[20] J. R. Quinlan, "Induction of decision trees," *Machine Learning,* vol. 1, no. 1, 1986.
[21] ——, "The effect of noise on concept learning," in *Machine Learning,* vol. II, R.S. Michalski *et al.*, Eds. Los Altos, CA: Morgan Kaufmann, 1986.
[22] ——, "Simplifying decision trees," *Int. J. Man-Machine Studies,* vol. 27, no. 3, pp. 221–234, 1987.
[23] ——, "Decision trees and multi-valued attributes," in *Machine Intelligence 11,* J. E. Hayes, D. Michie, and J. Richards, Eds. Oxford: Oxford University Press, 1988.
[24] E. M. Rounds, "A combined nonparametric approach to feature selection and binary decision tree design," *Patt. Recog.*, vol. 12, pp. 313–317, 1980.
[25] I. K. Sethi and G. P. R. Saravarayudu, "Hierarchical classifier design using mutual information," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-4, pp. 441–445, July 1982.
[26] P. H. Swain and H. Hauska, "The decision tree classifier: Design and potential," *IEEE Trans. Geosci. Electron.,* vol. GE-15, pp. 142–147, July 1977.
[27] Q. R. Wang and C. Y. Suen, "Analysis and design of a decision tree based on entroy reduction and its application to large character set recognition," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-6, pp. 406–417, July 1984.
[28] X. J. Zhou and T. S. Dillon, "A heuristic-statistical feature selection criterion for inductive machine learning in the real world," in *Proc. 1988 IEEE Int. Conf. Syst. Man Cybern.* (Beijing, China), 1988.
[29] ——, "Combining artificial intelligence with statistical methods for machine learning in the real world," in *Proc. 2nd Int. Workshop Artificial Intell. Stat.* (Fort Lauderdale, FL), 1989.
[30] ——, "A comparative study of heuristics for the construction of decision trees," to be published.

A Validity Measure for Fuzzy Clustering

Xuanli Lisa Xie and Gerardo Beni

***Abstract*—Cluster analysis has been playing an important role in solving many problems in pattern recognition and image processing. This correspondence presents a fuzzy validity criterion based on a validity function which identifies overall compact and separate fuzzy c-partitions without assumptions as to the number of substructures inherent in the data. This function depends on the data set, geometric distance measure, distance between cluster centroids, and more importantly on the fuzzy partition generated by any fuzzy algorithm used. The function is mathematically justified via its relationship to a well-defined hard clustering validity function: the separation index, for which the condition of uniqueness has already been established. The performance evaluation of this validity function compares favorably to that of several others. Finally, we have applied this validity function to color image segmentation in a computer color vision system for recognition of IC wafer defects which are otherwise impossible to detect using gray-scale image processing.**

***Index Terms*—Cluster validity, color vision, fuzzy clustering, IC wafer defect detection, image segmentation, pattern recognition, separate and compact clusters.**

Manuscript received August 15, 1989; revised December 21, 1990. This work was supported in part by the National Science Foundation under Grant 08421415.

The authors are with the Center for Robotic Systems, University of California, Santa Barbara, CA 93106.

IEEE Log Number 9142752.

I. Introduction

The engineering literature has paid very little attention to cluster validity issues [1], limiting the effort to present new clustering algorithms which perform reasonably well on a few data sets. In particular, the issue of validity for clustering of fuzzy data sets has been neglected (with few notable exceptions [2],[3]). On the other hand, if fuzzy cluster analysis is to make a significant contribution to engineering applications, much more attention must be paid to fundamental questions of cluster tendency. Recently, validity of fuzzy clustering has been discussed in applications to mixtures of normal distributions [4]. Also applications to distributed perception [5] have been proposed which rely in an essential way on good validity criteria for fuzzy clustering.

In the latter applications, separated sensors observe a common set of objects. They communicate to a central processor and not (perceptual) data (which, due to their size, cannot be transmitted in real time) [6] but decisions (which, due to their smaller bit size, can be transmitted in real time). In such cases, a fundamental decision is often the determination of the number of "objects" observed, i.e., the validity of the clustering procedure. Since higher level decisions by the central processor are based on the validity of these separated clustering procedures, it is essential that an efficient method is developed for fuzzy clustering validity.

Generally, the issue of cluster validity is a broad one and involves many questions. In view of the applications to distributed perception, in this correspondence, we focus on the validity of a partition. The answer is sought, as is generally accepted [1],[3], in measures of separation among clusters and cohesion within clusters.

The correspondence is organized as follows. In Section II, we review the fuzzy c-means clustering algorithm and some validity criteria related to our work for both fuzzy and hard clustering. Section III presents our new fuzzy validity function and its implementation strategy. Section IV contains the mathematical justification of the fuction and numerical comparisons to other validity fuctions. In Section V, we describe an application of our validity fuction to color image segmentation for recognition of defects in integrated circuit (IC) wafers.

II. Clustering Algorithm and Validity Criteria

Clustering is a tool that attempts to assess the relationships among patterns of the data set by organizing the patterns into groups or clusters such that patterns within a cluster are more similar to each other than are patterns belonging to different clusters. Many algorithms [2],[3],[7] for both hard and fuzzy clustering have been developed to accomplish this. An intimately related important issue is the "cluster validity" which deals with the significance of the structure imposed by a clustering method. Performance of many existing clustering algorithms are studied in [8]. Here we briefly review the fuzzy c-means clustering algorithm for later reference.

A. Fuzzy c-Means Clustering Algorithm

The fuzzy c-means (FCM) clustering algorithm (Bezdek [2]) is the fuzzy equivalent of the nearest mean "hard" clustering algorithm (Duda and Hart [9]), which minimizes the following objective function with respect to fuzzy membership μ_{ij} and cluster centroid V_i.

$$J_m = \sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^m d^2(X_j, V_i), \tag{1}$$

where

$$d^2(X_j, V_i) = (X_j - V_i)^T A (X_j - V_i). \tag{2}$$

A is a $p \times p$ positive definite matrix, p is the dimension of the vectors $X_j (j = 1, 2, \cdots, n)$, c is the number of clusters, n is the number of vectors (or data points), and $m > 1$ is the fuzziness index [2]. The FCM algorithm is executed in the following steps [2].

1) Initialize memberships μ_{ij} of X_j belonging to cluster i such that

$$\sum_{i=1}^{c} \mu_{ij} = 1. \tag{3}$$

2) Compute the fuzzy centroid V_i for $i = 1, 2, \cdots, c$ using

$$V_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^m X_j}{\sum_{j=1}^{n} (\mu_{ij})^m}. \tag{4}$$

3) Update the fuzzy membership μ_{ij} using

$$\mu_{ij} = \frac{\left(\frac{1}{d^2(X_j, V_i)}\right)^{\frac{1}{(m-1)}}}{\sum_{i=1}^{c} \left(\frac{1}{d^2(X_j, V_i)}\right)^{\frac{1}{(m-1)}}}. \tag{5}$$

4) Repeat steps 2) and 3) until the value of J_m is no longer decreasing.

The FCM algorithm always converges to strict local minima of J_m [2, p. 80] starting from an initial guess of μ_{ij}, but different choices of initial μ_{ij} might lead to different local minima.

B. Validity Criteria for Hard and Fuzzy Clustering

A well-established hard cluster validity criterion is the separation index D_1 (Dunn [12]) which identifies "compact, separate" (CS) clusters and is defined by

$$D_1 = \min_{1 \le i \le c} \left\{ \min_{i+1 \le j \le c-1} \left\{ \frac{\text{dis}(u_i, u_j)}{\max_{1 \le k \le c} \{\text{dia}(u_k)\}} \right\} \right\}, \tag{6}$$

where

$$\text{dia}(u_k) = \max_{X_i, X_j \in u_k} d(X_i, X_j), \tag{7}$$

$$\text{dis}(u_i, u_j) = \min_{X_i \in u_i, X_j \in u_j} d(X_i, X_j). \tag{8}$$

d is any metric induced by an inner product on R^p. The CS clustering of X is to be found by solving $\max_{2 \le c \le n} \left\{ \max_{\Omega_c} D_1 \right\}$, where Ω_c denotes the optimality candidates at fixed c. It is proved [12] that a hard c-partition of X contains c compact, separate (CS) clusters if $D_1 > 1$. Furthermore, there is at most one CS partition of X if $D_1 > 1$. The main drawback with direct implementation of this validity measure is computational since calculating D_1 becomes computationally very expensive as c and n increase. Another validity criterion which also measures compact and separate clusters is introduced by Davies and Bouldin [13]. Its major difference from D_1 is that it considers the average case by using the average error of each class. Jain and Moreau [14] also defined a method for cluster validity by using a bootstrap technique, that could be used with any clustering algorithm.

As a fuzzy clustering validity function Bezdek [15] designed the partition coefficient F to measure the amount of "overlap" between clusters.

$$F = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^2. \tag{9}$$

In this form F is inversely proportional to the overall average overlap between pairs of fuzzy subsets. In particular, there is no membership sharing between any pairs of fuzzy clusters if $F = 1$. Solving $\max_{c} \left\{ \max_{\Omega_c} \{F\} \right\}$ $(c = 2, 3 \ldots n-1)$ is assumed to produce valid clustering of the data set X. Disadvantages of the partition coefficient are the lack of direct connection to a geometrical property and its monotonic decreasing tendency with c. There are several other criteria in the literature which also measure the amount of fuzziness, such as classification entropy [16], proportion exponent [17], uniform data functional [18], nonfuzziness index [19], and information ratio [20]. Those criteria share a similar drawback with F, that is the lack of direct connection to the geometrical property of data set.

Gunderson [21] introduced a separation coefficient which takes into account geometrical properties. This validity criterion is designed to identify compact and separated clusters (which is similar to our goal). However, this method cannot be directly applied. It works on fuzzy clustering outputs by first converting them to hard ones. Since there are many ways one can convert fuzzy partitions to hard ones, this method shares the shortcomings of nonuniqueness of transferring from fuzzy partitions to hard partitions.

III. A Compact and Separate Fuzzy Validity Criterion

In this section, we define S as a fuzzy clustering validity function which measures the overall average compactness and separation of a fuzzy c-partition. We also present an implementation strategy for the use of this function.

A. Definition of a New Fuzzy Clustering Validity Function S

Consider a fuzzy c-partition of the data set $X = \{X_j; j = 1, 2, \cdots, n\}$ with $V_i (i = 1, 2, \cdots, c)$ the centroid of each cluster and $\mu_{ij} (i = 1, 2, \cdots, c, j = 1, 2, \cdots, n)$ as the fuzzy membership of data point j (also called vector j) belonging to class i.

Definition 1: $d_{ij} = \mu_{ij} \|X_j - V_i\|$, is called the *fuzzy deviation* of X_j from class i.

Note that $\|\cdot\|$ is the usual Euclidean norm. Thus d_{ij} is just the Euclidean distance between X_j and V_i weighted by the fuzzy membership of data point j belonging to class i.

Definition 2: $n_i = \sum_{x\ j} \mu_{ij}$ is the *fuzzy number* of vectors in or fuzzy cardinality of class i.

Note that $\sum_{x\ i} n_i = n$, where n is a "hard" number, e.g., the total number of data points in X. In the extreme case, when the partition is hard, n_i becomes exactly the number of vectors in class i.

Definition 3: For each class i, the summation of the squares of fuzzy deviation of each data point, denoted by σ_i, is called the *variation* of class i, that is: $\sigma_i = \sum_{x\ j} (d_{ij})^2 = (d_{i1})^2 + (d_{i2})^2 + \cdots + (d_{in})^2$. The summation of the variations of all classes, denoted by σ, is called the *total variation* of data set X with respect to the fuzzy c-partition, i.e., $\sigma = \sum_{x\ i} \sigma_i = \sum_{x\ i} \sum_{x\ j} (d_{ij})^2$.

Note that σ_i and σ depend on the data set, but more importantly they depend on the fuzzy c-partition, i.e., μ_{ij}'s and V_i's. A better c-partition should result in smaller σ. These values are not normalized, and they depend on how we choose our coordinate system. For example, if the fuzzy c-partition is obtained by using the fuzzy c-means algorithm with $m = 2$, the value of σ will be equal to the c-means objective function J_2 in (1).

Definition 4: The ratio, denoted by π, of the total variation to the size of the data set, that is, $\pi = (\sigma/n)$, is called the *compactness* of the fuzzy c-partition of the data set.

The value π measures how compact each and every class is. The more compact the classes are, the smaller π is. π is a function of the distribution characteristics of the data set itself, and more importantly a function of how we divide the data points into clusters. But it is independent of the number of data points. For a given data set, a smaller π indicates that we have reached a partition with more compact clusters, thus indicating a better partition. Gath and Geva [4] introduced fuzzy hypervolume which is the probability weighted total variation. This validity measure can identify ellipsoidal clusters and overlapped clusters. By incorporating covariance into the distance matrix A in (2), π can also identify ellipsoidal clusters.

Definition 5: The quantity $\pi_i = (\sigma_i/n_i)$ is called the *compactness of class i.*

Since n_i is the number of vectors in class i, σ_i/n_i is the average variation in class i. We have defined the compactness of fuzzy c-partition in terms of total variation and number of vectors. After defining π_i, we have some alternative ways to define the compactness of the fuzzy c-partition, such as: $\pi = (\sum_x {}_i\pi_i)/c$, i.e., the average compactness of each class; or $\pi = \max\pi_i$, i.e., the worst case. It can be shown that both ways have similar effect to Definition 4.

Definition 6: $s = (d_{\min})^2$ is called the *separation* of the fuzzy c-partition, where $d_{\min}$ is the minimum distance between cluster centroids, i.e.,

$$d_{\min} = \min_{i,j} \|V_i - V_j\|.$$

A larger s indicates that all the clusters are separated.

Definition 7: The *compactness and separation validity function S* is defined as the ratio of compactness π to the separation s, i.e., $S = \pi/s$.

After substituting for π and s, we get $S = (\sigma/n)/(d_{\min})^2$. A smaller S indicates a partition in which all the clusters are overall compact, and separate to each other. Thus, our goal is to find the fuzzy c-partition with the smallest S.

S can be explicitly written as

$$S = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ij}^2 \|V_i - X_j\|^2}{n \min_{i,j} \|V_i - V_j\|^2}. \quad (10)$$

We note that the definition of S is independent of the algorithm used to obtain μ_{ij}. Thus it is not internal to the clustering algorithm. For the FCM algorithm with $m = 2$, S can be shown to be

$$S = \frac{J_2}{n * (d_{\min})^2} \quad (11)$$

which is very easy to calculate. More importantly, minimizing S corresponds to minimizing J_2, which is the goal of FCM. The additional factor in S is $(d_{\min})^2$, which is the separation measurement. The more separate the clusters, the larger $(d_{\min})^2$, and the smaller S. Thus, the smallest S indeed indicates a valid optimal partition. If the fuzzy clustering algorithm used is to optimize some very different J, one may wish to modify the compactness measure so that minimizing S is compatible with minimizing J. For example, in (1) $m \neq 2$; then we can substitute ${\mu_{ij}}^2$ by ${\mu_{ij}}^m$ in (10).

We note, however, that S is still monotonically decreasing when c gets very large and close to n. One thing we can do is to impose an *ad hoc* punishing function [22] to eliminate this decreasing tendency. How to choose this function is not discussed here. Nevertheless, we shall see that even without a punishing function the validity function S provides a well defined method to solve the validity problem.

There are some existing validity criteria in the literature which measure compact and separate clustering. The separation coefficient in [21] considers the worst case, whereas S considers the total average case. Furthermore, the separation coefficient cannot be directly applied to fuzzy clustering as mentioned before. In [13], Davies and Bouldin introduce a hard partition validity criterion R. It is roughly related to S by $R = S/c$ if S is used for hard partitions. However, from our experience, S/c has a strong decreasing tendency as c increases.

B. Minimization of S and Implementation Strategy

Since smaller S means a more compact and separate c-partition, we assume that the minimum S partition is the most valid. Thus, a heuristic strategy to use S as a validity function is as follows. Using any fuzzy clustering algorithm, find one or more optimal c-partitions of the data set X for each $c = 2, 3, \cdots, n-1$. Let Ω_c denote the optimality candidates at each c; then the solution of

$$\min_{2 \le c \le n-1} \left\{ \min_{\Omega_c} S \right\}$$

is assumed to yield the most valid fuzzy clustering of the data set X.

Once we have defined the validity function S, our implementing strategy can be summarized into the following pseudo algorithm.

1) Initialize $c \leftarrow 2, S^* \leftarrow \infty, c^* \leftarrow 1$;
2) Initialize fuzzy membership μ_{ij};
3) Use any stable fuzzy clustering algorithm to update centroids V_i and μ_{ij};
4) Do convergence test; if negative goto 3;
5) Compute function S;
6) If $S < S^*, S* \leftarrow S, c^* \leftarrow c$;
7) If optimal candidate not found, goto 2;
8) $c \leftarrow c + 1$, if c =stop-value, stop;
9) Goto 2;

Steps 2–4 are the fuzzy c-partition algorithm. For FCM, (4) and (5) can be used. The convergence test can be $(J_m)_{q+1} - (J_m)_q \epsilon$ (e.g., 0.001), where q is an iteration index and J_m is as in (1). With $m = 2$, the S can be easily calculated as in (11). In step 2, the initial values of μ_{ij} can be assigned randomly and then normalized to satisfy $\sum_i \mu_{ij} = 1$ for all j. There is another way to initialize μ_{ij} for $c > 2$ in [23]: we do not discuss it here.

A problem of implementation is that S will have a tendency to eventually decrease when c is very large. So, the value of S is meaningless when c gets close to n. Fortunately, this is not a serious problem since in practice the feasible number of clusters c is much smaller than the number of data points n. Thus we can use the following three heuristic methods to determine the stop-value of step 8.

First, as mentioned in Section III-A, we can use a punishing function which imposes on S to counter this decreasing tendency. In Dunn [22], the "normalization and standardization of a validity function" is a simple example of the idea of punishing function.

The second method is that of plotting the optimal value of S for $c = 2$ to $n-1$, then selecting the starting point of monotonically decreasing tendency as the maximum c to be considered. Let $c_{\max}$ denote such a c; then, we find c by solving $\min_{2 \le C \le C_{\max}} \left\{ \min_{\Omega_c} s \right\}$.

The third way is application dependent. For most applications we do not need to compute S for very large c. It is almost always the case that c at the stop-value is $\ll n$. In this instance, we can either choose the maximum c according to preknowledge or, e.g., let $c_{\max} = n/3$ which very likely would not reach the starting point of the decreasing tendency.

IV. Mathematical and Numerical Justifications

We have already defined the new validity function and given an implementation strategy to use this function. In this section, we will mathematically justify this new fuzzy validity function via its relationship to a well-established hard partition validity measure and give a numerical example.

A. Uniqueness and Global Optimality of the c-Partition

The separation index D_1 (proposed by Dunn [12]) is a hard c-partition clustering validity criterion. If $D_1 > 1$, unique compact and separated hard clusters have been found. This result turns out to be useful also for fuzzy clustering validity. In fact, we may expect that if the data set X really has distinct substructure, i.e., hard clusters, a fuzzy partitioning algorithm should produce relatively hard memberships μ_{ij} and small total variations. We can prove that if the optimal solution D_1 becomes sufficiently large, the optimal validity function S will be very small, which means that a unique c-partition has been found. The proof of this is as follows.

Definition 8: Let $\mu_{ij}(i = 1, \cdots, c; j = 1, \cdots, n)$ be the membership of any fuzzy c-partition. The corresponding hard c partition of μ_{ij} is defined as ω_{ij}: for $j = 1, 2, \cdots, n : \omega_{ij} = 1$ if $i = \text{argmax}_i\{\mu_{ij}\}$; $\omega_{ij} = 0$ otherwise.

Theorem 1: For any $c = 2, \cdots, n-1$, let S be the overall compact and separated validity function of any fuzzy partition, and D_1 be the separation index of the corresponding hard partition; then we have

$$S \leq \frac{1}{(D_1)^2}.$$

Proof: Let the fuzzy c-partition be an optimal partition of the data set $X = \{X_j; j = 1, 2, \cdots, n\}$ with $V_i(i = 1, 2, \cdots c)$ the centroids of each class u_i, and μ_{ij} the fuzzy membership of the data points X_j belonging to class u_i. The total variation σ_{opt} of the optimal fuzzy c-partition is defined in Definition 3. Thus, the total variation σ_h of the corresponding hard c partition is

$$\sigma_h = \sum_i \sum_{X_j \in u_i} \|X_j - V_i\|^2.$$

From the definitions of σ_{opt} and σ_h above, we can get

$$\sigma_{\text{opt}} \leq \sum_i \sum_{X_j \in u_i} \|X_j - V_i\|^2.$$

Suppose that the centroid V_i is inside the boundary of cluster i for $i = 1$ to c. Then

$$\|X_j - V_i\|^2 \leq \text{dia}^2(u_i)$$

for $X_j \in u_i$, where $\text{dia}(u_i)$ is defined in (7) . We thus have

$$\begin{aligned}\sigma_{\text{opt}} &\leq \sum_i \sum_{X_j \in u_i} \text{dia}^2(u_i)\\ &\leq \sum_i n_i\, \text{dia}^2(u_i)\\ &\leq n \max\left\{\text{dia}^2(u_i)\right\}.\end{aligned}$$

We also have that $(d_{\min})^2 \geq \min\{\text{dis}^2(u_i, u_j)\}$, where $\text{dis}(u_i, u_j)$ was defined in (8); thus

$$\frac{\sigma_{\text{opt}}}{n * (d_{\min})^2} \leq \frac{\max_i \left\{\text{dia}^2(u_i)\right\}}{\min_{i,j}\left\{\text{dis}^2(u_i, u_j)\right\}}.$$

Using (6) and (10), we get

$$S \leq \frac{1}{(D_1)^2}.$$

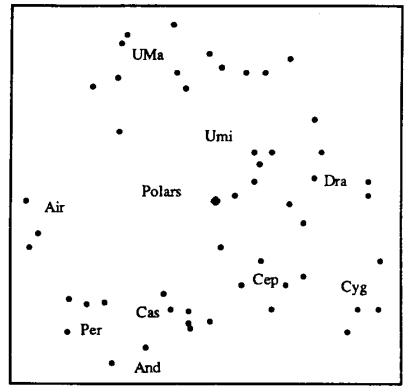


Fig. 1. Nine visually apparent starfield clusters classified by astronomers for fifty-one bright stars near Polaris.

Evidently, S becomes arbitrarily small as D_1 grows without bound. As mentioned, it has been proved by Dunn [12] that if $D_1 > 1$ the hard c-partition is unique. Thus, if the data set has a distinct substructure and the fuzzy partition algorithm has found it, then the corresponding $S < 1$.

There are some standard data sets in the literature that have been widely used to verify validity criteria. In Section IV-B, we apply our validity function S to one of this data sets, and compare its performance to existing validity criteria.

B. Clustering Validity Function S and Partition Coefficient F

Both the validity function S and the partition coefficient F are fuzzy validity criteria. Thus we can use the functions directly for fuzzy cluster validity. But there are differences between the two. F is inversely proportional to the overall average overlap between any pairs of fuzzy subsets, whereas S is proportional to the overall average compactness and separation. F lacks a direct connection to some property of the data themselves. On the other hand, S is directly related to the geometric properties of the data set X, the distance measure on R^p and the locations of cluster centroid. We give an example to compare the results obtained from F and S.

Example 1: Fig. 1 depicts a well-known set data X of 51 points corresponding to 51 bright stars near Polaris projected onto the plane of the Celestial Equator [21]. Considering position and light intensity, astronomers have grouped those stars into nine visual clusters as shown in Fig. 1. This data set has caused difficulty for fuzzy clustering algorithms based on position alone since it has chain-like and unequal population substructure. For example, the graph-theoretic method [24] could be successfully applied to a 60-points superset of X using only (x, y)-coordinates on the equatorial plane. In contrast, the fuzzy c-means did quite poorly with a 48-point subset of X [2].

Several validity methods have been applied to X, such as the partition entropy H [2], the separation coefficient G [21] and the partition coefficient F [2]. However, by directly using these methods, F and H yield $c^* = 2, G$, which measures hard partitions, yields $c^* = 3$. Gunderson in [21] has used a second application of G and obtained $c^* = 10$.

We processed the data set X (x-y coordinates only) using the FCM algorithm with $m = 2, \epsilon = 0.0001$, $\|\cdot\|$ = Euclidean, and $c = 2, 3, \cdots 17$. The values of S and F for each c are listed in Table I.

The minimum of S in Table I is $c^* = 8$, the second smallest values yield $c^* = 9$, and $c^* = 10$; all results are close to the number

TABLE I
VALUES OF VALIDITY FUNCTION F AND S FOR EXAMPLE 1

No. of Clusters c	Partition Coefficient F	CS Function S
2	0.72*	0.26
3	0.68	0.12
4	0.63	0.18
5	0.62	0.13
6	0.61	0.12
7	0.60	0.20
8	0.62	0.10*
9	0.63	0.11*
10	0.63	0.11*
11	0.62	0.16
12	0.63	0.13
13	0.63	0.14
14	0.63	0.13
15	0.62	0.12
16	0.63	0.27
17	0.63	0.17

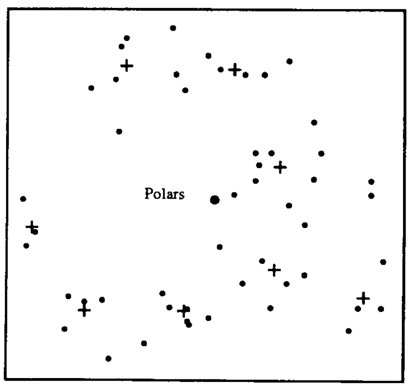


Fig. 2. Fuzzy partition of eight clusters obtained from S. Note: + indicates the centroids of each cluster.

of visual clusters. In contrast the maximum of F is at $c^* = 2$. F decreases to a minimum at $c = 7$, then progresses to $c = 17$ at a nearly constant low value. For the partition of $c^* = 8$ indicated by S, observe (Fig. 2) from the cluster centroids (indicated by symbol '+') that five of them are located in the visual clusters; and for $c^* = 9$ and $c^* = 10$, seven are located in the visual clusters. The fuzzy partition $c^* = 9$ is shown in Fig. 3.

We notice that the star cluster "And" is not an independent cluster according to our results in the partition of either $c^* = 8$ or $c^* = 9$. Also the "UMa" is split into two different clusters. This is expected since considering only the x-y geometric location of X, the distance of two stars of "And" is not close enough and well separated to form an independent cluster while "UMa" has the geometric property to form two clusters. The partitions for $c^* = 8$ and $c^* = 9$ are reasonable for our data set X. Since intensity is not taken into account, we should not expect to obtain clusters exactly like Fig. 1.

Since the validity function S is a fuzzy partition measurement, we expect "good" clusters at more than one value of c. Which one is more suitable should be determined from prior knowledge about the

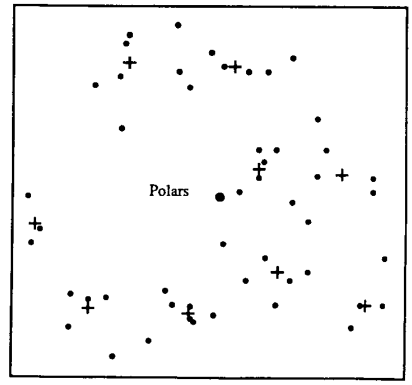


Fig. 3. The fuzzy partition of nine clusters obtained from S. Note: + indicates the centroids of each cluster.

data [1]. In our case, the visual clusters (Fig. 1) could be a reference, but not necessarily the best one to use.

Thus, the validity function S has suggested a substructure reasonable for our data set. In contrast, the partition coefficient F, which yields $c^* = 2$, does not correspond well with the actual substructure. Even if we take into account that for fuzzy functional (such as F and S) we do not need to identify the very best solution but we can choose (depending on prior knowledge and/or application) among the best few, F still yields unsatisfactory results [1) $c^* = 2$, 2) $c^* = 3$; see Table I]. In contrast, S yields [1) $c^* = 8$, 2) $c^* = 9, 10$] which are all close to the actual substructure. It is also worth noting that effective dynamic range for S is very wide.

V. APPLICATION TO COMPUTER COLOR VISION

Cluster analysis has been playing an important role in solving many problems in pattern recognition and image processing. For example, it is used for feature selection in Jain and Dubes [25] and for image segmentation for range image in Hoffman and Jain [26]. Image segmentation is a very critical step in image processing because errors at this stage influence feature extraction, classification, and interpretation at later stages.

In this section, we describe an application of our clustering criterion to color image segmentation for recognition of defects in integrated circuit (IC) wafers. The features of IC wafers are inherently colorful because of the interference effects taking place on the thin films which make up the IC structures [27]. Certain classes of IC defects can be detected by the use of colors which are otherwise not possible to detect in gray-scale image processing [27]. Various IC patterns manifest different colors due to the varying thicknesses in their structure.

In particular, we are interested in color ring defect recognition. A color ring defect is formed by a particle on the IC wafer causing a nonuniform thin film thickness surrounding the particle. The interference of different light wave lengths forms several cocentered color rings. The maximum number of rings among the colors reflects the size of the defect. Our task is to segment the color ring defect image and find the number of colors in the image and number of rings strongly formed in each color.

Ideally, one would expect an image to have regions of distinct colors separated by well-defined boundaries. However, in practice there is always some source of fluctuation or noise which imparts some uncertainty to the image. The noise in the image could be due to the degradations occurring in the process of image capture (sensor

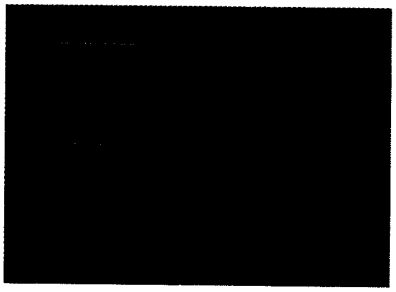


Fig. 4. Picture of color ring image for Example 2.

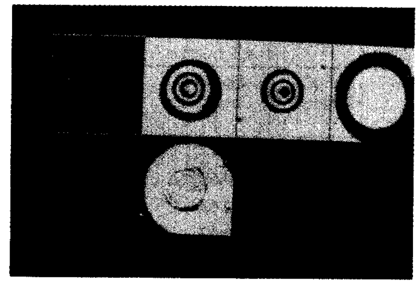
noise, variation of intensity of the light, error due to digitization, etc.). In any case, the actual image is quite complicated and it is not possible to label distinct regions in the image without using the segmentation.

In a color image, intensity of each pixel is represented by RGB, the three primary colors, and each pixel is viewed as a point in this three-dimensional color space. The image segmentation of a color image is to partition the image into regions or segments such that pixels belonging to a region are more similar to each other than pixels belonging to different regions. This can be done by using clustering methods to group the pixels in the color space into clusters. Consequently, a clustering criterion is used and it directly affects segmentation results.

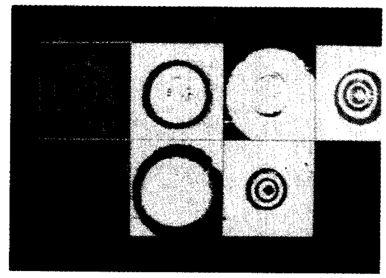
For the color ring defect problem, Barth [28] used a clustering method together with an appropriate distance threshold to successfully detect some defects. However, the performance of the method depends very much on the choice of the distance threshold. An improper choice of the threshold may lead to erroneous partitions. Furthermore, the right choice of such a threshold is unknown *a priori*. Another concern is the potential of the method to give a very large number of partitions which may not be very useful for solving problems such as the "ring" defect problem.

Here, we describe the use of our clustering validity criterion to the segmentation of a color ring defect image taken from real sample (showing in Fig. 4). The image is of 512×480 pixels. Since each defect occupies a very small part of the image, a focusing-of-attention strategy [28] is employed, that is, we only segment a small part of the image which contains one color ring defect. In such a way, the size of data to be processed can be largely reduced and computing time saved. Notice that there are some noises in the image. To reduce the noise effect, a threshold for pixel density in color space is used. Only those colors with density larger than the threshold are processed by using the clustering algorithm. The choice of this threshold does not essentially affects the results. The remaining data points are assigned to the nearest cluster. After segmentation, all pixels in each cluster are assigned the color value of the centroid of that cluster.

Example 2: Fig. 4 shows an image with three color ring defects. The color structure in this image is not so obvious to human eyes, which is usually the case for most color images (which makes computerized cluster analysis necessary). We focus our attention on the left defect inside window, whose color structure is relatively clear. This part of the image needs to be segmented in order to find the number of colors existing in this defect as well as the number of color rings formed in each color. In the area inside the window in Fig. 4, the human guesses that the number of distinct color rings could be four, five, or six. Two-hundred and sixty-seven distinct data points in the color space are obtained from the image in the attention window, using a density threshold of 6. Fig. 5 shows the results of



(a)



(b)

Fig. 5. Segmented image of 4–5 partitions for Example 2.

4 to 5 partitions with each segmented color displayed in a separate window. In Fig. 5(a), the lower left window shows the segmented image for the window above it.

Although the image in the attention area does not seem to have a very clear color structure at first glance, the results from validity function S in Fig. 6 indicate four cluster partitions as the best partition by quite a clear margin. A careful re-examination of the image tells us that this result is quite rational. The three rings of dark green and three rings of yellow in the center contrast quite clearly with the outer large orange ring and background. The 4-partition result in Fig. 5(a) clearly captured this character of the image. The second-best partition is the 5 partition, in which the orange color of the 4 partition is split in two: the inner smaller one a little bit close to the yellow end and the outer larger one a little bit close to the brown end. Although the three segmented colors of the 3–partition result are quite distinct (which means centroids of the three clusters are separate in the color space), the second color is clearly a mixture of the dark green and the background and is visually not a correct color. This means that the corresponding cluster is not compact, leading to a validity value larger than that of the 4 partitions. In the 2 partition, the orange color and the yellow color are further mixed, yielding another incompact cluster and even larger validity value. On the other hand, in the 6–9 positions, some colors (at least one) are split into two or more quite similar colors, resulting in nonseparate clusters, hence, larger validity value.

In addition, we give the partition results by using F shown in Fig. 6. F identifies the 2–partition as the best solution, and the validity value keeps decreasing until $c = 7$. It is clear that the best and second-best solutions indicated by F do not correctly reflect the actual color structure in the image. We also note that F has a very monotonically decreasing tendency as c increases.

VI. Conclusion

The issue of fuzzy cluster validity is still an open problem.

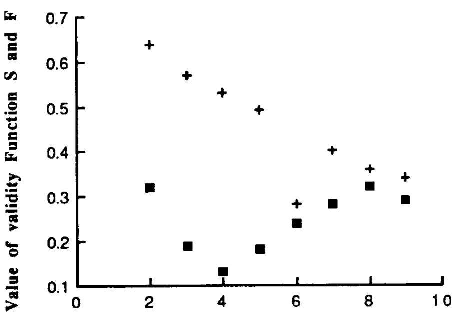


Fig. 6. Validity value of S and F for the segmented image in Example 2. *Note:* • indicates validity value of S; + indicates validity value of F.

More developments are expected before it can be effectivly used in applications. Any new validity function needs to satisfy the following requirements: 1) It has intuitive meaning; 2) it is easy to compute; 3) it is mathmaticaly justifiable. The fuzzy validity function S introduced in this correspondence has the above three features. In addition, its numerical performance compares well with existing validity functions. We also specifically applied this validity function to color image segmentation for IC ring defect detection. However, more numerical tests are needed, and this validity function only measures compact and separate clusters, as defined.

ACKNOWLEDGMENT

The authors wish to thank D. Chen for many valuable discussions and suggestions and Dr. M. Barth and Y. Xu for useful comments on clustering in color vision. We also thank B. Anderson and Prof. S. Hackwood for making the color vision system available and Prof. A. K. Jain for providing many useful suggestions.

REFERENCES

[1] A. K. Jain, *Handbook of Pattern Recognition and Image Processing.* New York: Academic, 1988.
[2] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms,* Plenum, New York, 1981.
[3] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data.*Englewood Cliffs, NJ: Prentice-Hall, 1988.
[4] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. 11, no 7, July 1989.
[5] X. L. Xie and G. Beni, "Distributed hierarchical decision fusion with cluster validity," in *Proc. IEEE Int. Conf. Syst. Man Cybern.* (Los Angeles, CA), 1990, p. 515.
[6] A. M. Al-Bassiouni, "Optimal signal processing in distributed sensor systems," Ph.D. dissertation, Naval Postgraduate School, Monterey, CA, 1987.
[7] A. Kandel, *Fuzzy Techniques in Pattern Recognition.* New York: Wiley, 1982.
[8] E. Backer and A. K. Jain, "A clustering performance measure based on fuzzy set decomposition," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-3, no 1, Jan. 1981.
[9] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis.* New York: Wiley, 1973.
[10] N. E. Day, "Estimating the components of a mixture of normal distributions," *Biometrika* vol. 56, pp. 463–474, 1969.
[11] J. H. Wolfe, "Pattern clustering by multivariate mixture analysis," *Multivar. Behav. Res.,* vol. 5, pp. 329–350, 1970.
[12] J. C. Dunn, "Well separated clusters and optimal fuzzy partitions," *J. Cybern.,* vol. 4, pp. 95–104, 1974.
[13] D. I. Davies and D. W. Bouldin, "A cluster seperation measure," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-1, no. 2, Apr. 1979.
[14] A. K. Jain and J. V. Moreau, "Bootstrap technique in cluster analysis," *Patt. Recog.,* vol. 20, no. 5, pp. 547–568, 1987.
[15] J. C. Bezdek, "Numerical taxonomy with fuzzy sets," *J. Math. Biol.,* vol. 1, pp. 57–71, 1974.
[16] J. B. Bezdek, "Mathematical models for systematics and taxonomy," in *Proc. 8th Int. Conf. Numerical Taxonomy,* G. Estabrook, Ed. San Francisco, CA: Freeman, 1975, pp. 143–166.
[17] M. P. Windham, "Cluster validity for fuzzy clustering algorithms," *Fuzzy Sets Syst.,* pp. 177–185, 1981.
[18] ——, "Cluster validity for fuzzy c-means clustering algorithm," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-4, no 4, July 1982.
[19] G. Libert and M. Roubens, "New experimental results in cluster validity of fuzzy clustering algorithms," in *New Trends in Data Analysis and Applications,* J. Janssen, J.-F. Macrotorchino, and J.-M. Proth, Eds. Amsterdam, The Netherlands: North-Holland, 1983, pp. 205–218.
[20] M. P. Windham, H. Bock, and H. F. Walker, "Clustering information from convergence rate," in *Proc. 2nd Conf. Int. Federation Classification Soc.* (Washington, DC), 1989, p. 143.
[21] R. Gunderson, "Application of fuzzy ISODATA algorithms to star tracker pointing systems," in *Proc. 7th Triennial World IFAC Cong.* (Helsinki, Finland), 1978, pp. 1319–1323.
[22] J. C. Dunn, "Indices of partition fuzziness and detection of clusters in large data sets," in *Fuzzy Automata and Decision Processes.* New York: Elsevier, 1977.
[23] X. Xie and G. Beni, "A new fuzzy clustering validity criterion and its application to color image segmentation," to be presented at the *IEEE int. Symp. Intell. Contr.* (Arlington, VA), Aug. 13–15, 1991.
[24] R. F. Ling, "Cluster analysis," Ph.D. dissertation, Yale Univ., New Haven, CT, 1971.
[25] A. K. Jain and R. Dubes, "Feature definition in pattern recognition with small sample size," *Patt. Recog.,* vol. 10, pp. 85–97, 1978.
[26] R. Hoffman and A. K. Jain, "Segmentation and Classification of range images," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-9, pp. 608–620, 1987.
[27] S. Parthasarathy, D. Wolfe, E. Hu, S. Hackwood, and G. Beni, "A color vision system for film thickness determination," in *Proc. IEEE Conf. Robotics Automat.* (Raleigh, NC), 1987, p. 515.
[28] M. Barth, "Robotic attentive sensing and its application to integrated circuit wafer inspection," Ph.D. dissertation, Univ. California, Santa Barbara, Aug. 1989.

A New Parameterization of Digital Straight Lines

M. Lindenbaum and J. Koplowitz

***Abstract*—A 1-1 correspondence is established between digital straight lines which start at a fixed point and a simple set of quadruples of integer parameters. Such a representation by parameters is useful for enumeration. An $O(N)$ algorithm is given for determining the parameters from the digital line, as well as $O(\log N)$ algorithms for transforming between these parameters and the parameters suggested by Dorst and Smeulders.**

***Index Terms*— Chain codes, digital straight lines, discrete geometry, farey series, line drawings.**

Manuscript received April 15, 1988. This work was supported by a Lady Davis Fellowship from the Department of Electrical Engineering, Technion-Israel Institute of Technology.

M. Lindenbaum is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel.
J. Koplowitz is with the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY 13676.

IEEE Log Number 9101449.