A Possibilistic Approach to Clustering

Raghu Krishnapuram, *Member, IEEE*, and James M. Keller, *Senior Member, IEEE*

Clustering methods have been used extensively in computer vision and pattern recognition. Fuzzy clustering has been shown to be advantageous over crisp (or traditional) clustering in that total commitment of a vector to a given class is not required in each iteration. Recently fuzzy clustering methods have shown spectacular ability to detect not only volume clusters, but also clusters which are actually "thin shells," i.e., curves and surfaces. Most analytic fuzzy clustering approaches are derived from the fuzzy C-means (FCM) algorithm. The FCM uses the probabilistic constraint that the memberships of a data point across classes sum to 1. The constraint was used to generate the membership update equations for an iterative algorithm. The memberships resulting from FCM and its derivatives, however, do not always correspond to the intuitive concept of degree of belonging or compatibility. Moreover, the algorithms have considerable trouble in noisy environments. In this paper, we cast the clustering problem into the framework of possibility theory. Our approach differs from the existing clustering methods in that the resulting partition of the data can be interpreted as a possibilistic partition, and the membership values may be interpreted as degrees of possibility of the points belonging to the classes, i.e., the compatibilities of the points with the class prototypes. We construct an appropriate objective function whose minimum will characterize a good possibilistic partition of the data, and we derive the membership and prototype update equations from necessary conditions for minimization of our criterion function. We illustrate the advantages of the resulting family of possibilistic algorithms with several examples.

I. Introduction

Clustering has long been a popular approach to unsupervised pattern recognition [1]. Fuzzy clustering has been shown to be advantageous over crisp (or traditional) clustering in that total commitment of a vector to a given class is not required in each iteration. It has become more attractive with the connection to neural networks [2]–[4], and with the increased attention to fuzzy clustering [5], [6]. One of the major factors that influence the determination of appropriate groups of points is the "distance measure" chosen for the problem at hand. In fact, recent advances in fuzzy clustering have shown spectacular ability to detect not only hypervolume clusters, but also clusters which are actually "thin shells," i.e., curves and surfaces [8]–[13].

Most analytic fuzzy clustering approaches are derived from Bezdek's fuzzy C-means (FCM) algorithm [14]. The FCM algorithm and its derivatives have been used very successfully in many applications, particularly those (such as pattern classification and image segmentation) in which the final goal of the task is to make a crisp decision. The FCM uses the probabilistic constraint that the memberships of a data point across classes must sum to 1. This constraint came from generalizing a crisp C-partition of a data set, and was used to generate the membership update equations for an iterative algorithm based on the minimization of a least-squares type of criterion function. The constraint on memberships used in the FCM algorithm is meant to avoid the trivial solution of all memberships being equal to 0, and it does give meaningful results in applications where it is appropriate to interpret memberships as probabilities or degrees of sharing. For example, the unsupervised fuzzy partition–optimum number of clusters (UFP-ONC) algorithm, introduced by Gath and Geva [5], interprets the membership values as likelihood estimates with impressive results. However, since the memberships generated by this constraint are relative numbers, they are not suitable for applications in which the memberships are supposed to represent "typicality," or compatibility with an elastic constraint [15], [16]. In this paper, we reformulate the fuzzy clustering problem so that clustering methods can be used to generate memberships that have a typicality interpretation.

The popularity of fuzzy set methods in fields such as control and rule-based reasoning is due to the fact that they are able to represent ill-defined classes and concepts in a natural way. In Zadeh's formulation of fuzzy set theory, the representation of such ill-defined classes or concepts is achieved by means of membership functions defined over the appropriate domain of discourse [15]–[17]. These memberships are absolute (i.e., not relative), and denote degrees of belonging or typicality. In other words, in such applications, the membership value of a point in the domain of discourse in a fuzzy set does not depend on its membership values in other fuzzy sets defined over the same domain of discourse. Zimmermann and Zysno have shown through empirical studies [18] that a good model for membership functions that model vague concepts or classes is

$$u(x) = \frac{1}{1 + d(x, x_0)}, \tag{1}$$

where $d(x, x_0)$ is the distance of a point x in the domain of discourse from the prototypical member x_0 of the class. In other words, in this formulation, membership values are solely a function of the "distance" of a point from a prototypical member [18]. The FCM algorithm and its derivatives are not really suitable for generating such membership functions from training data, since they do not generate memberships that can be interpreted as degrees of compatibility. At the same time, the lack of good methods to generate membership functions automatically from training data remains a serious problem in

Manuscript received May 1, 1992; revised December 1, 1992.

This work was supported by NASA through a grant from the University of Houston, Clear Lake (RICIS project, No. SE 42).

The authors are with the Department of Computer and Electrical Engineering, University of Missouri, Columbia, MO 65211.

IEEE Log Number 9208811.

many fuzzy set applications, including multicriteria decision making, rule-based reasoning, and control [18]–[20].

The following simple examples illustrate the problems associated with the probabilistic constraint used in the FCM algorithm as related to the interpretation of the resulting memberships. Fig. 1(a) shows a situation containing two clusters. The FCM would produce very different membership values in cluster 1 for the points A and B, even though they are equally typical (i.e., equidistant from the prototypical member—the fuzzy centroid) of this class. This problem arises from the constraint on the memberships, which forces point B to give up some membership in cluster 1 in order to increase its membership in cluster 2. Similarly, point A and point C may have equal membership values in cluster 1, even though point C is more typical than point A. In other words, in the FCM algorithm, the membership of a point in a class is a relative number, and it depends on the membership of the point in all other classes, thus indirectly on the total number of classes itself. That is, the final membership of point B reflects a "sharing" of B between the two clusters. While this may not be a problem in some applications such as pattern classification, it is not always appropriate for many fuzzy set applications. Fig. 1(b) represents another situation where there are two intersecting circular shell clusters. In this case, point A is a "good" member of both clusters, whereas point B is a "poor" member. Here again, the probabilistic constraint in the FCM updates would force a membership of 0.5 in the two clusters onto both point A and point B. These membership assignments are again counterintuitive in the sense of "compatibility to the prototype." If one wishes to interpret the final memberships at the degree of sharing, then the value of 0.5 is, in fact, appropriate. Fig. 1(c) shows another situation with a pair of clusters and two outlying points A and B. Intuitively, point A, being an outlier, should not have a high degree of membership (in the sense of "belonging" or "typicality") in either cluster. Point B should have an even smaller membership in either cluster, because it only vaguely represents either one of them. Yet, the FCM assigns a membership of 0.5 in the two clusters to both of them. Thus, not only are the membership values unrepresentative of the degree of belonging, but also they cannot distinguish between a moderately atypical member and an extremely atypical member. The above situations arise because the constrained memberships cannot distinguish between "equal evidence" and "ignorance." More recent theories such as belief theory [21] and possibility theory [22], [23] have tried to ameliorate this problem.

There may be another important motivation for using possibilistic memberships in clustering in some situations. Clustering methods have been used in situations where the number of subgroups present in the data are both known and unknown. (For example, in engineering pattern recognition applications, the number of classes is usually known). Like all unsupervised techniques, clustering (crisp or fuzzy) suffers from the presence of noise in the data. Since most distance functions are geometric in nature, noise points, which are often quite distant from the primary clusters, can drastically influence the estimates of the class prototypes and, hence, the final partition

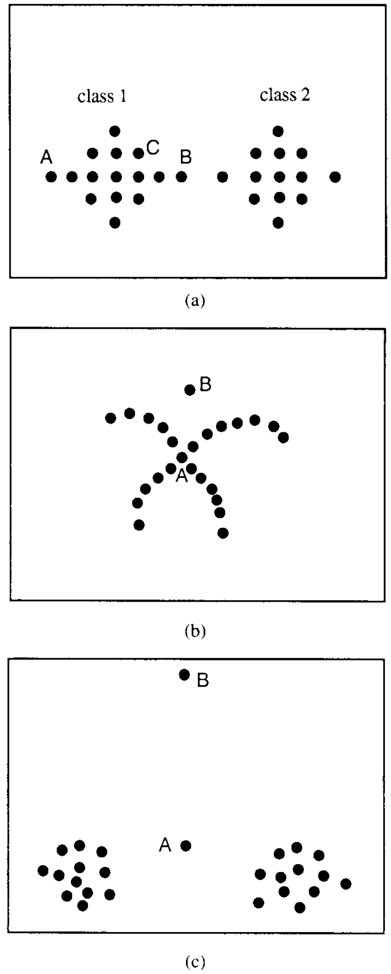


Fig. 1. (a) Example of a data set with two clusters in which the memberships generated by the FCM for points A and B are different, even though they are equally typical of class 1. Points A and C will have similar membership values even though they are not equally typical of the class. (b) Example of a data set with two intersecting clusters in which the memberships in both clusters of points A and B resulting from the FCM are about 0.5, even though point A is a "good" member of both clusters and point B is a "poor" member of both clusters. (c) Example of a data set with two noise points, A and B, in which the FCM memberships of these points in both clusters are about 0.5, even though point B is much less representative of either cluster than point A.

and the resulting classification. While this is not really a drawback when nothing is known about the nature and number of subgroups in the data set, it can be a serious problem in situations where one wishes to generate membership functions from training data. Fuzzy clustering methods ameliorate this problem when the number of classes is greater than 1, since the noise points tend to have somewhat smaller membership values in all the classes. However, this difficulty still remains in the fuzzy case, since the memberships of unrepresentative (or noise) points can still be significantly high. In fact, if there is only one "real" cluster present in the data, there is essentially no difference between the crisp and fuzzy methods. The prototype parameters (such as the center and orientation) and properties of the cluster (such as hypervolume) can be greatly affected by the noise in the data, and this is problematic in applications where such parameters and properties are important, such as automated visual inspection.

One way to improve the performance of the FCM algorithm and its derivatives in the presence of noise is to consider cluster validity. By employing a suitable validity measure, one may conclude that the best partition of the data set in Fig. 1(c) is obtained for the number of clusters $C = 3$, where the noise points are clubbed into one separate cluster. This would give more meaningful membership values for the noise points in the two good clusters. However, validity-based methods are computationally expensive since one needs to perform the clustering for a range of C values. Hence, they are not attractive for use in situations where memberships in a known number of classes need to be generated from training data. Moreover, reliable validity measures are quite difficult to define, and the number of clusters that optimizes a particular validity may not always be "correct" [8]. In any case, validity-based methods still do not give us membership values that can be interpreted as degrees of compatibility.

Another approach to the noise problem is to include a noise cluster [24], as has been suggested by Dave. In this approach, all points are considered to be equidistant from an amorphous noise cluster, and this distance has a relatively high value compared with the distances of the "good" points to the cluster prototypes. Thus, noise points, which are generally far away from "good" clusters, are attracted to the noise cluster. This approach is adequate if one's only concern is to reduce the effect of noise. However, using a single value for the distance of the noise cluster from all points may be too restrictive if the cluster sizes vary widely in the data set. Moreover, this approach still suffers from the drawbacks arising from the constrained memberships, and would produce membership values which are not suitable for applications in which they are required to be interpreted as degrees of typicality. This modified algorithm, for example, would produce '"asymmetric" memberships very similar to those of the original FCM in the situation shown in Fig. 1(a). It would also assign a membership of about 0.5 in both classes to point A in Fig. 1(b). Finally, since in this method one is forced to introduce a noise class, the interpretation of the resulting memberships would be somewhat less intuitive than those of the FCM when in fact there is no noise in the data. In summary, this approach and other approaches that use cluster validity are effective in certain situations but cannot correct the problem of relative memberships.

One established technique for robust estimation of parameters in the literature is to multiply the squared distances in the criterion function by weights that have an inverse relationship to the distances of the points from the prototype [25]–[28]. If the membership values have a typicality interpretation, then they can similarly be made to serve as weights in the criterion function. However, fuzzy memberships do not have such a simple inverse relation with the distance to the prototype, since they are relative numbers.

In Section II, we show how the probabilistic constraint on the sum of memberships can be relaxed to facilitate a possibilistic interpretation of the memberships. In Section III, we modify the objective function used in the FCM algorithm so that its minimum will characterize a good possibilistic partition of the data, and we derive the membership and prototype update equations from the necessary conditions for minimization of our criterion function. These equations lead to a new family of possibilistic clustering algorithms. In Section IV, we illustrate the behavior of the resulting family of possibilistic algorithms (with regard to the interpretation of the generated membership values) with several examples. As will be seen, our alternative formulation of the clustering problem, in which memberships are interpreted as degrees of typicality, automatically solves the noise problem. Finally, Section V gives the summary and conclusions.

II. Possibilistic Memberships

Let U denote a fuzzy partition matrix generated by the FCM algorithm. Then the elements u_{ij} of U satisfy the following conditions [14]:

$$u_{ij} \in [0,1] \text{ for all } i \text{ and } j,$$

$$0 < \sum_{j=1}^{N} u_{ij} < N \text{ for all } i, \text{ and}$$

$$\sum_{i=1}^{C} u_{ij} = 1 \text{ for all } j. \tag{2}$$

Here, u_{ij} is the grade of membership of the feature point $\boldsymbol{x}_j$ in cluster β_i, C is the number of classes, and N is the total number of feature points. In what follows, the symbol β_i will be used to denote the ith cluster and its prototype, since the prototype contains the parameters that characterize the cluster. The last condition in (2) is originally due to Ruspini [29], and it confines the memberships to lie on the hyperplane defined by $\sum_{i=1}^{C} u_{ij} = 1$. The partitions generated by validity and noise cluster approaches discussed in Section I still satisfy the conditions in (2) when all classes are considered, and satisfy the following conditions when only the good clusters are taken into account:

$$u_{ij} \in [0,1] \text{ for all } i \text{ and } j,$$

$$0 < \sum_{j=1}^{N} u_{ij} < N \text{ for all } i, \text{ and}$$

$$\sum_{i=1}^{C} u_{ij} \leq 1 \text{ for all } j. \tag{3}$$

The last condition in (3) confines the memberships to lie on the negative side of the hyperplane defined by $\sum_{i=1}^{C} u_{ij} = 1$. However, if the memberships are to be interpreted as degrees of compatibility or possibility [15], the conditions in (2) and (3) on the sum of memberships are too restrictive [30], since they give rise to relative numbers dependent on C. We now show that one can recast the clustering problem into the framework of possibility theory to overcome this drawback.

Following Zadeh [15]–[17], we consider each class prototype as defining an elastic constraint, with a point's membership denoting the degree to which this constraint must be "stretched" to match the point. In this interpretation, the β_i represent C fuzzy classes, and u_{ij} is the degree of compatibility of the feature point $\boldsymbol{x}_j$ with the prototypical member of β_i, or the possibility of feature point $\boldsymbol{x}_j$ belonging to class β_i [15]. If the classes represented by the clusters are thought of as a set of fuzzy subsets defined over the domain of discourse $\boldsymbol{X} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$, then there should be no constraint on the sum of the memberships. The only real constraint is that the assignments do really represent degrees of compatibility or possibility values; i.e., they must lie in the interval [0,1]. This is achieved by changing the conditions in (2) to

$$u_{ij} \in [0, 1] \text{ for all } i \text{ and } j,$$

$$0 < \sum_{j=1}^{N} u_{ij} \leq N \text{ for all } i, \text{ and}$$

$$\max_i u_{ij} > 0 \text{ for all } j. \tag{4}$$

The last condition in (3) simply ensures that the set of C fuzzy classes "covers" the domain of discourse $\boldsymbol{X}$. The resulting possibilistic C-partition defines C distinct (uncoupled) possibility distributions (and the corresponding fuzzy subsets) over the universe of discourse of the set of feature points. Thus, this interpretation leads to intrinsically fuzzy or possibilistic memberships, in the sense that the memberships are not "hard" even when there is only one class in the data set. This interpretation of membership values is in keeping with the concept of membership functions in most fuzzy set theory applications [2], [18], [19], [31]. Our intention in this paper is to bridge the gap between the "hyperplane constrained" [14] memberships present in fuzzy clustering algorithms, and the interpretation of these values as degrees of belonging or compatibility with each independent cluster—the more common fuzzy set theory concept of membership [15]–[19], [22], [23].

III. Possibilistic Clustering Algorithms

The original FCM formulation minimizes the objective function given by

$$J(L, \boldsymbol{U}) = \sum_{i=1}^{C} \sum_{j=1}^{N} (u_{ij})^m d_{ij}^2 \text{ subject to } \sum_{i=1}^{C} u_{ij} = 1 \text{ for all } j. \tag{5}$$

In (5), $L = (\beta_1, \cdots, \beta_C)$ is a C-tuple of prototypes, d_{ij}^2 is the distance of feature point $\boldsymbol{x}_j$ to prototype β_i, N is the total number of feature vectors, C is the number of classes, and $\boldsymbol{U} = [u_{ij}]$ is a $C \times N$ matrix, called the fuzzy C-partition matrix [14], satisfying the conditions in (2). Here, u_{ij} is the grade of membership of the feature point $\boldsymbol{x}_j$ in cluster β_i, and $m \in [1, \infty)$ is a weighting exponent called the fuzzifier.

Simply relaxing the constraint in (5) produces the trivial solution, i.e., the criterion function is minimized by assigning all memberships to 0. Clearly, one would like the memberships for representative feature points to be as high as possible, while unrepresentative points should have low membership in all clusters. The objective function which satisfies our requirements may be formulated as

$$J_m(L, \boldsymbol{U}) = \sum_{i=1}^{C} \sum_{j=1}^{N} (u_{ij})^m d_{ij}^2 + \sum_{i=1}^{C} \eta_i \sum_{j=1}^{N} (1 - u_{ij})^m, \tag{6}$$

where η_i are suitable positive numbers. The first term demands that the distances from the feature vectors to the prototypes be as low as possible, whereas the second term forces the u_{ij} to be as large as possible, thus avoiding the trivial solution. The choice of η_i will be discussed later.

Theorem: Suppose that $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\}$ is a set of feature vectors, $L = (\beta_1, \cdots, \beta_C)$ is a C-tuple of prototypes, d_{ij}^2 is the distance of feature point $\boldsymbol{x}_j$ to the cluster prototype $\beta_i (i = 1, \cdots, C; j = 1, \cdots, N)$, and $\boldsymbol{U} = [u_{ij}]$ is a $C \times N$ matrix of possibilistic membership values. Then $\boldsymbol{U}$ may be a global minimum for $J_m(L, \boldsymbol{U})$ only if $u_{ij} = [1 + (d_{ij}^2/\eta_i)^{\frac{1}{m-1}}]^{-1}$. The necessary conditions on the prototypes are identical to the corresponding conditions in the FCM and its derivatives.

Proof: In order to derive the necessary conditions and the membership updating equations, we first note that the rows and columns of U are independent of each other. Hence, minimizing $J_m(L, \boldsymbol{U})$ with respect to $\boldsymbol{U}$ is equivalent to minimizing the following individual objective function with respect to each of the u_{ij} (provided that the resulting solution lies in the interval [0,1]):

$$J_m^{ij}(\beta_i, u_{ij}) = u_{ij}^m d_{ij}^2 + \eta_i (1 - u_{ij})^m. \tag{7}$$

Differentiating (7) with respect to u_{ij} and setting it to 0 leads to the equation

$$u_{ij} = \frac{1}{1 + \left(\dfrac{d_{ij}^2}{\eta_i} \right)^{\frac{1}{m-1}}}. \tag{8}$$

It is obvious from (8) that u_{ij} lies in the desired range. It is also apparent that the memberships in (8) satisfy the conditions in (4). Since the newly added second term in the objective function is independent of the prototype parameters and the distance measure, the derivative of our new criterion function with respect to those parameters will be identical to that for the FCM or the appropriate generalization. Hence, the necessary conditions on the prototypes for optimization are the same as those in the theorems for the FCM-type algorithms [6], [7], [9], [12]–[14], [32]. QED

Thus, in each iteration, the updated value of u_{ij} depends only on the distance of $\boldsymbol{x}_j$ from β_i, which is an intuitively pleasing result. From the standpoint of "compatibility with the prototype," the membership of a point in a cluster should be determined solely by how far it is from the prototype of the class, and should not be coupled with its location with respect to other classes. In effect, this formulation allows optimal membership solutions to lie in the entire unit hypercube rather than restricting them to the hyperplane given by $\sum_{i=1}^{C} u_{ij} = 1$. The updating of the prototypes depends on the distance measure chosen, and will proceed exactly the same way as in the case of the FCM algorithm and its derivatives, as will be explained shortly. We would also like to point out the similarity between the memberships in (8) and the form of the membership function in (1), proposed by Zimmermann and Zysno [18].

Equation (8) defines a possibility distribution (membership) function for cluster β_i over the domain of discourse consisting of all feature points $\boldsymbol{x}_j$. We denote this distribution by Π_i. The value of m determines the fuzziness of the final possibilistic C-partition and the shape of the possibility distribution. When $m \to 1$, the membership function is hard, and when $m \to \infty$, the memberships are maximally fuzzy. A value of 2 for m (which seems to give good results in practice) yields a very simple equation for the membership updates. Fig. 2 shows a plot of the membership values resulting from (8) as a function of the normalized distance d_{ij}^2/η_i. From the figure, it can be seen that a variety of shapes for membership functions can be generated from this model.

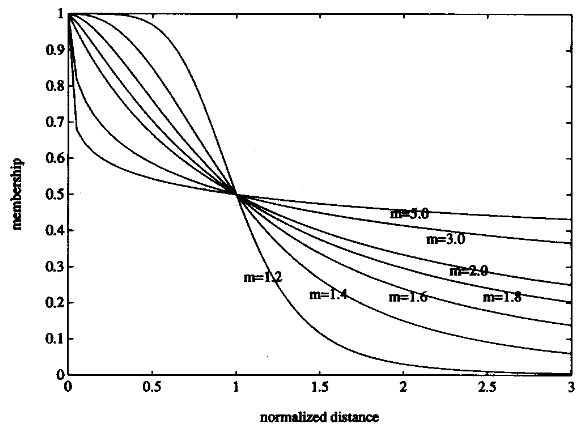


Fig. 2. Plot of the membership functions that can be generated by the possibilistic clustering algorithm.

The value of η_i determines the distance at which the membership value of a point in a cluster becomes 0.5 (i.e., "the 3 dB point"). Thus, it needs to be chosen depending on the desired "bandwidth" of the possibility (membership) distribution for each cluster. This value could be the same for all clusters if all clusters are expected to be similar. In general, it is desirable that η_i relate to the overall size and shape of cluster β_i. Also, it is to be noted that η_i determines the relative degree to which the second term in the objective function is important compared with the first. If the two terms are to be weighted roughly equally, then η_i should be of the order of d_{ij}^2. In practice we find that the following definition works well:

$$\eta_i = K \frac{\sum_{j=1}^{N} u_{ij}^m d_{ij}^2}{\sum_{j=1}^{N} u_{ij}^m}. \tag{9}$$

This choice makes η_i proportional to the average fuzzy intracluster distance of cluster β_i. Typically K is chosen to be 1. The following rule may also be used:

$$\eta_i = \frac{\sum_{x_j \in (\Pi_i)_\alpha} d_{ij}^2}{|(\Pi_i)_\alpha|}, \tag{10}$$

where $(\Pi_i)_\alpha$ is an appropriate α-cut of Π_i. In this case, η_i is the average intracluster distance for all of the "good" feature vectors (those vectors whose memberships are greater than or equal to α).

The value of η_i can be fixed for all iterations or it may be varied in each iteration. When η_i is varied in each iteration, care must be exercised, since it may lead to instabilities. (Note that the proof of the above theorem assumes that the η_i are constant.) Our experience shows that the final clustering is quite insensitive to large (an order of magnitude) variations in the values of η_i, although the final shapes of the Π_i do depend on the exact values of η_i. Thus, the best approach is to compute approximate values for the η_i based on an initial fuzzy partition using (9), and after the algorithm converges, recompute more accurate values for the η_i using (10) and run the algorithm for the second time. The second run typically converges in a couple of iterations, and it is only necessary if the actual values of class memberships are required. If only the relative degrees of within cluster strength are needed (for example, to generate parameters for the cluster or to produce a hard partition), then this final step can be omitted. The second pass through the algorithm with refined values for η_i allows the resultant memberships in a noisy environment to be nearly identical to those obtained in a noise-free state. Any value of α between 0.1 and 0.4 seems to yield consistent results.

There is another advantage to estimating the values of η_i based on the intracluster distances as in (9). Since the value of η_i determines the zone of influence (or bandwidth) of a cluster, clusters with larger values of η_i will have more mobility as the iterations proceed, since they "see" more points. This means that clusters that are largely empty and have large average intracluster distances have a greater degree of freedom to move compared with well-formed compact clusters as the iterations proceed. This eventually allows poorly formed clusters to settle down in denser areas of the data set.

When the nature of the clusters is known, values for η_i may also be fixed *a priori*. For example, in the case of line or hyperplane clustering algorithms such as the fuzzy C varieties (FCV) algorithm, in which the clusters are expected to be thin lines or planes, the values for η_i may be set approximately equal to the square of the expected thickness of the lines or planes. The same rule of thumb may be used for more general shell clustering algorithms such as the adaptive C shells algorithm [10] and the fuzzy C plano-quadric algorithm [32]. In the case of the Gustafson–Kessel algorithm [6], [7], a good choice for the value of η_i would be $V^{2/n} = |\boldsymbol{F}|^{1/n}$, where V is the expected hypervolume of the cluster, $\boldsymbol{F}$ is the covariance matrix, and n is the dimensionality of the feature space.

We propose a family of possibilistic clustering algorithms whose general form is as follows.

The Possibilistic Clustering Algorithm
Fix the number of clusters C; fix $m, 1 < m < \infty$;
Set iteration counter $l = 1$;
Initialize the possibilistic C-partition $\boldsymbol{U}^{(o)}$;
Estimate η_i using (9);
Repeat
Update the prototypes using $\boldsymbol{U}^{(l)}$, as indicated below;
Compute $\boldsymbol{U}^{(l+1)}$ using (8);
Increment l;
Until $(\|\boldsymbol{U}^{(l-1)} - \boldsymbol{U}^{(l)}\| < \varepsilon)$;
{The remaining part of the algorithm is optional and is to be used only when the actual shape of the generated possibility distribution is important}
Reestimate η_i using (10);
Repeat
Update the prototypes using $\boldsymbol{U}^{(l)}$, as indicated below;
Compute $\boldsymbol{U}^{(l+1)}$ using (8);
Increment l;
Until $(\|\boldsymbol{U}^{(l-1)} - \boldsymbol{U}^{(l)}\| < \varepsilon)$;

The updating of the prototypes depends on the distance measure chosen. Different distance measures lead to different algorithms. If the distance is an inner product induced norm metric as in the case of the FCM algorithm, i.e., if

$$d_{ij}^2 = (\boldsymbol{x}_j - \boldsymbol{c}_i)^T \boldsymbol{A}_i (\boldsymbol{x}_j - \boldsymbol{c}_i),$$

where $\boldsymbol{c}_i$ is the center of cluster β_i, updating of the prototype is achieved by [14]

$$\boldsymbol{c}_i = \frac{\sum_{j=1}^{N} u_{ij}^m \boldsymbol{x}_j}{\sum_{j=1}^{N} u_{ij}^m}. \tag{11}$$

This gives us the possibilistic C-means (PCM) algorithm. If the distance measure is the scaled Mahalanobis distance [6], [7], i.e., if

$$d_{ij}^2 = |\boldsymbol{F}_i|^{1/n} (\boldsymbol{x}_j - \boldsymbol{c}_i)^T \boldsymbol{F}_i^{-1} (\boldsymbol{x}_j - \boldsymbol{c}_i),$$

where $\boldsymbol{F}_i$ is the fuzzy covariance matrix of cluster $\beta_i = (\boldsymbol{c}_i, \boldsymbol{F}_i)$, then the center $\boldsymbol{c}_i$ is still updated using (11), and the fuzzy covariance matrix is updated using

$$\boldsymbol{F}_i = \frac{\sum_{j=1}^{N} u_{ij}^m (\boldsymbol{x}_j - \boldsymbol{c}_i)(\boldsymbol{x}_j - \boldsymbol{c}_i)^T}{\sum_{j=1}^{N} u_{ij}^m}. \tag{12}$$

This gives us the possibilistic Gustafson–Kessel (PGK) algorithm. In the case of spherical shell clusters, one possible distance measure [9] is

$$d_{ij}^2 = d^2(\boldsymbol{x}_j, \beta_i) = (\|\boldsymbol{x}_j - \boldsymbol{c}_i\|^{1/2} - r_i)^2,$$

where $\boldsymbol{c}_i$ is the center and r_i is the radius of cluster β_i, and the approximate updating of the prototypes may be accomplished by [13]

$$\boldsymbol{p}_i = -\frac{1}{2}(\boldsymbol{H}_i)^{-1} \boldsymbol{w}_i, \tag{13a}$$

where

$$\boldsymbol{p}_i = \begin{bmatrix} -2\boldsymbol{c}_i \\ \boldsymbol{c}_i^T \boldsymbol{c}_i - r_i^2 \end{bmatrix} \qquad \boldsymbol{H}_i = \sum_{j=1}^{N} u_{ij}^m \begin{bmatrix} \boldsymbol{x}_j \\ 1 \end{bmatrix} [\boldsymbol{x}_j^T, 1], \text{ and}$$

$$\boldsymbol{w}_i = 2 \sum_{j=1}^{N} u_{ij}^m [\boldsymbol{x}_j^T \boldsymbol{x}_j] \begin{bmatrix} \boldsymbol{x}_j \\ 1 \end{bmatrix}. \tag{13b}$$

The resulting algorithm may be called the possibilistic C-spherical shells (PCSS) algorithm. A possibilistic C quadric shells (PCQS) algorithm [12], [32] may also be defined similarly.

The results of the possibilistic algorithms do depend on initialization, just as any clustering technique. In possibilistic algorithms, the clusters do not have a lot of mobility, since each data point sees only one cluster at a time rather than all the clusters simultaneously. Therefore, a reasonably good initialization is required for the algorithms to converge to the global minimum. Any suitable (hard or fuzzy) clustering algorithm may be used to obtain the initial partition. The FCM algorithm or one of its derivatives provides an excellent choice for initialization of the corresponding possibilistic algorithms, and the estimates of η_i using the resulting fuzzy partition are quite adequate. However, when there is too much noise, the estimates of η_i computed from the fuzzy partition are not very good. This is because the FCM assigns relatively high memberships to noise points in good clusters. To overcome this problem, in highly noisy situations, a good approach seems to be to use the estimated η_i only for a couple of iterations in the possibilistic algorithm, reestimate them using (9) or (10), and then use them until convergence.

IV. Examples of Possibilistic Clustering

In this section, we show several examples of possibilistic clustering to illustrate the ideas presented in the previous section. We first present a simple example to provide insights into the possibilistic approach. We then present more realistic examples, and compare the performance of the possiblistic algorithms with those of the corresponding hard and fuzzy algorithms.

The first example involves two well-separated clusters of seven points each. In this case, the hard C-means algorithm (HCM), the FCM algorithm, and the PCM algorithm all give the same final crisp partition, shown in Fig. 3(a). The crisp partitions for the FCM and PCM are obtained by assigning each feature vector to the cluster in which it has the highest membership. Ties are broken arbitrarily. The cluster centers in all three cases are the same. The membership values for the FCM and PCM cases are shown in Table I. The feature vectors are numbered in the order in which they would be encountered in the top to bottom, left to right scan of the image shown in

TABLE I
MEMBERSHIPS AND CENTERS RESULTING FROM THE FCM AND PCM VALUES FOR THE NOISE-FREE DATA SET SHOWN IN FIG. 2(a)

	Fuzzy *C*-Means		Possibilistic *C*-Means	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2
1	0.996	0.004	0.632	0.007
2	0.004	0.996	0.007	0.632
3	0.988	0.012	0.300	0.005
4	0.997	0.003	0.631	0.006
5	1.000	0.000	1.000	0.007
6	0.996	0.004	0.632	0.008
7	0.980	0.020	0.300	0.009
8	0.020	0.980	0.009	0.300
9	0.004	0.996	0.008	0.632
10	0.000	1.000	0.007	1.000
11	0.003	0.997	0.006	0.631
12	0.012	0.988	0.005	0.300
13	0.996	0.004	0.632	0.007
14	0.004	0.996	0.007	0.632
Centers	(60.0, 150.0)	(140.0, 150.0)	(60.0, 150.0)	(140.0, 150.0)

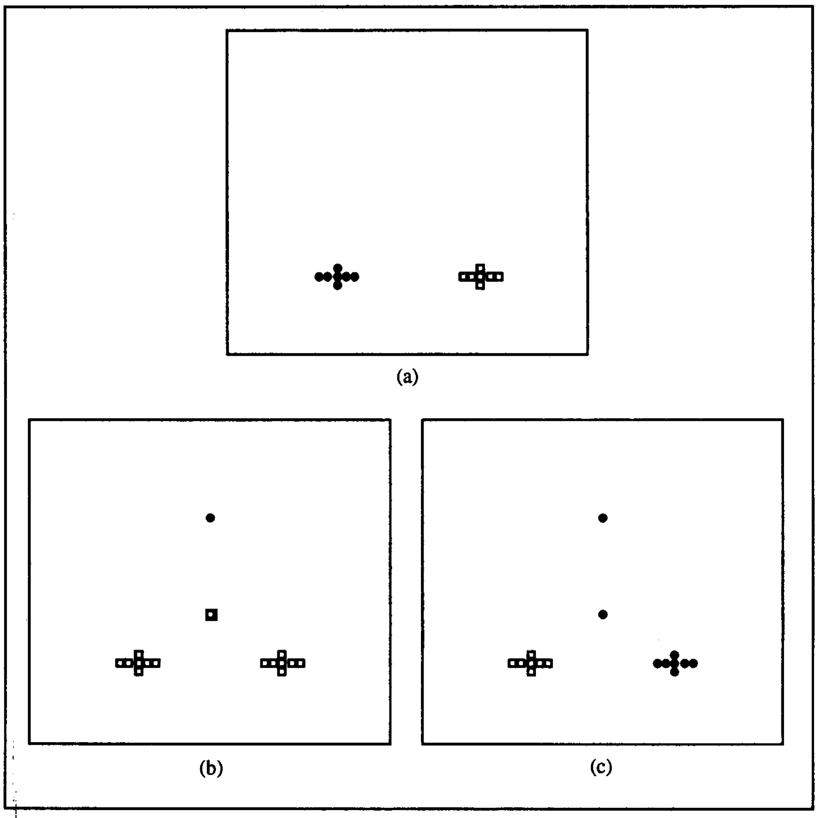
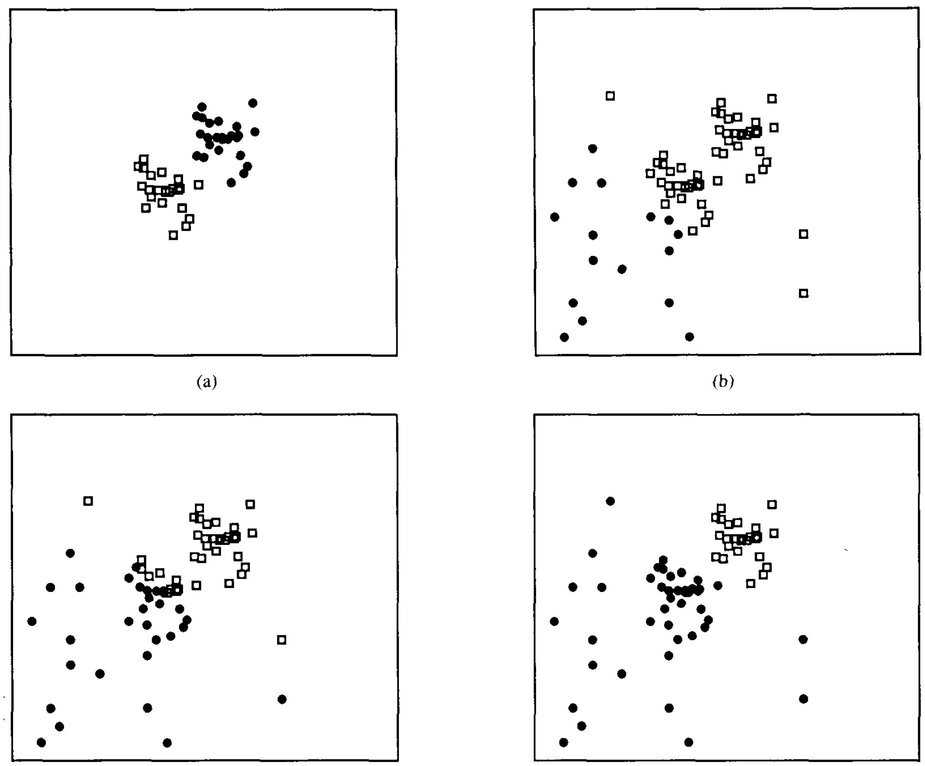


Fig. 3. Results on a simple data set: (a) the crisp partition resulting from the HCM, FCM and PCM algorithms; (b) the crisp partition from the HCM algorithm when noise is added; (c) the crisp partition from the FCM algorithm and the PCM algorithm when noise is added.

Fig. 3(a). Note the asymmetry of membership values entries (3 and 7 for example) in the fuzzy case compared with the possibilistic case. This asymmetry gets far worse if the two clusters are moved toward each other. It can also be seen that the FCM memberships are almost hard (i.e., close to 1 or 0) in every case. This may be desirable if a hard partition is required, but not if one needs to differentiate between close and far members within each cluster. On the other hand, the PCM algorithm provides more graded membership values, and these membership values are more in keeping with one's intuitive notion of typicality. Note that the farther away the feature vector is to the typical member (i.e., the prototype), the smaller the membership. As we noted in the previous section, the rate of fall of the membership values depends on the choice of

TABLE II
MEMBERSHIPS AND CENTERS RESULTING FROM THE FCM AND PCM VALUES FOR THE NOISY DATA SET SHOWN IN PART (c) OF FIG. 3

	Fuzzy C-Means		Possibilistic C-Means	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2
1	0.500	0.500	0.004	0.004
2	0.500	0.500	0.017	0.017
3	0.999	0.001	0.636	0.007
4	0.001	0.999	0.007	0.636
5	0.977	0.023	0.299	0.005
6	0.989	0.011	0.626	0.006
7	0.996	0.004	1.000	0.007
8	0.994	0.004	0.644	0.008
9	0.985	0.015	0.307	0.009
10	0.015	0.985	0.009	0.307
11	0.004	0.996	0.008	0.644
12	0.004	0.996	0.007	1.000
13	0.011	0.989	0.006	0.626
14	0.023	0.977	0.005	0.299
15	0.985	0.015	0.634	0.007
16	0.015	0.985	0.007	0.634
Centers	(62.8, 145.9)	(137.2, 145.9)	(60.0, 150.0)	(139.9, 150.0)



(a) (b) (c) (d)

Fig. 4. Results on a data set generated by a Gaussian random number generator: (a) the crisp partition resulting from the HCM, FCM, and PCM algorithms; (b) the crisp partition from the HCM algorithm when noise is added; (c) the crisp partition from the FCM algorithm when noise is added; (d) the crisp partition from the PCM algorithm when noise is added.

η_i. However, this has virtually no effect on the final crisp clustering. We chose to keep the computation of η_i the same for all examples, and this was done as explained in the previous section.

Parts (b) and (c) of Fig. 3 show the final crisp partition obtained from the HCM algorithm and the FCM and PCM algorithms, respectively, when two noise points are added to the set of feature vectors shown in Fig. 3(a). The HCM algorithm actually puts the farthest noise point as one cluster, and lumps all the rest into another cluster (although this may depend on initialization). The crisp partitions of the FCM and PCM are identical; however, the membership values and the cluster centers obtained are considerably different, as can be seen in Table II. The first two entries in the table

TABLE III
THE ESTIMATES OF CENTERS USING THE HCM, FCM, AND PCM ALGORITHMS FOR FIG. 4

	Hard *C*-Means	Fuzzy *C*-Means	Possibilistic *C*-Means
No Noise	(107.9, 72.6) (77.8, 103.0)	(107.7, 71.8) (77.8, 102.6)	(106.4, 73.0) (79.1, 100.9)
With Noise	(93.6, 88.9) (42.8, 139.8)	(100.4, 80.5) (59.7, 126.5)	(100.3, 79.3) (82.6, 97.1)

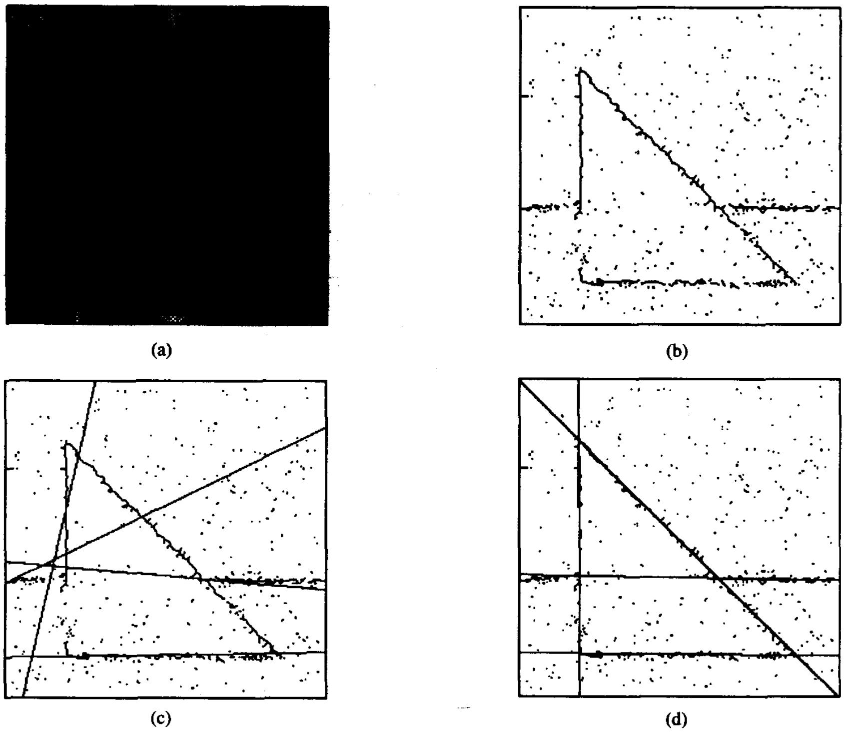
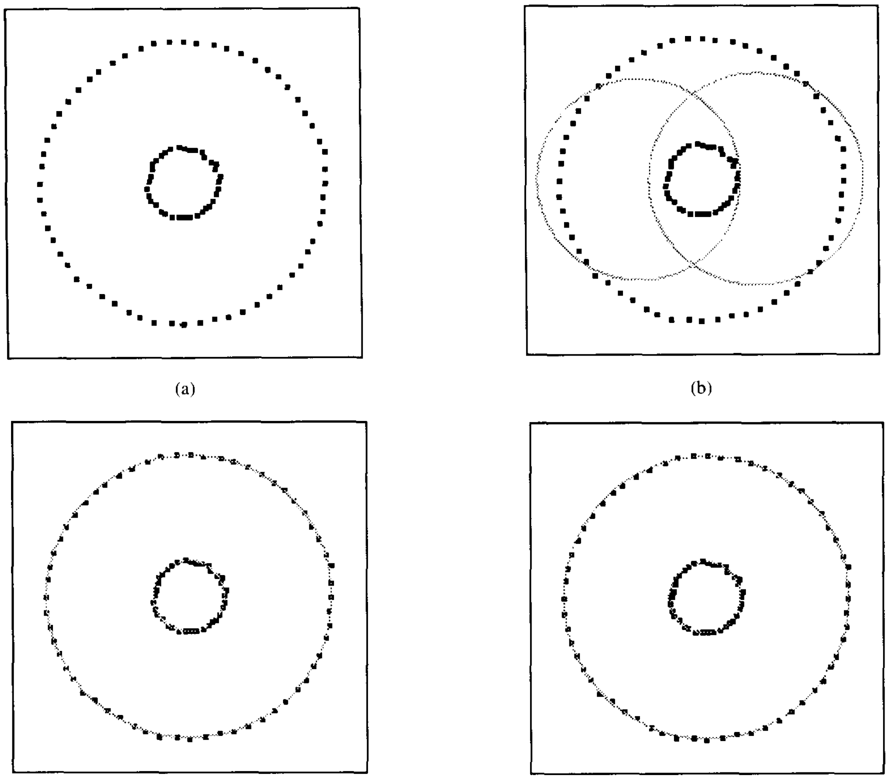


Fig. 5. Estimation of parameters of lines (The lines generated from the estimated prototype parameters are superimposed on the original data set): (a) image of a triangular fractal region; (b) edges obtained; (c) parameters obtained with the FGK algorithm; (d) parameters obtained with the PGK algorithm.

correspond to the two noise points, and the FCM algorithm gives approximately equal memberships of 0.5 in both clusters for the noise points. This significantly affects the estimates of the cluster centers, as can be seen in Tables I and II. The PCM algorithm, on the other hand, gives very low memberships for the two noise points in each cluster, and the farther point has a lesser membership than the closer one, as desired. As a result, the cluster centers are virtually unchanged. The membership values of the points in each of the clusters are also essentially unchanged in spite of the addition of the noise points. In fact, the memberships will not change even if an entire new cluster of feature points is added to the data set. This is desirable, especially if the clustering algorithm is to be used to estimate membership distribution functions for the various classes, to be used for processing at higher levels, as in computer vision [33]. The results of Dave's modified algorithm would also produce very similar memberships for the noise points, although they would still produce asymmetric relative memberships for the remaining points.

Fig. 4 shows a more realistic example with two classes. Each class has 25 feature vectors in the noise-free case. These points were generated with a Gaussian random number generator. Fig. 4(a) shows the clustering obtained by the HCM, FCM, and PCM algorithms. The crisp partitions are identical. Parts (b), (c), and (d) of Fig. 4 show the crisp partition resulting from the HCM, FCM, and PCM algorithms when the data from class 2 (the lower class) are noisy. The crisp partition by the HCM is miserable, and the crisp partition by the FCM is not satisfactory either. The performance of the PCM is quite acceptable. The cluster centers for the three methods for the noise-free and noisy cases are shown in Table III. As can be seen, the center estimates are poor in the cases of the hard and fuzzy *C*-means algorithms.

Fig. 5 depicts an example involving linear clusters. Fig. 5(a) shows a synthetic fractal image containing a triangular region with two rectangular regions as background, all with different fractal characteristics, and Fig. 5(b) shows the edges detected by estimating the fractal characteristics in the neighborhood of each pixel in different directions [34]. As can be seen, the edges are ill defined, and there are many noise points. However, these results are much better than those derived through intensity-based algorithms. Parts (c) and (d) of the



(a) (b) (c) (d)

Fig. 6 Estimation of parameters of circles when no noise is present (The circles generated from the estimated prototype parameters are superimposed on the original data set): (a) original data set; (b) parameter estimates obtained with the HCSS algorithm; (c) parameter estimates obtained with the FCSS algorithm; (d) parameters obtained with the PCSS algorithm.

figure show the clustering of the edge data by the fuzzy G–K (FGK) algorithm [6], [7] and the possibilistic G–K (PGK) algorithm. The results of the hard G–K algorithm are worse than those shown in Fig. 5(c). For the possibilistic version, any value of η_i between 50 and 500, which is the approximate hypervolume (length $\times$ width) of the lines, yields good results. The final estimates of the prototypes are shown superimposed on the original data set. As can be seen, the result of the FGK algorithm is poor for such a noisy data set. However, the result of the PGK algorithm is quite good.

The fourth example involves the detection of circles. Fig. 6(a) shows the original data set with two circles. Parts (b), (c), and (d) of the figure show the results of the hard C spherical shells HCSS [13], fuzzy C spherical shells (FCSS) [9], [13], and possibilistic C spherical shells (PCSS) algorithms, respectively, with the final estimates of the prototypes superimposed on the original data set. As can be seen, the results of the HCSS algorithm are very poor. (They correspond to a local minimum.) The FCSS and PCSS algorithms give identical results in this case. Fig. 7 shows the results of the same data set when noise is added. The performance of the HCSS algorithm is again unacceptable. The FCSS algorithm performs better; however, the estimates of the centers and the radii suffer from the presence of noise. The results of the PCSS algorithm are virtually unaffected by noise. Any value of η_i in the range 1–10, which is the approximate thickness of the circles, yields good results in this case.

The greatest difference between the FCM-based and PCM-based algorithms is for the case where there is but one cluster in the data set. In this case there is essentially no difference between the FCM-based methods and hard methods. Figs. 8 and 9 illustrate this idea. Fig. 8 shows the estimates of the prototype parameters for a noisy line when the FCV and PCV algorithms are used. The estimates of the FGK are severely affected by noise. Fig. 9 shows a similar result for a noisy circle generated by detecting the edges of a circular fractal region embedded in a different fractal background.

V. Conclusions

In this paper, we have presented a possibilistic approach to objective-function-based clustering. We argue that the existing fuzzy clustering methods do not provide appropriate membership values for applications in which memberships are to be interpreted as degrees of compatibility or possibility. This is due to the fact they use an inherently probabilistic constraint, which gives rise to relative numbers. As a result, membership of a feature vector in a cluster depends not only on where the feature vector is located with respect to the cluster, but also on how far away it is with respect to other clusters. Thus, this "conservation of total membership law" forces

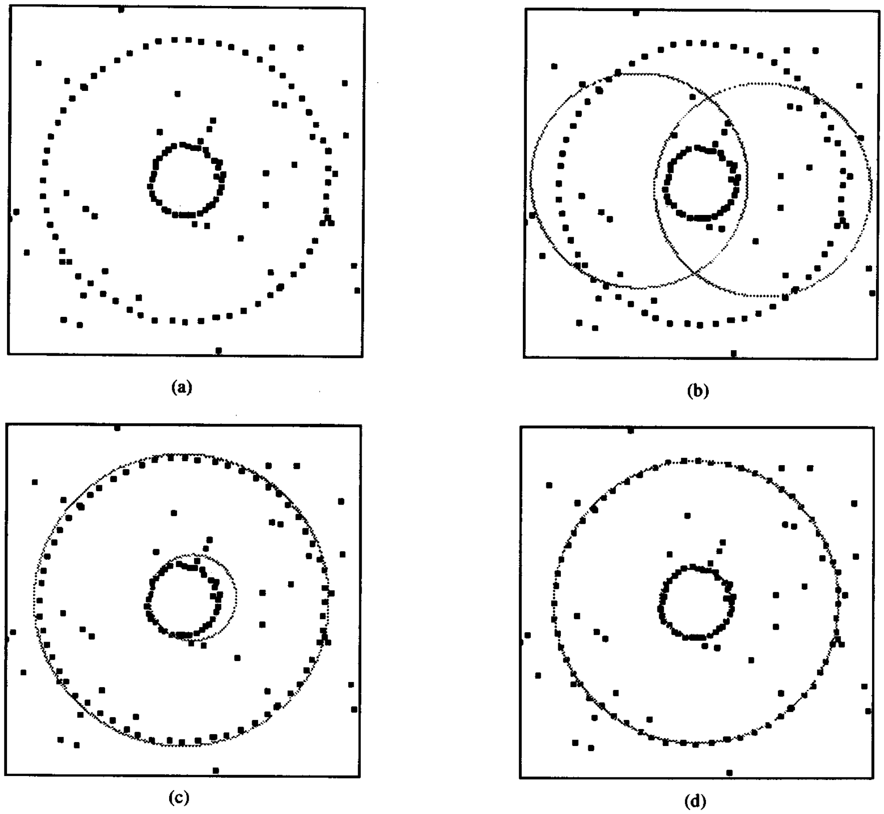


Fig. 7. Estimation of parameters of circles in noise (The circles generated from the estimated prototype parameters are superimposed on the original data set): (a) original data set; (b) parameter estimates obtained with the HCSS algorithm; (c) parameter estimates obtained with the FCSS algorithm; (d) parameters obtained with the PCSS algorithm.

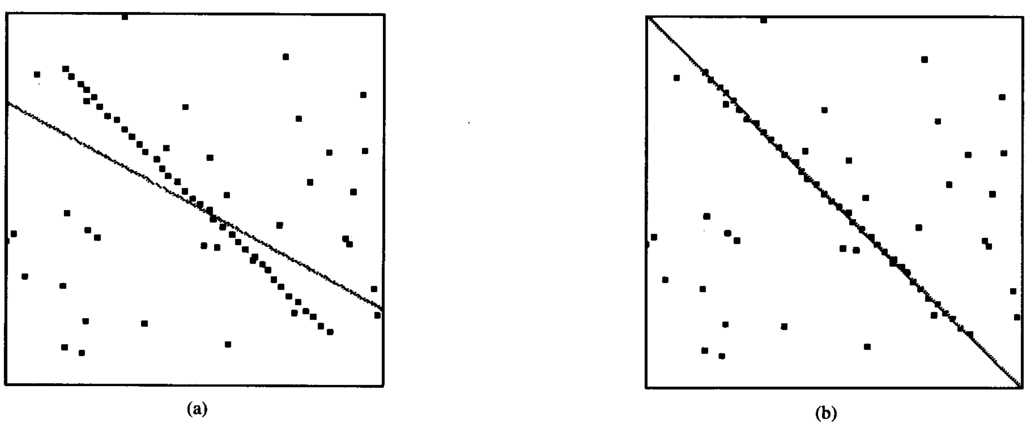


Fig. 8. Estimation of prototype parameters in noise when only one cluster is present: (a) line parameters obtained with the FGK algorithm; (b) line parameters obtained with the PGK algorithm.

the memberships to be spread across the classes, and makes them dependent on the number of clusters present. While this is desirable in situations where the memberships are meant to indicate probabilities or degrees of sharing, the resulting membership values cannot always distinguish between good members and poor members, even when there is no noise. This situation arises because probabilistic membership values cannot distinguish between "equally likely" and "unknown." On the other hand, if one takes the possibilistic view that the membership of a feature vector in a class has nothing to do with its membership in other classes, then we can use the modified clustering methods to generate membership distributions that model vagueness. Our possibilistic approach to clustering is based on this idea.

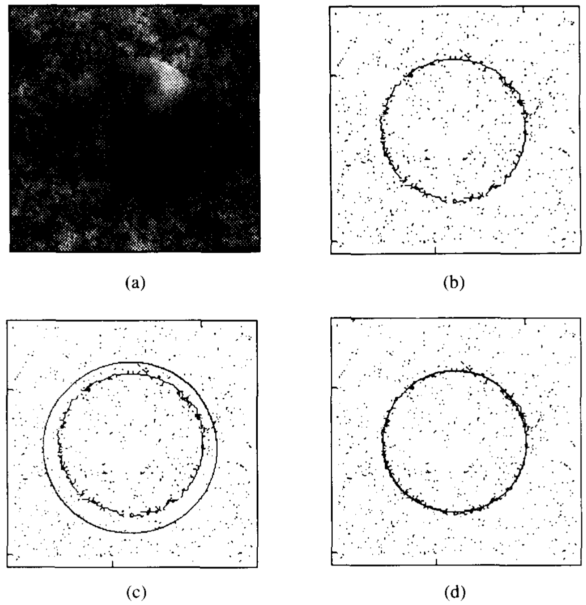


Fig. 9. Estimation of prototype parameters in noise when only one cluster is present: (a) image of a circular fractal region "moon in clouds"; (b) edges obtained; (c) circle parameters obtained with the FCSS algorithm; (d) circle parameters obtained with the PCSS algorithm.

Since our membership functions correspond more closely to the notion of typicality, the resulting algorithms are naturally more immune to noise. Noise points will have low degrees of compatibility in all clusters, which makes their effect on the clustering negligible. Also, our approach is intrinsically fuzzy, since the memberships are not "hard" even when there is only one class in the data set. This makes it compatible with the fuzzy set theory notion of membership functions. The partition of the data can be interpreted as a possibilistic partition, and the membership values may be interpreted as possibility values, or degrees of typicality of the points in the classes. The possibilistic C-partition defines C distinct (uncoupled) possibility distributions (and the corresponding fuzzy sets) over the universe of discourse of the set of feature points. Therefore, the family of algorithms we propose can be used to estimate possibility distributions directly from training data. Currently there are only a few algorithms to estimate possibility distributions directly from training data, other than those that do so by converting probabilities to possibilities [37], [38]. This conversion does not yield appropriate results when the FCM-based memberships are used, since the FCM memberships do not have a frequency interpretation, and since the memberships have already lost the distinction between "equally highly likely" and "equally highly unlikely." The possibilistic approach has the added advantage of being a natural mechanism to assign "fuzzy labels" to training data for use in more sophisticated pattern recognition algorithms. Finally, we would like to point out that the possibilistic algorithms may be viewed as a generalization of the weighted least-squares approaches [25] and robust parameter estimation methods [26], which have been used with good results in computer vision [27], [28].

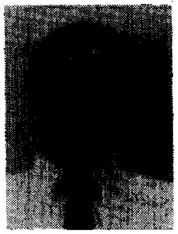
ACKNOWLEDGMENT

We are grateful to our students H. Frigui and O. Nasraoui, without whose suggestions and assistance the simulation experiments would not have been possible. We would also like to thank the anonymous reviewers for their helpful suggestions in the revision of this manuscript.

REFERENCES

[1] R. Dubes and A. K. Jain, *Algorithms for Clustering Data.* Englewood Cliffs, NJ: Prentice-Hall, 1988.
[2] B. Kosko, *Neural Networks and Fuzzy Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1992.
[3] T. Huntsberger and P. Ajjimarangsee, "Parallel self-organizing feature maps for unsupervised pattern recognition," *Int. J. General Systems,* vol. 16, pp. 357–372, 1990.
[4] J. Bezdek, E. C. K. Tsao, and N. Pal, "Kohonen clustering networks," in *Proc. First IEEE Conf. Fuzzy Systems* (San Diego), Mar. 1992, pp. 1035–1043.
[5] I. Gath and A. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 11, pp. 773–781, July 1989.
[6] *Fuzzy Models for Pattern Recognition,* J. C. Bezdek and S. K. Pal, Eds. New York: IEEE Press, 1992.
[7] R. Krishnapuram and C.-P. Freg, "Fitting an unknown number of lines and planes to image data through compatible cluster merging," *Pattern Recognition,* vol. 25, no. 4, pp. 385–400, 1992.
[8] R. N. Dave, "New measures for evaluating fuzzy partitions induced through C-shells clustering," in *Proc. SPIE Conf. Intelligent Robots and Computer Vision X: Algorithms and Techniques* (Boston), Nov. 1991, pp. 406–414.
[9] R. N. Dave, "Fuzzy-shell clustering and applications to circle detection in digital images," *Int. J. General Systems,* vol. 16, pp. 343–355, 1990.
[10] R. N. Dave and K. Bhaswan, "Adaptive C-shells clustering and detection of ellipses," *IEEE Trans. Neural Networks,* vol. 3, no. 5, pp. 643–662, 1992.
[11] J. C. Bezdek and R. H. Hathaway, "Numerical convergence and interpretation of the fuzzy c-shells clustering algorithm," *IEEE Trans. Neural Networks,* vol. 3, pp. 787–793, Sept. 1992.
[12] R. Krishnapuram, H. Frigui, and O. Nasraoui, "New fuzzy shell clustering algorithms for boundary detection and pattern recognition," in *Proc. SPIE Conf. Intelligent Robots and Computer Vision X: Algorithms and Techniques* (Boston) Nov. 1991, pp. 458–465.
[13] R. Krishnapuram, O. Nasraoui, and H. Frigui, "The fuzzy C spherical shells algorithm: A new approach," *IEEE Trans. Neural Networks,* vol. 3, no. 5, pp. 663–671, 1992.
[14] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms.* New York: Plenum Press, 1981.
[15] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems,* vol. 1, pp. 3–28, 1978.
[16] L. Zadeh, "The theory of approximate reasoning," in *Machine Intelligence,* vol. 9, J. Hayes, D. Michie, and L. Mikulich, Eds. New York: Halstead Press, 1979, pp. 149–194.
[17] L. Zadeh, "Fuzzy sets," *Information and Control,* vol. 8, pp. 338–353, 1965.
[18] H.-J. Zimmerman and P. Zysno, "Quantifying vagueness in decision models," *European J. Operational Res.,* vol. 22, pp. 148–158, 1985.
[19] J. Dombi, "Membership function as an evaluation," *Fuzzy Sets and Systems,* vol. 35, pp. 1–21, 1990.
[20] H. Takagi, "Fusion technology for fuzzy theory and neural network: survey and future directions," in *Proc. Int. Conf. Fuzzy Logic and Neural Networks* (Iizuka, Japan), July, 1990, pp. 13–26.
[21] G. Shafer, *A Mathematical Theory of Evidence.* Princeton, NJ: Princeton University Press, 1976.
[22] D. Dubois and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty.* New York: Plenum Press, 1988.
[23] G. Klir and T. Folger, *Fuzzy Sets, Uncertainty, and Information.* Englewood Cliffs, NJ: Prentice-Hall, 1988, chap. 4.
[24] R. N. Dave, "Characterization and detection of noise in clustering," *Pattern Recognition Lett.,* vol. 12, no. 11, pp. 657–664, 1992.
[25] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision,* vol. 1. Reading, MA: Addison Wesley, 1992, chap. 11.
[26] X. Zhuang, T. Wang, and P. Zhang, "A highly robust estimator through partially likelihood function modeling and its application in computer vision," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 14, pp. 19–35, Jan. 1992.

[27] D. G. Lowe, "Fitting parametrized three-dimensional models to images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 441–450, May 1991.
[28] P. Whaite and F. P. Ferrie, "From uncertainty to visual exploration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 1038–1049, Oct. 1990.
[29] E. Ruspini, "A new approach to clustering," *Information and Control*, vol. 15, pp. 22–32, 1969.
[30] D. Dubois and H. Prade, "Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions," *Fuzzy Sets and Systems*, vol. 40, no. 1, pp. 143–202, 1991.
[31] A. Kandel, *Fuzzy Mathematical Techniques and Applications*. Reading, MA: Addison-Wesley, 1986.
[32] R. Krishnapuram, H. Frigui, and O. Nasraoui, "A fuzzy clustering algorithm to detect planar and quadric shapes," in *Proc. North American Fuzzy Information Processing Society Workshop* (Puerto Vallarta, Mexico), Dec. 1992, pp. 59–68.
[33] J. Keller and R. Krishnapuram, "Fuzzy set methods in computer vision," in *An Introduction to Fuzzy Logic Applications in Intelligent Systems*, R. Yager and L. Zadeh, Eds. Boston: Kluwer Academic Publishers, 1992, pp. 121–145.
[34] J. Keller and B. Goldstein, "One variable fractal feature analysis for edge detection," submitted to *IEEE Trans. Image Processing*.
[35] E. Diday, A. Schroeder, and Y. Ok, "The dynamic clusters method in pattern recognition," in *Proc. Int. Federation for Information Processing Congress*. 1974, pp. 691–697.
[36] G. S. Sebestyen, *Decision-Making Processes in Pattern Recognition*. New York: Macmillan, 1962.
[37] D. Dubois and H. Prade, "Unfair coins and necessity measures: Towards a possibilistic interpretation of histograms," *Fuzzy Sets and Systems*, vol. 10, pp. 15–20, 1983.
[38] B. Bharathi Devi and V. V. S. Sarma, "Estimation of fuzzy memberships from histograms," *Information Sciences*, vol. 35, pp. 43–59, 1985.



Raghu Krishnapuram (S'82–M'87) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 1978. He was with Bush India, Bombay, for a year, where he participated in developing electronic audio entertainment equipment. He later joined the staff of Bharat Electronics Ltd., Bangalore, India, manufacturers of defense equipment. He obtained the M.S. degree in electrical engineering from Louisiana State University, Baton Rouge, in 1985 and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, in 1987.

He is currently an Associate Professor in the Electrical and Computer Engineering Department at the University of Missouri, Columbia. In 1993, he will be visiting the European Laboratory for Intelligent Techniques Engineering (ELITE), Aachen, Germany, as a Humboldt Fellow. His research encompasses many aspects of computer vision and pattern recognition. His current interests include applications of fuzzy set theory, morphology, and neural networks.

James M. Keller (M'79–SM'92) received the Ph.D. in mathematics in 1978. He has had faculty appointments in the Bioengineering/Advanced Automation Program, the Research Reactor, and the Electrical and Computer Emgineering Department at the University of Missouri-Columbia, where he currently holds the rank of Professor. He is also the E. A. Logan Research Professor in the College of Engineering. His research interests include computer vision, pattern recognition, fractal geometry for natural scene analysis, and the modeling and management of uncertainty.

Dr. Keller is currently president of the North American Fuzzy Information Processing Society (NAFIPS). He is an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS and the *International Journal of Approximate Reasoning* and is on the editorial board of the *Journal of Intelligent and Fuzzy Systems*