ADAPTIVE HYBRID PICTURE CODING

AFOSR-TR- 87 - 1652

INTERIM SCIENTIFIC REPORT

by

Richard A. Jones
Principal Investigator

Department of Electrical Engineering
University of Arkansas
Fayetteville, Arkansas 72701

November 30, 1986

Prepared for the

AIR FORCE OFFICE
OF SCIENTIFIC RESEARCH

87 11 3 388

Under Grant No. AFOSR 84-0322

AD-A187 586

DTIC
ELECTE
NOV 1 9 1987
S D
H

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
UNCLASSIFIED	

2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT
	Approved for public release, distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
	AFOSR·TR· 87·1652

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
University of Arkansas		AFOSR/NE

6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
Fayetteville Arkansas 72701	Bldg 410 Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING / SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
SAME AS 7A		AFOSR-84-0322

8c. ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
SAME AS 7B	61102 F	2305	B3	

11. TITLE (Include Security Classification)
Adaptive Hybrid Picture Coding

12. PERSONAL AUTHOR(S)
Dr Jones

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim Report	FROM 30 Sep85 TO 30Sep86		

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In order to recognize 3-dimensional objects in 2-dimensional scenes a shape description, complete enough to determine the 3-dimensional object, mus be recovered fromthe scene. Although it is possible to recover some 3-dimensional information, for instance shape from shade (21) or shape from texture (13), in general the information recovered will be a 2-dimensional representation of the object. Many methods are availabe for interpreting 3-dimensional objects in 2-dimensional arrayed range images (3), 20; however this is considered as a somewhat different problem, since depth information can be derived from the scene with little or no ambiguity.

UNCLASSIFIED

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	**UNCLASSIFIED**

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
DR C LEE GILES	(202) 767-4931	NE

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

INTERIM SCIENTIFIC REPORT

Grant No. 84-0322

ADAPTIVE HYBRID PICTURE CODING

Richard A. Jones
Principal Investigator

DTIC
COPY
INSPECTED
6

University of Arkansas
Department of Electrical Engineering
Fayetteville, Arkansas 72701

Accession For		
NTIS GRA&I	✓	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution/		
Availability Codes		
Dist	Avail and/or Special	
A-1		

30 November 1986

TABLE OF CONTENTS

PART ONE

RECOGNITION OF PARTIALLY OCCLUDED 3-DIMENSIONAL OBJECTS
USING INFORMATION FROM CONCAVITIES

I INTRODUCTION

In order to recognize 3-dimensional objects in 2-dimensional scenes a shape description, complete enough to determine the 3-dimensional object, must be recovered from the scene. Although it is possible to recover some 3-dimensional information, for instance shape from shade [21] or shape from texture [13], in general the information recovered will be a 2-dimensional representation of the object. Many methods are available for interpreting 3-dimensional objects in 2-dimensionally arrayed range images [3], [20]; however this is considered as a somewhat different problem, since depth information can be derived from the scene with little or no ambiguity. The representation recovered from regular 2dimensional intensity sceneswill consist of a mapping from an object-centered coordinate space to a viewer-centered coordinate space [2]. This mapping represents any reorientation of the 3-dimensional object within its coordinate system which results in a change with regard to the viewer.

Currently there are two approaches to forming an object description under these conditions. In the first approach [4], a 3-dimensional representation of all the objects to be recognized is stored. Then when a 2-dimensional scene is to be analyzed, the 3-dimensional model is transformed to obtain its 2-dimensional equivalent. The corresponding 2-dimensional versions are then matched against those extracted from the scene. The main difficulty with this approach is the amount of upfront processing required to compute an adequate number of transformations to required to compute an adequate number of transformations to represent the possible object views.

In the second approach, the transformed object representations are calculated a-priori and the corresponding 2-dimensional descriptions stored for the scene

analysis algorithm. Obviously the 2-dimensional descriptions must be compact enough so that a sufficient number of views can be stored without exceeding memory restrictions. Fourier descriptors [18],[19], and boundary curvature [26] have been used to describe 3dimensional objects in this manner however, they are inadequate when conditions of occlusion and boundary noise respectively are allowed. In this paper, a method for 3-dimensional object identificationsimilar to the two above but with much greater immunity to conditions of occlusion and boundary noise is presented. The method is believed to be applicable in many industrial and military systems where efficient and reliable identification is necessary.

In section II, the concepts of shape space and shape vector are introduced. The formalization of these concepts provides a basis upon which shape contours can be analyzed, even in the context of incomplete knowledge (occlusion). The underlying structure for constructing the shape vector is the critical point. In section III a new method for defining critical points is described, which in many cases offers improved performance versus those defined on the basis of curvature. In section IV a procedure for extracting these new critical points from descrete data is presented, and in section V the corresponding shape identification algorithm or cognitive stage is described. Finally, in section VI the procedures ability to recognize partially obscured, complex 3-dimensional objects in 2-dimensional scenes is demonstrated.

II SHAPE SPACE AND FEATURE VECTORS

In order to distinguish one shape from another, or a partial shape as a part of a whole shape, some mechanism which examines the relationship between shape measurements is needed. One such mechanism which is ideal for this application is a shape space [24]. By introducing a shape space defined on the measurements, relationships between these measurements can be quantified and thereby compared.

The basic element of shape space is the shape vector $_j^kZ$, where j indicates the jth set of measurements from the kth shape. If there are K shapes to be analyzed, then k is integer valued from 1 to K. The range of j depends on the total number of measurements possible on a given shape k. The shape vector $_j^kZ$ consists of a measurement vector, $_j^kM$, normalized by a scalar factor $_j^ks$ referred to as the size variable. Thus,

$$_j^kZ = {_j^kM} \, / \, {_j^ks} \qquad\qquad 1$$

The measurement vector consists of I measurements between predetermined points on the shape $_kS$, represented notationally by

$$_j^kM = (\, _j^km_1 \, , \, _j^km_2 \, , \, \dots \, , \, _j^km_i \, , \, \dots \, , \, _j^km_I \,) \qquad\qquad 2$$

Combining (1) and (2) yields the final form of the shape vector

$$_j^kZ \; = \; (\frac{_j^km_1}{_j^ks_1} \, , \frac{_j^km_2}{_j^ks_2} \, , \, \dots \, , \frac{_j^km_i}{_j^ks_i} \, , \, \dots \, , \frac{_j^km_I}{_j^ks_I} \,) \qquad\qquad 3$$

For notational purposes it is convenient to rewrite (3) as

$$_j^kZ = (\, _j^kz_1 \, , \, _j^kz_2 \, , \, \dots \, , \, _j^kz_i \, , \, \dots \, , \, _j^kz_I \,) \qquad\qquad 4$$

where each $_j^kz_i$ represents the ith measurement component normalized by the size variable. When constructing the shape vector two important properties must be preserved: 1) all measurements are made between a set C of predefined points, 2)

3

the shape vector is independent of the size variable. If these conditions are met, then any two shapes which have the same shape vector are said to be equal with respect to the measurements.

The Measurement Vector

The exact nature of the measurement vector is described next. Let each element of C be represented by the ordered pair $c_n = (x_n, y_n)$, where x_n and y_n are the spatial coordinates in R^2. The displacement vector between any critical point c_n and any reference point $c' = (x', y')$ is given as

$$c_n \pm c' = (x_n \pm x', y_n \pm y')$$ 5

the normal Euclidean definition for displacement in R^2. The resulting vector yields a displacement with respect to the origin. With this definition in mind, consider I arbitrary critical points which constitute the jth measurement vector formed from C. Now form a reference point, $_j\bar{c} = (_j\bar{x}, _j\bar{y})$, by calculating the centroid of the I points as

$$_j\bar{x}_c = \frac{1}{I} \sum_{i=1}^{I} {_jx_i} ; \qquad _j\bar{y}_c = \frac{1}{I} \sum_{i=1}^{I} {_jy_i}.$$ 6

Then form the components of the measurement vector by calculating the displacement between each critical point and the feature centroid. Mathematically this is given by

$$_jm_i = (_jx_i - _j\bar{x}_c, _jy_i - _j\bar{y}_c) ; \qquad i=1 \text{ to } I.$$ 7

This in effect translates all measurements to the origin of R^2, in other words, the

4

measurement vector is an I-tuple of vectors originating from the origin of R^2. The relationships described by (6) and (7) are demostrated graphically in figure 1.

Rotation, Translation and the Measurement Vector

It is a desirable goal to remove the effects of rotation, translation and scaling from the shape representation or to at least account for them. An examination of these effects for the first two transformations on the measurement vector is now discussed (the effects of scaling are closely related to the definition of the size variable and will be described later).

Since the rotation R_θ, of the shape by an angle θ is linear [10], the rotation of the arbitrary set $\{c_1, c_2, c_3, \ldots, c_I\}$ by an angle θ about the origin results in a rotation of the m_i's about the origin plus a displacement to the centroid of the rotated feature. This can be shown mathematically as

$$_jc_i' = R_\theta\left\{_jc_i\right\} = R_\theta\left\{_jm_i + _j\bar{c}\right\} = R_\theta\left\{_jm_i\right\} + R_\theta\left\{_j\bar{c}\right\} \quad 8$$

where the latter two rotations are about the origin. The latter displacement term in (8) will be factored out by the subtraction in (7). Thus the angle of rotation between the rotated measurement vector and the original measurement vector can be calculated for any component i as

$$_j\theta = \text{SIGN}(\theta)\left[\cos^{-1}\left\{\frac{(_jm_i, _jm_i')}{\|_jm_i\| \|_jm_i'\|}\right\}\right] \quad 9$$

where $(\bullet, \bullet)$ is the standard inner product and $\|\bullet\|$ is the standard norm for vectors

in R^2. From this it can be seen that the only difference between measurement vectors from the rotated set of critical points and the unrotated set is a rotation θ about the origin.

Similarly, a translation T, of the set $\{c_1, c_2, c_3, \dots, c_I\}$ by a vector c_t can be written as

$$_j c_i' = T\left\{_j c_i\right\} = T\left\{_j m_i +_j \bar{c}\right\} = _j m_i +_j c + c_t \qquad 10$$

However, $_j c + c_t$ is just a new centroid which will be factored out in (7) leaving only the original m_i's. Therefore, translation has no effect on the measurement vector.

The Size Variable and the Effects of Scaling

The second condition required for the formulation of the shape vector is that the shape vector must be independent of the size variable. This can be achieved by making the size variable a function of the measurement vector. One such fuction can be formulated as

$$_j s = \sum_{i=1}^{I} ||_j m_i || \qquad 11$$

This formulation can be interpretted as the spatial energy contained in the measurement vector. With this value as the size variable the sum of all spatial energy in any shape vector always equals 1.

As a result, the effects of scale change can be removed. As an example, consider the previously mentioned set of arbitrary points. If they are scaled by a factor α, their measurement vector will become

6

$$_jM' = (\alpha_j m_1, \alpha_j m_2, \alpha_j m_3, \alpha_j m_4, \alpha_j m_5, \alpha_j m_6) \qquad 12$$

which has total energy given by

$$E_t = \| \alpha \| \sum_{i=1}^{I} \| _j m_i \|. \qquad 13$$

This is equivalent to the size variable for $_jM$ multiplied by α. When the corresponding shape (feature) vector is calculated the scale factor α will be canceled by the size variable.

GENERALIZED DESCRIPTION FOR IRREGULAR 2-DIMENSIONAL SHAPES

Since initially only 2-dimensional shapes are considered the domain for a shape, called S, is restricted to R^2. The shape S consists of a bounded, simple, closed curve B [22] and the area enclosed by this curve, called the interior I. In other words, $S = B \cap I$. The shape description will consist of a finite set of measurements extracted from S. These measurements will be made between a set of points C, which are referred to as critical points. The set of critical points is a subset of B. In order to define this set C, another structure called a concavity tree must first be described.

According to Sklansky [24], the concavity tree is a structure for describing simply connected silhouettes. If the shape S is convex , then the concavity tree is trivially the shape S itself. In this work, it is assumed that the shape is non-convex, this will be the case in most applications where irregular objects are to be processed. The concavity tree is described for the shape S by the following

7

procedure.

1) Compute the convex hull of the shape S, called S^c. This is defined operationally by

$$S^c = C_h(S). \qquad\qquad 14$$

The convex hull is the smallest convex set which contains S.

2) Subtract the original shape S from its convex hull S^c. This is equivalent to finding the intersection of S^c with the complement of S ($\bar{S}$). The set formed by $S^c \cap \bar{S}$ is most likely a disconnected subset of S^c. Consider each connected element of this subset as a branch of the concavity tree (subshape). Denote each branch by S_i, which indicates the ith component of $S^c \cap \bar{S}$. Each S_i will have the same properties as S.

3) To find the next level of the concavity tree, replace S by S_i and S_i by S_{ij}. Repeat the above procedure for each level 1 branch of the concavity tree. The index j will denote the branches of the concavity tree at level 2.

4) To complete the concavity tree, continue with the above procedure progressing from $\{S_i\}$ to $\{S_{ij}\}$ to $\{S_{ijk}\}$ ect. until all residual subshapes from the second step are convex. The set of shapes $\{S_i, S_{ij}, S_{ijk}, ...\}$ for all i,j,k,... forms the concavity tree. Figure 2 shows an arbitrary shape S and its resulting concavity tree.

Since the concavity tree is based on the relationship between a shape (subshape) and its convex hull there is only one unique tree for each shape (subshape). However it is possible for this tree to have a countably infinite number of levels and an uncountable infinite number of branches, for example see Koch curves [65]. In this work the class of shapes examined will be resticted to those which are sufficiently smooth as to have only a finite number of levels and branches, and thereby a finite number of critical points. This restiction still includes all shapes found in real applications and all but the most abstract found in any application.

Now that the concavity tree has been described, the definition for the critical

points can be considered. Let the set C represent the critical points for a shape S and $\Psi(\bullet)$ the operation which yields C from S,

$$C = \Psi(S). \qquad 15$$

The set C is equivalent to the union of all critical points extracted from all branches of the concavity tree of S, i.e.,

$$C = \bigcup_{\substack{\text{all levels} \\ n}} C_n \qquad 16$$

where,

$$C_1 = \bigcup_{\substack{\text{all branches} \\ i}} \Psi(S_i),$$

$$C_2 = \bigcup_{\substack{\text{all branches} \\ j}} \Psi(S_{ij}),$$

$$C_3 = \bigcup_{\substack{\text{all branches} \\ k}} \Psi(S_{ijk}),$$

$$\text{etc...} \qquad 17$$

The operation $\Psi(\bullet)$ for any shape (subshape) must now be described. Consider any branch of the concavity tree $(S_{ijk...})$ and find its intersection with the boundary B, so that

$$B_{ijk...} = B \cap S_{ijk...} \qquad 18$$

This intersection defines a line segment for each branch of the concavity tree. The

collection of all such line segments formed at each level of the tree is the basis for the critical point sets C_n. The critical points are finally defined as the endpoints of the $B_{ijk...}$'s. Figure 3 illustrates the boundary segments and critical points for the first level concavities of the shape in figure 2. Finally, the operation $\Psi(\bullet)$ can be summed up by the following hierachy of point set relationships.

$$C = \Psi(S),$$

$$= \bigcup_{\substack{\text{all levels} \\ n}} C_n,$$

$$= \bigcup_{\substack{\text{all levels} \\ n}} \left\{ \bigcup_{\substack{\text{all branches} \\ m}} \Psi(S_{ijk...m}) \right\},$$

$$= \bigcup_{\substack{\text{all levels} \\ n}} \left\{ \bigcup_{\substack{\text{all branches} \\ m}} \left\{ \text{endpoints} [B_{ijk...m}] \right\} \right\},$$

$$= \bigcup_{\substack{\text{all levels} \\ n}} \left\{ \bigcup_{\substack{\text{all branches} \\ m}} \left\{ \text{endpoints} [B \cap S_{ijk...m}] \right\} \right\}. \qquad 19$$

These relationships, between all allowable shapes S with their boundaries and their concavity trees, are unique and well defined. Therefore, the relationships between any shape S and its corresponding critical point set C are also unique and well defined.

Although these relationships are unique and well defined, their exists two special cases which can lead to some ambiguity if not handled properly. The first case occurs when two branches from the same level of the concavity tree, $S_{ij...mn}$ and $S_{ij...mn}'$, have the following property:

$$S_{ij...mn} \cap S_{ij...mn}' = \text{a single point} \qquad 20$$

but the intersection of the open sets formed by $S_{ij...m} - S_{ij...m}'$ is empty. When this happens the subsequent boundary segments associated with the two branches of the concavity tree, $S_{ij...mn}$ and $S_{ij...mn}'$, will in fact be a single line segment. To remove any ambiguity, the single point described in (20) will be used to partition this line segment into two seperate ones. This single point will be called "critical" and will be considered as two seperate points, each with singular associations to $S_{ij...mn}$ and $S_{ij...mn}'$ respectively.

The second case is similar to the first except involves an overlapping of critical points from different levels of the concavity tree. In other words, the sets C_n are not mutually exclusive. This will not be a problem as long as the membership information within the C_n's is retained along with spatial information in R^2. Consequently, the location of critical points within R^2 and the knowledge of the levels and branches of the concavity tree from which they originated are of equal importance.

As a rule of thumb, each branch of the concavity tree will generate two critical points. Each of these critical points can be represented by a triplet (x,y,l), where x and y represent the location of the points in R^2 and l indicates the level of the concavity tree at which the points were generated; x and y are real valued numbers; l is integer valued.

Since initially, only shapes which are invariant in R^2 are allowed three transformations must be considered. The three transformations - scaling, rotation and translation - are all well known mappings from R^2 to R^2, and are represented symbolically by $S_\alpha(S)$, $R_\theta(S)$ and $T(S)$ respectively. They are defined mathematically in the following. The functions $S_\alpha(S)$, $R_\theta(S)$, $T(S)$ map a shape S

into a shape S' such that all points (x,y), elements of S are mapped to points in S'
by

$$S_\alpha \left\{ (x,y) \right\} = (\alpha x , \alpha y) = (x' , y'), \qquad\qquad 21$$

$$R_\theta \left\{ (x,y) \right\} = (x \cos\theta - y \sin\theta , x \sin\theta + y \cos\theta) = (x' , y'), \qquad 22$$

$$T \left\{ (x,y) \right\} = (x + x_t , y + y_t) = (x' , y'), \qquad\qquad 23$$

where α, θ, x_t, $y_t \in R$, $(x,y) \in S \subset R^2$, $(x',y') \in S' \subset R^2$.

If α is restricted to being positive and finite, θ to being greater than zero and less
than two pi, and x_t, y_t to being finite, then all three mappings are not only onto but
one-to-one [10].

Using the definitions for the transforms of scale, rotation and translation the
following theorem can be formulated.

Theorem: the set of critical points C obtained for an arbitrary shape S which has
undergone one of the transformations S_α, R_θ or T is identical to the set of critical
points for the untransformed shape, subjected to the same transform. This can be
stated mathematically as:

$$\Psi (S_\alpha (S)) = S_\alpha (\Psi (S)),$$

$$\Psi (R_\theta (S)) = R_\theta (\Psi (S)),$$

$$\Psi (T (S)) = T (\Psi (S)). \qquad\qquad 24$$

The proof of this theorem appears in the appendix.

In general compositions of $S_{\alpha}(\bullet)$, $R_{\theta}(\bullet)$ and $T(\bullet)$ are not commutative because of the nonlinearity of $T(\bullet)$; therefore their order must be preserved. However, combinations involving only $S_{\alpha}(\bullet)$ and $R_{\theta}(\bullet)$ are commutative [10].

The next question to be addressed involves the invariance of critical point determination when only partial shape information is available. this problem can be divided into two simpler and similar type problems, namely, truncated shape representation and composite shape representation. In the first case, it is desirable to determine a source shape S from a given shape S' where S'⊂S and B'∩B=∅. In the second case, it is desirable to determine two or more source shapes $_iS$ from S' where S'= $\bigcup_{all\ i}$ $_iS$ and B'∩$_iB$=∅ for all i.

In order to accomplish either of the above task it is necessary for a subset of critical points from each source shape to remain unaltered in S'. Since the critical points are fundamentally linked to the concavity tree of the source shape, this is equivalent to requiring that a portion of the concavity tree remain unaltered.

In the case of truncation, the portion of the concavity tree which is distorted can be identified by examining the convex hulls of the whole and partial shapes, or S and S' respectively. Those subshapes or branches of the concavity tree which do not intersect a portion of $(S^c-S^{c'})$ remain unaltered. An example of this relationship is illustrated in figure 4 for an arbitrary shape S and subshape S'. From this it can be seen that each branch of the concavity tree acts as a generating node for all subsequent branches in that portion of the tree. Any node which remains unaltered will generate sub-level branches identical to those of the original tree. A comparison between the branches of the original shape S from figure 4 and its corresponding truncated shape S' is shown in figure 5. The branches inscribed by squares are

13

identical in both trees.

In the case of composite shapes, the effects are more complicated. Not only are branches of the original tree truncated, but new branches may be created and others augmented. This occurs because 1) the union of the convex hulls of the combinig shape is not necessarily convex, thereby creating new concavities and 2) concavities from one shape may overlay interior areas of another. These possibilities are illustrated in figure 6 for a composition of the shape used previously in figure 4.

In general any branch of the concavity tree which intersects a portion of the overlap region or intersects a concavity which includes the difference between $S^{c'}$ and $\underset{\text{all } i}{\cup} {}_iS^c$ will be distorted. In certain cases it is possible for certain branches, a subset of the distorted portion of the tree, to remain unaltered. However the conditions under which this occurs are complex and specific to the particular shape.

In the next section, the gap between the theoretical shape description developed in the previous section and a discrete implementation for digitized shapes is bridged. The effects of finite boundary representation, quantization noise and additive observation noise are also considered.

III CRITICAL POINTS FROM REAL DATA

The definition of critical points advanced in the previous section can be restated as follows: critical points are the endpoints of the line segments formed by the intersection of a shape's boundary B with the branches (subshapes) of its concavity tree. If the shape boundary is continuous and can be observed without error, then these critical points can be exactly determined for all combinations of scaling, rotation and translation. However, in the discrete case the shape is usually represented by a finite, integer-valued sequence. This sequence is almost always in

14

the form of a chain code or a thinned boundary code [12]. In this work it is assumed that the sampled boundary has been properly preprocessed so that it meets the discrete definitions of being closed and single-connected. These definitions are described in [22] along with various techniques for processing images which guarantee that these conditions are met.

With these requirements in mind, let the processed shape be represented by the sequence $\{x_i, y_i\}_{i=1}^{N}$, which corresponds to the quantized estimate of the shape boundary. The coordinates of the sequence, x_i and y_i, should be eight-connected and integer-valued. Because only boundary information is available the previous definition for critical points will have to be reinterpreted.

The general procedure for extracting critical points is illustrated by the flow chart in figure 7. At the front end of the system is a spatial filter which is chosen to minimize the effects of the quantization and observation noises. At the next stage, the convex hull information about the shape is obtained. A decision process is then required to identify and confirm critical points. Depending on the output of the decision stage, the shape is either segmented into the next level of branches of its concavity tree or the procedure is terminated. This procedure is recursive, with each pass around the loop corresponding to the processing of one level of the concavity tree.

The purpose of the spatial filter is to reduce the effects of the quantization noise and the additive noise in the boundary sequence. The type of filter chosen should be optimized for the statistics of the noise processes and the source shapes [1]. In general these statistics depend on the imaging equipment and the set of source shapes, and cannot be determined globally. In fact, they are most likely nonstationary random processes. On the other hand, if the filter is made robust, good results can still be obtained. Therefore, an averaging filter is chosen.

The output of the filter will be a real-valued sequence $\{x^a, y^a\}_{i=1}^{N}$, with the same

15

number of elements as the input sequence. By increasing the length of the filter window the effects of the noise can be made arbitrarily small. However, the smoothing effects may cause smaller features of the shape to be eliminated. In many cases, this may be considered advantageous because a simultaneous reduction in feature processing and increased noise immunity is obtained.

The next step is to find those points from the sequence $\{x^a_i, y^a_i\}_{i=1}^N$ which are elements of the boundary of the convex hull of the digitized version of S. This boundary is equivalent to the Minimum Perimeter Convex Polygon (MPCP) enclosing S. It can be calculated using a modified version of the routine described by Freeman and Shapira [7]. This modified version is adapted to process real data instead of integer data. It requires a longer running time but still iterates to a final solution in all cases. The actual convex hull is not required, only the perpendicular distance from each point of the shape boundary to the MCPC. This distance is referred to as the difference sequence $\{d_i\}_{i=1}^N$.

The procedure for determining the critical points from the difference sequence is now described. As a first step, an interpretation of the definition of critical points must be applied in the context of the difference sequence. Since the difference sequence measures the distance between the boundary of S and the boundary of S^c, it can be used to find the intersection defined in (18). This intersection for the discrete case corresponds to those points of $\{d_i\}_{i=1}^N$ which equal zero. The endpoints (critical points) of the associated line segments are just those points where the difference sequence goes from a zero value to a nonzero value. If the shape boundary is observable without error, then the critical points can be determined exactly; however, this is not the case when applied to any discrete data. Therefore, the

16

critical points will be determined by finding certain transition regions in the difference sequence. This will compensate for most of the errors due to observation noise as well as errors incured by the resampling of shapes which have undergone one or more of the transformations discussed previously. Of these t! ~ee, only the transformation of scale seriously effects the difference sequence (the maximum error resulting from the other two transformations is limited to $2/\sqrt{2}$ times the quantization grid size). Since the spatial energy contained in the shape is directly proportional to the area of the quantization grid which it covers, any reduction in scale by a factor α results in a simultaneous reduction in signal to noise ratio of α^2. Because of this, certain compromises must be made in the performance of the critical point extraction algorithm, specifically, it will be designed to operate over a limited range of scale factors. To be more precise, the resulting procedure tries to predict the behavior of the critical points over the design range of scale factors. This reduces to a more basic problem of distinguishing concavities formed by the noise processes from those concavities due to the shape structure. The exact procedure for this is described next.

The input shape for extracting critical points for the shape dictionary is evaluated at α equal to one (the maximum scale factor). This corresponds to the highest signal to noise ratio expected for the shape. The first step is to isolate a set of potential critical points. This is accomplished by comparing the difference sequence to a threshold, τ_1. When the absolute value of the difference sequence is greater than τ_1 a concavity is formed. The points inside the interval $[0, \tau_1]$ closest to the position where the τ_1 threshold is crossed are considered as potential critical points and added to the critical point list. These points are in direct correspondence with the endpoints of the line segments in (18). They determine certain subsequences of $\{d_i\}_{i=1}^N$ called concavity sequences. The concavity sequences $\{d_i^c\}$ are the sequential portions of $\{d_i\}_{i=1}^N$ which are greater than τ_1. Associated with each concavity

17

sequence are a pair of critical points, one at each end. The first threshold simply identifies possible critical points. A second threshold is used to evaluate the most probable underlying cause of a concavity.

The second threshold τ_2, which is greater than τ_1, is used to eliminate critical points associated with small concavitities from the list. It also helps remove most concavities which are formed strictly from observation noise. For each sequence of points within a concavity, $\{d_i^C\}$, the maximum absolute value for the d_i^C's are found.

If this value, called d_{max}^C, is less than τ_2 then the critical points associated with the concavity in question are eliminated from the critical points list. Since it is impossible to distinguish small shape concavities from those produced by noise with just one sample, the alternative is to eliminate them all. The reasoning being that the total number of shape concavities eliminated will be small compared to the number of noise produced concavities. This can be justified theoretically by assuming that the average maximum depth of the actual shape concavities is much greater than the maximum depth of the noise concavities. This is illustrated in figure 8 for a generalized representative shape. This type of decision is analogous to the classical maximum-likelihood decision criteria [16]. Note that the depth of the true concavities can be considered as the signal energy, and the depth of the noise related concavities as the noise energy. From figure 8 it can be seen that for appropriate thresholds τ_1 and τ_2 most of the effects of the noise can be eliminated, provided the signal to noise ratio is sufficient. The matching algorithm, to be described later in this chapter, is designed to be very tolerant of missed critical points as long as they are listed in the dictionary. However, it is susceptible to error if unexpected critical points are found in the problem text. Therefore, the threshold τ_2 is biased by an amount τ_b when the shape dictionary is being prepared. The direction of the bias is in the direction of higher false alarm rates for concavities due to noise.

Now that the majority of potential critical points have been isolated, it is

18

necessary to find a method of predicting their behavior in the problem text over the design range of scale factors. The effects of changes in α on the $\{d_i^C\}$'s falls into two categories: 1) concavities found at the maximum value of α may collapse below the τ_2 threshold and their associated critical points eliminated from the problem text; 2) segments of the difference sequence, which are outside the interval $[0,\tau_2]$ for large α, may cross the τ_2 threshold and create new critical points. Of these two possibilities, the first case has been observed to happen most frequently. This can be explained as follows. Because the depth of a concavity is determined from two extremal points of opposite direction, namely, the outer most points on the shape boundary contained in the convex hull and the inner most shape points corresponding to the maximum depth of a concavity, the smoothing effect of the spatial filter converges these two points approximately twice as fast as distances based on only one of the extremal points, i.e. case two.

In order to predict the occurrence of these two changes, two additional thresholds are introduced. Their purpose is to assign a measure to regions of the difference sequence which may cause one of the above changes. This measure, called the confidence number, will be assigned individually to each critical point in the shape dictionary. The confidence number indicates the percentage of the scale factor range in which a particular critical point is detectable as well as the order in which it enters (or leaves) the problem text.

The first of these two thresholds, τ_3, is used to identify candidate critical points which may be detected in the problem text for smaller values of α. If any sequential portion of $\{d_i^C\}$ for any concavity has the property that

$$\tau_2 < |d_i^C| \leq \tau_3 \qquad\qquad 25$$

for all d_i^C in that group, then a potential critical point is indicated. The two points

19

where the sequence $\{d_i^c\}$ indicated by (25) crosses into the interval $[\tau_2, \tau_3]$ are added to the shape dictionary and assigned a confidence number, $\#_c$, given by

$$\#_c = \frac{|\tau_3 - d_{min}^c|}{|\tau_2 - \tau_3|} \qquad 26$$

where d_{min}^c is the minimum absolute value of the difference sequence in region $[\tau_2, \tau_3]$. An example of this situation is illustrated in part a of figure 9.

The second new threshold, τ_4, is used to measure the proportion of scale factors at which a particular concavity, and its associated critical points, can be detected. For each concavity sequence with d_{max}^c greater than τ_2 assign a confidence number according to the following:

if $\quad d_{max}^c > \tau_4, \qquad$ then $\qquad \#_c = 1;$

else $\qquad\qquad\qquad \#_c = \frac{|d_{max}^c - \tau_2|}{|\tau_4 - \tau_3|} \qquad 27$

This situation is illustrated in part b of figure 9. Note that τ_4 is generally twice as large as τ_3 and that $\tau_3 > \tau_2 > \tau_1$.

At this point the algorithm for extracting critical points and assignment of confidence numbers for a single level of the concavity tree has been described. To extend the procedure to the rest of the concavity tree, the recursive nature of the algorithm must be defined. Since the concavity sequences correspond to the next level of concavity tree branches, they can be processed individually in the same manner as the original shape. The only modification required is the reversal of the value for s in (9) at each new level. New critical points are added to the shape dictionary as previously described. Confidence numbers are assigned using the same criteria as before with the following exception: concavities which have confidence numbers less

20

than 1, have all sublevel critical points assigned a confidence number calculated as a product of the current level confidence numbers with the values determined according to (26) and (27). The process is propogated through the concavity tree until no new points are found. Figure 10 and 11 show the difference sequence and critical points respectively from an arbitrary shape for levels 1, 2 and 3 of the concavity tree respectively.The resulting dictionary page contains the following information extracted sequentially from the shape boundary for each critical point: 1) its x and y coordinate values, 2) its confidence number, 3) the level of the concavity tree at which it was found.

Shape Dictionary Calculation for 3-Dimensional Objects

Consider an arbitrary, rigid object as shown in figure 12, where θ_x, θ_y and θ_z describe angles of rotation in the three axes of the rectangular, object-centered coordinate system. The z-axis of the object-centered coordinate system is restricted to parallel alignments with the z'-axis of the viewer-centered coordinate system. This restriction results in the following interpretations for motion attributed to the object. The imaging plane shown in figure 12 is the space in which the object is projected

1) Translation of the object (or the object-centered coordinate system) with respect to the z'-axis results in a scale change in the imaging plane.

2) Translation of the object (or object-centered coordinate system) along either or both of the x' and y' axes results in a translation in the imaging plane.

3) Rotation of the object by an angle θ_z about the z-axis of the object-centered coordinate system results in a rotation by angle θ_z in the imaging plane.

4) Rotation by the angles θ_x and θ_y in the x and y axis of the object-centered coordinate system results in a new projection of the 3-dimensional object in

21

the imaging plane

The first three transformations are the same transformations which have been shown not to effect the critical point/feature vector formation procedure. Therefore, only the fourth transformation will necessitate a modification in the procedure previously described for 2-dimensional shape recognition.

The necessary modifications will result in an expanded shape dictionary. Now instead of one set of critical points for each shape, there will be several sets; each representing a different rotation about object-centered axes x and y. Therefore the following procedure is suggested for assembling the 3-dimensional shape dictionary.

1) With the object at its minimum distance from the z' plane (maximum scale factor) obtain a projection of the object in the viewer imaging plane with θ_x and θ_y at zero degrees.

2) Next, obtain the outer contour of the projected shape and extract its critical points using the procedure detailed earlier. Store as the first dictionary page for this shape.

3) Increment θ_x by θ_{inc} (a divisor of 180^0) and repeat steps 2 and 3 to obtain the next page in the dictionary for this shape. Continue incrementing θ_x until a rotation of 180^0 has been reached.

4) Reset θ_x and increment θ_y by θ_{inc}. Keeping θ_y fixed at the new value, repeat steps 1 thru 3 until θ_y has been incremented to 180^0.

The shape dictionary will now contain $(180^0 / \theta_{inc})^2$ views of the 3-dimensional object. The other views (for θ_x and θ_y between 180^0 and 360^0) are mirror images of those already calculated, and may be extracted from the existing dictionary by processing the critical points in reverse order.

22

The choice of θ_{inc} directly effects the accuracy and completeness with which the 3-dimensional object is described as well as the amount of memory needed (as θ_{inc} is reduced the accuracy of the description is increased along with the amount of memory required to store the dictionary). With regard to the accuracy of the 2-dimensional shape description, an unintended but benificial side effect is obtained from the extraction of the critical point's using the concavity tree. As the 3-dimensional object is rotated, various sections of the object become visible. The projections to these sections onto the imaging plane manifest themselves as new concavities in the corresponding 2-dimensional silhouette. Depending on the direction of rotation, these new concavities are either entering or leaving the concavity tree. Since this is the essence of the information measured by the confidence number, additional accuracy is gained. This is especially useful when the actual angle of rotation for the problem text is not a multiple of θ_{inc}.

Critical Point Determination for the Problem Text

When the boundary of an unknown shape is obtained a critical point list (problem text) must be configured for it. The procedure by which this is done is quite similar to the one described in the previous section; however, no confidence number is calculated. This is due to the fact that an exact value of the scale factor α is unknown. As a consequence only those portions of the difference sequence partitioned by the thresholds τ_1 and τ_2 are used to determine concavities. These concavities may be formed in three ways which are illustrated in figure 13, parts a, b and c. In the first case the τ_1 and τ_2 thresholds are both crossed in forming a concavity; for this case, the critical points for this concavity are chosen as the points nearest to threshold τ_1 inside the interval $[0,\tau_1]$. In the second case only the τ_2 threshold is crossed; therefore, the nearest points to τ_2 inside the interval $[0,\tau_2]$ are chosen as

critical. In the last case, only threshold τ_1 is crossed; therefore, no critical points are chosen since the concavity is considered too small to register. As a final result the problem text contains the following information extracted sequentially from the unknown shape's boundary for each critical point: 1) its x and y coordinate values, 2) the level of the concavity tree at which it was detected.

IV COGNITIVE STAGE

Once a problem text has been defined it must be matched against all the dictionary pages (shapes). Since the algorithm is required to recognize partial shapes as a part of a larger shape, features must be formed and matched accordingly. The basis for comparison is the shape or feature vector described previously. It consists of measurements made from I arbitrary critical points selected from the shape dictionary or problem text. As a first step, the process by which each feature vector is selected must be made non-arbitrary. Four general conditions are suggested for choosing the individual critical points; they are as follows:

1) The set of all feature vectors should give near equal representation to
   all parts of the shape.

2) The total number of feature vectors is kept as small as possible.

3) Each feature vector represents a feature likely to appear under certain
   conditions of translation, scale or rotation.

4) Each feature represents a localized characteristic of the shape.

The conditons for the representative feature set can be met by combining the concavity tree information with the confidence number information contained in the shape dictionary. The procedure for this is as follows. Let each pair of critical

24

points associated with a given concavity act as seed points for a feature vector. Choose the two most reliable neighboring critical points to each side of the seed points to complete the sixtuple. This method guarantees at least one feature vector for each branch of the concavity tree. Furthermore, it restricts each feature to three adjacent concavities for a given scale factor. Finally, a linear growth rate is achieved - if there are N critical points then there will be N/2 feature vectors. The method for selecting the most reliable, neighboring critical points is described next. Recall that the confidence number not only indicates the percentage of scale factors for which a critical point is detectable but also the order in which they are detected. For example, if $\#_a$ and $\#_{b'}$ represent the confidence numbers for two critical point pairs - a and b respectively, and $\#_a > \#_b$; then if pair b is detected, pair a should also be detected. Therefore, for a given critical point pair with confidence number $\#_c$ choose the two nearest neighboring critical points which have confidence numbers greater than or equal to $\#_c$. When considering the problem text, information concerning scaling is unavailable; therefore, the simplest scheme possible is derived, i.e, each consecutive grouping of six critical points forms a feature vector.

Let the feature vectors from the shape dictionary be denoted by $^{d_k}_m Z$ where k indicates the page of the dictionary and m the mth feature vector within page k. Similarly denote the nth feature vector of the problem text by $^p_n Z$.

Step 1) Associated with each feature vector of the dictionary and the problem text is a sixtuple called the level code vector. Each component of the level code vector is integer valued and indicates the level of the concavity tree at which the corresponding critical point was detected. The symbol for the level code vector associated with feature vector $^{d_k}_m Z$ ($^p_n Z$) is $^{d_k}_m l$ ($^p_n l$). For feature vectors to match it is necessary that their level code vectors also match. Therefore, in the first stage of the procedure all of the level code vectors from the shape dictionary page and the problem text are cross matched. The distance defined by

25

$$d_1(m,n) = \sum_{i=1}^{I} | \,{}_m^{d_k}1_i - {}_n^{p}1_i \,|$$ 28

is the basis for each level code vector comparison. A value of zero for $d_1(m,n)$ is a necessary condition for two feature vectors to match. This distance is calculated for all $m \in M_k$ and all $n \in N$. This cross referenced set is referred to as NXM_k. Each pair (m,n) such that $d_1(m,n)=0$, is added to the reduced feature set called NXM_k'. Since $d_1(m,n)$ requires only $(2*I-1)$ integer operations, compared to $(21*I-2)$ real valued operations for a complete feature vector comparison, computational overhead is significantly reduced.

Step 2) For each pair of features associated with the elements of the reduced set NXM_k' calculate a pair of feature vectors using the definitions methods outlined in chapter III. Next evaluate the distance $d_i(m,n)$ defined by

$$d_i(m,n) = \left| \, \| \,{}_m^{d_k}z_i \, \| - \| \,{}_n^{p}z_i \, \| \, \right|$$ 29

for each component of the feature vector pair. In the noiseless case the distance $d_i(m,n)$ is zero for all components of the feature vectors m and n if they match. Since this is never the case a finite positive distance τ_m must be allowed when evaluating (29). Therefore, if $d_i(m,n) < \tau_m$ for $i = 1$ to I, then the feature vectors are said to match. The set of matching feature vector pairs from this stage form the new set NXM_k''. It should be mentioned at this point that a zero value for all components i with respect to the distance $d_i(m,n)$ is not a sufficient condition for a match, since only the magnitude information from each component is considered in (29). On the other hand the angle information is examined in the next phase of the algorithm and errors rarely occur due to this in the final evaluation.

Step 3) Encoded within each group of matching feature vectors is information about

the scale and rotation parameters of the problem text, α and θ respectively. If a noiseless case is assumed, then the α parameter is equal to the ratio of the two size variables and the θ parameter is equal to the angle between the feature vectors. Since perfect knowledge of the feature vectors is never obtained the values of α and θ can only be estimated. The equations for the estimates of α $(\hat{\alpha})$ and θ $(\hat{\theta})$ are given by (30) and (31) respectively.

$$\hat{\theta}_{m,n} = \sum_{i=1}^{6} \left\| {}^{d_k}_{m}z_i \right\| \bullet \theta_i$$

where

$$\theta_i = \cos^{-1} \left\{ \frac{\left({}^{d_k}_{m}z_i \, , \, {}^{p}_{n}z_i \right)}{\left\| {}^{d_k}_{m}z_i \right\| \, \left\| {}^{p}_{n}z_i \right\|} \right\} \qquad 30$$

and

$$\hat{\alpha}_{m,n} = \frac{{}^{p}_{n}S}{{}^{d_k}_{m}S} \qquad 31$$

These estimates are obtained for each pair of matching feature vectors in the set NXM''_k. Note that the estimate in 30 is weighted by the magnitude of the individual components so that large errors in small components do not overly bias the estimate.

Step 4)   By this point there are only a few possible matches remaining, the problem now is to pick the best match of the remaining few, or to reject all of them if they are false matches. Since feature vectors from the same shape should all have the same scale factor and angle of rotation, clusters in the estimates of α and θ should be formed. The following procedure is used to define such a cluster. Let p be the index on set NXM'' with values from 1 to nm and form the vectors $[\hat{\alpha}_1 \, ,..., \hat{\alpha}_p \, ,..., \hat{\alpha}_{nm}]^T$ and $[\hat{\theta}_1 \, ,..., \hat{\theta}_p \, ,..., \hat{\theta}_{nm}]^T$ from the estimates in step 3.   Now

27

calculate the two circulant difference matrices T_α and T_θ and normalize them by τ_α and τ_θ so that

$$T_\alpha = \frac{\begin{bmatrix} \hat{\alpha}_1 - \hat{\alpha}_1 & \cdots & \hat{\alpha}_1 - \hat{\alpha}_{nm} \\ & \cdots \ \hat{\alpha}_p - \hat{\alpha}_p \ \cdots & \\ \hat{\alpha}_{nm} - \hat{\alpha}_1 & \cdots & \hat{\alpha}_{nm} - \hat{\alpha}_{nm} \end{bmatrix}}{\tau_\alpha} \qquad 32$$

and

$$T_\theta = \frac{\begin{bmatrix} \hat{\theta}_1 - \hat{\theta}_1 & \cdots & \hat{\theta}_1 - \hat{\theta}_{nm} \\ & \cdots \ \hat{\theta}_p - \hat{\theta}_p \ \cdots & \\ \hat{\theta}_{nm} - \hat{\theta}_1 & \cdots & \hat{\theta}_{nm} - \hat{\theta}_{nm} \end{bmatrix}}{\tau_\theta} . \qquad 33$$

The values τ_α and τ_θ represent the maximum allowable intra-cluster distances for the estimates of parameters α and θ respectively. Next calculate T_D the total difference matrix between all estimates as

$$T_D = \sqrt{(T_\alpha^2) + (T_\theta^2)} \qquad 34$$

where each operation is performed element wise. T_D will be an nm by nm difference matrix normalized by $\sqrt{\tau_\alpha^2 + \tau_\theta^2}$. To find the clusters find the matrix formed by setting all elements of $T_D > 1$ to zero and all others to 1. Now each row (or column) contains a 1 for each estimate that is clustered with the indexing estimate. In general there should be only one cluster if there is a match; however, the above procedure may also produce several subclusters of the larger cluster. In order to pick the best cluster, the number of terms and the strength of the

28

individual matches between each feature vector should be considered. The following decision scheme is chosen to select the final cluster.

1) Total the number of members in each cluster, call this number GCNT

2) Multiply each group member by the weighting factor

$$1 - \sum_{i=1}^{I} d_i(m,n)$$

this indicates the degree of match between each feature vector in the cluster. Sum this number and call it GSUM

3) Add the value $\tau_w \cdot (GCNT - 1)$ to GSUM and find the average value of this total for each group; $GAVG = (\tau_w \cdot (GCNT - 1) + GSUM)/GCNT$

4) Pick the cluster for which GAVG is maximum.

To get the final estimates of α and θ, find the average of the individual $\hat{\alpha}_i$ and $\hat{\theta}_i$ for the chosen cluster. If there is no cluster the above procedure reduces to picking the individual feature vector with the highest degree of match.

Unlike α and θ the translation of a whole shape is not equivalent to the same translation in each of its feature vectors (recall that translation is not a linear operation). Therefore, scale and rotation information is required to estimate z_t (the translation vector between the centroid of the dictionary shape and the centroid of the problem shape which have been matched). To estimate z_t the matched shape dictionary is scaled and rotated with respect to its centroid using the final estimates obtained for α and θ. The rotated and scaled versions of the dictionary coordinates are calculated as

$$_m^{d_k}x' = \bar{x}_k + \hat{\alpha} \left[\cos(\hat{\theta}) \, (_m^{d_k}x - \bar{x}_k) + \sin(\hat{\theta}) \, (_m^{d_k}y - \bar{y}_k) \right]$$

$${}_m^{d_k}y' = \bar{y}_k + \hat{\alpha}\left[\cos(\hat{\theta})\,({}_m^{d_k}y - \bar{y}_k) - \sin(\hat{\theta})\,({}_m^{d_k}x - \bar{x}_k)\right] \qquad 35$$

where $m = 1$ to M_k and $(\,x_k\,,\,\bar{y}_k\,)$ is the centroid of the dictionary. Note if the dictionary is prepared so that its centroid is $(0,0)$ then (35) can be reduced to

$${}_m^{d_k}x' = \hat{\alpha}\left[\cos(\hat{\theta})\,({}_m^{d_k}x) + \sin(\hat{\theta})\,({}_m^{d_k}y)\right]$$

$${}_m^{d_k}y' = \hat{\alpha}\left[\cos(\hat{\theta})\,({}_m^{d_k}y) - \sin(\hat{\theta})\,({}_m^{d_k}x)\right] \qquad 36$$

Since translations from the same shape should be equal (except for certain cases of symmetry) a clustering of translation parameters, similar to that just described for α and θ, is expected for the estimates of x_t and y_t. Therefore, an additional step can be implemented to further reduce the possibility of false alarms and large errors in the estimate of z_t.

Consider the final cluster from which α and θ were obtained. Calculate the values $_c x_t$ and $_c y_t$, the translation coordinates between each matched feature vector from the cluster, where

$${}_c x_t = \frac{1}{6}\sum_{i=1}^{6} {}_c x_i^t$$

$${}_c y_t = \frac{1}{6}\sum_{i=1}^{6} {}_c y_i^t \qquad 37$$

where

$$c x_i^t = {}_c^p x_i - {}^{d_k} x_i,$$

$$c y_i^t = {}_c^p y_i - {}^{d_k} x_i, \qquad 38$$

with

c = the index over the cluster obtained from the estimates of α and θ

     (c=1 to C),

i = the index on the individual components of each feature vector from the

     cluster,

x'= the x coordinate from the transformed dictionary page,

y'= the y coordinate from the transformed dictionary page.

Next form the two vectors $[\ _1 x_t \ ,..., \ _c x_t \ ,..., \ _C x_t \ ]^T$ and $[\ _1 y_t \ ,..., \ _c y_t \ ,..., \ _c y_t \ ]^T$ from the values obtained from (37). Then calculate the two circulant difference matrices T_x and T_y normalized by τ_x and τ_y (the maximum allowable intra-cluster distance for x_t and y_t) as

$$T_x = \frac{\begin{bmatrix} _1 x_t - _1 x_t & \cdots & & \cdots & _1 x_t - _C x_t \\ & \cdots & _c x_t - _c x_t & \cdots & \\ _C x_t - _1 x_t & & \cdots & & _C x_t - _C x_t \end{bmatrix}}{\tau_x} \qquad 39$$

$$T_y = \frac{\begin{bmatrix} _1 y_t - _1 y_t & \cdots & & \cdots & _1 y_t - _C y_t \\ & \cdots & _c y_t - _c y_t & \cdots & \\ _C y_t - _1 y_t & & \cdots & & _C y_t - _C y_t \end{bmatrix}}{\tau_y}. \qquad 40$$

Next calculate T_z the total normalized translation difference, as

31

$$T_z = \sqrt{(T_x)^2 + (T_y)^2} \qquad\qquad 41$$

where each operation is performed element wise. T_z can be checked for clustering in the same manner used previously with $\hat{\alpha}$ and $\hat{\theta}$. In general all elements of T_z are expected to cluster; however, there is a small probability (dependent on the values of τ_x and τ_y) that they will not. In this case, the non-clustering terms are discarded and the final estimates of x_t and y_t are obtained by averaging over the remaining cluster elements.

V RESULTS

A 3-dimensional shape dictionary was constructed. It consisted of two airplanes, an F-15 and an F-4. A value of 15 degrees was chosen for θ_{inc} and a total of 20 views generated for each plane. The range of views covered ± 30 degrees pitch and 0 to -45 degrees roll. The boundary curves were obtained by thresholding a 512X512 pixel image of each view to obtain a binary representation. A contour was then obtained from this image and processed for critical points. The contours obtained for each plane are shown in figure 14.

A value of 13 was chosen for the window length of the averaging filter used in the first stage of the critical point extraction routine. The values used for the segmenting thresholds were $\tau_1 = -.3$ and $\tau_2 = -1.2$. The values of the thresholds used to determine the confidence numbers were $\tau_3 = -2.4$ and $\tau_4 = -4.8$ respectively. This corresponds to a design range of .5 to 1 for the scale factor α. A bias value of τ_b equal to .48 was calculated based on a maximum deviation due to grid rotation of $(\sqrt{2}-1)^* \tau_2$. A maximum normalized distance of $\tau_m = .04$ is allowed between feature vector components. Maximum cluster distances of $\tau_\alpha = \pm.075$ and $\tau_\theta = \pm 5$ degrees were

32

selected for the scale, rotation cluster portion of the algorithm. Translation parameter clustering was implemented with a maximum scattering of ± 10 pixels for both τ_x and τ_y. A cluster weight of $\tau_w = .1$ was chosen for the cluster decision procedure.

The problem texts were then extracted from the scene shown in figure 15. This scene contains three planes, an F15, an F-4 and a HARRIER, placed at arbitrary orientations within the range covered by the shape dictionary. The HARRIER was used as a control to check the algorithm's ability to discriminate non-dictionary shapes from dictionary shapes. The video camera has been moved away from the planes to provide a scale change and adverse lighting has been employed to add uncertainty in the determination of the planes' boundaries. The three problem texts shown in figure 16 were extracted from the scene using binary thresholding. They were then matched with the shape dictionary. From figure 16 it can be seen that only partial shape information is recovered for each plane (the shadows are interpretted as part of the object information).

Each problem text was cross-matched with the two objects in the shape dictionary. The results of the match are demonstrated in figure 17. The white contours superimposed over the original scene indicate the view from the dictionary matched to each problem text with its estimated scale, rotation and translation parameters. The first problem text is identified as an F-15 at -15 degrees pitch and -15 degrees roll. The second problem text was identified as an F-4 at 0 degrees pitch and -30 degrees roll. The third problem text was not recognized as a member of the shape dictionary. The procedure generated no false alarms.

VI DISCUSSION

If additional identification power is required a second stage can be implememted to

33

eliminate portions of any composite shape problem text as they are identified, thereby reducing the composite problem text to a truncated problem text which has a higher recognition rate. Also false alarm rates can be significantly reduced by implementing a similar procedure to compare boundary points once a possible match has been indicated. It is also possible to extend the feature set chosen for each dictionary shape thus reducing the probability of a miss. Unfortunately this results in an increase in match times required for each dictionary shape. This increase in time will be linearly related to the increase in the feature vector set. Miss rates may also be reduced by relaxing the maximum distance allowed between feature vector components, This is recommended only if the previously discussed verification procedure is implememted because relaxation also increases the rate of false alarms.

If a sequence of dynamic images (scenes with objects in motion) is being analyzed, it is believed that the overall recognition rates will improve over the entire course of the sequence. This is because the amount of occlusion will not remain constant during the entire sequence, thus revealing more object information in the totality of the sequence than in any one frame.

VII CONCLUSIONS

The definition of critical points using the concavity tree structure and the corresponding identification algorithm described in this paper have been shown to offer the following features and advantages when the identification of 3-dimensional shapes in 2-dimensional scenes is considered. The theoretical definition of critical points using the concavity tree structure of a shape is invariant to transforms of scale, rotation and translation within the 2-dimensional scene. In the actual implementation, invariance to the latter two transforms is maintained and

invariance over a specific range of scale factors is obtained. The critical point definition is reasonably stable under conditions of shape occlusion and partial framing, such that partial shape information is usually sufficient to identify a shape. The algorithm requires only one pass per concavity tree branch to extract the critical points. These critical points are the same regardless of starting point. The critical points are concentrated in areas of high shape dynamics, i. e. the highest areas of shape quantification occur in those regions of the shape of highest qualitative interest. The algorithm is considered robust in the sense that , the more complex a shape becomes the higher the probability of identifying it in a scene. This is due to the fact that more critical points are allocated to the more complex shapes. The algorithm is flexible in the sense that it can be modified to accomodate restrictions on memory or computation. also as a byproduct of the matching procedure, estimates of the scale, rotation and translation parameters of an identified shape are automatically generated.

APPENDIX

The proof of the theorem in section III follows. It has already been established that the operation $\Psi(\bullet)$ is a point set relationship between a shape (subshape) and its concavity tree, or to be more exact a shape (subshape) and its convex hull. Therefore it is sufficient to show that this relationship is maintained under the three transformations. The convex hull S^C of a shape S is defined mathematically as the set of all points r such that

$$ r = \lambda p + (1-\lambda) q ; \qquad 0 < \lambda < 1 \qquad\qquad a1 $$

where p,q are points in S. This includes all points on all line segments joining every pair of points in S. For points r, p, q, in R^2 (a1) may be rewritten as

$$ (r_x, r_y) = \lambda (p_x, p_y) + (1-\lambda) (q_x, q_y) , \qquad\qquad a2 $$

where $r_x = \lambda p_x + (1-\lambda) q_x$; $r_y = \lambda p_y + (1-\lambda) q_y$; $0 < \lambda < 1$. The effect of each transformation is to map individual points of S into S', therefore the definition of $S^{C'}$ is simply the above definition with p,q replaced by p',q'. Consider first the convex hull of a shape under the transformation $S_\alpha(\bullet)$. It becomes the set of all points r' such that

$$ r' = \lambda p' + (1-\lambda) q' ; \qquad 0 < \lambda < 1 $$

or

$$ (r_x', r_y') = \lambda (\alpha p_x , \alpha p_y) + (1-\lambda) (\alpha q_x , \alpha q_y) , \qquad\qquad a3 $$

where

$$ r_x' = \lambda \alpha p_x + (1-\lambda) \alpha q_x ; $$
$$ r_y' = \lambda \alpha p_y + (1-\lambda) \alpha q_y . \qquad\qquad a4 $$

the terms in (a4) can be rearranged so that

$$r_x' = \alpha \left[\, \lambda\, p_x + (1-\lambda)\, q_x \,\right] ;$$

$$r_y' = \alpha \left[\, \lambda\, p_y + (1-\lambda)\, q_y \,\right] \qquad \text{a5}$$

which is equivalent to the transformation $S_\alpha(\bullet)$ on the whole line segment, or

$$r' = S_\alpha \left\{\, \lambda\, p + (1-\lambda)\, q \,\right\}. \qquad \text{a6}$$

Taken over the entire set S, this can be rewritten as

$$C_h(\, S_\alpha\, (\, S\,)) = S_\alpha\, (\, C_h(\, S\,)) , \qquad \text{a7}$$

which can be interpretted as "the convex hull of the scaled shape S is equal to the scaled convex hull of S. A similar argument can be used for the transformations $R_\theta(\bullet)$ and $T(\bullet)$ with (a4), (a5) and (a6) replaced by (a8), (a9), (a10) and (a11), (a12), (a13) for $R_\theta(\bullet)$ and $T(\bullet)$ respectively. Specifically,

$$(r_x',\ r_y') = \lambda\, (\, p_x \cos\theta - p_y \sin\theta\, ,\ p_x \sin\theta + p_y \cos\theta\,) +$$
$$(1-\lambda)\, (q_x \cos\theta - q_y \sin\theta\, ,\ q_x \sin\theta + q_y \cos\theta\,)$$

where,

$$r_x' = \lambda\, (\, p_x \cos\theta - p_y \sin\theta\,) + (1-\lambda)\, (\, q_x \cos\theta - q_y \sin\theta\,) ;$$
$$r_y' = \lambda\, (\, p_x \cos\theta + p_y \sin\theta\,) + (1-\lambda)\, (\, q_x \sin\theta - q_y \cos\theta\,) \qquad \text{a8}$$

which can be rewritten as,

$$r_x' = \left[\, \lambda\, p_x + (1-\lambda)\, q_x \,\right] \cos\theta - \left[\, \lambda\, p_y + (1-\lambda)\, q_y \,\right] \sin\theta ;$$
$$r_y' = \left[\, \lambda\, p_x + (1-\lambda)\, q_x \,\right] \sin\theta + \left[\, \lambda\, p_y + (1-\lambda)\, q_y \,\right] \cos\theta \qquad \text{a9}$$

and finally,

$$r' = R_\theta \left\{\, \lambda\, p + (1-\lambda)\, q \,\right\}. \qquad \text{a10}$$

37

Similarly,

$$(r_x', r_y') = \lambda (p_x + x_t , p_y + y_t) + (1-\lambda) (q_x + x_t , q_y + y_t)$$

where,

$$r_x' = \lambda (p_x + x_t) + (1-\lambda) (q_x + x_t) \; ;$$
$$r_y' = \lambda (p_y + y_t) + (1-\lambda) (q_y + y_t) \; ; \qquad \text{a11}$$

which can be rewritten as,

$$r_x' = \lambda p_x + \lambda x_t + q_x - \lambda q_x + x_t - \lambda x_t = [\lambda p_x + (1-\lambda) q_x] + x_t$$
$$r_y' = \lambda p_y + \lambda y_t + q_y - \lambda q_y + y_t + \lambda y_t = [\lambda q_y + (1-\lambda) q_y] + y_t$$

$$\text{a12}$$

and finally,

$$r' = T \left\{ \lambda p + (1-\lambda) q \right\}. \qquad \text{a13}$$

These arguments can be summed up, as in (a7) for the scaling transform, by (a14) and (a15) for the rotation and translation transforms respectively.

$$C_h(R_\theta (S)) = R_\theta (C_h(S)) . \qquad \text{a14}$$
$$C_h(T (S)) = T (C_h(S)) . \qquad \text{a15}$$

It has been shown that the relationship between the convex hull of a shape with the shape itself is preserved under the defined transformations. The only other operations involved with the determination of C from S are intersections. However, since $S_\alpha(\bullet)$, $R_\theta(\bullet)$ and $T(\bullet)$ are restricted to be one-to-one, all intersections are preserved by the nature of the transforms, i.e., any two points which coincide in the image of the transforms will also coincide in the preimage of transforms. Therefore the relationships in (24) are true.

38

BIBLIOGRAPHY

[1] Anderson, Brian D. O. and Moore, John B., Optimal Filtering, Prentice-Hall, Inc., 1979.

[2] Ballard, Dana H. and Sabbah, Daniel, "Viewer Independent Shape Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-5, No. 6, November, 1983, pp. 653-659.

[3] Bhanu, Bir, Representation and Shape Matching of 3-D Objects," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-6, No. 2, March, 1984, pp. 137-155.

[4] Brooks, R. A., "Symbolic Reasoning Among 3-D Models and 2-D Images," Artificial Intelligence, Vol. 17, 1981, pp. 285-348.

[5] Duda, Richard O. and Hart, Peter E., Pattern Classification and Scene Analysis , John Wiley and Sons, Inc., 1973.

[6] Freeman, H., "Shape Description Via the Use of Critical Points," Pattern Recognition, Vol. 10, pp. 159-166.

[7] Freeman, H. and Shapira, R., "Determining the Minimum-Area Encasing Rectangle for an Arbitrary Closed Curve," Communications of the ACM, Vol. 18, No. 7, July, 1975, pp. 409-413.

[8] Fu, King-Sun, Syntactic Pattern Recognition and Applications, Prentice-Hall, Inc., 1982.

[9] Gonzalez, Rafael C. and Thomason, Michael G., Syntactic Pattern Recognition, Addison-Wesley Publishing Co., Inc., 1978.

[10] Guggenheimer, Heinrich W., Differential Geometry, Dover Publishing, Inc., 1977.

[12] Koplowitz, Jack, "On the Performance of Chain Codes for Quantization of Line Drawings," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-3, No. 2, March, 1981, pp. 180-185.

[13] Mandelbrot, Benoit B., The Fractal Geometry of Nature, W. H. Freeman and Co., Inc., 1977.

[14] Marr, David, Vision, W. H. Freeman and Company, 1982.

[15] Martin, A. Fischler and Bolles, Robert C., " Perceptual Organization and Curve Partitioning," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 1, January, 1986, pp. 100-105.

[16] Melsa, James L. and Cohn, David L., Decision and Estimation Theory, McGraw-Hill Book Co., Inc., 1978.

[17] Miller, Billy Keith and Jones, Richard A., "Reliable Formation of Feature Vectors for 2-D Shape Representation," Proceedings of the 2nd International Technical Symposium on Optical and Electro-Optical Applied Science and Engineering, Cannes, France, December, 1986.

[18] Mitchell, O. Robert, Kuhl, Frank P. Grogan, Timothy A., and Charpentier, Didier J., "A Shape Extractions and Recognition System," SOUTHEON/82 Orlando, Florida, 1982, pp. 4.1.1-4.1.4.

[19] Mitchell, O. Robert, Reeves, Anthony P., and Grogan, Timothy A., "Algorithms and Architectures for Global Shape Analysis in Time Varying Imagery," SPIE Proceedings 360, Robotics and Industrial Inspection, San Diego, California, August, 1982.

[20] Oshima, Masaki, and Sharai, YoShiaki, "Object Recognition Using Three-Dimensional Information," IEEE Transactions on Pattern Analysis and Machine Intelligence," Vol. PAMI-5, No. 4, July, 1983, pp. 353-361.

[21] Pentland, Alex P., Local Analysis of the Image: Limitations and Uses of Shading," IEEE Computer Society Proceedings of the Workshop on Computer Vision: Representation and Control, August, 1982, pp. 153-161.

[22] Serra, J., Image Analysis and Mathematical Morphology, Academic Press, Inc., 1982.

[23] Sklansky, J., "Measuring Concavity on a Rectangular Mosaic," IEEE Transactions on Computers, Vol. C-21, No. 12, December, 1972, pp. 1355-1364.

[24] Tejwani, Yogendra J. and Jones, Richard A., "Machine Recognition of Partial Shapes Using Feature Vectors," IEEE Transactions on Systems, Man, and

Cybernetics, Vol. SMC-15, No. 4, July, 1985, pp. 504-516.

[25] Turney, Jerry L., Mudge, Trevor N., and Volz, Richard A., "Recognizing Partially Occluded Parts," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-7, No. 4, July, 1985, pp. 410-421.

[26] Wallace, Timothy P., Mitchell, Robert Owen, and Fukunaga, Keinosuke, "Three-Dimensional Shape Analysis Using Local Shape Descriptors," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-3, No. 3, May, 1981, pp. 310-322.

[27] Wang, Y. F., Magee, M. J., and Aggerwal, J. K., "Matching Three-Dimensional Objects Using Silhouettes," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-6, No. 4, July, 1984, pp. 513-518.

[28] Zabelle, G. Stephen and Koplowitz, Jack, "Fourier Encoding of Closed Planar Boundaries," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-7, No. 1, January, 1985, pp. 98-102.

[29] Zahn, Charles T. and Roskies, Ralph Z., "Fourier Descriptors for Plane Closed Curves," IEEE Trans. on Computers, Vol. C-21, No. 3, March, 1972, pp. 269-281.
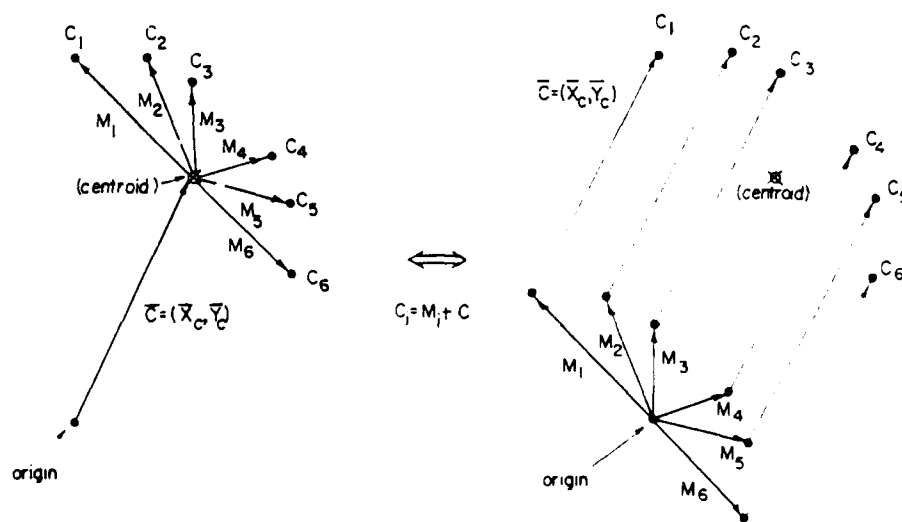
Figure 1 - Feature Vector Representation ($\therefore = 6$)

Original Shape

Level 1 Concavities

Level 2 Concavities

Level 3 Concavities
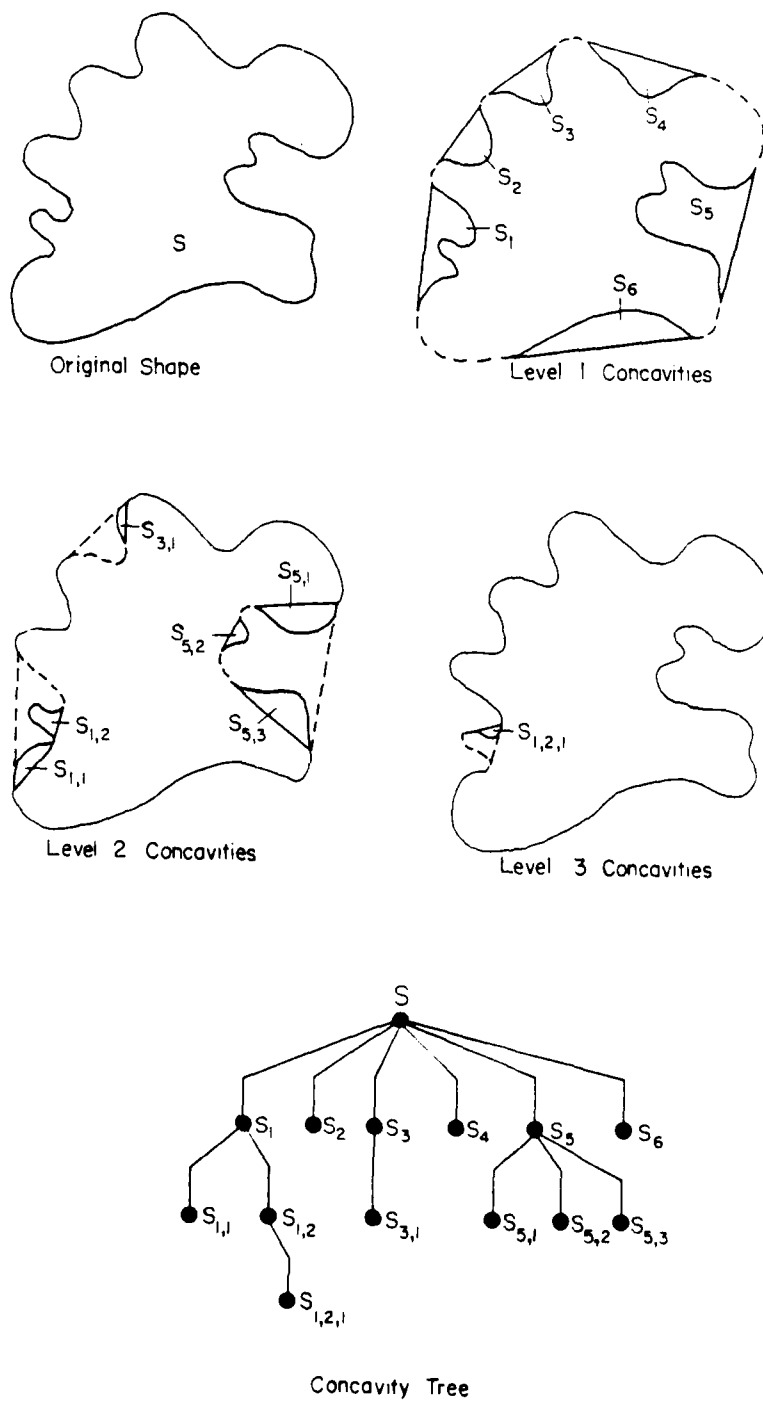
Concavity Tree

Figure 2 - Concavity Tree for Arbitrary Shape S
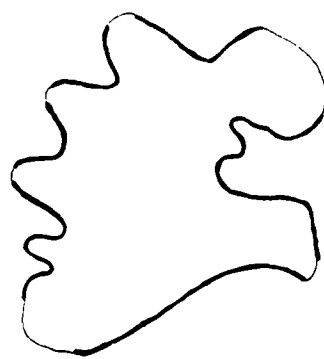
Boundary Segment for
Level I Concavities
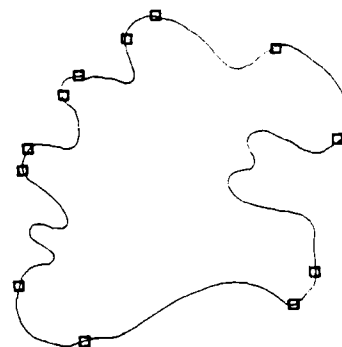
Critical Points for
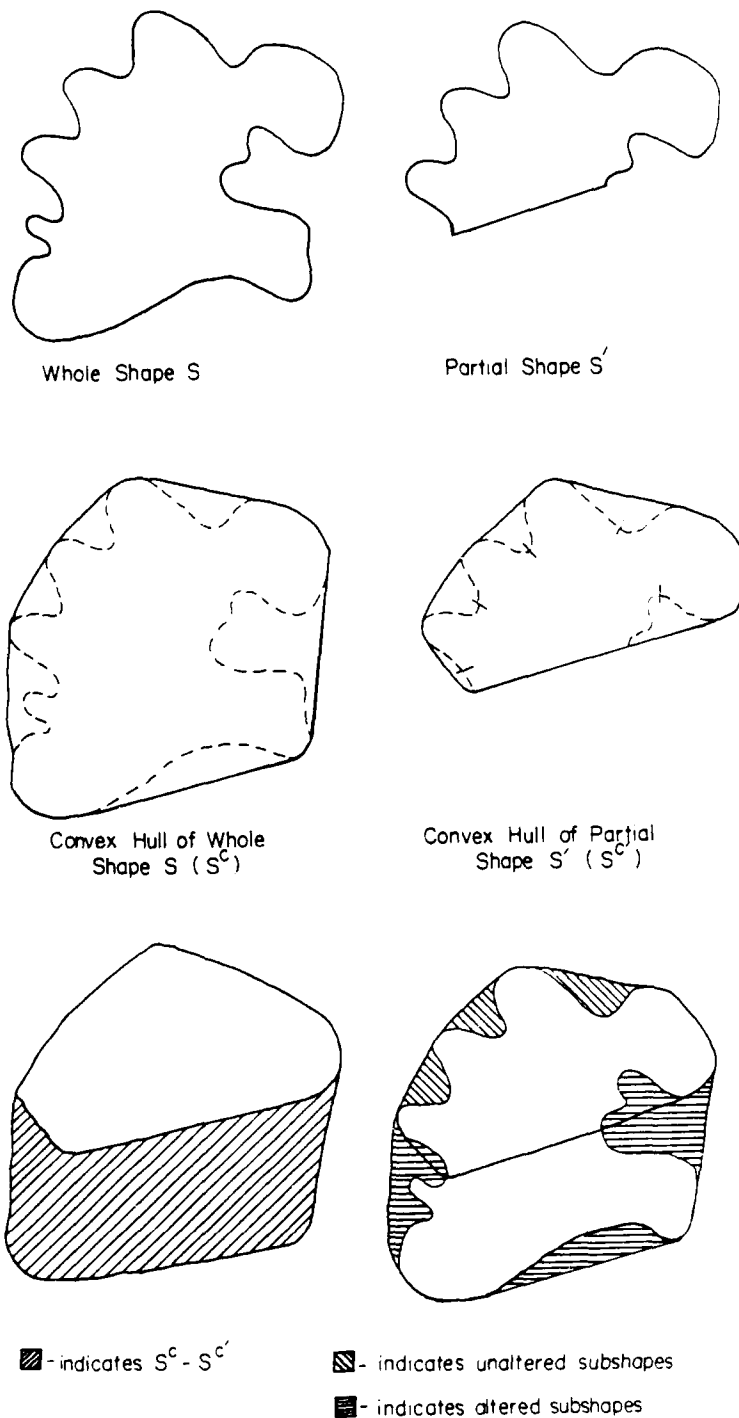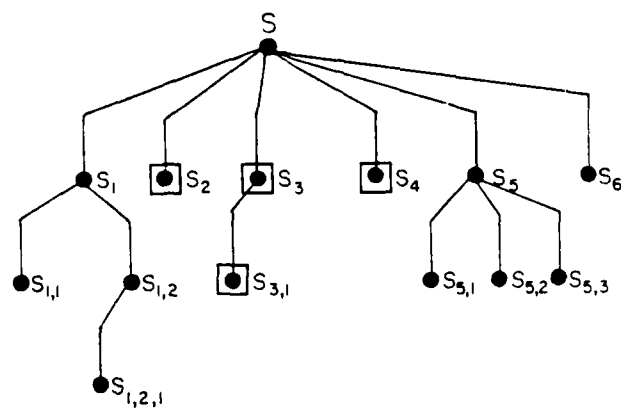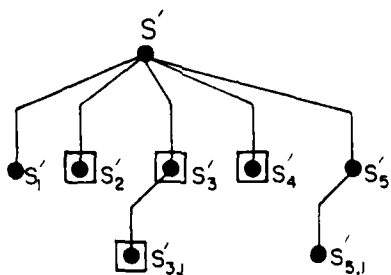Level I Concavities

Figure 3 - Boundary Segmentation at Level 1 for Arbitrary Shape S

Whole Shape S

Partial Shape S′

Convex Hull of Whole
Shape S (S^c)

Convex Hull of Partial
Shape S′ (S$^{c′}$)

▨ - indicates S^c - S$^{c′}$

◩ - indicates unaltered subshapes

◪ - indicates altered subshapes

Figure 4   - Concavity Tree for Partial Shape S′

Concavity Tree of Whole Shape

Concavity Tree of Partial Shape

☐ - indicates branchs of the two trees which
are identical

Figure 5 - Concavity Tree Comparison for Whole and Partial Shapes S and S'

46

$$S' = {}_1S \cup {}_2S$$
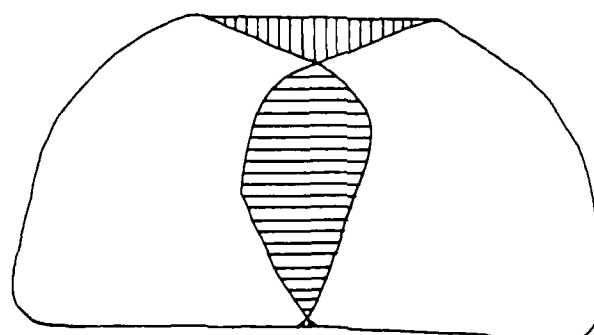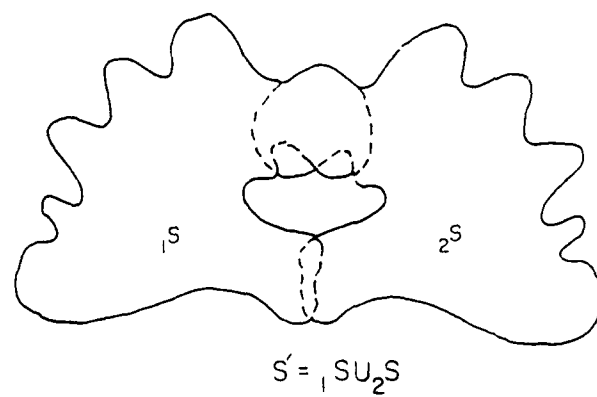
▥ indicates new concavities ($S^{c'} - (\underset{\text{all } i}{\cup} {}_iS^c)$)

▤ indicates overlap regions ($\underset{\text{all } i}{\cup} \cap \underset{\text{all } i}{} {}_iS^c$)

Figure 6 - Regions of Distortion in Convex Hull Due to Shape Composition
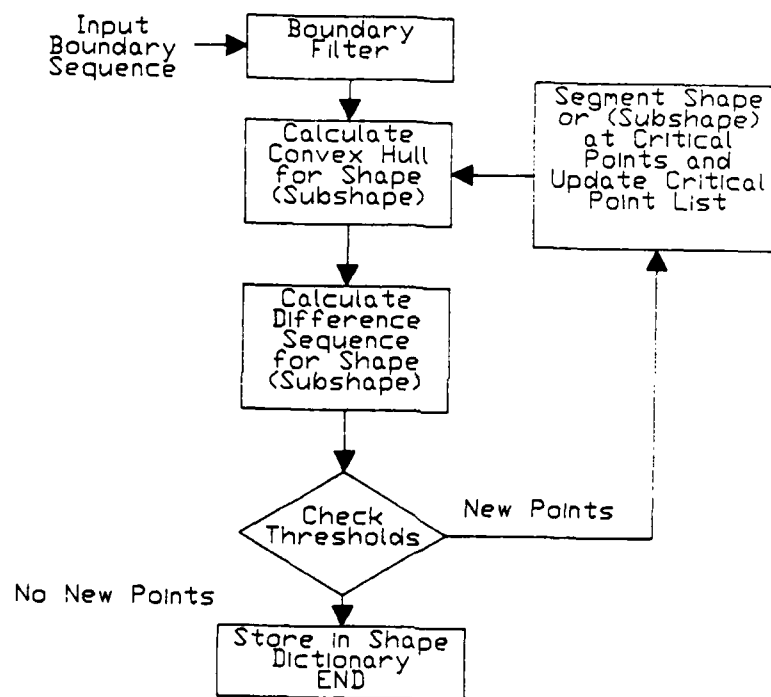
Figure 7 - Flow Diagram for Critical Point Extraction Procedure

48
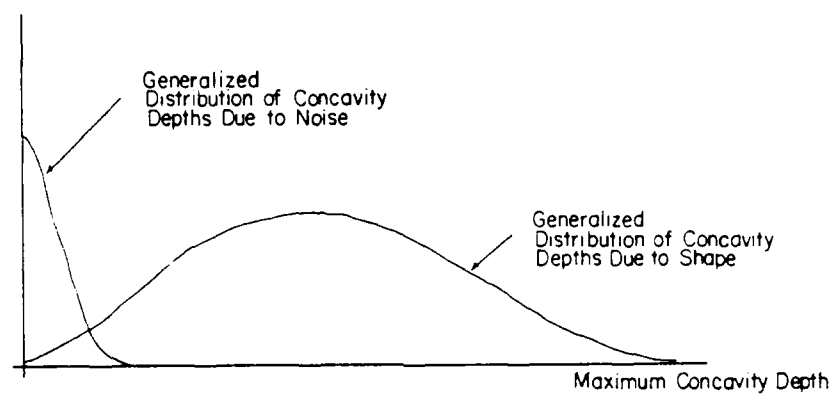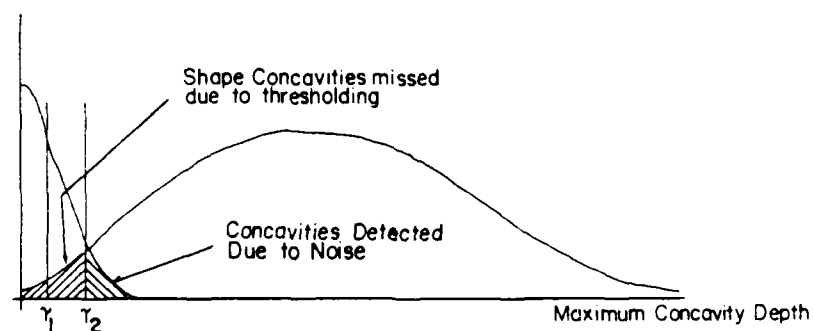
Figure 8 - Concavity Thresholding

(a)   Case 1

□ indicates critical points
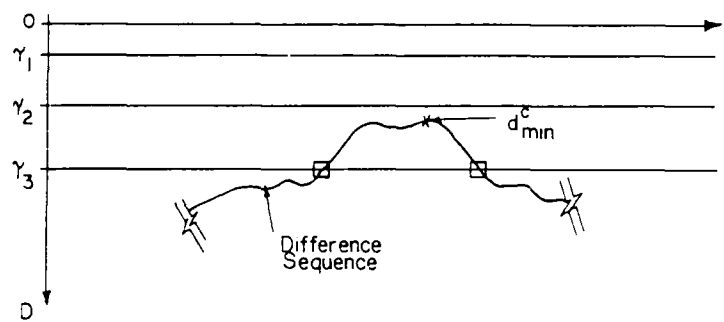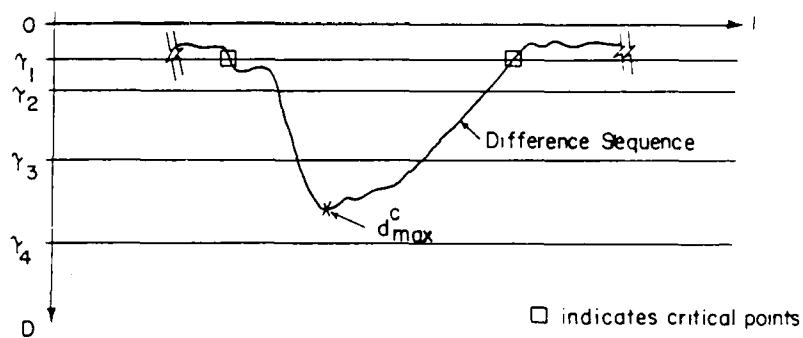
(b) Case 2

Figure 9   - Potential Concavities
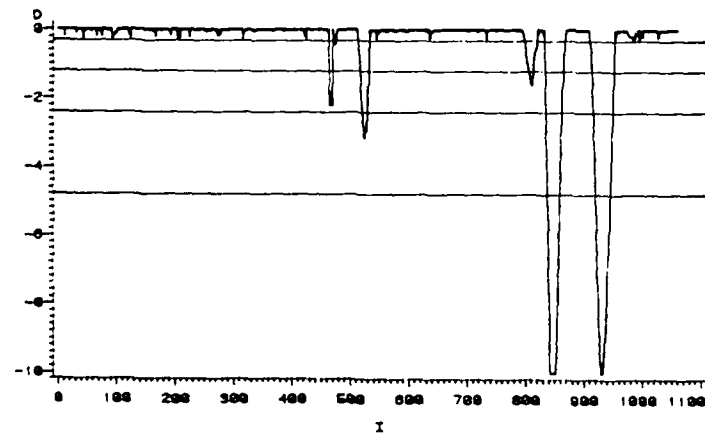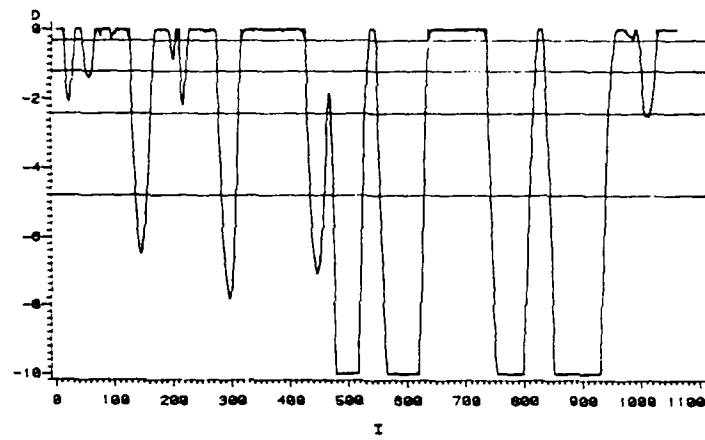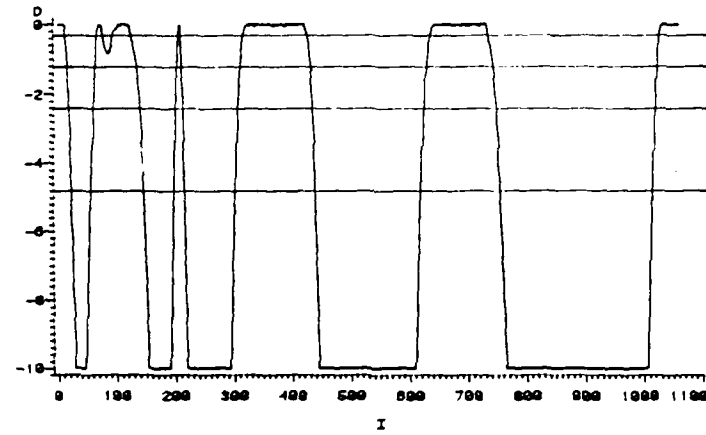
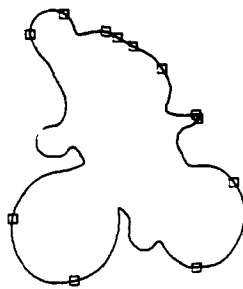Figure 10 - Levels 2 and 3 Difference Sequences

51
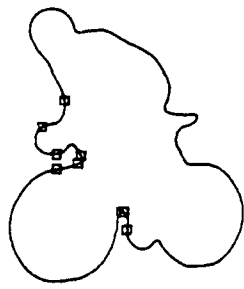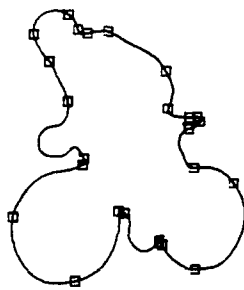
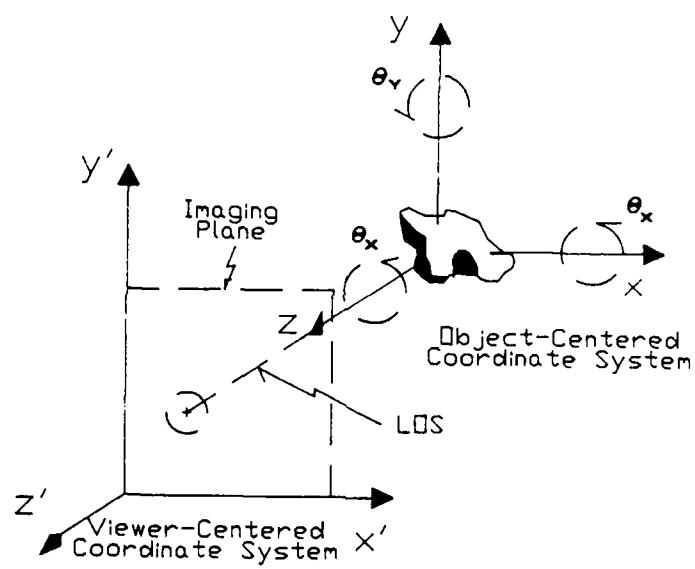Figure 11 – Critical Points for Levels 1, 2 and 3

52

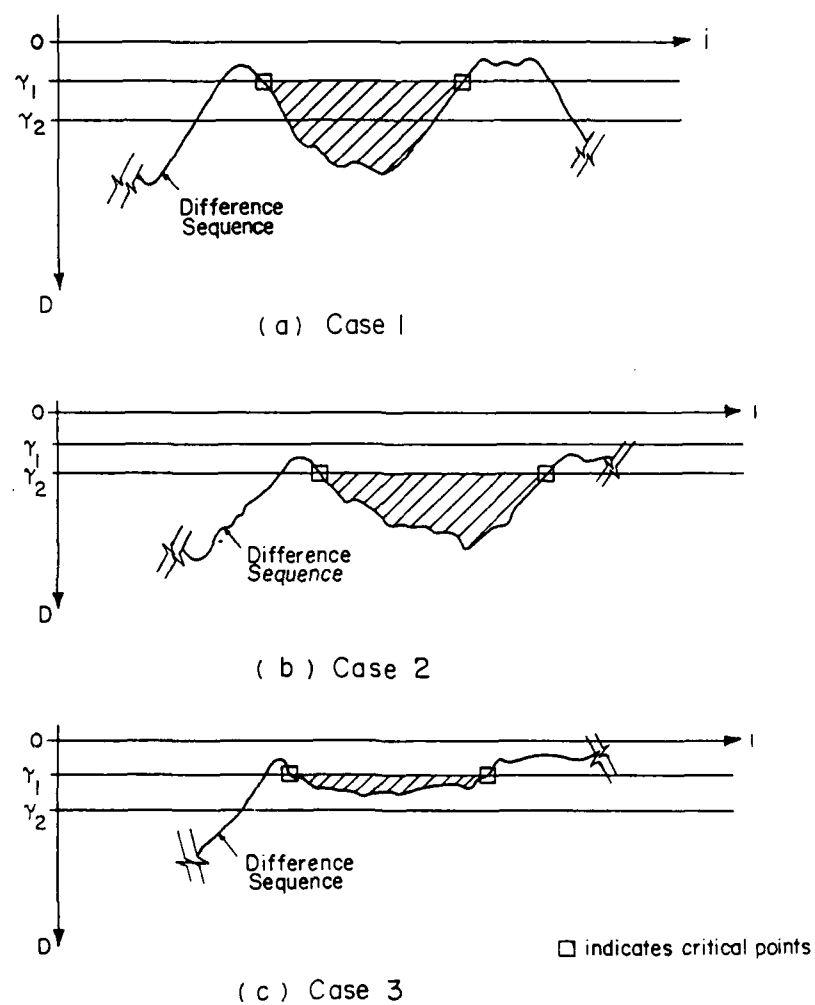Figure 12 - 3-Dimensional Object Orientation
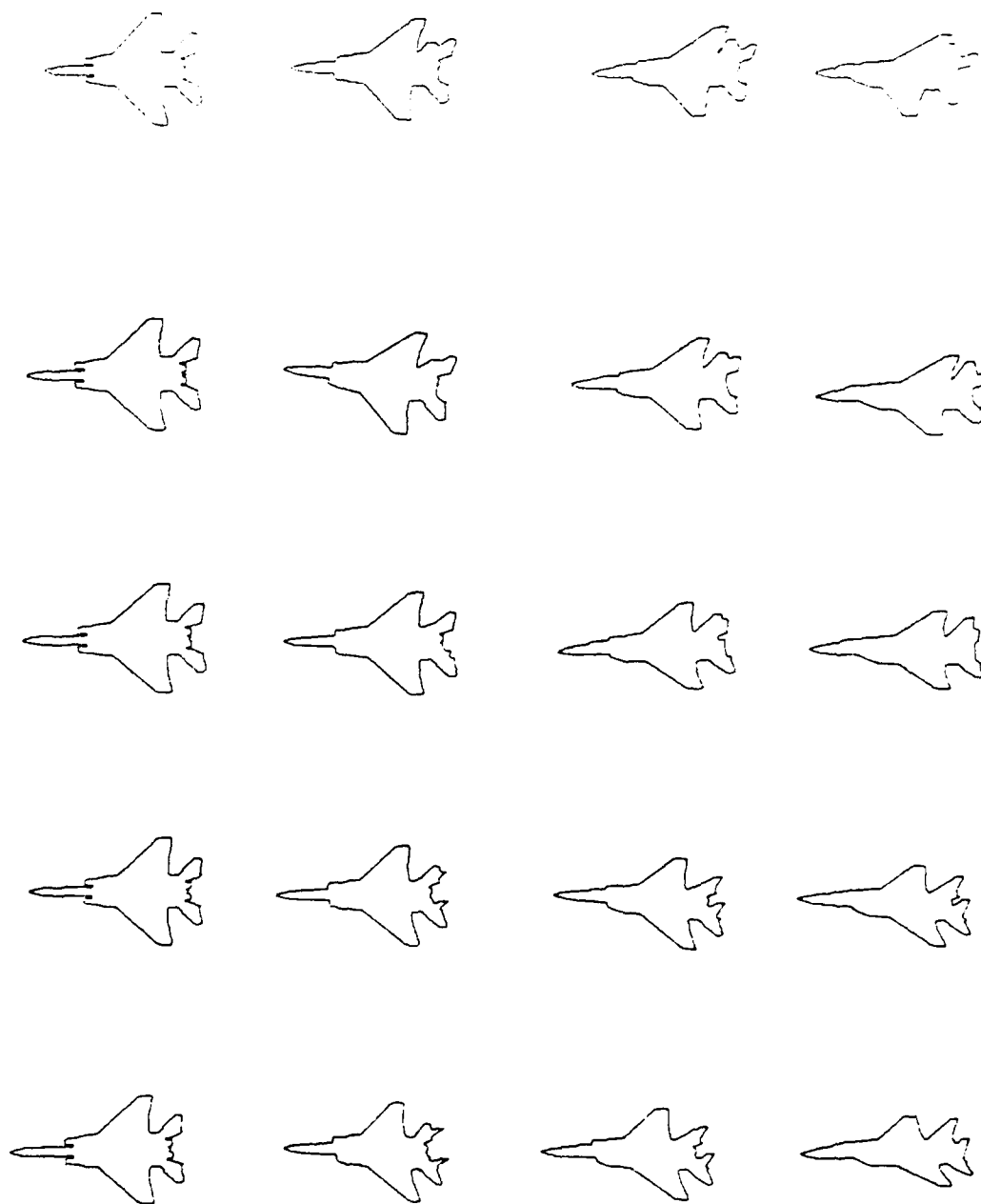
Figure 13 — Types of Concavities

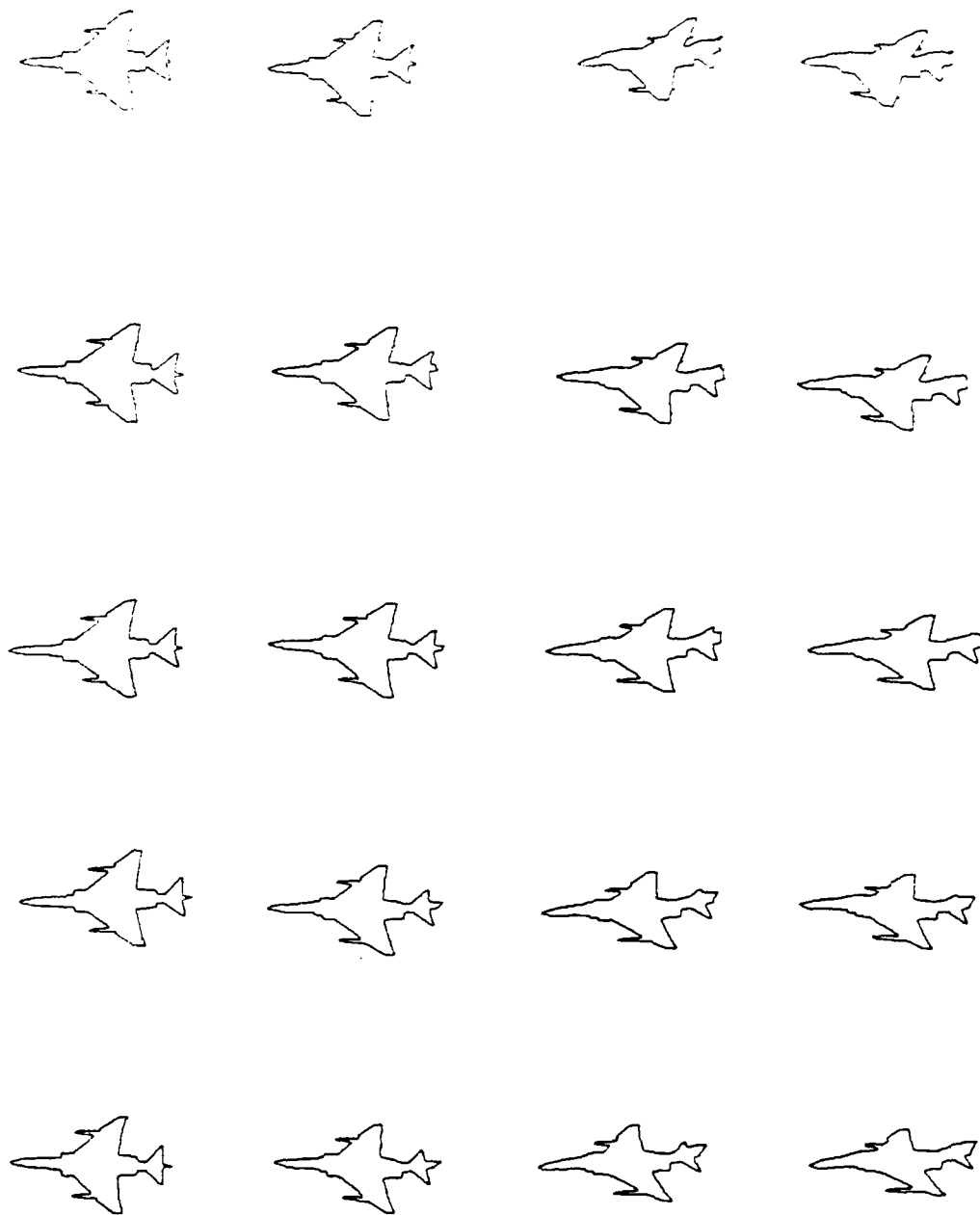Figure 14. — 3-Dimensional Object Stationary (6).

Figure 14-b - 3-Dimensional Object Dictionary - F4
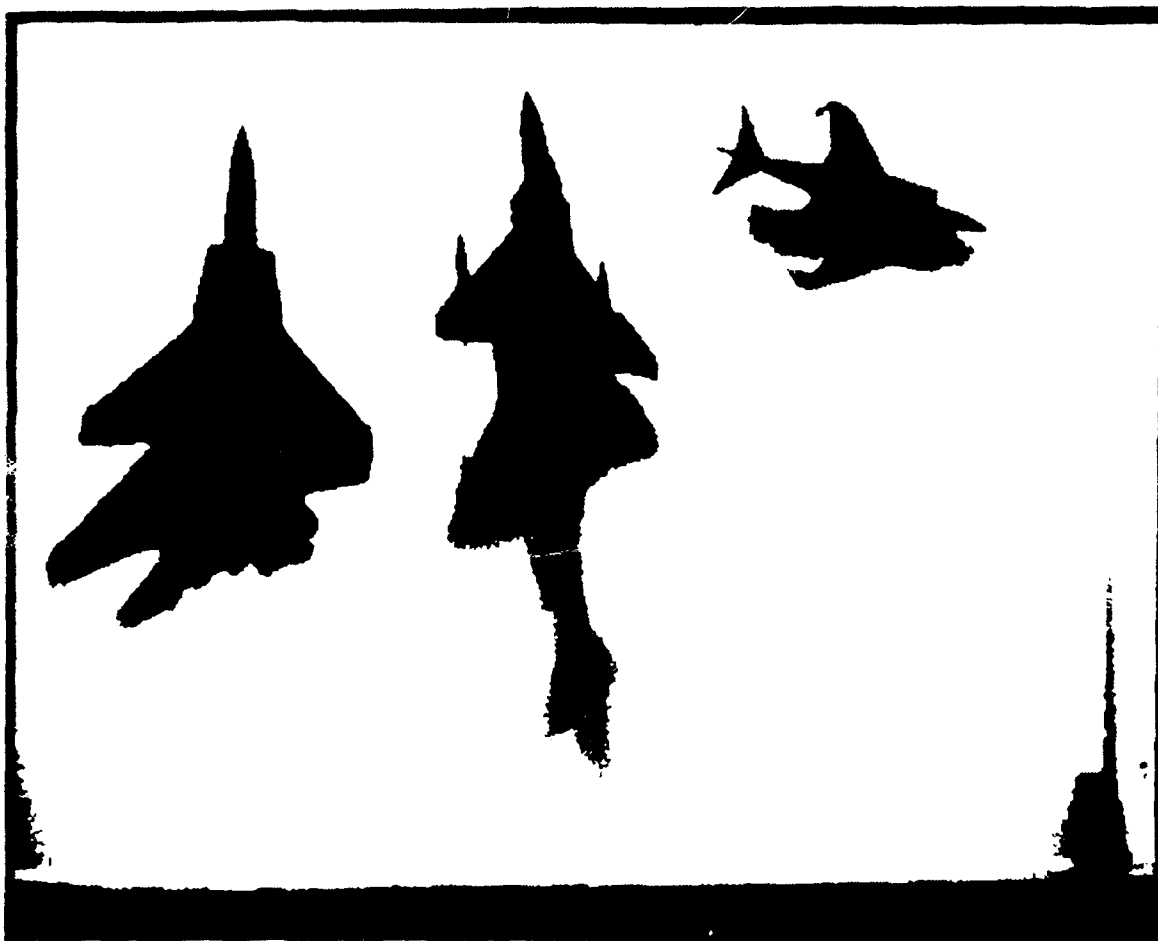
Figure 15 - Unknown Scene

57

Figure 16 - Problem Texts Extracted from the Scene

Figure 17 - Scene with Identified Objects

A Minimax Risk Quantizer for Noisy Sources

I.   Introduction

Classical quantizer design theory has focused on
determining the optimal quantization mapping which
minimizes the expected value of a distortion measure
between the quantizer input and output. These techniques
usually require *a priori* information concerning the
statistics (probability distribution, 1st and 2nd moments,
etc.) of the signal to be quantized. The need for
quantizing sources with incomplete knowledge of the source
statistics is well recognized, and a number of efforts
[1][6][12] have been made to address the problem. Bath
and VandeLinde [1] have presented the fundamental
development for minimax quantization of signals with
distributions from a unimodal generalized moment
constrained class. The minimax quantizer is said to be

that quantizer with a maximum distortion between the
quantizer input and output less than that of any other
quantizer for the given distribution class.

Minimum risk quantization theory [4] has been applied
to the problem of quantizing a source signal corrupted by
independent additive noise (figure 1) and has shown the
promise of using risk theory to determine an optimal
quantization scheme. This approach also uses considerable
a priori information concerning the source and noise. The
general risk theory analysis however, allows the use of a
number of estimation schemes to solve the corruptive noise
problem. Minimax estimation can be applied to the
quantization problem to produce a quantizer which
guarantees a maximum risk less than that of any other
quantizer. In section II the general quantization problem
is stated. Section III follows with a description of the
risk theory approach to quantization. Section IV develops
the minimax risk quantization theory for an additive noise
corrupted source with parallels drawn to the foundation
provided by Bath and VandeLinde. Section V describes some
simulation results followed by conclusions.

II.   General Quantization

Most of the prior work in quantizer design

63

characterize the quantizer mapping as an estimator of the quantizer input. A decision is made regarding within which region of the domain an input sample falls, and a representative value is determined which best estimates elements of that region. Denote the quantizer domain space X, and the finite range space Y. Then the quantizer mapping, $y = q(x)$, produces $y \in Y$, $\forall x \in X$. The design of a quantizer consists of completely specifying $q(x)$ and Y, when X is known. The mapping is surjective since every element in Y is a possible outcome of $q(x)$, $\forall x \in X$. Since Y is finite, call N the number of elements in Y. Define an index set on the elements of Y so that $Y = \{y_1, \cdots, y_N\}$. The domain can be partitioned into N disjoint sets, $X_i = \{x \mid q(x) = y_i, \ x \in X\}$, $i = 1, \cdots, N$, with $\bigcup_i X_i \equiv X$.

As mentioned above, it is common practice to require that a representative value y estimate the sample x. To this end, Y and $q(x)$ are defined so that y closely approximates x. This will require that each subset X_i be a connected subset with $y_i \in X_i$. The use of risk theory for quantization analysis will describe the quantization process in such a fashion that the membership requirement, $y_i \in X_i$, may no longer be applicable and will not be required.

III. Risk Quantization

The problem of quantizing noisy sources, as
illustrated in figure 1, requires new techniques to
determine optimal quantization schemes. Here an
independent identically distributed (i.i.d.) information
source has been corrupted by an i.i.d. additive noise
source, and it is required to quantize the resulting noisy
signal. The classical quantizer design procedure yields
the quantizer mapping which produces range elements that
best estimate the noisy input samples -- which may produce
unsatisfactory results. The problem is caused by the
perception of the quantizer as an input estimator. By
changing this perception, it will be shown that the
problem may be solved using risk and estimation theory.

Risk quantization is based upon the state space
representation of figure 2. Here the source, θ, is mapped
through some probabilistic transition mechanism,
$P_{X|\theta}(x|\theta)$, from the parameter space Θ to an observation
space X. The quantizer mapping $q(x)$ is the decision rule
which determines a representative element y from the
decision space Y which estimates the source, θ, based upon
the observation $x \in X$. Since the quantizer maps onto the
decision space Y, and the quantizer range must consist of
a finite number of elements, the elements of Y may be

65

indexed so that y_i, $i=1,\ldots N$, and $\bigcup_i y_i \equiv Y$ as before.

Although the observation x is an element of the set X_i, y_i representing the observation set will not be required to be an element of that set. This is because the probabilistic transition (due to the noise) might corrupt the original signal source such that the best estimate does not lie within the observation set X_i. The concept of quantization as estimation is the key to risk quantization, with the specific requirement of source estimation rather than input estimation. This will provide the necessary quantization schemes for noise corrupted systems, while incorporating the classical quantizer definition of input estimation for the noiseless case.

To determine an estimation rule, some form of quality measure is required. This may be given by a cost function $C(\theta, y)$ over $\Theta \times Y$ which represents the cost associated with estimating the parameter θ by decision y, based on observation x, $\forall \theta \in \Theta$, $x \in X$, $y \in Y$. The cost is generally expressed as a function of the difference between the parameter and its estimate,

$$C(\theta, y) = C(e = \theta - y) \tag{1}$$

With the specified cost function, the risk is defined as the expected value of the cost:

$$R = E[C(\theta, y=q(x))] \qquad (2)$$

Estimation theory techniques may be employed to determine
the quantization mapping $q(x)$ which minimizes the risk
function.

IV. Minimax Risk Quantization

The risk theory approach has been applied to the
additive noise problem depicted in figure 1 for well
defined *a priori* source and noise distributions to obtain
a minimum risk quantization solution [4]. It is shown in
[4] that the risk approach is a viable one to quantizer
design for such systems, and is equivalent to the
classical result for noiseless systems.

It is often likely, however, that the quantizer
design must be based on incomplete knowledge of the signal
statistics. For example, a problem may arise when a
quantizer is needed for a source where the statistics are
fairly well known, but various unknown noise distributions
can corrupt the source signal. Likewise, the noisy
characteristics of a sensor may be well defined, but the
exact probability distribution of the source signal to be
quantized with the sensor unknown. These types of
problems suggest the use of minimax estimation to

67

determine the quantization mapping. Minimax estimation seeks to minimize the maximum possible distortion by the estimator, and is directly applicable to the risk approach to quantizer design.

This is demonstrated by examining the problem of an additive noise source using risk theory and minimax estimation. Specifically, the source statistics and probability distribution will be assumed known *a priori*. The exact distribution of the noise, however, will be assumed to be unknown, but belonging to a generalized moment constrained class of distributions, C. A risk analysis of the system will be applied using minimax estimation to produce a minimax risk quantization scheme for the system. Further, it will be shown that the analysis is reversible and may be applied to the additive noise problem when the noise probability distribution is well defined, but the knowledge of the source distribution is incomplete.

A minimax estimator t^* is one which guarantees a maximum risk no greater than that for any other estimator [11]. That is,

$$\sup_{\theta} \mathcal{R}_{t^*}(\theta) \leq \sup_{\theta} \mathcal{R}_{t}(\theta) \tag{3}$$

Likewise, the minimax quantizer will be defined as one which guarantees the minimum maximal risk for all quantizers. Minimax quantization theory has been

68

developed for the classical quantizer problem of noiseless

sources by Bath and VandeLinde [1]. The approach

developed will be applied in the risk theory development

of an appropriate quantizer for the additive noise

problem.

Known Source Distribution

Consider the sequence of independent, identically

distributed random variables $\{\theta_k\}$, from the known

cumulative probability distribution function (c.p.d.f) F_θ.

The source sequence has been corrupted by a noise sequence

$\{N_k\}$, with unknown c.p.d.f.s $\{F_{N_k}\}$, to produce the

observation sequence $\{X_k\}$ to the quantizer.

The sequence $\{F_{N_k}\}$, while unspecified, are

constrained to belong to the set C of all possible

c.p.d.f.s with whatever *a priori* information of the noise

characteristics is available. The set to be considered

here is the set of c.p.d.f.s belonging to the generalized

moment constrained class. These distributions are

required to have a generalized moment less than or equal

to some finite constant. This may be viewed as a

restriction of the noise power to be finite.

The minimax risk is defined as:

$$R^* = \min_{q \in \mathcal{Q}_N} \max_{F_N \in C} R(q, F_\theta, F_N) \qquad (4)$$

where $R(q, F_\theta, F_N)$ is the risk for an N-level quantizer q

from the set of all possible N-level quantizers $\mathcal{Q}_N$ and the

c.p.d.f.s F_θ, F_N, described above. The minimax risk

quantizer, q^*, is that quantizer which provides the

minimax risk R^*. The properties of the minimax risk

quantizer, analogous to the minimax quantizer, are:

1) $\forall F_N \in C$, $\nexists R_{q^*} > R^*$

> where R^* is the absolute maximal risk for
> quantizer q^*.

2) $\nexists q \in \mathcal{Q}_N$, $q \neq q^* \ni \max_{F_N \in C} R(q, F_\theta, F_N) < R^*$

> where q^* guarantees a maximum risk no
> greater than the maximum risk
> of any other quantizer $q \in \mathcal{Q}_N$.

The quantizer mapping, $q(x) \in \mathcal{Q}_N$, maps an $x \in X_i$ to a

quantizer level $y_i \in Y$, $i = 1, \cdots, N$. For the scalar

quantizer, $X_i = [x_i, x_{i+1})$, where $x_1 = -\infty$, $x_{N+1} = \infty$, and

the quantizer thresholds comprise the set $\{x_i\}$,

$i = 1, \cdots, N+1$. Note that the set of quantizers $\mathcal{Q}_N$ does not

require the quantizer level y_i be an element of Y_i.

The risk for any particular quantizer mapping may be defined as:

$$R(q, F_\Theta, F_N) = \int_\Theta \int_N C(\Theta, q(x)) \, dF_N \, p_\Theta(\Theta) \, d\Theta \qquad (5)$$

where the inner integral is a Lebesgue-Stieltjes integral over the sequence of noise c.p.d.f.s $\{F_{N_k}\}$. The order of integration is changed to yield:

$$R(q, F_\Theta, F_N) = \int_N \int_\Theta C(\Theta, q(x)) \, p_\Theta(\Theta) \, d\Theta \, dF_N \qquad (6)$$

The purpose of the interchange is to facilitate the minimax operation procedure that occurs later.

It is apparent that if the system is noiseless, the risk function is precisely that of the classical quantizer distortion function. The minimax quantizer then is the same as the results of Max [9] since the source distribution is assumed known.

$$R^* = \inf_q \int_\Theta C(\Theta, q(x)) \, p_\Theta(\Theta) \, d\Theta, \quad \Theta = x$$

$$\Longrightarrow \quad R^* = \inf_q \int_\Theta C(\Theta, q(\Theta)) \, p_\Theta(\Theta) \, d\Theta \qquad (7)$$

The cost function used is restricted to some bounded distortion measure, $d(\Theta, q(x))$, between the source and quantizer output:

72

$$C(\theta, q(x)) = \begin{cases} d(\theta, q(x)), & d(\theta, q(x)) \leq L \\ L, & d(\theta, q(x)) > L. \end{cases} \qquad (8)$$

The distortion measure $d(\theta, q(x))$ is required to be even, continuous, monotonic strictly increasing in $|\theta - q(x)|$, and zero for perfect estimation. The bound on the cost implies no additional penalty for a distortion larger than the limit, L.

The generalized moment constraint is given by some function $\rho(n)$ satisfying the same conditions as those for the distortion function such that:

$$\int_N \rho(n) \; dF(n) \leq c. \qquad (9)$$

The constraint function $\rho(n)$ simply implies that the noise signal power is bounded. The distortion and constraint functions considered here are:

$$d(\theta, q(x)) = |\theta - q(x)|^n \qquad (10)$$

$$\rho(n) = |n|^m \qquad m \geq n \geq 1.$$

The conditions of bounded cost and those imposed on $d(\theta, q(x))$ and $\rho(n)$ are required for the Lagrange minimization of the next section.

The problem then, is to determine the minimax risk quantizer q^* for the worst case c.p.d.f., F_N^*, which achieves:

73

$$\inf_{q \in \mathcal{Q}_N} \sup_{F_N \in C} \mathcal{R}(q, F_\theta, F_N) = \mathcal{R}(q^*, F_\theta, F_N^*). \tag{11}$$

The set $\mathcal{Q}_N$ is the set of all symmetric quantizers where the levels are symmetric about the source mean, and the thresholds are symmetric about the observation mean. The set C is the set of c.p.d.f.s belonging to the generalized moment class, and $\mathcal{R}$ is the risk function using the cost function $C(\theta, q(x))$. To determine the minimax risk quantizer, it first necessary to find the maximum risk due to a particular quantizer mapping $q \in \mathcal{Q}_N$, then determine which quantizer provides the minimum of all of the maximum risks.

It has been shown [1] that the set C is a weak* compact subset of a normed Banach vector space (N.B.V.) of normalized functions of bounded variation. Also with any linear functional on N.B.V.$[0, \infty]$ (such as $\mathcal{R}(q, F_\theta, F_N)$), which is weak* continuous in F_N, then:

$$\forall\, q \in \mathcal{Q}_N, \ \exists\, F_N^* \in C, \ \mathcal{R}^*(q) \ni:$$
$$\mathcal{R}^*(q) = \mathcal{R}(q, F_\theta F_N^*) = \max_{F_N \in C} \mathcal{R}(q, F_\theta, F_N) \tag{12}$$

Furthermore, a method for determining a minimax quantizer via a constrained minimization in a Lagrange multiplier space ($\mathbb{R}^2$) through the use of the Lagrange duality theorem has been developed [1]. The technique will be

74

paraphrased here to determine a minimax risk quantizer.

Let $C' = \{F_N \in N.B.V.[0,\infty]\}$ where F_N exhibits the following properties:

1) F_N is nonnegative, monotonically nondecreasing,

2) $\int_N dF_N \leq 1$,

3) $\int_N \rho(n)\ F_N(n) \leq c$.

Define the convex set $\overline{C} = \{\ F_N \in N.B.V.[0,\infty]\ \}$, where F_N satisfies 1). Now define a convex functional

$$G:\ N.B.V.[0,\infty] \longrightarrow \mathbb{R}^2,$$

$$G = \begin{bmatrix} \int\ dF_N - 1 \\ \int \rho\ dF_N - c \end{bmatrix}$$

which represents the necessary constraints for F_N to be a c.p.d.f. and have a generalized moment constraint. This implies:

$$C' = \{\ F_N \in \overline{C}:\ G(F_N) \leq 0\ \}$$

By the Lagrange duality theorem [8], an expression for the maximum risk due to the quantization mapping $q(x)$ may be found:

75

$$R^*(q) = \sup_{F_N \in C} R(q, F_\theta, F_N)$$

$$\sup_{F_N \in C} R(q, F_\theta, F_N) = \min_{\lambda_1, \lambda_2 \geq 0} \sup_{F_N \in \overline{C}} \left\{ R(q, F_\theta, F_N) \right.$$

$$\left. - \lambda_1 \left[\int dF_N - 1 \right] - \lambda_2 \left[\int \rho \, dF_N - c \right] \right\} \tag{13}$$

where the inner maximization is achieved for some worst

case distribution, $F_N^* \in \overline{C}$, and the outer minimization by

λ_1^* and λ_2^*. If the cost weighting function is bounded,

$R^*(q)$ is finite and equation (6) can be used to replace

$R(q, F_\theta, F_N)$ of equation (13) to yield:

$$R^*(q) = \min_{\lambda_1, \lambda_2 \geq 0} \left[(\lambda_1 + \lambda_2 c) + B(\lambda_1, \lambda_2) \right] \tag{14}$$

where

$$B(\lambda_1, \lambda_2) = \max_{F_N \in \overline{C}} \int_N I(n) \, dF_N(n) \tag{15}$$

$$I(n) = \int_\Theta C(\theta - q(x)) \, p_\theta(\theta) \, d\theta - \lambda_1 - \lambda_2 \rho(n) \tag{16}$$

The minimization of equation (14) need only be

considered when $B(\lambda_1, \lambda_2)$ is finite. This will only occur

when $I(n)$ is nonpositive. If $I(n)$ is positive at any

point, a sequence of noise random variables $\{N_k\}$ with

c.p.d.f.s $\{F_{N_k}\}$ could occur such that as k gets large, the

integration tends to infinity.

If $I(n)$ is required to be nonpositive, then the maximization of equation (15) over all possible $F_N \in \overline{C}$ implies:

$$B(\lambda_1, \lambda_2) = 0. \qquad (17)$$

Since $F_{N_k} =$ constant is an element of $\overline{C}$ for $\{N_k\}$, it follows that if $I(n)$ is made nonpositive, the maximum possible integration of equation (15) is zero.

The minimax risk for any quantizer $q \in Q_N$ and the specified conditions then reduces to:

$$\mathcal{R}^*(q) = \min_{\substack{\lambda_1, \lambda_2 \geq 0 \\ I(n) \leq 0}} (\lambda_1 + \lambda_2 c) \qquad (18)$$

Figures 3 through 6 are a series of curves which depict the function $I(n)$ for various quantizers with $p_\Theta(\theta) \sim N(0,1)$, and distortion bounds of $L = 0.25, 0.75$. These curves show the effects of the distortion bound on the function, and that it is quantizer dependent. Note the discontinuity near the quantizer threshold.

The minimax risk quantizer q^* then, is that quantizer which produces:

$$\mathcal{R}^* = \min_{q \in Q_N} \min_{\substack{\lambda_1, \lambda_2 \geq 0 \\ I(n) \leq 0}} (\lambda_1 + \lambda_2 c) \qquad (19)$$

This result is similar to that obtained for minimax

quantization by Bath and VandeLinde for the generalized
moment constrained class of quantizer input distributions.
It differs primarily in the function $I(n)$ which
incorporates the cost of estimating the source by the
quantization mapping and the *a priori* knowledge of the
clean source probability distribution function, $F_\theta(\theta)$.

The minimization of equation (14) over all
nonnegative λ_1, λ_2, and nonpositive $I(n)$ implies the
maximum of $I(n)$ is zero. As can be seen in equation
(16), the role played by λ_1 in $I(n)$ is merely that of a
bias term. This implies that the minimization of equation
(14) can be determined by solving:

$$\lambda_1^*(\lambda_2) = \max_n \left[\int_\Theta C(\theta, q(x)) \, p_\theta(\theta) \, d\theta - \lambda_2 \rho(n) \right] \qquad (20)$$

and minimizing over all possible $\lambda_2 \geq 0$. Maximization of
$I(n)$ with respect to λ_2 and $\lambda_1^*(\lambda_2)$ for a specific
quantizer $q \in Q_N$ will then yield the minimum λ_2 and the
requisite λ_1 for $\max_{\lambda_1, \lambda_2 \geq 0} I(n) = 0$.

Known Noise Distribution

The minimax risk theory is easily modified for the
converse problem. That is, the problem considered thus far

is that of quantizing an i.i.d. source, with known probability distribution, corrupted by a sequence of independent additive noise random variables each of uncertain distribution. If the noise sequence is i.i.d. with known c.p.d.f. F_N, but the source sequence $\{\theta_k\}$ is only known to be distributed with the c.p.d.f. sequence $\{F_{\theta_k}\}$ from the set C, then the risk may be written:

$$R(q, F_\theta, F_N) = \int_\theta \int_N C(\theta, q(x)) \, p_N(n) \, dn \, dF_\theta \qquad (21)$$

The minimax risk quantizer development for this problem parallels the development for the previous case. A similar solution is described as the minimax risk quantizer q^* which yields the minimum maximal risk, $R^*(q)$, where:

$$R^*(q) = \min_{\substack{\lambda_1, \lambda_2 \geq 0 \\ I(\theta) \leq 0}} (\lambda_1 + \lambda_2 c) \qquad (22)$$

and

$$I(\theta) = \int_N C(\theta - q(x)) \, p_N(n) \, dn - \lambda_1 - \lambda_2 \rho(\theta) \qquad (23)$$

The same restrictions apply to the distortion and constraint functions, $d(\theta, q(x))$ and $\rho(\theta)$ as before. Likewise,

$$\lambda_1^*(\lambda_2) = \max_{\theta} \left[\int_N C(\theta, q(x)) \, p_N(n) \, dn - \lambda_2 \rho(\theta) \right] \qquad (24)$$

and the minimax risk quantizer q^* is specified by a procedure analogous to that of the previous section.

Figures 7 through 10 show the effects due to the constraint on the noise power and the distortion bound L on the function $I(\theta)$. These appear roughly similar to $I(s)$ in [1], differing through the knowledge of the noise for $I(\theta)$.

Note that in the noiseless case, this risk definition coincides precisely with the distortion function of an unknown input sequence used by Bath and VandeLinde. The minimax risk quantizer then is identical to the minimax quantizer developed there for the classical quantizer distortion measure.

V. Results and Conclusions

The minimax risk quantizer determination procedure, is described by the following algorithm. No attempt has been made to optimize the computation procedure.

1) Select $\lambda_2 \geq 0$.

2) Determine the optimal λ_1 for the specified

80

λ_2: $\lambda_1^*(\lambda_2)$.

3) Repeat step 2) over all $\lambda_2 \geq 0$ to determine

   the minimum of $(\lambda_1^*(\lambda_2) + \lambda_2 c)$. This is the

   maximum risk for the quantizer q: $\mathcal{R}^*(q)$.

4) Repeat steps 1)-3) over all quantizers

   $q \in \mathcal{Q}_N$ to determine the minimum of the

   maximum risks: $\mathcal{R}^* = \min_{q \in \mathcal{Q}_N} \mathcal{R}^*(q)$.

5) The minimax risk quantizer is that

   quantizer q^* which yields the minimax risk:

   $\mathcal{R}^*(q^*) = \mathcal{R}^*$.

It appears from the plots of $I(n)$ that the areas of
maxima occur at or within a δ - neighborhood of the
quantizer thresholds. A simulation has been performed
based on this observation with the results appearing in
tables 1 and 2. These results were determined by sampling
$I(n)$ in the region surrounding each threshold. The
quantizer step size was fixed at 0.01 and the accuracy of
the Lagrange minimization is such that the error in the
estimate is at most 1×10^{-7}. The minimax risk quantizer
with specific quantizer thresholds were found, and the
minimum of these selected as the minimax risk quantizer.

Comparisons of the minimax risk quantizers to the

Max-Lloyd and minimum risk quantizers (G.S.+G.N.) are
shown at the bottom of tables 1 and 2. The mean square
power constraint, c, is fixed such that for the Gaussian
noise case, the Signal-to-Noise Ratio (SNR) is 8 dB. This
value is shown in table 1. At 8 dB, the thresholds of all
three are nearly identical, with the levels for the
minimum risk quantizer inside (closer to the mean) those
of the Max-Lloyd quantizer, and the minimax risk quantizer
levels just slightly inside those of the minimum risk
quantizer. The constraint c is fixed for table 2 so that
for the Gaussian noise case, the SNR is 4 dB. At 4 dB,
the outer threshold of the minimax risk quantizer has
migrated well outside of the other two quantizers, which
may be from the Max-Lloyd and minimum risk quantizers
assumption of Gaussian densities, while the minimax risk
quantizer only considers constrained noise power. The
levels of the minimax risk quantizer are now between those
of the Max-Lloyd and minimum risk quantizers. This may be
interpreted from the standpoint of the assumptions also.
The minimum risk quantizer assumes both the source and
noise are Gaussian, so it arrives at closely compacted
estimates for the levels. The minimax risk quantizer, on
the other hand, only assumes the source is Gaussian for
this example, and the noise power constrained -- therefore
the estimates are outside those of the minimum risk

quantizer (for this example). The minimax risk quantizer compensates for the corrupting influence of the noise, whereas the Max-Lloyd quantizer does not. For this reason, the estimates for the quantizer levels of the minimax risk quantizer are then placed within those of the Max-Lloyd quantizer.

References

[1]   Bath, W.G., and VandeLinde, V.D., "Robust Memoryless Quantization for Minimum Signal Distortion," IEEE Trans. on Info. Theory, Vol. IT-28, No. 2, March 1982, pp.296-306.

[2]   Berger, T., Rate Distortion Theory, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971.

[3]   Cook, M.K., "A Risk Theory Approach to Quantizer Design," Ph.D. dissertation, Univ. of Arkansas, 1987.

[4]   Cook, M.K., and Jones, R.A., "A Risk Theory Approach to Quantizer Design," submitted to IEEE Trans. on Info. Theory.

[5]   Kanaya, F., and Oishi, S., "Rate Risk Function Theory," IEEE Int. Symposium on Info. Theory, 1985.

[6]   Kazakos, D., "New Results on Robust Quantization," IEEE Trans. on Comm., Vol. COM-31, No. 8, Aug. 1983, pp. 965-974.

[7] Lloyd, S.P., "Least Squares Quantization in PCM," IEEE Trans. on Info. Theory, Vol. IT-28, No. 2, March 1982, pp. 129-137.

[8] Luenberger, D.G., Optimization by Vector Space Methods, John Wiley and Sons, Inc., New York, 1969.

[9] Max, J, "Quantizing for Minimum Distortion," IRE Trans. on Info. Theory, Vol. IT-6, March 1960, pp. 7-12.

[10] Melsa, J.L., and Cohn, D.L., Decision and Estimation Theory, McGraw-Hill Book Co., New York, 1974.

[11] Mood, A.M., Graybill, F.A., and Boes, D.C., Introduction to the Theory of Statistics, McGraw-Hill Book Co., New York, 1974.

[12] Morris, J.M., and VandeLinde, V.D., "Robust Quantization of Discrete-Time Signals with Independent Samples," IEEE Trans. on Comm., Vol. COM-22, No. 12, Dec. 1974, pp. 1897-1902.

[13] Royden, H.L., Real Analysis, Macmillan Publishing Co., Inc., 1968.

[14] Van Trees, H.L., Detection, Estimation and Modulation Theory, Part I, John Wiley & Sons, Inc., New York, 1968.

TABLE 1

Minimax Quantizer Determination
(N = 4, Squared-Error Cost, $p_\theta \sim N(0,1)$)

$\mu_N = 0,$ L = 0.25,   SNR = 8dB $\Rightarrow$ c = 0.1585

x_1	y_0	y_1	Minimax-Risk
0.90	0.41	1.23	0.09931
0.91	0.41	1.24	0.09854
0.92	0.41	1.24	0.09780
0.93	0.41	1.24	0.09708
0.94	0.41	1.24	0.09638
$\longrightarrow$ 0.95	0.42	1.24	0.09571 $\longleftarrow$
0.96	0.42	1.28	0.09743
0.97	0.42	1.31	0.09921
0.98	0.41	1.31	0.09875
0.99	0.41	1.31	0.09818
1.00	0.41	1.31	0.09762
1.01	0.41	1.32	0.09709
1.02	0.41	1.32	0.09657
1.03	0.41	1.32	0.09606
1.04	0.41	1.36	0.09807
1.05	0.41	1.39	0.09973

Max-Lloyd		Minimum Risk[*]		Minimax Risk	
x_i	y_i	x_i	y_i	x_i	y_i
0	0.4874	0	0.4774	0	0.42
1.056	1.625	1.056	1.298	0.95	1.24

[*]Minimum Risk for Gaussian Sources.

TABLE 2

Minimax Quantizer Determination
$(N = 4$, Squared-Error Cost, $p_\theta \sim N(0,1))$

$\mu_N = 0$, $L = 0.25$,   SNR $= 4$dB $\Rightarrow$ $c = 0.398$

x_1	y_0	y_1	Minimax-Risk
1.27	0.41	1.51	0.12232
1.28	0.41	1.53	0.12372
1.29	0.41	1.56	0.12500
1.30	0.41	1.56	0.12447
1.31	0.41	1.56	0.12395
1.32	0.41	1.57	0.12343
1.33	0.41	1.57	0.12294
1.34	0.41	1.57	0.12254
$\longrightarrow$ 1.35	0.41	1.57	0.12198 $\longleftarrow$
1.36	0.41	1.58	0.12348
1.37	0.41	1.60	0.12492
1.38	0.41	1.60	0.12450
1.39	0.41	1.60	0.12408
1.40	0.41	1.60	0.12367
1.41	0.41	1.61	0.12326

Max-Lloyd		Minimum Risk[*]		Minimax Risk	
x_i	y_i	x_i	y_i	x_i	y_i
0	0.5354	0	0.3830	0	0.41
1.160	1.785	1.160	1.277	1.35	1.57

[*]Minimum Risk for Gaussian Sources.
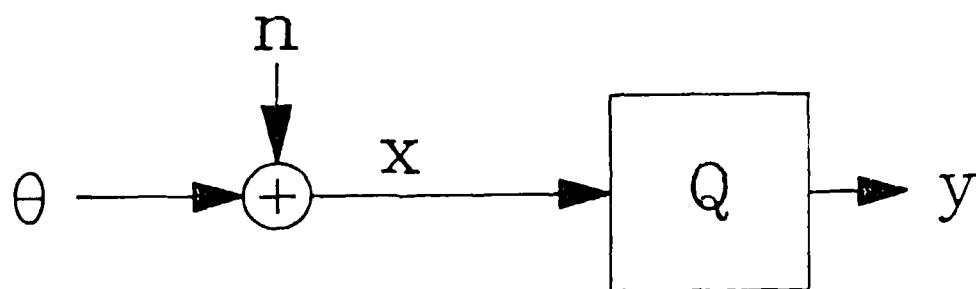
Additive Noise Problem

Figure 1
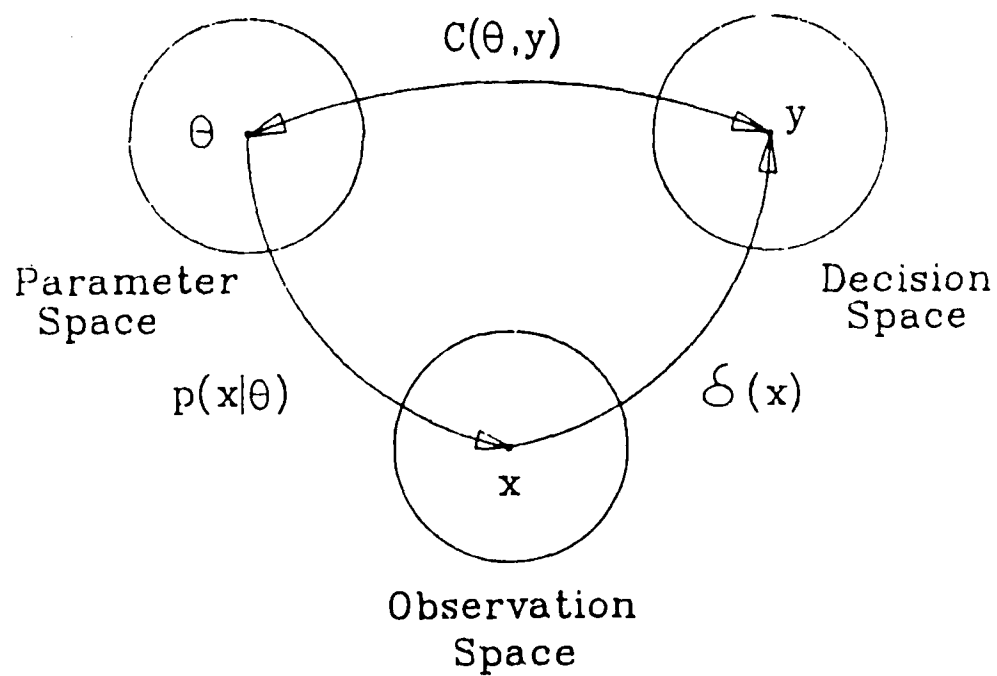
Estimation Problem

Figure 2

Figure 3

PLOT OF I(N)
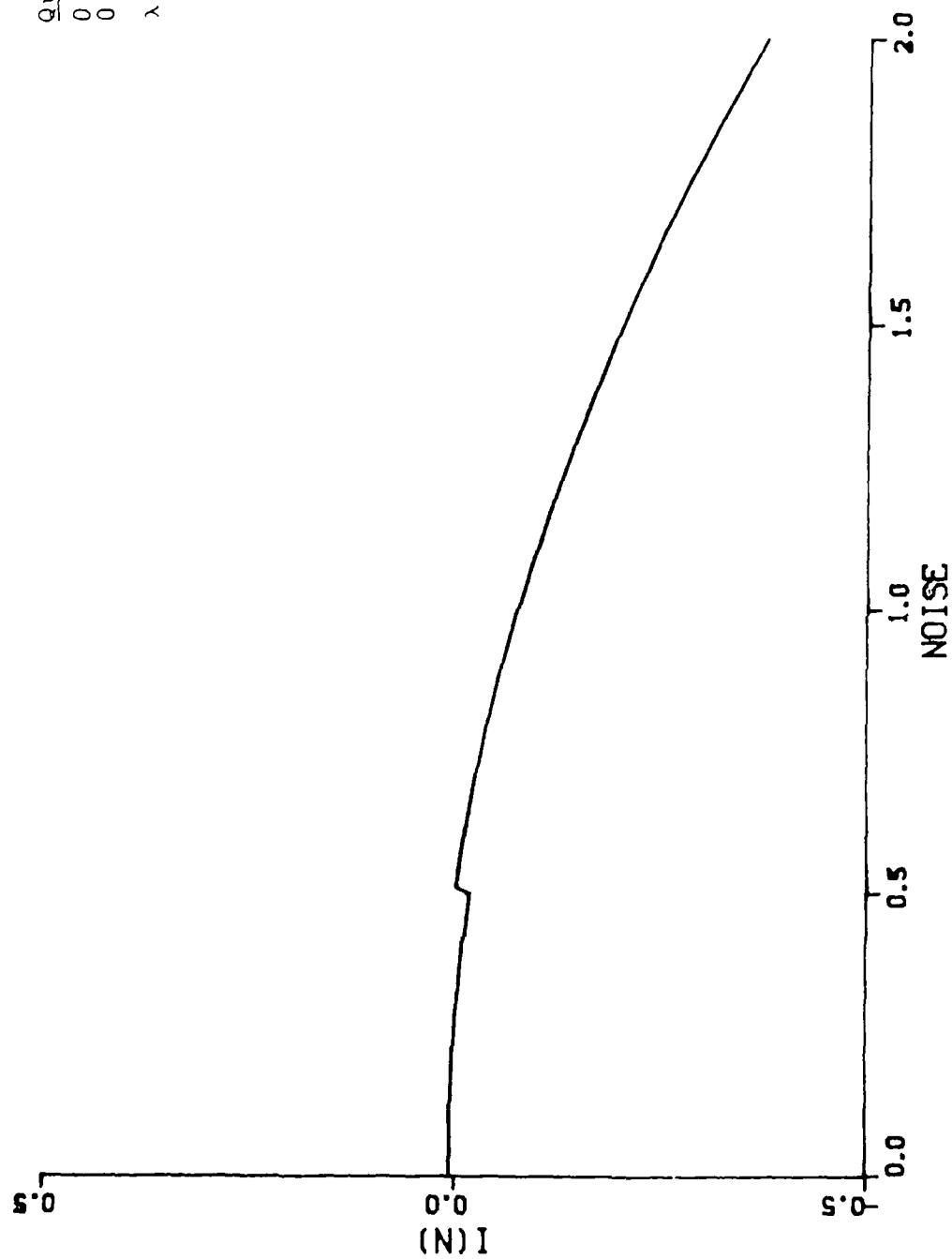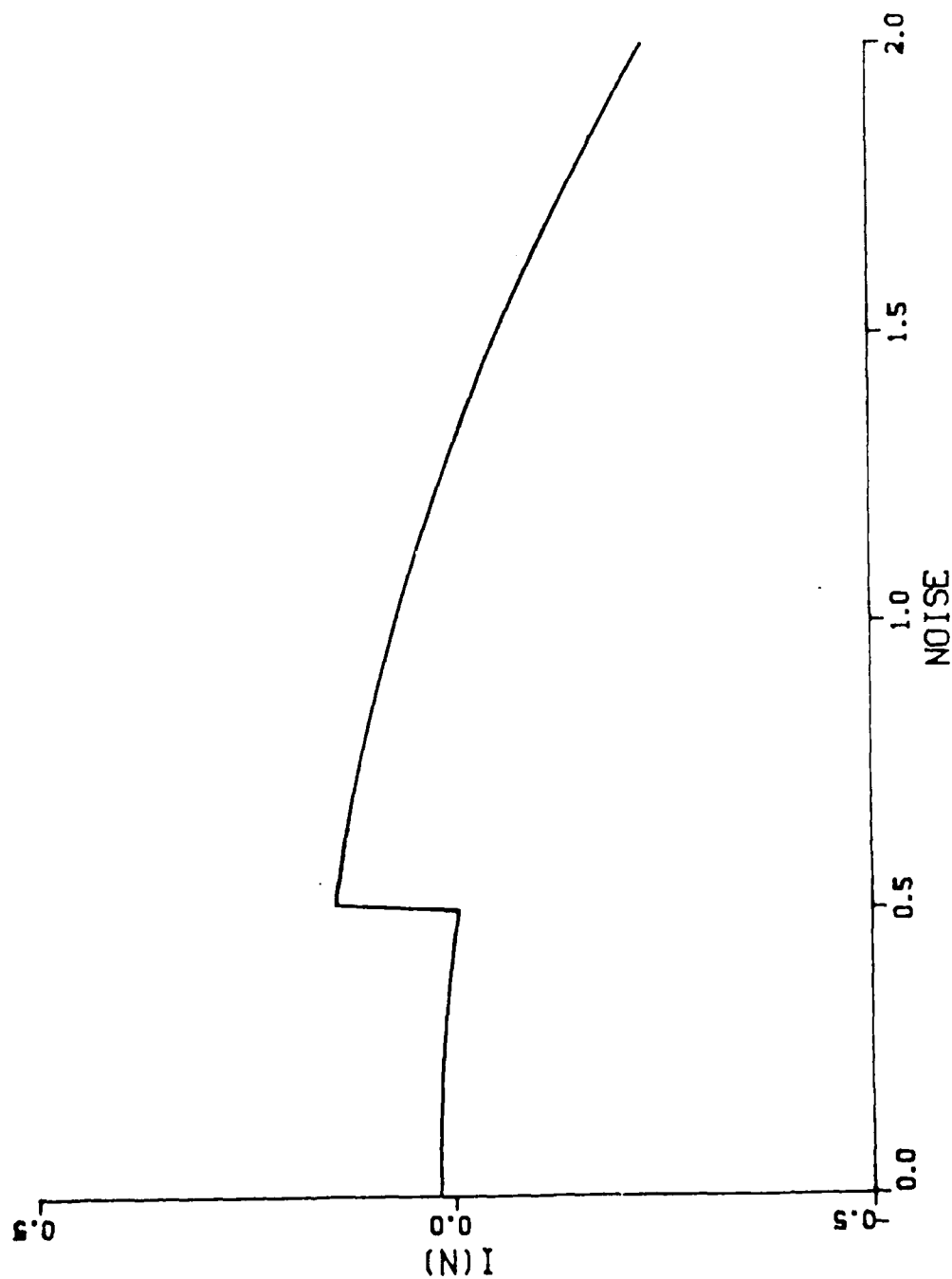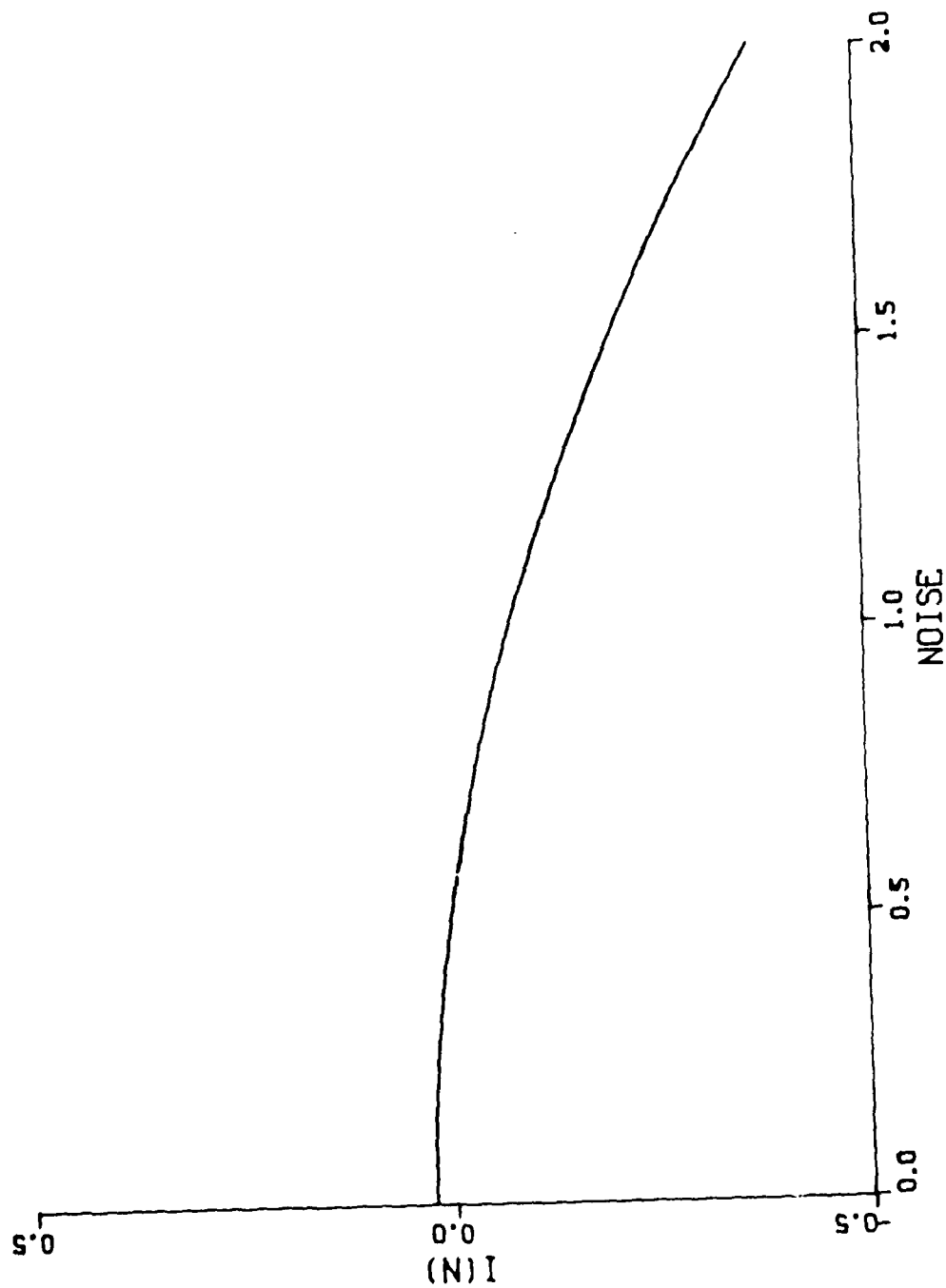
Figure 4

PLOT OF I(N)

Figure 5

PLOT OF I(N)
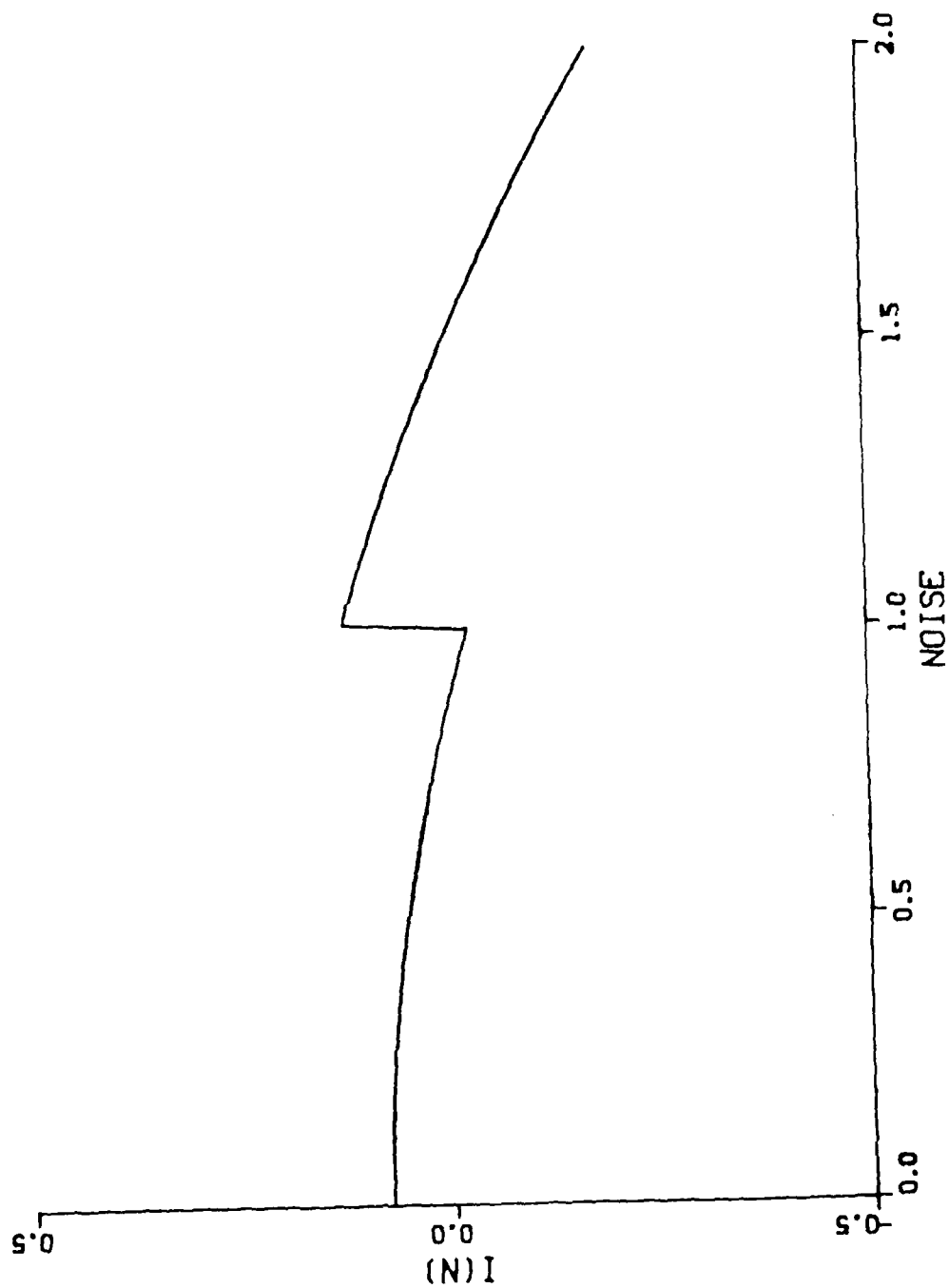
Figure 6

PLOT OF I(N)

Figure 7

PLOT OF I(THETA)

Quantities:
0.   0.5
1.   1.5
$\lambda_2 = 0.05$
$\sigma_N = 0.25$
1.   0.25

I(THETA)

0.5   0.0   -0.5
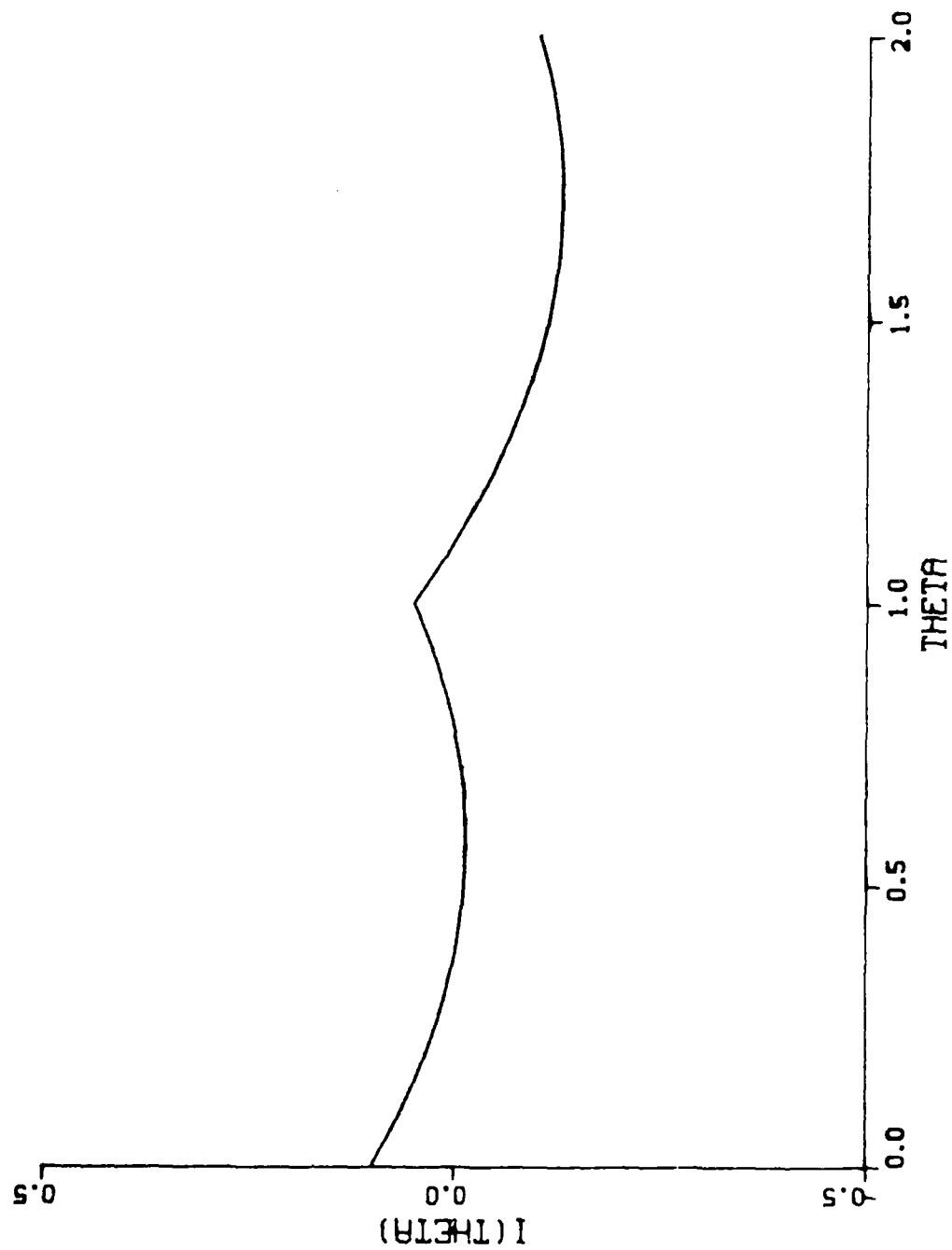
0.0   0.5   1.0   1.5   2.0

THETA

Figure 8

PLOT OF I(THETA)

Quantities
0.   0.5
1.   1.5

$\lambda_2 = 0.05$

$\sigma_y = 1.$

L   0.25

0.5

1.0   1.5   2.0
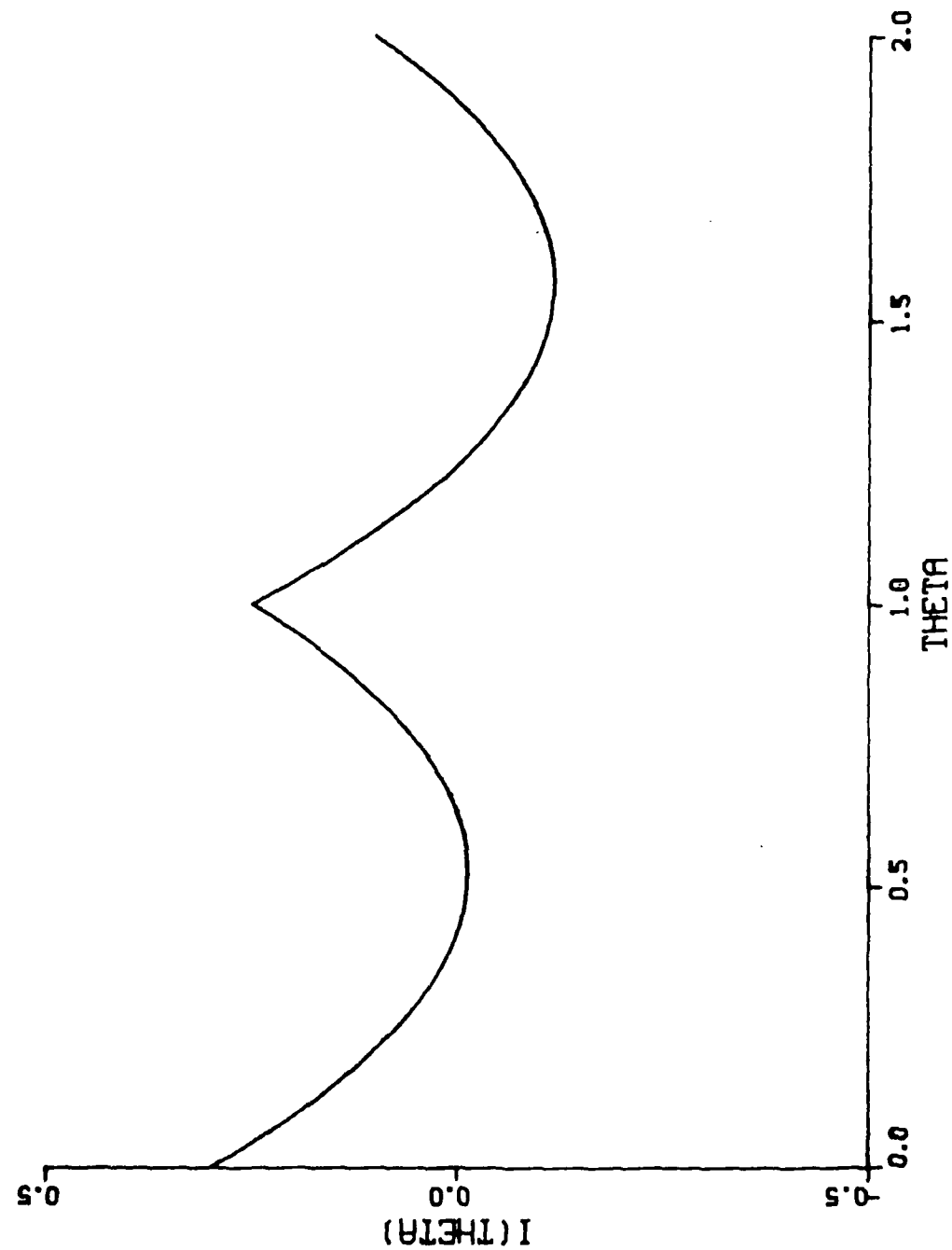THETA
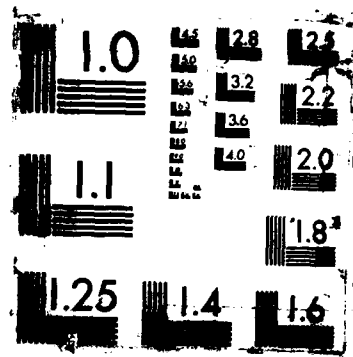
Figure 9

PLOT OF I(THETA)
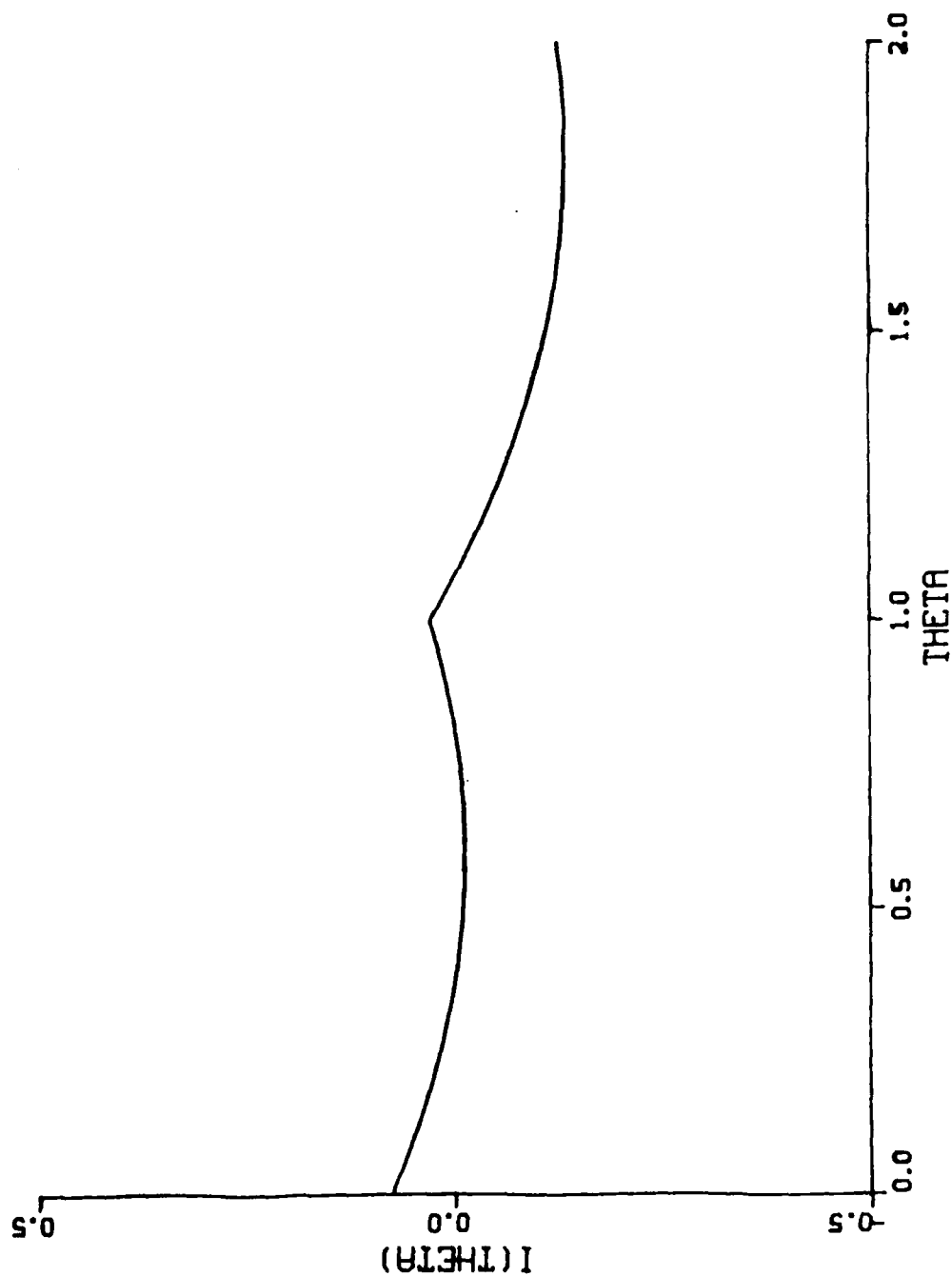
Quantizer

0.   0.5
1.   1.5

$\lambda_2 = 0.05$

$\sigma_N = 0.25$

L   0.75

THETA

I(THETA)

95

Figure 10

PLOT OF I(THETA)

Quantities:
0.   0.5
1.   1.5

$\lambda_2 = 0.05$

$\sigma_N = 1.$

$L = 0.75$

THETA

I(THETA)

0.0   0.5   1.0   1.5   2.0

-0.5   0.0   0.5

END

DATE
FILMED

FEB.
1988