<table>
<thead>
<tr>
<th>REPORT DOCUMENTATION PAGE</th>
<th>READ INSTRUCTIONS BEFORE CompleTING FORM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. REPORT NUMBER</td>
<td></td>
</tr>
<tr>
<td>0203</td>
<td></td>
</tr>
<tr>
<td>2. GOMY ACCESSION NO.</td>
<td></td>
</tr>
<tr>
<td>A22.1117.4</td>
<td></td>
</tr>
<tr>
<td>3. RECIPIENT'S CATALOG NUMBER</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>4. TITLE (and subtitle)</td>
<td>Monotone equivariant segmentation techniques</td>
</tr>
<tr>
<td>5. TYPE OF REPORT &amp; PERIOD COVERED</td>
<td>Theory: 1</td>
</tr>
<tr>
<td>6. PERFORMING ORG REPORT NUMBER</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>7. AUTHOR(S)</td>
<td>M. F. Janowitz</td>
</tr>
<tr>
<td>8. CONTRACT OR GRANT NUMBER</td>
<td>N00014-77-C-0001</td>
</tr>
<tr>
<td>9. PERFORMING ORGANIZATION NAME AND ADDRESS</td>
<td>University of Massachusetts</td>
</tr>
<tr>
<td>Amherst, MA 01003</td>
<td></td>
</tr>
<tr>
<td>10. PROGRAM ELEMENT PROJECT TAC</td>
<td>Area &amp; Work Unit Numbers</td>
</tr>
<tr>
<td>Area 5: Unit 121405</td>
<td></td>
</tr>
<tr>
<td>11. CONTROLLING OFFICE NAME AND ADDRESS</td>
<td>Procuring Contract Officer</td>
</tr>
<tr>
<td>Office of Naval Research</td>
<td></td>
</tr>
<tr>
<td>Arlington, VA 22217</td>
<td></td>
</tr>
<tr>
<td>12. REPORT DATE</td>
<td>Sep. 1982</td>
</tr>
<tr>
<td>14. SECURITY CLS (of this report)</td>
<td>Unclassified</td>
</tr>
<tr>
<td>15. NUMBER OF PAGES</td>
<td>16</td>
</tr>
<tr>
<td>16. DISTRIBUTION STATEMENT (of this Report)</td>
<td>APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED</td>
</tr>
</tbody>
</table>

17. DISTRIBUTION STATEMENT (if the abstract entered in Block 25, if different from Report):

DTIC ELECTED NOV 9 1982

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Image segmentation, thresholding, cluster analysis, monotone equivariance

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

An earlier model for nonhierarchical clustering is extended so that it encompasses the ordinal image segmentation problem. A theoretical discussion within the model is provided for certain segmentation techniques. The discussion is carried out within the framework of a Bernoulli process. The techniques are illustrated on real and simulated data.
1. Introduction. An order theoretic model for hierarchical cluster analysis was introduced in [5] and further developed in [6], [7], and [8]. The model was broad enough to include earlier models due to Jardine and Sibson [11] and Matula [12]. The purpose of the present paper is to show that the model also includes ordinal non-hierarchical clustering, and in particular may be viewed as a model for ordinal image segmentation. Section 2 contains some background material from the theory of partially ordered sets, while section 3 relates this material to the problems of image segmentation. The material is approached via a simple statistical model in section 4, and section 5 relates the present model to certain other segmentation techniques. The final section of the paper outlines how the theory of section 3 may be used to construct a segmentation program, and illustrates the operation of this program on real data.
2. Background material from the theory of partially ordered sets.

A basic familiarity with the theory of partially ordered sets and lattices will be assumed. Despite this, it will be convenient to specifically develop certain notions here. Unless otherwise specified, all partial orders will be denoted \( \leq \), with \( \subseteq \) reserved for set inclusion. The symbol \( P(X) \) will represent the Boolean algebra formed by the subsets of the set \( X \), when ordered by set inclusion. Unions and intersections will be represented as usual by the symbols \( \cup \) and \( \cap \), and it turns out to be convenient to let \( R \) denote the nonnegative real numbers ordered in the usual manner.

Let \( P, Q \) be partially ordered sets. A mapping \( \phi: P \rightarrow Q \) is called isotonous if \( p_1 \leq p_2 \) in \( P \) implies that \( \phi(p_1) \leq \phi(p_2) \) in \( Q \). It is said to be residuated in case it is isotonous and there is an isotonous mapping \( \phi^*: Q \rightarrow P \) such that \( p \leq \phi^*(q) \) and \( q \geq \phi(q) \) hold for every \( p \in P, q \in Q \). This mapping \( \phi^* \) is called the residual mapping associated with \( \phi \), and is uniquely determined by \( \phi \). Likewise, \( \theta \) is completely determined by \( \phi^* \), and for a given residual mapping \( \theta \), it will often be convenient to let \( \theta^* \) denote the residuated mapping with which \( \theta \) is associated. Let \( \text{Res}(P, Q) \) denote the set of all residuated mappings of \( P \) into \( Q \), writing \( \text{Res}(P) \) in place of \( \text{Res}(P, P) \); dually, let \( \text{Res}^*(Q, P) \) denote the set of all residual mappings of \( Q \) into \( P \), with \( \text{Res}^*(Q) \) having its expected meaning. The reason that residuated and residual mappings crop up is very simple. It was shown in [5], Lemma 4.1,
p. 60 that a residual mapping \( C \) from \( R \) into \( P(X) \), where \( X \) is a finite nonempty set, may be characterized by the following three conditions:

1. \( C \) is isotone
2. There is an element \( h \) of \( R \) such that \( C(h) = X \).
3. Given \( h \in R \), there corresponds a positive real number \( \varepsilon \) such that \( C(h) = C(h+\varepsilon) \).

If \( X \) happens to be the set of all 2-element subsets of the set \( P \), then this says that \( C \) being residual is equivalent to \( C \) being a \textit{numerically stratified clustering} in the sense of Jardine and Sibson [11], p. 61.

The theory of residuated mappings is rather extensively developed in [3], and the reader should refer to that source, should further information be needed.

3. The image segmentation problem. The underlying input data may be thought of as a function \( F \) defined on a bounded closed rectangle and taking values either in \( R \) or in \( R^p \) for some positive integer \( p \). The values of \( F \) might represent brightness, temperature, or some other attribute of the rectangle. The function \( F \) is transformed by a sensing device into a function \( F \) having the same range as \( F \), but whose domain is \( X = (1,2,\ldots,m) \times (1,2,\ldots,n) \) for suitable positive integers \( m \) and \( n \). The idea here is to divide the original rectangle into
finitely many subregions by means of a rectangular grid, and on each such subregion, let \( F \) somehow summarize the values of \( I \). The actual input data that is presented to the observer then is not the original function \( F \), but rather the summary function \( F \) defined on \( X \); alternately, the input data may be thought of as an \( m \times n \) matrix taking values in \( R \) or \( R^n \). Unfortunately, the process of going from \( F \) to \( F \) has associated with it some noise, so \( F \) represents a distorted view of the actual data. The idea of the segmentation process is to recapture the principal regions represented by \( F \). A more precise and detailed description of the assumptions relating to this are contained in [9] and [10], for that reason they will not be repeated here. We should mention though that our attention will be restricted to the case where \( I \) takes values in \( R \), with the multifeature situation to be investigated in a later paper.

Let us see how to embed the segmentation problem into the model for hierarchical clustering that was presented in [5]. Think of the input data either as a mapping \( D: X \rightarrow R \), where \( X = \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\} \), or as an \( m \times n \) matrix \( A \) with entries from \( R \). The idea is to somehow represent the information provided by \( A \) as a matrix \( B \) (possibly of a size smaller than \( A \)) that has fewer distinct levels of entries. To properly place this problem in an order theoretic setting, recall ([5], p. 61) that \( D \) extends naturally to a residuated mapping \( D^p: P(X) \rightarrow R \) by letting
$D^*(M) = (D(m): m \in M)$ for all subsets $M$ of $X$. If $D^*$ denotes
the unique residual mapping associated with $D^*$, we have established
a bijection between mappings from $X$ into $R$ and residual mappings
from $R$ into $P(X)$. Thus the image segmentation problem may be
viewed as the study of transformations of $\text{Res}^*(R, P(X))$ into
$\text{Res}^*(R, P(Y))$ where $Y \subseteq X$. Transformations such as this will be
called segmentation methods. This places the segmentation problem
squarely within the framework of the model described in [5].

For input data having only ordinal significance, we would
like to describe the class of segmentation methods that should
properly be used. In order to do this, we need the notion of
compatibility ([5], p. 68). Let $F$ be a segmentation method and
$\rho$ a residual mapping on $R$. To say that $F$ is $\rho$-compatible is to
say that for every $C \in \text{Res}^*(R, P(X))$, $F(C \cdot \rho) = F(C) \cdot C$. For
data having only ordinal significance, one should insist on
$\rho$-compatibility for all order automorphisms of $R$ and this leads
to a class of segmentation techniques that are called monotone
equivariant. These are characterized ([7], Theorem 1, p. 149) as
follows: Let $C \in \text{Res}^*(R, P(X))$. To say that $h (h \cdot R)$ is a
splitting level of $C$ is to say that $h$ is the image of some
subset of $X$ under the residuated mapping $C^*$ associated with $C$.
If $0 < h_1 < \ldots < h_q$ denote the splitting levels of $C$, then to say
that the segmentation method $F$ is monotone equivariant is to say
that:
(i) every splitting level of $F(C)$ is a splitting level of $C$.
(ii) the sequence $F(0) < F(h_1) < \ldots < F(h_t)$ depends only upon the sequence $C(0) < C(h_1) < \ldots < C(h_t)$ and is independent of the actual values of the $h_i$'s.
(iii) conditions (i) and (ii) hold for every $C \in \text{Res}^\times(R, P(X))$.

One can associate with each subset $T$ of $\text{Res}^\times(R)$ the collection $\alpha(T)$ of all segmentation methods that are $\alpha$-compatible for every $\alpha \in T$; conversely, for any set $S$ of segmentation methods, one can associate the set $\beta(S)$ of all $\alpha \in \text{Res}(R)$ such that $F$ is $\alpha$-compatible for every $F$ in $S$. Naturally, for this to be meaningful, one must think of a segmentation method as a mapping $F: \text{Res}^\times(R, P(X)) \rightarrow \text{Res}^\times(R, P(Y))$ where $Y$ is a fixed subset of $X$.

In any event, the mappings $(\alpha, \beta)$ define a Galois correspondence in the sense of [2], p. 124 between the residual mappings on $R$ and the segmentation methods on $X$. It generally turns out that the Galois closed objects (i.e., those objects that appear as images of $n$ or $s$) of such correspondences have some significance. We have already seen that the monotone equivariant cluster methods appear as the image under $\alpha$ of the set of order automorphisms of $R$, so they are Galois closed. F. Baulieu [1] has among many other things characterized all closed classes that are contained in this one. Here is a description of these classes:

(S1) **Flat methods.** $F(h)$ depends only upon $C(h)$.

(S2) **Semiflat methods.** $F(h)$ depends upon $C(h)$ and $C(0)$. 
(S3) **Divisive methods.** $F(h_i)$ depends upon $C(h_i), C(h_{i+1}), \ldots, C(h_t)$ where $h_t$ is the highest splitting level of $C$.

(S4) **SF(1) methods.** $F(0)$ depends only upon $C(0)$; for $h_i \neq 0$, $F(h_i)$ depends on $C(h_i), C(h_{i+1}), \ldots, C(h_t)$.

(S5) **Monotone equivariant methods.**

The related sets of residual mappings, (T1) corresponding to (S1), are:

(T1) $\text{Res}(R)$

(T2) The set of all residual mappings $u$ on $R$ for which $u(0) = 0$.

(T3) Those residual maps whose range is of the form $\{h: h \leq k\}$ for some fixed element $k$ of $R$; i.e., whose range is a principal filter of $R$.

(T4) Those residual maps whose range is the union of $\{0\}$ with a principal filter of $R$.

(T5) The set of all injective residual mappings on $R$.

The simplest of the closed classes of segmentation methods is of course the flat methods, and it is to this class that we now direct our attention. Needless to say, such methods are easy to describe and easy to implement. We are given a fixed finite nonempty set $\mathcal{X}$ and an increasing sequence

$$\mathcal{X}_1 \subseteq \mathcal{X}_2 \subseteq \cdots \subseteq \mathcal{X}_t = \mathcal{X}.$$
of subsets of $X$, where $M_1 = C(h_1)$, and wish to produce a sequence

$$N_1 \supseteq N_2 \supseteq \ldots \supseteq N_t = X$$

of subsets of $X$ that somehow summarize or better represent the underlying picture than does the input data to the problem. In that $N_t$ depends only upon $M_1$, this amounts to defining an isotone mapping $\gamma$ on $\mathcal{P}(X)$.

Having recognized the need to define an isotone mapping on $\mathcal{P}(X)$, it is appropriate to decide upon reasonable candidates for these mappings. The idea is to attach some spatial significance to the decision as to whether a given point belongs to the output of the mapping. In other words, the decision as to whether $x$ belongs to $\gamma(M)$ should be based upon all points in some small region surrounding $x$. This may be precisely stated by saying that the mapping $\gamma$ shall be point-based in that for each $x \in \gamma(M)$ there is a subset $S(x)$ of $X$ containing $x$ such that:

(PB1) $\gamma(x) \neq \emptyset$ and $\gamma(S(x)) \neq \emptyset$.

(PB2) For $A \subseteq S(x)$, if $\gamma(A) \neq \emptyset$, then $x \in \gamma(A)$.

(PB3) For $M \subseteq X$, $x \in \gamma(M)$ if and only if $x \in \gamma(M \setminus S(x))$.

It is immediate that for $M \subseteq X$, $\gamma(M) = \bigcup_{x \in \gamma(M)} S(x)$. So that the output will be independent of the polarity of the input image, it sometimes turns out to be useful to insist that $\gamma$ preserve complements in that

(PB4) $\gamma(M) = \gamma(X \setminus M) \cap \gamma(X \setminus \gamma(M))$. 
The local version of (PB4) is the content of Theorem 1.

**Theorem 1.** For a point-based isotone mapping $\gamma$ on $\mathcal{P}$, the axiom (PB4) is equivalent to the assertion that for each subset $A$ of $N(x)$, exactly one of $\gamma(A)$ and $\gamma(N(x) \setminus A)$ will be nonempty.

**Proof:** Let $\gamma$ satisfy (PB4). If $x \in (A) \cap (N(x) \setminus A)$, then $x \notin (N(x) \setminus A)$, contrary to (PB4), for the converse implication, assume that $\gamma$ is point-based and that for $A \subseteq N(x)$, exactly one of $\gamma(A)$ and $\gamma(N(x) \setminus A)$ is nonempty. If now $y \in (X \setminus N)$, then $y \notin (X \setminus N) \cap (N(y))$, so $y \notin \gamma(N(y))$, and consequently, $y \notin \gamma(N)$. Conversely, $y \notin \gamma(N) \rightarrow y \notin \gamma(N \setminus (N(y)))$, so $y \in (N \setminus (N(y)))$.

Now let $\gamma$ be point-based with an associated family of neighborhoods $N(x)$, $x \in \gamma(X)$. If $M(x)$ is a second such family of neighborhoods, it is clear that so also is $M(x) \cap N(x)$. In view of this, there is no harm in assuming

(PB3) If $M(x)$ satisfies (PB1) through (PB3), then $N(x) \subseteq M(x)$.

Such a minimal family of subsets of $X$ will be called the system of neighborhoods associated with $\gamma$. Unless otherwise specified, when we speak of a point-based isotone mapping $\gamma$, it will always be assumed that the family $(N(x) : x \in \gamma(X))$ represents this system of neighborhoods.
Definition. The point-based isotone mapping \( \gamma \) is said to be 
frequency-defined if there exist positive integers \( j \) and \( k \) such 
that for each \( x \in \gamma(I) \),

(i) \( \#N(x) = k \), and

(ii) \( x \leq N(M) \) for \( M \leq X \) if and only if \( j \leq \#(M \cdot N(x)) \).

Here \( \#A \) denotes the cardinality of the set \( A \).

Theorem 2: Let \( \gamma \) be a point-based, complement-preserving 
isotone mapping on \( P(X) \). Suppose that the regions \( N(x) \) all have 
the same cardinality, and that for each \( x \in \gamma(I) \), \( A_x \) has minimal 
cardinality among those subsets \( A \) of \( N(x) \) for which \( x \leq \gamma(A) \).

necessary and sufficient conditions for \( \gamma \) to be frequency-based 
are that \( k = \#N(x) \) be odd, and that \( \#A_x = (1+2)/2 \) for all 
\( x \in \gamma(I) \).

Proof: Let \( \gamma \) be frequency-based. If \( \#A_x = k/2 \), then 
\( \#(N(x) \cdot A_x) = k/2 \), and there is a subset \( B \) of \( N(x) \cdot A_x \) such that 
\( \#B = \#A_x \). Since \( \#(B) = q \), this produces a contradiction. If \( k 
\) is even, taking \( B \subseteq N(x) \) with cardinality \( k/2 \) produces a similar 
contradiction. Thus \( k \) is odd and \( \#A_x = k/2 \). If \( \#A_x = (1+k)/2 \), 
we may take \( B \subseteq A_x \) so that \( \#B = (1+k)/2 \). Since \( \#(N(x) \cdot B) = \#B \), 
this tells us that \( \gamma(B) = \gamma(N(x) \cdot B) \), contrary to (184). Thus 
\( \#A_x = (1+k)/2 \). Suppose conversely that \( k = \#N(x) \) is 
odd and that \( \#A_x = (1+k)/2 \). We are to show that ...
is frequency-based. By construction of $A_x$, $x \cdot \gamma(B)$ for $B : N(x)$ implies that $yB \cdot x = (1+k)/2$. If on the other hand, $yB = (1+k)/2$, then $y(N(x):B) < (1+k)/2$, so $\gamma(N(x):B) = 1$, and consequently, $x \cdot \gamma(B)$.

One is often interested in the extent to which a segmentation algorithm blurs an image. The merging of clusters that should not be merged is one of the ways in which such blurring occurs. To illustrate this concretely, let us consider a frequency-defined mapping that operates on 3 by 3 neighborhoods centered on points.

Suppose $X$ has 5 rows and 5 columns, and that $M, N$ are the regions denoted by 1's, 2's in Fig. 2(a). In Fig. 2(b), the members of $(M), \gamma(M)$ are denoted by 1's, 2's, while in Fig. 2(c), the members of $(M:N)$ are indicated by 3's. All remaining entries are 0. Notice that $(M:N) \cdot \gamma(M) \cdot \gamma(N)$. In connection with this, we shall say

\[
\begin{array}{ccc}
1 & 1 & 1 \\
1 & 1 & 1 \\
0 & 0 & 0 \\
2 & 2 & 2 \\
2 & 2 & 2 \\
\end{array}
\begin{array}{ccc}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\end{array}
\begin{array}{ccc}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\end{array}
\]

(a) (b) (c)

Fig. 2 Illustration of blurring. See text for explanation.

that $M, N$ are $\gamma$-separated for the isotone point-based mapping $\gamma$, in case there is no set of the form $N(x)$ that intersects both $M$ and $N$. We then have
Theorem 3. Let \( \gamma \) be a point-based isotone mapping on \( P(X) \). If \( M, N \) are \( \gamma \)-separated, then \( \gamma(M \cup N) = \gamma(M) \cup \gamma(N) \). 

Proof: Evidently, we need only show that \( \gamma(M \cup N) \supseteq \gamma(M) \cup \gamma(N) \). Accordingly, let \( x \in \gamma(M \cup N) \), so \( (M \cup N) \cap \gamma^{-1}(x) \neq \emptyset \). Since \( M, N \) are \( \gamma \)-separated, we must have \( (M \cup N) \cap \gamma^{-1}(x) = M \cap \gamma^{-1}(x) \) or \( M \cap \gamma^{-1}(x) = \emptyset \), whence \( x \in \gamma(M) \cup \gamma(N) \).

Along these lines, the mapping \( \gamma \) will be called a join homomorphism in case \( \gamma(M \cup N) = \gamma(M) \cup \gamma(N) \) for all subsets \( M, N \) of \( X \). There is a dual notion of meet homomorphism, and to say that \( \gamma \) is a homomorphism is to say that it is both a join and a meet homomorphism. The next theorem shows that these conditions are extremely powerful, and will only be met in trivial situations.

Theorem 4. Let \( \gamma \) be a point-based isotone mapping on \( P(X) \). Then (1a) \( \rightarrow \) (1b), (2a) \( \leq \) (2b) and (3a) \( \rightarrow \) (3b).

(1a) \( \gamma \) is a join homomorphism.

(1b) For each \( x \in \gamma(X) \) there is a subset \( A_x \) of \( N(x) \) such that \( x \in \gamma(M) \) if and only if \( x \cap A_x = \emptyset \).

(2a) \( \gamma \) is a meet homomorphism.

(2b) Corresponding to each \( x \in \gamma(X) \), there is a subset \( B_x \) of \( N(x) \) such that \( x \in \gamma(M) \) if and only if \( x \cap B_x = \emptyset \).

(3a) \( \gamma \) is a homomorphism.
(3b) For each \( x \in \gamma(X) \) there corresponds an element \( \gamma_x \) of \( N(x) \) such that \( x \cdot \gamma(M) \) if and only if \( \gamma_x \cdot M \).

Proof: (1a) \( \Rightarrow \) (1b). If \( \gamma \) is a join homomorphism, let \( C \) be the union of all subsets of \( N(x) \) that are mapped to \( x \) by \( \gamma \), and note that \( \gamma(C) = \phi \). Observing that \( \gamma(M \cdot N(x)) = \phi \) if and only if \( M \cdot N(x) \subseteq C \), we may now take \( A_x = N(x) \cdot C \).

(1b) \( \Rightarrow \) (1a) If \( x \cdot \gamma(N) \), then \( M \cdot A_x = \phi \), so \( M \cdot A_x \) or \( N \cdot A_x = \phi \). Hence \( x \cdot \gamma(M) \cup \gamma(N) \).

(2a) \( \Rightarrow \) (2b) If \( B_x \) is the intersection of all subsets \( C \) of \( N(x) \) for which \( \gamma(C) = \phi \), then \( x \cdot \gamma(C) \) for every \( \gamma(C) \), so \( x \cdot \gamma(B_x) \).

(2b) is now clear.

(2b) \( \Rightarrow \) (2a) is obvious, as is (3b) \( \Rightarrow \) (3a).

(3a) \( \Rightarrow \) (3b) If \( B_x \) is defined as in (2), then \( \gamma(B_x) = \phi \), but \( \gamma(C) = \phi \) for every proper subset \( C \) of \( B_x \). It is immediate that \( B_x \cdot (\gamma_x) \), and (3b) now follows.

Remark 5. Noting that if \( \gamma \) in the theorem is complement-preserving, then it is a join homomorphism if and only if it is a meet homomorphism, it follows that for such mappings the six conditions of the theorem are mutually equivalent.

If \( M \subseteq X \) is acted upon by a rigid motion to produce \( N \), one would naturally want that same rigid motion acting upon \( \gamma(M) \) to produce \( \gamma(N) \). A careful formulation of this idea shows that this
is too much to hope for. To see why, we need the concept of a translation on $X$: this is a mapping of the form $T_{p,q}$ where $p$ and $q$ are fixed integers and $T_{p,q}(i,j) = (i+p, j+q)$. Unless $p = q = 0$, the domain and range of $T_{p,q}$ will be proper subsets of $X$. We agree to call the point-based isotone mapping translation-invariant provided it satisfies:

**(PB6)** If $N(x)$ is contained in the domain of the translation $T$, and if $y = T(x)$, then $N(y) = T(N(x))$, and for $A \subseteq N(x)$, $x \cdot \gamma(A)$ if and only if $y \cdot \gamma(T(A))$.

**Remark 6.** If $\gamma$ is a point-based, translation-invariant isotone mapping on $P(X)$, and if $M \subseteq$ domain (1) with $M \cdot T(M) \subseteq \gamma(X)$, it is true that $T(M) = \gamma(T(M))$. This may fail, however, if either $M$ or $T(M)$ is not contained in $\gamma(X)$. This is illustrated in Fig. 3. Here $X = S \times S$, where $S = \{1,2,3,4,5\}$. The mapping $\gamma$ is defined by looking at 3 by 3 neighborhoods centered on the points of $X$, and saying that for $A \subseteq N(x)$, $x \cdot \gamma(A)$ iff $A \supseteq 5$. This choice of $\gamma$ satisfies (PB1) through (PB6), and is even frequency-defined. Let $x = (2,2)$ and $T = T_{1,1}$, so that $T(x) = (3,3)$. Then Fig. 3(a) shows $N(x)$, Fig. 3(b) shows $T(N(x)) = N(T(x))$, while $\gamma(N(x))$ and $\gamma(T(N(x)))$ are displayed in (c) and (d). The reason that $T(\gamma(N(x))) \neq \gamma(T(N(x)))$ is that $T(N(x))$ is contained in $\gamma(X)$, while $N(x)$ is not.
Remark 7. For translation-invariant mappings \( \gamma \) on \( P(X) \), the conditions expressed in Theorems 2, 3 and 4 can be simplified, as one need only state them for a single member of \( P(X) \). The details of this are left to the reader. We also mention without proof that versions of the above results are true for atomistic orthomodular lattices ([2], p. 53).

Remark 8. The actual construction of a point-based isotone mapping on \( P(X) \) can now be easily understood. One chooses a system of neighborhoods \( N(x) \) for points \( x \) in some subset \( Y \) of \( X \), and defines a family \( (\gamma_x)_{x \in Y} \), where \( \gamma_x \) is an isotone mapping on \( N(x) \) such that:

1. \( \gamma_x(\emptyset) = \emptyset \), and \( \gamma_x(N(x)) = \{x\} \).

If one wants to produce a complement-preserving mapping, then one also wants

2. If \( A \subseteq N(x) \), then exactly one of \( \gamma_x(A) \) and \( \gamma_x(N(x) \setminus A) \) is non-empty.
The mapping $\gamma$ is now defined by the rule $x \mapsto \gamma(M)$ if and only if $\gamma_x(M \cup N(x)) = \{x\}$. This is the technique that will be used for the remainder of the paper.

Remark 9. Though it might not necessarily make sense to use point-based mappings that are not frequency-defined, it is easy to construct examples of them. To see this, let $N(x)$ be a 3 by 3 region centered on $x$. For $A \subseteq N(x)$, let $\gamma_x(A) = \{x\}$ if at least 2 out of the 3 rows of $N(x)$ each has at least 2 members of $A$; otherwise, let $\gamma_x(A) = \emptyset$. Do this for each $x$ that is not in an outer row or column of $X$, and use the construction of Remark 8 to define a mapping $\gamma$ on $P(X)$. The result is a point-based, complement-preserving isotone mapping that is not frequency-defined. Clearly this example is representative of an entire class of such mappings. Other candidates for $\gamma$ are pattern-defined mappings. These put $x$ in $\gamma_x(A)$ if $A$ contains a subset of a desired type. For 3 by 3 neighborhoods, such desired types might include sets such as

\[
\begin{array}{ccc}
1 & 1 & 1 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{array}
\quad \quad \quad
\begin{array}{ccc}
1 & 1 & 0 \\
1 & 0 & 0 \\
0 & 0 & 0
\end{array}
\]

4. Underlying statistical considerations. We have just seen that the construction of a flat segmentation method involves the
definition of an isotone mapping on \( P(X) \). If the input data were correct, then the most reasonable choice for such a mapping might be the identity map or a map of the form \( M \rightarrow Y \) for some subset \( Y \) of \( X \). The idea though is to use the isotone mapping \( \mathcal{R} \) to produce an estimate of the true data that is in some sense better than the one provided by the input data. Specifically, let us assume that the data represents a subset \( M^* \) of \( X \), but that the input data produced \( M \) instead. One wants to estimate \( M^* \) by \( \mathcal{R}(M) \). As in the last section, it is assumed that \( X = \{1,2,...,m\} \times \{1,2,...,n\} \). Let \((N(x)) \) be a system of neighborhoods of points \( x \) in some subset \( Y \) of \( X \). To say that a point \((i,j)\) is an interior point of the set \( M^* \) will be to say that for \( x = (i,j) \), \( N(x) \subseteq M^* \); a similar definition applies to the complement of \( M^* \). If \( N(x) \) intersects both \( M^* \) and its complement, then \( x \) will be called a boundary point of \( M^* \). Thus \( X \) is divided into 4 regions: (1) the interior of \( M^* \), (2) the interior of \( X \setminus M^* \), (3) the boundary of \( M^* \), and (4) those points which do not have neighborhoods (i.e., \( X \setminus Y \)). This is concretely illustrated in Fig. 4. Here we take \( m = 10 \), \( n = 7 \), and \( N(x) \) a 3 by 3 region centered on \( x \). \( Y \) would then consist of all points not in an outer row or column of \( X \). The members of \( M^* \) are denoted by 1's in Fig. 4(a) with those of \( X \setminus M^* \) represented by 0's. The resulting 4 regions are displayed in Fig. 4(b) and are numbered as above.
A crude model may be constructed by assuming that membership in
M has probability $p \ (p > 0.5)$ of providing a correct estimate of
membership in $M^*$, and that membership in $X \setminus M$ has a probability
$q \ (q > 0.5)$ of providing a correct estimate for membership in $M$. Let
us also assume that these probabilities are each independent, bearing
in mind that in practice this is sometimes not so. Realistically,
one should also notice that boundary points will have a lower
probability of correct classification than do interior points, but
this is an issue that will be temporarily ignored. The gain of an
isotone mapping $\gamma$ is defined to be the sum of the probability of
correct classification by $\gamma$ for an interior point of $M^*$ and an
interior point of $X \setminus M^*$, and is denoted $G(\gamma)$. Until further notice,
it will be assumed that $\gamma$ is a point-based, translation-invariant
isotone mapping on $P(X)$. It will prove convenient for a fixed $N(x)$
to let $T_\gamma = \{A: A \subseteq N(x), x \in \gamma(A)\}$ and $F_\gamma = \{A: A \subseteq N(x), \gamma(A) = \emptyset\}$.
Letting $A^i$ denote a typical subset of $\mathcal{N}(x)$ having cardinality $i$, we have then have

$$G(y) = \varepsilon(p^{i-1} \cdot A^i \cdot T_y) + \varepsilon((1-q)^{i-1} \cdot A^i \cdot F_y),$$

where $k = \#N(x)$. In order to see the role of frequency-defined mappings in this model, we present

**Lemma 10.** Let $y$ be complement-preserving with $k = \#N(x)$ odd, and let $y'$ be the frequency-defined, point-based isotone mapping having the same neighborhood system as $y$. Then $G(y) \leq G(y')$.

**Proof:** Since $y$ is complement-preserving, $\#T_y = \#F_y = 2^{k-1}$. The same is true for $y'$. Thus a bijection $\phi$ may be defined on $P(N(x))$ such that $\phi$ maps $T_y$ onto $T_{y'}$ and $F_y$ onto $F_{y'}$. We may even take $\phi$ to be the identity map on $(T_y \cap T_{y'}) \cup (F_y \cap F_{y'})$.

For $A \in T_y \cap F_{y'}$, $\#A < j = (1+k)/2$, while $\#\phi(A) > j$, so $\#A < \#\phi(A)$. A similar argument shows that for $A \in F_y \cap T_{y'}$, $\#A > \#\phi(A)$. Using $\phi$ to match the terms of $G(y)$ with those of $G(y')$, it is clear that each term of $G(y)$ is no larger than the corresponding term of $G(y')$, so $G(y) \leq G(y')$.

**Theorem 11.** If $p = q$, and if $y$ and $y'$ have the same system $N(x)$ of neighborhoods with $y'$ frequency-defined, then $G(y) \leq G(y')$. 
Proof: In view of Lemma 10, we need only produce a complement-preserving point-based isotone mapping \( \gamma'' \) with \( N(x) \) as neighborhoods such that \( G(\gamma) < G(\gamma'') \). This will be accomplished by making use of Remark 8. Choose some fixed \( N(x) \), and consider subsets \( A \) of \( N(x) \). If \( x \in A \cdot \gamma(N(x) \setminus A) \) with \( \cdot j \), take \( \gamma''(A) = \{ x \} \) and \( \gamma''(N(x) \setminus A) = \emptyset \). If \( \gamma(A) = \emptyset = \gamma(N(x) \setminus A) \) with \( \cdot A \cdot j \), do the same. Otherwise, let \( \gamma''(A) = \{ x \} \) if \( x \cdot \gamma(A) \) and \( \gamma''(A) = \emptyset \) if \( \gamma(A) = \emptyset \). We need to show that \( \gamma'' \) is isotone. The proof will be by contradiction. Suppose then that \( A \subset B \subset N(x) \) with \( \gamma''(A) \subset \gamma''(B) \). Then \( \gamma''(A) = \{ x \} \) and \( \gamma''(B) = \emptyset \). By the construction of \( \gamma'' \), we must either have \( x \cdot \gamma(A) \) or \( \#A \geq j \).

Case 1. \( x \cdot \gamma(A) \). Then also \( x \cdot \gamma(B) \), so if \( \gamma''(B) = \emptyset \), then \( x \cdot \gamma(B) \subset \gamma(N(x) \setminus B) \) and \( \#B < j \). Since \( A \subset B \) implies \( N(x) \setminus A > N(x) \setminus B \), this puts \( x \) in \( \gamma(N(x) \setminus A) \), whence \( \gamma''(A) = \emptyset \), a contradiction.

Case 2. \( \#A \geq j \). Then also \( \#B \geq j \). The only way for \( \gamma''(B) = \emptyset \) is for \( \gamma(B) = \emptyset \), so \( \gamma(A) = \emptyset \) and consequently \( \gamma''(A) = \emptyset \), again a contradiction.

This then establishes that \( \gamma'' \) is isotone. The construction of Remark 8 can now be used to extend \( \gamma'' \) to a point-based, complement-preserving isotone mapping on \( P(X) \) having the same neighborhood system as \( \gamma \). We must still show that \( G(\gamma) < G(\gamma'') \). The changes in these sums occur from members of \( (T_{\gamma} \cap F_{\gamma'} \cup (T_{\gamma'} \setminus F_{\gamma}) \). If \( A^i \in T_{\gamma} \cap F_{\gamma'} \), then \( i < j \), and the term corresponding to \( A^i \) in
G(y) is \( p^i(1-p)^{k-i} \), while the corresponding term in \( G(y') \) is \( p^{k-i}(1-p)^i \). In that \( i \cdot k \cdot i \) and \( p > 0.5 \), this represents an increase. Similarly, if \( A^i \cdot T, F \), then \( i \cdot j \), and again the term from \( G(y') \) represents the corresponding term from \( G(y) \).

When \( p = q \), the above theorem shows that the "gain" from a flat segmentation method can be maximized by using a point-based isotone mapping that is complement-preserving and frequency-defined. This amounts to saying that such mappings maximize the probability of correct classification of interior points, as this probability is just half the gain. With this thought in mind, we shall direct our attention to mappings of the above type defined on \( k \) by \( k \) neighborhoods of points with \( k \) an odd integer. The \( j/k^2 \) rule with \( j = (1+k^2)/2 \) will be the unique such mapping on a \( k \) by \( k \) neighborhood, and the 3/5 rule will refer to the unique \( \gamma \) defined on the region consisting of \( x \) and the 4 points that are immediately to its North, South, East and West. These rules were introduced in a slightly different context in [9], and some discussion given to the probability of correct classification of points in the case of interior points as well as boundary points. These probabilities are all based on an underlying binomial distribution of \( N(x) \), and will not be repeated here.

There is another item worth mentioning, however. To improve the classification of interior points, one simply enlarges the size of the \( k \) by \( k \) region. This is fine except that as \( k \) increases
the number of boundary points also increases, and as was shown in [9], the probability of misclassification of boundary points can actually increase when the \( j/k^2 \) rule is used. This is illustrated by some simulations in Table 1. The segmentations are done on the data of Fig. 6(b) corrupted by various types of additive noise with the indicated signal to noise ratio. The theoretical probability of misclassification of interior points is compared to the actual probability based on 10 trials, and this is compared to the probability of misclassification in the entire picture. Note how the ratio of the probability for misclassification of interior points over that of the entire picture changes as the value of \( k \) increases from 5 to 9 to 25.

All of this suggests that there may be occasions when one wants to use a flat segmentation method other than that provided by the \( j/k^2 \) rule. For 3 by 3 regions, one idea might be to define \( \gamma \) on \( \text{MAX} \) as follows: \( \gamma(A) = |x| \) if \( \#A \geq 6 \) or if \( \#A = 4 \) or \( 5 \) and \( x \neq A \). Naturally this will increase the probability of misclassification of interior points, but may do a slightly better job near the boundary of a region. This is illustrated in Table 2, where this rule is compared to the \( 5/9 \) rule for regions containing various mixtures of members of \( M \) and its complement. Notice the rather dramatic increase on probability of correct classification near the boundary as evidenced by the box relating to \( \#M = 5 \). A comparison of the two methods on real data is displayed in Fig. 5.
Table 1. Simulated data based on 10 trials. Entries in the table represent probabilities of misclassification. The column labelled SD refer to the standard deviation of this probability, and the "Ratio" column is simply the observed probability for interior points divided by the observed probability for the entire picture. The columns headed "Actual Image" refer to Fig. 6 with additive gaussian noise having the indicated signal to noise ratio. The columns apply to the identical noise on the interior of a single region. The signal to noise ratio is the difference in grey levels that characterize the input set from its background divided by the standard deviation of the noise.

<table>
<thead>
<tr>
<th>Noise Ratio</th>
<th>Rule</th>
<th>Actual Image</th>
<th>Interior Points</th>
<th>Ratio of Interior/Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.25</td>
<td>None</td>
<td>.2697</td>
<td>.0203</td>
<td>.2820</td>
</tr>
<tr>
<td></td>
<td>3/5</td>
<td>.1607</td>
<td>.0290</td>
<td>.1401</td>
</tr>
<tr>
<td></td>
<td>5/9</td>
<td>.1233</td>
<td>.0227</td>
<td>.0783</td>
</tr>
<tr>
<td></td>
<td>13/25</td>
<td>.0947</td>
<td>.0146</td>
<td>.0102</td>
</tr>
<tr>
<td>1.00</td>
<td>None</td>
<td>.3100</td>
<td>.0199</td>
<td>.3178</td>
</tr>
<tr>
<td></td>
<td>3/5</td>
<td>.2218</td>
<td>.0336</td>
<td>.1874</td>
</tr>
<tr>
<td></td>
<td>5/9</td>
<td>.1784</td>
<td>.0271</td>
<td>.1221</td>
</tr>
<tr>
<td></td>
<td>13/25</td>
<td>.1381</td>
<td>.0272</td>
<td>.0283</td>
</tr>
<tr>
<td>0.67</td>
<td>None</td>
<td>.3810</td>
<td>.0306</td>
<td>.3857</td>
</tr>
<tr>
<td></td>
<td>3/5</td>
<td>.3015</td>
<td>.0524</td>
<td>.2930</td>
</tr>
<tr>
<td></td>
<td>5/9</td>
<td>.2629</td>
<td>.0578</td>
<td>.2374</td>
</tr>
<tr>
<td></td>
<td>13/25</td>
<td>.2100</td>
<td>.0527</td>
<td>.1210</td>
</tr>
<tr>
<td>0.50</td>
<td>None</td>
<td>.4170</td>
<td>.0265</td>
<td>.4068</td>
</tr>
<tr>
<td></td>
<td>3/5</td>
<td>.3564</td>
<td>.0447</td>
<td>.3469</td>
</tr>
<tr>
<td></td>
<td>5/9</td>
<td>.3276</td>
<td>.0611</td>
<td>.3026</td>
</tr>
<tr>
<td></td>
<td>13/25</td>
<td>.2739</td>
<td>.0669</td>
<td>.1987</td>
</tr>
<tr>
<td>0.25</td>
<td>None</td>
<td>.4501</td>
<td>.0282</td>
<td>.4544</td>
</tr>
<tr>
<td></td>
<td>3/5</td>
<td>.4153</td>
<td>.0513</td>
<td>.4150</td>
</tr>
<tr>
<td></td>
<td>5/9</td>
<td>.3987</td>
<td>.0652</td>
<td>.3890</td>
</tr>
<tr>
<td></td>
<td>13/25</td>
<td>.3619</td>
<td>.0744</td>
<td>.3222</td>
</tr>
<tr>
<td>n</td>
<td># of M</td>
<td>Prob. Correct</td>
<td>5/9 rule</td>
<td>Modified 5/9 rule</td>
</tr>
<tr>
<td>-----</td>
<td>---------------</td>
<td>---------------</td>
<td>--------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>.9</td>
<td>99.80</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.8</td>
<td>97.55</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.7</td>
<td>90.50</td>
<td>3.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.6</td>
<td>75.50</td>
<td>5.04</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>.9</td>
<td>97.55</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.8</td>
<td>97.45</td>
<td>2.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.7</td>
<td>83.65</td>
<td>3.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.6</td>
<td>68.55</td>
<td>4.01</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>.9</td>
<td>98.50</td>
<td>1.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.8</td>
<td>99.50</td>
<td>2.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.7</td>
<td>77.25</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.6</td>
<td>64.45</td>
<td>4.62</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>.9</td>
<td>91.50</td>
<td>2.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.8</td>
<td>79.25</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.7</td>
<td>67.35</td>
<td>5.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.6</td>
<td>59.75</td>
<td>5.49</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>.9</td>
<td>2.05</td>
<td>3.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.8</td>
<td>62.25</td>
<td>4.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.7</td>
<td>57.70</td>
<td>5.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.6</td>
<td>55.70</td>
<td>5.12</td>
</tr>
</tbody>
</table>

Table 2. M is a subset of a 3 by 3 rectangle with the indicated cardinality. The Prob. Correct refers to the probability that the actual input data provides a correct estimate of membership in M. The entries in the columns labelled "Mean" refers to the mean percentage of probability of correct classification based on 20 trials, and the "SD" columns refer to the standard deviations of these means. See text for an explanation of the modified 5/9 rule.
Another approach to modifying a frequency-based mapping is to use some sort of pattern-based mapping. A way of implementing this is via the iterated 3 by 3 mean filter. This filter is based on a 5 by 5 neighborhood, and can either be viewed as a weighted filter on this neighborhood or as a 3 by 3 mean filter applied twice. (See [10] for a discussion of this filter.). Table 3 provides a comparison of the following 5 segmentation techniques: (1) original data; (2) 5/9 rule; (3) 5/9 rule but using iterated 3 by 3 mean for 4 or 5 members of M in the 3 by 3 neighborhood; (4) iterated 3 by 3 mean; (5) 13/25 rule. By Theorem 2, the 13/25 rule has the highest probability of correct classification for interior points, but as evidenced by Table 3, other segmentation techniques can give better overall performance. To see why this is so, one might notice that the 13/25 treats equally the following two data sets:

| 0 0 0 0 0 | 0 0 1 0 0 |
| 0 1 1 1 0 | 0 0 0 1 0 |
| 0 1 1 1 0 | 1 0 0 1 0 |
| 0 1 1 1 0 | 0 0 1 1 1 |
| 0 0 0 0 0 | 1 1 0 0 0 |

since they each have exactly 9 entries of 1. Yet the first data set is much more likely to reflect x being in the set M. On the other hand, the iterated 3 by 3 mean would produce values of 0.61 and 0.58 for these sets, thus putting x in M for the first but not the second data set.
<table>
<thead>
<tr>
<th>Data</th>
<th>Error</th>
<th>Mean and SD of percentage incorrect classification</th>
<th>Iterated 5 by 3 mean</th>
<th>13/3 rule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>data 5/9 rule modified 5/9 rule</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>U8</td>
<td>17.0 3.8 2.2 2.1 2.0 1.9 1.8 1.7 1.6 1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U15</td>
<td>15.6 17.9 17.8 17.7 17.6 17.5 17.4 17.3 17.2 17.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U25</td>
<td>15.3 16.5 16.4 16.3 16.2 16.1 16.0 15.9 15.8 15.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>U8</td>
<td>16.3 3.2 2.3 2.2 2.1 2.0 1.9 1.8 1.7 1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U15</td>
<td>15.7 16.5 16.4 16.3 16.2 16.1 16.0 15.9 15.8 15.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U25</td>
<td>15.5 25.5 15.4 15.3 15.2 15.1 15.0 14.9 14.8 14.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>U8</td>
<td>18.3 7.8 4.9 4.8 4.7 4.6 4.5 4.4 4.3 4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U15</td>
<td>33.1 21.3 16.1 15.4 15.0 14.6 14.2 13.8 13.4 13.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U25</td>
<td>40.4 32.5 28.2 27.3 26.4 25.5 24.6 23.7 22.8 21.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>U8</td>
<td>16.1 6.9 5.4 5.1 4.8 4.5 4.2 3.9 3.6 3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U15</td>
<td>32.0 18.6 14.0 13.5 13.0 12.5 12.0 11.5 11.0 10.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U25</td>
<td>39.3 30.7 26.3 25.0 23.7 22.4 21.1 19.8 18.5 17.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>GS</td>
<td>16.6 3.0 2.0 1.9 1.8 1.7 1.6 1.5 1.4 1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G10</td>
<td>34.7 14.4 9.3 8.0 7.6 7.2 6.8 6.4 6.0 5.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GB</td>
<td>57.1 25.0 16.9 16.0 15.1 14.2 13.3 12.4 11.5 10.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>GS</td>
<td>16.3 3.3 2.4 2.3 2.2 2.1 2.0 1.9 1.8 1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G10</td>
<td>30.0 15.0 8.5 6.9 5.0 3.9 2.8 1.8 0.9 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GS</td>
<td>36.4 19.6 15.0 14.3 13.6 13.0 12.3 11.6 10.9 10.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>GS</td>
<td>18.8 6.7 4.8 4.8 4.7 4.6 4.5 4.4 4.3 4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G10</td>
<td>31.4 18.1 14.0 13.9 13.5 13.1 12.7 12.3 11.9 11.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GS</td>
<td>37.5 25.0 21.7 21.3 20.9 20.6 20.2 19.8 19.4 19.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>GS</td>
<td>16.3 6.2 4.7 4.7 4.6 4.5 4.4 4.3 4.2 4.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G10</td>
<td>30.8 17.8 14.5 13.8 13.1 12.4 11.7 11.0 10.3 9.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GS</td>
<td>36.9 25.1 20.9 20.2 19.5 18.8 18.1 17.4 16.7 16.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Percentage incorrect classification based on 10 trials. See text for explanation of rules. Error U8 refers to a uniform error distribution taking values from -8 to 8. Error GS refers to a Gaussian error distribution with expected value 0 and standard deviation 5. The other error entries are defined similarly. Data set A is the image indicated by 1's in Fig. 6(a) and B is its complement. Data C is the image of Fig. 6(b), and D is its complement.
Fig. 5. Top portion is original image. Next portion is result of S/9 rule, and bottom is result of modified S/9 rule.
Fig. 6(a)
5. **Some monotone equivariant segmentation techniques.** Many commonly used spatial techniques for image enhancement lead to monotone equivariant segmentation algorithms. To illustrate this, consider the histogram modification techniques described in [3], pp. 127-136. If the data are first rank ordered, and after application of the technique, the output is labeled by the original data values, the resulting technique is easily seen to be monotone equivariant. For example, if the original data values are 2.3, 5, 8.7, 12 and 20, the rank-ordering would produce ranks of 1 through 5. If a histogram modification technique produces output levels of 3 and 5, and if this output is then labeled as 8.7 and 20 (the values of the original data corresponding to these ranks), then the resulting technique is monotone equivariant.

We described in [9] a segmentation technique that is also monotone equivariant. It will be convenient to briefly describe it here. The input data is an \( m \) by \( n \) array \( A \) of nonnegative integers. The program consists of the following parts:

A. **Prefiltering.** A version of a \( k \) by \( k \) mean or median filter is applied to smooth the data, and the output is rounded to the nearest integer to form a matrix \( B \).

B. **Thinout.** A frequency count is made of the values appearing in \( B \). Those values that occur with frequency less than some threshold are deleted and the points corresponding to them reassigned to the closest valid remaining data value. Alternately, the \( k \) highest occurring values can be retained. The resulting matrix is denoted \( C \).
C. Suppose matrix C has data values \( j_1, j_2, \ldots, j_k \). The 3 by 3 dispersion of value \( j_1 \) is defined as follows: For each point in C having value \( j_1 \), look at a 3 by 3 neighborhood centered on that point, and calculate the number of points in that neighborhood that do not have value \( j_1 \). Compute the average over all points having value \( j_1 \) for which 3 by 3 neighborhoods can be found. This is the dispersion for \( j_1 \). Continue the process for \( j_2, j_3, \ldots, j_k \).

D. Various options exist for deleting 1 or more of the data values having the highest level dispersion. The simplest is to just delete the highest level value. Other options might include deleting all values whose dispersion exceeds 0.9 on the first pass, and then dropping this cutoff down by increments of 0.1 until a stopping criterion is reached. Various options also exist for the reassignment of points whose value has been deleted. They proceed on both a global and local basis. Cluster means can be computed for each valid cluster, and points assigned either to the nearest mean, or by means of a Bayes decision rule distance. Alternatively, this can be done on a global basis, and all points having a given deleted value can be assigned to the nearest valid cluster. These techniques are called ANIAR and ANIARU. A second type of technique involves assignment of points according to whether their labels are above or below the averages of the next valid clusters that are above or below them. This can be done on either a local or global basis, and the resulting techniques are ANEXT and ANEXTU. These techniques as they have been described
are not monotone equivariant. If, however, the input data are first rank ordered, and the output is labelled according to which input value corresponds to which rank, then the resulting techniques do become monotone equivariant. An illustration of the operation of the algorithm occurs in Table 4. The input data consists of a mixture of 2 normal distributions, one with expected value 50 and the other with expected value 30, each having a standard deviation of 4. The data entries are rounded to the nearest integer, and a 3 by 3 mean prefilter is used. This prefilter tries to remove outliers by deleting the highest and lowest entries from each neighborhood, and outputting the average of the remaining 7 members. The "nondirected dispersion" shown in this table is identical to the version of dispersion we described earlier. The "directed dispersion" refers to the average of the maximum number of points in the subregions to the North, South, East and West of the center point that lie in regions other than that of the center point. The idea here is that the directed dispersion would distinguish between boundary points to a region and noise. Notice the dramatic drop in the values of the dispersion when the proper number of regions is found. The segmentation algorithm was ANEAR, and the input data was arranged so that one population was in the upper half of the input matrix, and the other in the lower half.
MIXTURE OF 2 NORMAL DISTRIBUTIONS WITH EXPECTED VALUES OF 20 AND 30. OUTPUT OF NON-DIRECTED AND DIRECTED DISPERSIONS AT VARIOUS STAGES OF THE CLUSTERING PROCESS

Table 4.
7. **Construction of a segmentation algorithm.** A program based on the earlier material will now be described and illustrated. The input data is as in Section 3. The program itself has two phases:

**Phase 1.** Find levels at which to slice the data.

**Phase 2.** Make appropriate slices and use a j/k rule to increase the probability of correct classification.

As was noted in section 4, the Phase 2 portion of the algorithm is a flat segmentation technique. The program can be used as a cleaning algorithm, or it is suited for the determination of the principal regions of a digital picture.

**Description of Phase 1.** 6 methods of determining the slice levels are built into the program.

**Method 1.** The mean and standard deviation of the input data are determined. The slices are taken at \( M-2S, M-S, M, M+S, \) and \( M+2S, \) where \( M \) is the mean and \( S \) the standard deviation. If this gets one outside the data range, then the extreme values are modified so that they lie between the minimum and maximum values of the input.

**Method 2.** A frequency count is taken of the levels that appear in the input data, and slices are taken at relative maximum points of this distribution.

**Method 3.** This uses relative minima of the dispersion.

**Method 4.** This is the same as method 3, except that if \( A \) represents the vector of distinct data values, then for \( A[i] \), one is interested in the average number of points that do not lie in \( A[i-1], A[i] \) or \( A[i+1] \). One then takes the relative minimum points.
Method 5. This too is similar to Method 3, except that the output is the average variance in \( t \) by \( t \) neighborhoods centered on points having a specified value. The slice levels again are the relative minima.

Method 6. User specifies slice levels.

B. Description of Phase 2. Having arrived at a list of slice levels, one now chooses a \( j/k^2 \) rule. The actual slices are constructed so that they are halfway between the slice input levels, and the resulting slices are given the labels of the slice inputs. Thus if one wanted slices at 17, 20, 22, 26, and 30, one would actually look at the increasing sequence of sets \( M_1 \subseteq M_2 \subseteq M_3 \subseteq M_4 \subseteq M_5 \), where \( M_i = \{ x : < k_i \} \) with \( k_1 = 18.5, k_2 = 21, k_3 = 24, k_4 = 28, \) and \( k_5 = \) the maximum value of \( X \). This forms 5 clusters, and they are assigned the values 17, 20, 22, 26 and 30. Alternately, they could be labeled by their means.

The actual implementation of this process is very simple. For input data \( X \), one looks at the lowest actual slice level \( k_1 \). A Boolean matrix is formed by letting 1 denote points whose value is \( < k_1 \) and 0 elsewhere. If a \( j/k^2 \) rule is used, one then looks at \( k \) by \( k \) regions. If the number of 1's in such a region is \( \geq j \), the output is 1; otherwise, it is 0. The output for \( k_2 \) is added to this, and the process continues until the supply of actual slice levels is exhausted. Computation time may be reduced by observing that if at a given stage a 1 is produced, then all
succeeding stages must also produce a 1, so there is no reason to consider neighborhoods involving that point. This of course is simply a reflection of the fact that a \( j/k^2 \) rule represents an isotone mapping on \( P(X) \). The actual implementation can be further optimized by first applying a \( k \) by \( l \) median filter, and applying the Phase 1 techniques to the output of this filter.

We close by considering some examples involving real data. The data are from two sources: the Westinghouse FLIR tape and an infrared weather tape that was supplied by the Air Force Geophysical Laboratory at Hanscom AFB. Plates 1-5 illustrate the 5 methods of selecting slice levels that were explained earlier in this section, and Plate 6 shows the output of the SIMENT program on the same data set. Though there are some minor differences shown between the various \( j/k^2 \) rules used, all of the methods did a good job of showing the central object of interest - a tank. One should bear in mind that the shades of grey in the pictures serve only to distinguish regions. They do not necessarily represent actual temperature levels of the regions. It is also of interest to note the difference in resolution between the 13/25 rule and the other rules that are used. This is especially apparent in Plates 14 and 17.
REFERENCES


Plate 2. Same as Plate 1, except that slice levels are obtained by Method 2.
Plate 3. Same as Plate 1, except that some levels are obtained by Method 3.

Plate 4. Same as Plate 1, except that some levels are obtained by Method 4.
Plate 5. Same as Plate 1, except that slice levels are obtained by Method 5.

Plate 6. Output of segment program, with a second FLIR tape. Top row represents original data, with A, B, and C in green letter, output of threshold program. Bottom row represents output of S, M, and C output levels. The segmentation is better with a FLIR output of level 8, constructed 3 by 3 dispersion matrix.
Plate 7. Portion of the SLAM Restinsonne of plate 2. Slave program output using Method 3. Percent of representative output is as follows: iterated 3 by 3 mean filter, 75% code, 0 code and 0 code rule.

Plate 8. Data and analysis to be included in the explanation of Method.
Plate 9. Portion of Record tape. Slice procedure used using Method 5 (as under Plate 1). See back explanation of Method 5.
Plate 11. Same as Plate 6, except that data is from Record 68.

Plate 13. Same as Plate 6, except that data is from Record 12.

Plate 14. Portion of weather picture showing Gulf Stream and some cloud cover. Slice program using Method 5. Format follows that of Plate 1.