of light; the probability that N instead of the average value $\langle N \rangle$ photons reach a unit area per second from a Poisson distribution.* The spatial and temperal characteristics of noise, as well as its magnitude, can be characterized by its power spectrum which is the Fourier Transform of the autocorrelation function of the irradiance fluctuations. Noise is characterized as white, i.e. frequency independent, when intensity fluctuations at two arbitrarily close points are uncorrelated. The photon noise incident upon the retina is white with power spectrum $\langle N \rangle$. Evidently the noise in the pseudo image that reaches the brain is no longer white; the retinal image forms the pseudo image through signals transmitted by retinal receptors and neural fibers shown by the very thin lines in Figure 1; two points in the pseudo image corresponding to points within the same retinal receptor are perfectly correlated; even if we could approximate the retinal receptors as a point detector the flux fluctuations in the pseudo image would not be uncorrelated because the complicated neural connections between the detectors introduce image degradations similar to optical degradations, causing the flux distribution at neighboring points to be correlated. The power spectrum of the noise processed by retinal receptors and reaching the brain is $\langle N \rangle H_r^2$ where H_r is the effective transfer function (i.e. the Fourier transform of the effective spread function) of the retinal receptors and neural interconnections.

At very low light levels considerably less than one photon per second reaches a retinal receptor, but we are still able to see because neural connections are used for extensive summation; the low light level spread function, illustrated on the right side of Figure 2, is wide and borders are very degraded, but large areas can still be recognized as patches of light or darkness.

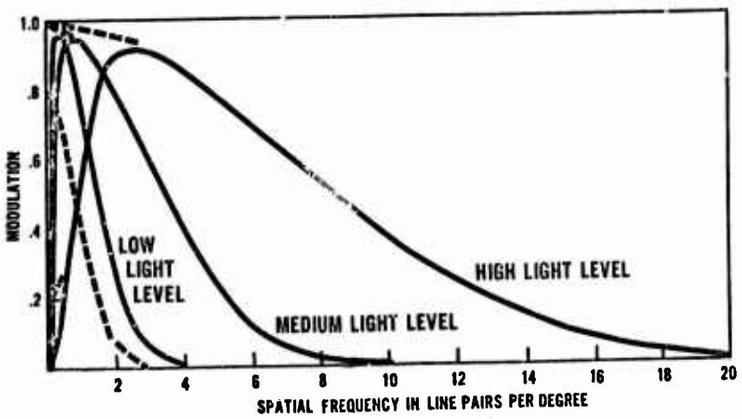
---

*A photon is a particle of light. We used the quantum mechanical notation for expectation value, the square brackets rather than a conventional notation for average because it is very clear on the slide; it does not imply a quantum mechanical treatment.



SPREAD FUNCTIONS
FIG. 2

At high light levels the neural connections are used to ensure maximal acuity; this is achieved by a positive excitation and negative inhibition spread function; the performance of the inhibition is to minimize the effect of the low frequency content of the image and can be compared to the contrast enhancement by filtering of low frequencies discussed in a different paper; inhibition is real and was experimentally demonstreated by the monitoring of electrical impulses from retinas; inhibition effects are noticeable in negative afterimages and some optical illusions and are extensively used by artists for border enhancement effects.



MODULATION TRANSFER FUNCTIONS OF THE EYE-BRAIN SYSTEM (GAUSSIAN APPROXIMATION)
FIG. 3

Figure 3 shows low, medium and high light level effective modulation transfer functions of the combined optical-neural system at three light levels. With an infinite field of view and a homogeneous retina the excitation part of the modulation transfer functions would be illustrated by the heavy lines in the high, but the dashed line approaching unity in the low frequency limit. The low frequency, high light level, discrepancy between a physiologically established transfer function shown by the dashed line (approaching .2) and the effective one shown by the solid lines (approaching 0) is due to the limited spatial sensitivity of the retina; the high light level inhibition spread function is illustrated by the dashed Gaussian curve with maximal amplitude of 0.8.

Experimentally the modulation transfer functions can be established from our contrast sensitivity as the inverse of the contrast necessary for a 50% detection probability of a target; in case the target is a sinusoidal wave pattern of infinite extent our contrast sensitivity would be proportional to our modulation transfer function of the combined optical-neural visual system, if our field of view were infinite and our retina of homogeneous sensitivity. In reality our limited field of view prevents us from observing extremely low spatial frequency sinusoidal wave patterns even if their contrast is very high. We define a phenomenological modulation transfer function that is proportional to our actual contrast sensitivity to sinusoidal wave patterns; it approaches zero in the zero spatial frequency limit.

$$P_D = \frac{1}{\sqrt{2\pi}} \int_{L-d}^{\infty} e^{-x^2/2}\, dx \quad \text{AND} \quad P_{fa} = \frac{1}{\sqrt{2\pi}} \int_{L}^{\infty} e^{-x^2/2}\, dx$$

WHERE

$$D^2 = \frac{1}{8\pi^3} \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega_x \int_{-\infty}^{\infty} d\omega_y \frac{I^2(\omega,\omega_x,\omega_y)}{S(\omega,\omega_x,\omega_y)}$$

We now give an outline of our binary decision theoretical criterium for the prediction of the detection probability P_D of a simple target, provided a false alarm probability P_{fa} is acceptable. The positive constant L establishes the acceptable false alarm probability; if L is zero or less, "wild guessing" is allowed while if L is of the order of 3 or more, false alarms are nearly eliminated. The quantity $I^2(\omega,\omega_x,\omega_y)$ is the square magnitude of the Fourier Transform of the degraded noise free signal irradiance distribution (target irradiance minus background irradiance) and $S(\omega,\omega_x,\omega_y)$ is the power spectrum of the noise. The quantity D, is a type of signal to noise ratio which yields the probability of detection for the given false alarm probability, according to binary decision theory, when the noise is Gaussian and additive and when some mild mathematical restrictions are imposed. For any reasonably large number <N> the noise closely approaches a Gaussian distribution so that the first condition applies.

Additive noise means that intensity fluctuations are assumed uniform over the image plane, this condition is never fully justified, but targets that are large and/or contrasty enough for the additive approximation to cause a serious error will be detected by a "half blind" man. No sophisticated decision theoretical model is necessary to predict the detection probability of targets when the additive noise approximation is invalid.

So we now elaborate the terms in the quantity D.

$$D^2 = \frac{C^2 t_e}{4\pi^2} \int_{\infty}^{\infty} d\omega_x \int_{-\infty}^{\infty} d\omega_y \frac{\langle N_B \rangle^2 Q^2(\omega,\omega_y) H_{op}^2(\omega_x,\omega_y) Hr^2(\omega_x,\omega_y)}{N' H_R^2(\omega_x,\omega_y) + N_D}$$

$$\text{where } N' = \beta \langle N_B \rangle^2 + \langle N_B \rangle$$

$$\text{and} \quad C = \left(\langle N_T \rangle - \langle N_B \rangle\right) / \langle N_B \rangle$$

In case the targets are stationary integration over the temporal frequency ω results in the glance time t_e. The term $Q^2(\omega_x,\omega_y)$ is the square magnitude of the Fourier transform of a target shape function, which for targets of uniform intensity if unity inside and zero outside the target boundaries. The terms $H_{op}^2(\omega_x,\omega_y)$ and $H_r^2(\omega_x,\omega_y)$ are the square magnitudes of the Fourier transforms of the optical and retinal spread functions respectively. The power spectrum of the noise consists of dark noise term N_D, and the photon induced noise N'. The constant β is very small; therefore, the term proportional to the square of the background intensity is only important at high light levels, where it describes saturation effects. At such light levels the signal to noise ratio approaches a limit which is independent of $\langle N_B \rangle$. At medium light levels the signal to noise ratio is proportional to the well known shot noise result $\langle N_B \rangle^{1/2}$. At very low light levels the dark noise term which describes random electrical impulses in our retinal receptors dominates and the signal to noise ratio is proportional to $\langle N_B \rangle$. The contrast, C, has the definition that is most advantageous to describe targets that are small with respect to the background area.

$$D^2 = \frac{a^2 b^2}{4\pi^2} \int_{-\infty}^{\infty} d\omega_x \int_{-\infty}^{\infty} d\omega_y \frac{\sin^2(\frac{1}{2}\omega_x a)}{(\frac{1}{2}\omega_x a)^2} \frac{\sin^2(\frac{1}{2}\omega_y b)}{(\frac{1}{2}\omega_y b)^2} \left[e^{-\frac{1}{2}\sigma^2(\omega_x^2+\omega_y^2)} - e^{-\frac{1}{2}\sigma^2 m^2(\omega_x^2+\omega_y^2)} \right]^2$$

WITH ASYMPTOTES

$$D^2 = \begin{cases} C^2 t_e \left(\frac{\langle N^2\rangle}{N' + N_D}\right) \frac{a^2 b^2}{4\pi\sigma^2} \left(1 + \frac{1}{m^2} - \frac{4}{m^2+1}\right) & \text{WHEN } a \ll \sigma\ b \ll \sigma \\ C^2 t_e \left(\frac{\langle N^2\rangle}{N' + N_D}\right) \frac{a^2 b}{2\sqrt{\pi}\,\sigma} \left(1 + \frac{1}{m} - \sqrt{\frac{8}{m^2+1}}\right) & \text{WHEN } a \ll \sigma\ b \gg m\sigma \\ \Sigma C^2 t_e \left(\frac{\langle N^2\rangle}{N' + N_D}\right) (a+b)\sigma\sqrt{m^2 - 1} & \text{WHEN } a \gg \sigma\ b \gg \sigma \end{cases}$$

These are asymptotic results for values of D of rectangular targets assuming white noise. The white noise approximation is consistent when phenomenological transfer functions are used. Closed form solutions are possible with Gaussian excitation and inhibition transfer functions; such an approximation was used by Shade[1]; it is a good one for high light levels; we elaborate on other spread functions for medium and low light levels elsewhere.[2] For the rectangle, $Q^2(\omega_x,\omega_y)$ is a sine function, that can be expanded into its Taylor series

$$a^2 \operatorname{sinc}^2[\tfrac{1}{2}\omega_x a] = 2\omega_x^{-2}[1-\cos(\omega_x a)] = 2\sum_{K=1}^{\infty} (-1)^K \frac{(\omega_x a)^{2K}}{2K!}$$

Term by term integration gives an absolutely convergent result; for very small rectangles the leading term is sufficient and the experimentally well established result that the detection probability is proportional to the target area emerges. Here σ and $m\sigma$ are half the resolution lengths of the excitation and inhibition spread functions respectively;* the factor m is of the order of eight. For large rectangles the series converges very slowly and a different expansion is advantageous

$$\left[\exp\left[-\tfrac{1}{2}\sigma^2(\omega_x^2+\omega_y^2)\right] - \exp\left[-\tfrac{1}{2} m^2\sigma^2(\omega_x^2+\omega_y^2)\right]\right]^2$$
$$= 2\exp\left[-\tfrac{1}{2}(m^2+1)\sigma^2(\omega_x^2+\omega_y^2)\right] \sum_{K=1}^{\infty}\sum_{e=0}^{K} (-1)^{K+1}\left[\tfrac{1}{2}\sigma^2\left[m^2-1\right]\right]^{2K} \frac{\omega_x^{4e}\omega_y^{(4K-4e)}}{(2K-2e)!\,2e!}$$

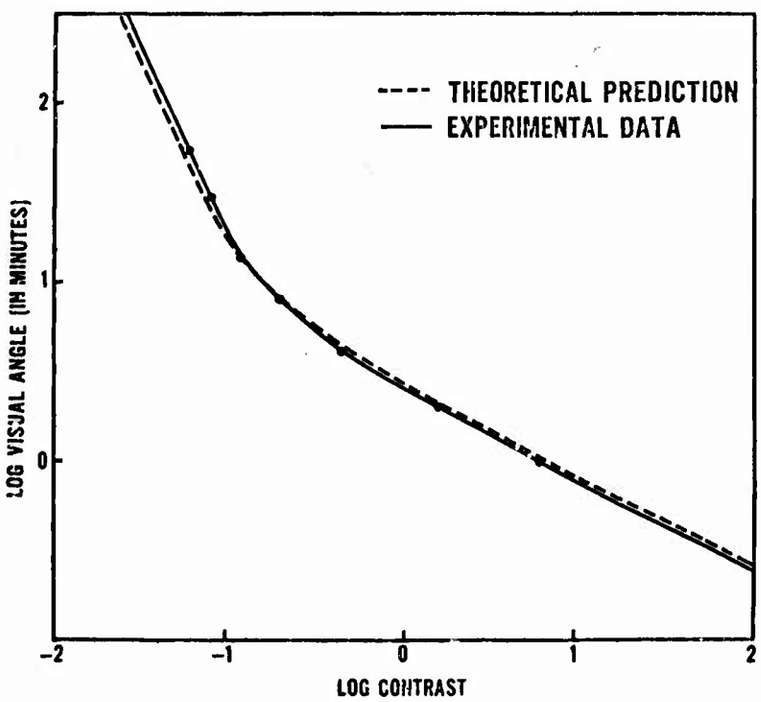
There are terms in the double expansion where e=k or e=0. These multiply the product of an exponential and a sine function; for large targets the sine function is so narrow that is multiplicative exponential can be approximated to unity in the main region of integration and a detection threshold that is proportional to the square root of the target circumference emerges. All terms where $0 < e < k$ are independent of the

---

* At a distance equal to the resolution length the Gaussain spread function is reduced to $e^{-1/2}$ its maximal value.

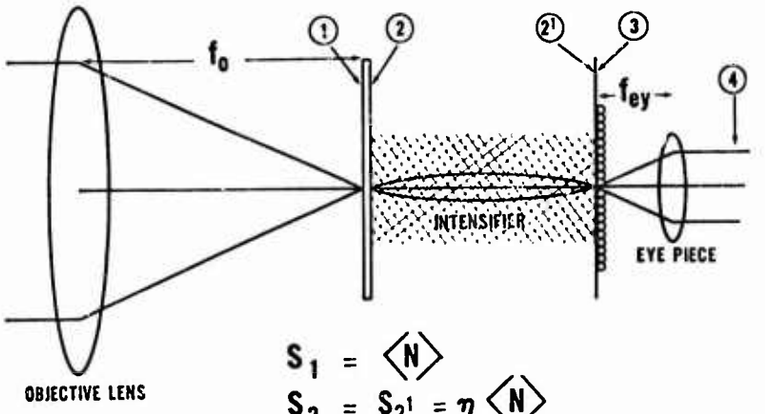
target size, or inversely proportional to target dimensions; they can be neglected for very large targets. The term Σ means the result of an infinite sum whose exact value depends on m but which is never far from unity.

Figure 4 shows the excellent agreement of our model with Blackwell's[3] short observation time thresholds for circular targets. Equally good results were achieved for rectangular targets also.



BLACKWELL'S 0.01 SEC. DETECTION TIME DATA.
FIG. 4

Having established that we can calculate detection thresholds in the absence of optical equipment, we now briefly examine the image intensifier case. Figure 5 is a schematic drawing of the intensifier tube. An objective lens of focal length f_o projects an image on a photocathode; under the impact of the photons the photocathode emits electrons with quantum efficiency η. The electron beam is focused onto



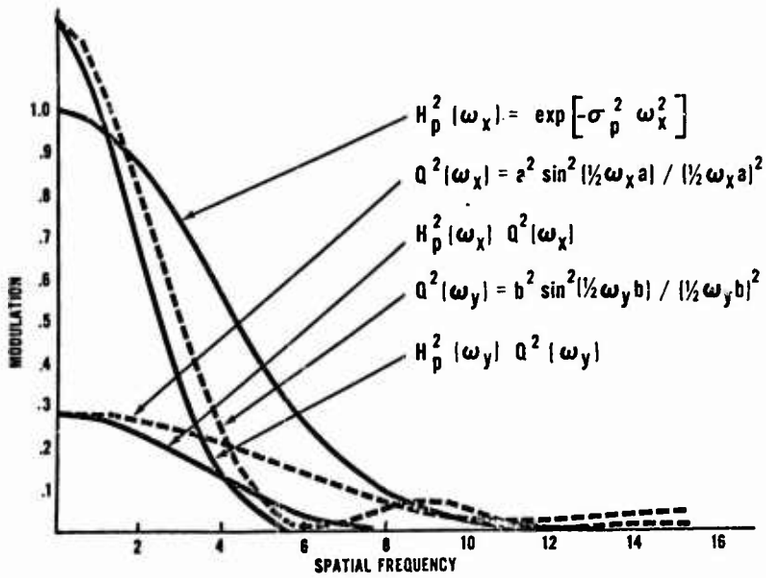
$$S_1 = \langle N \rangle$$
$$S_2 = S_{2'} = \eta \langle N \rangle$$
$$S_3 = \eta \langle N \rangle G^2 H_p^2 + \eta \langle N \rangle G$$
$$S_4 = \eta \langle N \rangle G^2 H_p^2 H_{ey}^2 + \eta \langle N \rangle G$$

SCHEMATIC DRAWING OF INTENSIFIER TUBE
FIG. 5

a phosphor which under the electron bombardment emits G photons. The picture on the phosphor is then viewed with an eyepiece of focal length f_{ey}. The main degradation takes place in the photoelectric stage. In other words, $H_p(\omega_x,\omega_y)$ is narrow with respect to the objective lens and eye piece transfer functions. The quantities S_1, S_2, S_3, and S_4 are the power spectra of the noise at the different stages. Unfortunately, the leading terms of the power spectra are proportional to G^2; therefore, the gain itself cannot lower the signal to noise ratio.

The detection criteria presented previously are directly applicable to image intensifiers. The noise introduced by the intensifier is added to that of the eye alone while the square magnitude of the transfer function of the device forms an additional factor in the numerator of the integrand used to determine D. Figure 6 shows a Gaussian approximation to the transfer function of an intensifier and the effect it has on target shapes functions of different size rectangle sides. Of course, in determining D, the eyeball noise and transfer function characteristics for output brightness of the intensifier must be used, and the magnification of the intensifier must be accounted for by appropriate "rescaling" of the spatial frequency of the signal and the various transfer functions. It is evident that the magnification of the intensifier is rather critical.

If this magnification is too small, the eye cannot detect much of the high frequency information contained in the signal. If the magnification is too great, those frequencies which are not seriously degraded by the intensifier might be presented to the eye in the region where the eye's frequency response is reduced due to inhibition effects. (Obviously, magnification will affect the search characteristics of intensifiers but this phenomena is not discussed in this paper.)



TARGET DEGRADATION BY INTENSIFIER TUBE
FIG. 6

Figure 7 is a sharp picture of a soldier holding a rifle to which an intensifier tube is mounted, and also shows a magnified detail. Our superior response to the high spatial frequency content of the image after its magnification can be noticed by observing the hair on the soldier's arms. Figure 8 shows the same entire scene and its detail, after equipment degradation was simulated on the negative. The degradation is much more objectionable in the presence of magnification.

As shown in Figure 5, the noise from image intensifiers is not white. Since the variance of the noise is given by

$$\sigma^2 = \int_{-\infty}^{\infty} d\omega_x \int_{-\infty}^{\infty} d\omega_y \int_{-\infty}^{\infty} d\omega\ S(\omega_x, \omega_y, \omega)$$

SHARP IMAGES

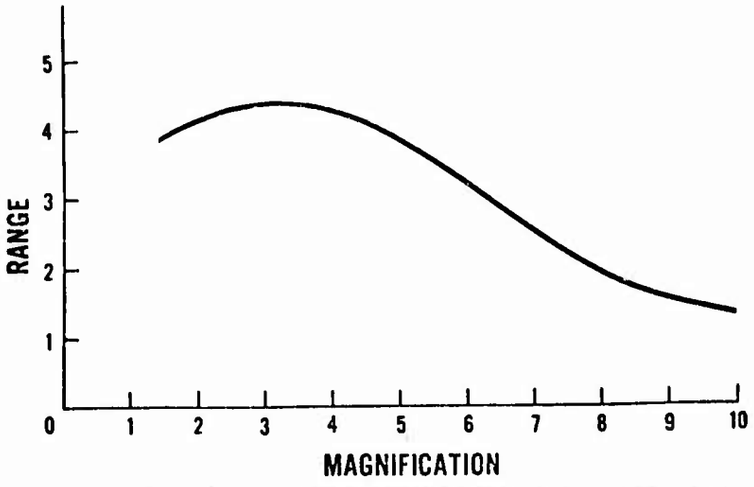
FIG 7

DEGRADED IMAGES

FIG 8

it is evident from the form of the equations for D, that noise having a frequency content much greater than the signal is less troublesome than a noise of equal variance having a frequency content roughly equal to that of the signal. Noise of high spatial frequency content is a common experience; it is not bothersome that TV pictures are made of dots provided these dots are not too large. The TV screen "snow" is most objectionable when its spatial content is lowest.

Figure 9 shows the result of a simple maximation. If the magnification of the intensifier is changed while all other significant variables are maintained constant (gain, transfer function referred to object space, signal to noise ratio) then range at which a man can be detected under starlight conditions varies as indicated in Figure 9. A maximum detection range occurs for a magnification of approximately 4. It is gratifying to point out that the actual magnification of the analyzed equipment is 4.



PREDICTION OF OPTIMUM MAGNIFICATION
FIG. 9

Most optimizations are difficult to perform because many parameters affect each other and it is not realistic to vary them individually. An important optimization is the gain; it cannot be altered without affecting transfer functions. An interesting property of our eye is its ability to observe the low frequency content of low light level images; high light levels are important for the detection of high spatial frequency content. If the high frequency content of the image is filtered out by the equipment, the advantage of high light levels disappears.

BIBLIOGRAPHY

[1] O. Shade, Sr., J Opt. Soc. Am 46 721-739 (1956)

[2] G. H. Kornfeld and W. R. Lawson submitted to J. Opt. Soc. Am.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* U.S. Army Research Office-Durham Box CM, Duke Station, Durhan, North Carolina 27706	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP N/a

3. REPORT TITLE

Proceedings of the 1970 Army Numerical Analysis Conference

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Interim Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*

6. REPORT DATE January 1971	7a. TOTAL NO. OF PAGES 329	7b. NO. OF REFS
8a. CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ARO-D Report 71-1 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*	

10. DISTRIBUTION STATEMENT

This document is subject to special export controls and each transmittal to foreign nationals may be made only with prior approval of the U. S. Army Research Office-Durham.

11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY Army Mathematics Steering Committee on behalf of the Office of the Chief of Research and Development

13. ABSTRACT

This is the technical report resulting from the 1970 Army Numerical Analysis Conference. It contains papers on the theory and foundations of pattern recognition, as well as papers on numerical analysis.

14. Key Words:

Pattern recognition
feature extraction
multi-image clustering
factor analysis
cubic spline subroutines
Kalman filters
strip charts
Mammalian pattern processing
night vision devices

DD FORM 1 NOV 68 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified
Security Classification