SELECTION OF ALPHA FOR ALPHA-HULL IN $\mathbb{R}^2$

DEBA PRASAD MANDAL and C. A. MURTHY*

Machine Intelligence Unit, Indian Statistical Institute, 203, Barrackpore Trunk Road, Calcutta 700 035, India

(*Received 23 March* 1994; *in revised form 5 July* 1995); *received for publication* 4 *November* 1996)

Abstract—For finding the shape of a planar set, Edelsbrunner, Kirkpatrick and Seidel introduced the concept of *α-hulls* as a natural generalization of convex hulls. While the α-hull is elegant and efficient to compute, it still suffers from a major drawback, i.e. the single parameter, namely α, must nevertheless be tuned. This paper deals with finding a way to overcome this drawback, i.e. we proposed here a selection criterion of α for α-hulls corresponding to a point set in $\mathbb{R}^2$. The selection criterion of α is based on the concept of minimum spanning trees and certain existing results. The effectiveness of the proposed selection criterion is demonstrated on some artificially generated data sets. The convergence (with sample size) of the α-hull, based on the proposed selection criterion for α, to the original pattern class has also been verified using symmetric difference, the Hausdorff metric, and a similarity metric.

Shape estimation   Convex hull   α-hull   Minimum spanning tree   Goodness of fit

1. INTRODUCTION

An important problem in pattern recognition is determining the shape of a pattern class from its sampled points. Once the shape is computed, some salient features of the class can then be extracted which are useful in making decisions about a course of action (e.g., identification, classification and pattern description) to be taken later. This will also reduce the storage requirement of the complete pattern class.

It may be noted that in most of the real life pattern recognition problems, the complete description of a pattern class is not known. Instead, a few sampled points (training samples) are usually available which are assumed to represent the class. Hence, determining the pattern class and its shape from sampled points (i.e. a set of training samples) is extremely useful.

There are various approaches described in the literature for determining the shape of a pattern class from sampled points.[1–8] Many of these approaches are concerned with the efficient construction of convex hulls for a set of points in the plane. Jarvis[1] presents several algorithms based on nearest neighbors that compute the shape of a finite point set. Akl and Toussaint[2] proposed a way of constructing a convex hull by identifying and ordering the extreme points of a point set, and the convex hull serves to characterize (in a rough way) the shape of such a set. Fairfield[3] has put forward a notion of the shape of a planar set based on the closest point voronoi diagram of the set.

A very elegant definition of the shape of a planar set was introduced in 1983 by Edelsbrunner *et al.*[6] They proposed a natural generalization of convex hulls which are referred to as *α-hulls*. The α-hull of a point set is based on the notion of generalized discs in the plane. The family of α-hulls includes the smallest enclosing circle, the set itself, and an essentially continuous set of enclosing regions in between these two extremes.

Edelsbrunner *et al.*[6] also define a combinatorial variant of the α-hull, called the *α-shape* of a planar set, which can be viewed as the boundary of the α-hull with curved edges replaced by straight edges. Unlike the family of α-hulls, the family of distinct α-shapes has only finitely many members. These provide a spectrum of progressively more detailed descriptions of the shape of a given point set. While the α-shape (hull) is elegant and efficient to compute, it still suffers from a major drawback, i.e. the single parameter, namely α, must nevertheless be tuned. In this connection, we would like to mention here that Worring and Smeulders[7] introduced the concept of the α-graph to find the shape of a point set. But they did not provide any criterion to decide the value of α.

This paper deals with finding a way to overcome the aforesaid drawback. In other words, we propose here a simple way to tune the parameter α of α-hulls (i.e. selecting automatically the right value of α) for a point set in $\mathbb{R}^2$. The proposed selection criterion of α is based on the concept of Minimum Spanning Trees (MST)[9,10] and certain existing results. Experimental results are shown to demonstrate the effectiveness of the selection criterion. The convergence (with sample size) of the α-hull, based on the proposed selection criterion for α, to the original pattern class has also been verified using symmetric difference, the Hausdorff metric, and a similarity metric.

In Section 2, some basic concepts are provided which are used later to justify the selection procedure. These

* Author to whom correspondence should be addressed.

include the definition of a pattern class, a description of α-hull and MST. Section 3 deals with the proposed selection procedure of α for α-hulls along with its justification. Experimental results along with verification of the convergence property are provided in Section 4. Section 5 deals with the conclusions.

2. SOME BASIC CONCEPTS

This paper deals with finding the "right value" of α using MST for determining α-hulls. We shall be dealing only with the sets in $\mathbb{R}^2$. Thus it becomes necessary to define formally the collection of subsets of $\mathbb{R}^2$ (pattern classes) under consideration in the present investigation. It is also necessary to mention here the basic concepts of α-hulls as well as the MST. These are furnished here.

2.1. *Pattern classes*

Generally, α-hulls need to be constructed for finite point sets which are subsets of $\mathbb{R}^2$. The α-hulls of finite point sets give rise to regions in $\mathbb{R}^2$. These regions are called pattern classes here. A formal definition of pattern classes in $\mathbb{R}^2$ is given below.

Definition 1. A set $\mathscr{A} \subseteq \mathbb{R}^2$ is said to be a pattern class[8] if

(i) $\mathscr{A}$ is path connected compact.
(ii) $\mathrm{cl}(\mathrm{Int}(\mathscr{A})) = \mathscr{A}$ [cl means *closure*, Int means *interior*],
(iii) $\mathrm{Int}(\mathscr{A})$ is path connected and
(iv) $\lambda(\delta\mathscr{A})=0$ where $\delta\mathscr{A}=\mathscr{A} \cap \mathrm{cl}(\mathscr{A}^c)$ and λ is the Lebesgue measure on $\mathscr{R}^2$.

The relevance of the properties (i)–(iv) of Definition 1 is provided in reference (8). Let $\mathbb{B}=\{\mathscr{A}: \mathscr{A}$ satisfies Definition 1$\}$. $\mathbb{B}$ is the collection of all classes in $\mathbb{R}^2$. Any $\mathscr{A} \subset \mathbb{B}$ is referred to as the pattern class in $\mathbb{R}^2$.

2.2. *α-hulls*

Conceptually, α-hulls[6] are a generalization of convex hulls of a planar set $\mathscr{S}$.

Definition 2. Let α be a sufficiently small but otherwise arbitrary positive real. The α-hull of $\mathscr{S}$ is the intersection of all closed discs with radius $1/\alpha$ that contain all the points of $\mathscr{S}$.

Definition 3. For arbitrary negative real α, the α-hull is defined as the intersection of all closed complements of discs (where these discs have radii $-1/\alpha$) that contain all the points of $\mathscr{S}$, i.e.

$$\alpha\text{-hull of } \mathscr{S} = \cap_{D(x,r) \subset \mathscr{S}} D^c(x,r),$$

where $D(x,r)$ is an open disc of radius r and center at x, and $r=-(1/\alpha)$.

The α-hulls with positive real α (as defined in Definition 2) are referred to as the positive α-hulls. On the other hand, the α-hulls with negative real α (as defined in Definition 3) are referred to as the negative α-hulls. Note that the 0-hull is the usual convex hull of the points. To bring the positive α-hulls, negative α-hulls and 0-hull under a single definition, Edelsbrunner *et al.*[6] introduced (as furnished below) the notion of generalized discs.

Definition 4. For arbitrary real value of α, a generalized disc of radius $1/\alpha$ is defined as follows:

(i) if $\alpha > 0$, it is a (standard) disc of radius $1/\alpha$;
(ii) if $\alpha < 0$, it is the complement of a disc of radius $-1/\alpha$, and
(iii) if $\alpha = 0$, it is a half plane.

Definition 5. The α-hull of a planar set $\mathscr{S}$ is defined as the intersection of all closed generalized discs of radius $1/\alpha$ that contain all the points of $\mathscr{S}$.

Thus, we have a family of α-hulls for α ranging from $-\infty$ to ∞. This family includes the smallest enclosing circle of $\mathscr{S}$ [when $\alpha = 1/\mathrm{radius}(\mathscr{S})$], the convex hull of $\mathscr{S}$ (for $\alpha=0$), and the set $\mathscr{S}$ itself (for α sufficiently small). While the selection of the single parameter α is very important for the procedure, the authors did not provide any criterion for the selection of α.

A combinatorial variant of the α-hull, called the α-shape of a planar set, was also defined by Edelsbrunner *et al.*[6] and it can be viewed as the boundary of the α-hull with curved edges replaced by straight edges. Although the α-shape is elegant and efficient to compute, the problem of selecting the value of α still persists.

Note. Observe that if $\alpha_1 < \alpha_2$ then α_1-hull of $\mathscr{S}_n \subseteq \alpha_2$-hull of $\mathscr{S}_n$ $\forall \alpha_1, \alpha_2$. Note also that 0-hull of $\mathscr{S}_n$ is nothing but the convex hull of $\mathscr{S}_n$. Thus

$$\alpha\text{-hull of } \mathscr{S}_n \supseteq \mathrm{Convex}(S_n) = 0\text{-hull of } \mathscr{S}_n, \quad \forall \alpha > 0.$$

In order to estimate non-convex sets, negative α-hulls provide a better alternative than the positive α-hull, since

$$\alpha\text{-hull of } \mathscr{S}_n \subseteq \mathrm{Convex}(S_n), \quad \forall \alpha < 0.$$

We are concentrating on negative α-hulls in this paper because these give non-convex sets and these are subsets of the convex hull. Thus, hereafter, in this paper "α-hull" refers to "negative α-hull" of a set of points in $\mathbb{R}^2$.

2.3. *Minimum spanning tree*

Before going to the definition of minimum spanning tree,[9,10] some terms of graph theory are reviewed. A *graph* consists of a set of *nodes* and a set of node pairs called *edges*. A *path* between two prescribed nodes is an alternating sequence of nodes and edges with the prescribed nodes as first and last elements, all other nodes distinct, and each edge linking the two nodes adjacent to it in the sequence. A *connected graph* has a path between any two distinct nodes. A *cycle* is a path beginning and ending with the same node. A *spanning tree* is a connected graph covering all the points and having no cycle. Note that there is a unique path between every two nodes in a spanning tree. An *edge weighted spanning tree* is a spanning tree with a real number (weight) assigned to each edge.

Definition 6. A minimum spanning tree is a spanning tree for which the sum of edge weights is minimum.

MSTs have the following two important properties that make them appropriate for applications:

(i) they connect all of the n nodes with $n-1$ edges and
(ii) the node pairs defining the edges represent points that tend to be close together (small distance or dissimilarity).

The first property follows the fact that MST is a spanning tree and the second from the requirement that the sum of the edge weights be a minimum.

In the present work, the set of nodes consists of the sample points in a given set $\mathscr{S}_n$. The weight of an edge is considered here as the Euclidean distance between the corresponding nodes. The sum of the edge weights of MST will be referred to as length of MST and this will be utilized (as will be seen in Section 3) in our proposed method for obtaining the value of α.

3. SELECTION OF α

We shall describe in this section the procedure for selecting the value of α for α-hulls with the help of a few theorems of Grenander.[11] We have visualized the problem as finding the consistent estimate of a set $\mathscr{S}$.

The α-hull of a planar set $\mathscr{S}$ provides its shape and it, by definition, is an uncountable set. It not only contains the set $\mathscr{S}$ but also many other points in $\mathbb{R}^2$. The α-hull of a finite set results in an uncountable set and consequently it provides the shape of $\mathscr{S}$.

We have basically visualized the problem of obtaining the shape of $\mathscr{S}$ as a set estimation problem. In the present work, an unknown set $\mathscr{A} \in \mathbb{B}$ is to be estimated on the basis of finitely many points $X_1, X_2, \ldots, X_n \in \mathscr{A}$. The definition of pattern class (Definition 1) has been utilized in this way. We shall also assume that $X_1, X_2, \ldots, X_n$ is a random sample from $\mathscr{A}$ following a continuous probability $\mathscr{P}_{\mathscr{A}}$ over $\mathscr{A}$, i.e. $\mathscr{P}_{\mathscr{A}}$ is the support of $\mathscr{A}$.

In the above formulation, $\mathscr{S}_n = \{X_1, X_2, \ldots, X_n\}$ is a planar set and it should be used in estimating the unknown set $\mathscr{A} \in \mathbb{B}$. Note that as n increases, $\mathscr{S}_n$ will be covering many parts of $\mathscr{A}$ and thus the value of α for $\mathscr{S}_n$ should depend on the sample size (n), i.e. α should be a function of n.

On the other hand, the value of α cannot be solely dependent on the number of points n of $\mathscr{S}_n = \{X_1, X_2, \ldots, X_n\}$. We shall give the reasons below.

Let $\mathscr{A}=[0,10^2]\times[0,10^2]$, $\mathscr{B}=[0,1]\times[0,1]$. Let T_{1000} be a set of 1000 points drawn uniformly from $\mathscr{A}$. Let U_{1000} be a set of 1000 points drawn uniformly from $\mathscr{B}$. Then note that the points are sparsely distributed in T_{1000} and densely (relative to T_{1000}) distributed in U_{1000}. It is easy to realize that the same value of α cannot provide good results for estimating both $\mathscr{A}$ and $\mathscr{B}$. If one particular α (<0) estimates $\mathscr{A}$ very closely using U_{1000} then it is likely to provide many isolated points for T_{1000} since the inter-point distances for T_{1000} are generally more than that of U_{1000}. Thus, it may be appropriate to consider α to be a function of not only the number of points (n) but also the inter-point distances of $\mathscr{S}_n$. We have considered here the length of MST (l_n) as a function of inter-point distances. Note that l_{1000} for U_{1000} is less than l_{1000} for T_{1000}. The proposed selection procedure for α reflects the aforementioned intuition. The necessary mathematical results of set estimation are given below.

Definition 7. Let $\mathscr{A} \in \mathbb{B}$ and the continuous probability measure on $\mathscr{A}$ be represented by $\mathscr{P}_{\mathscr{A}}$ (i.e. $\mathscr{P}_{\mathscr{A}}$ is the support of $\mathscr{A}$). Let $X_1, X_2, \ldots, X_n, \ldots$ be independent and identically distributed random vectors taking values in $\mathscr{A}$. Let $\mathscr{A}_n$ be an estimated set on the basis of $X_1, X_2, \ldots, X_n$. Then $\mathscr{A}_n$ is said to be a consistent estimate of $\mathscr{A}$ if

$$E_{\mathscr{A}}[\lambda(\mathscr{A}_n \triangle \mathscr{A})] \to 0 \text{ as } n \to \infty, \tag{1}$$

where λ is the Lebesgue measure in $\mathbb{R}^2$, $E_{\mathscr{A}}$ represents the expectation under the probability measure, $P_{\mathscr{A}}$, and $\triangle$ represents the symmetric difference [i.e. $\mathscr{B} \triangle \mathscr{C} = (\mathscr{B} \cap \mathscr{C}^c) \cup (\mathscr{B}^c \cap \mathscr{C})$].

Remarks. The symmetric difference ($\triangle$) between two sets represents those portions which are covered exactly by one of them but not both. If $\mathscr{A}_n$ is really a good estimator of $\mathscr{A}$, then the area of the symmetric difference [i.e. $\lambda(\mathscr{A}_n \triangle \mathscr{A})$] should tend towards zero as $n \to \infty$. Note that $\mathscr{A}_n$ is, in a sense, a random set since for different $X_1, X_2, \ldots, X_n$, $\mathscr{A}_n$ is different. Hence the average of $\lambda(\mathscr{A}_n \triangle \mathscr{A})$ [i.e. $E_{\mathscr{A}}(\lambda(\mathscr{A}_n \triangle \mathscr{A}))$] is considered in Definition 7 for a consistent estimator.

Theorem 1. Let $\mathscr{A} \in \mathbb{B}$ and $\mathscr{P}_{\mathscr{A}}$ be the uniform distribution on $\mathscr{A}$. Let $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to \infty$. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random vectors taking values in $\mathscr{A}$. Let

$$\mathscr{A}_n = \cup_{i=1}^{n}\{x \in \mathbb{R}^2 : |x - X_i| \le \varepsilon_n\}. \tag{2}$$

Then $\mathscr{A}_n$ is a consistent estimate of $\mathscr{A}$.

Proof. It can be found in reference (11).

Remarks. The above theorem has been stated for uniform distribution only. Note that the principles for selecting ε_ns above are stated in the theorem but not the exact choice of ε_n. It is intuitively clear that for the above stated $\mathscr{A}_n$, ε_n should tend towards zero as n tends to ∞. The restriction $n\varepsilon_n^2 \to \infty$ makes ε_n^2 proceed towards zero at a slower speed than as $n \to \infty$. This is because, if $\varepsilon_n \to 0$ at a higher speed, then $\mathscr{A}_n$ may tend towards just the collection of points $\{X_1, X_2, \ldots, X_n, \ldots\}$.

For practical problems, n is fixed and thus $\mathscr{S}_n = \{X_1, X_2, \ldots, X_n\}$ is a fixed set. ε_n should be chosen suitably for that specific n. Now, as mentioned in the beginning of the section, ε_n should not only be a function of n but also the inter-point distances between $X_1, X_2, \ldots, X_n$. That means that ε_n needs to be a random variable. The choice of the ε_ns has been stated below in the form of a theorem.

Theorem 2. Let $\mathscr{A} \in \mathbb{B}$ and $X_1, X_2, \ldots, X_n, \ldots$ be independent and identically distributed random vectors with continuous support $\mathscr{P}_{\mathscr{A}}$. Let l_n represent the sum of the edge weights of MST of $\mathscr{S}_n = \{X_1, X_2, \ldots, X_n\}$ where the edge weight is taken to be the Euclidean distance. Let

$$h_n = \sqrt{\frac{l_n}{n}}. \tag{3}$$

Then

(i) $h_n \to 0$ in probability and
(ii) $nh_n^2 = l_n \to \infty$ in probability.

Proof. It can be found in references (8,12).

Remarks. h_ns in the above theorem provide the necessary sequence for obtaining a consistent estimator of a set $\mathscr{A}$. Note that $h_n \to 0$ in probability and $nh_n^2 \to \infty$ in probability. Thus Theorem 1 has to be proved with this modification. Note also that $\mathscr{P}_{\mathscr{A}}$ in Theorem 1 is uniform. In practice, $\mathscr{P}_{\mathscr{A}}$ need not be uniform. Hence, this modification has been incorporated by the following theorem.

Theorem 3. Let $\mathscr{A} \subset \mathbb{B}$ and $X_1, X_2, \ldots, X_n$ be independent and identically distributed random vectors with continuous support $\mathscr{P}_{\mathscr{A}}$. Let

$$\mathscr{A}_{1n} = \cup_{i=1}^{n}\{x \in \mathbb{R}^2 : |x - X_i| \le h_n\}. \tag{4}$$

Then $\mathscr{A}_{1n}$ is a consistent estimate of $\mathscr{A}$.

Proof. It can be found in reference (8).

Theorem 4. Let $\mathscr{A} \in \mathrm{B}$ and $\mathscr{P}_{\mathscr{A}}$ be the uniform distribution on $\mathscr{A}$. Let for every $Z \subseteq \mathbb{R}^2$,

$$\beta_\mu(Z) = \cap_{D_\mu(y) \subseteq Z^c} D^c_\mu(y), \tag{5}$$

where $D_\mu(y)$ denotes an open disc with radius μ and center at y. Let $A_{\mu_n} = \beta_{\mu_n}(S_n)$, where $\mu_n \to 0$ as $n \to \infty$. Then for any positive ε, we can find a μ_n such that

$$\mathscr{P}\{\lambda(\mathscr{A}_{\mu_n} \triangle \mathscr{A}) < \varepsilon\} > 1 - \varepsilon \quad \text{for sufficiently large } n.$$

Proof. It can be found in reference (11).

Note. We wanted to take the μ_n in Theorem 4 as h_n in Theorems 2 and 3, because $h_n \to 0$ as $n \to \infty$ in probability. Note that $\beta_\mu(Z)$ in Theorem 4 is nothing but the $1/\mu$-hull of Z.

Proposed criterion for the selection of α. We consider the value of α for α-hull as

$$\alpha = -\frac{1}{h_n}. \tag{6}$$

Observe that h_n has been shown to give good results in one method of set estimation [Theorem 3].

A solution to the problem of obtaining the contour of a set of points in $\mathbb{R}^2$ is given below:

(i) Obtain the value of α using equation (6) and correspondingly find the α-hull of $\mathscr{S}_n = \{x_1, x_2, \ldots, x_n\}$.
(ii) Obtain the boundary of the estimated set.

In this section, we have furnished a selection criterion of α [equation (6)] for α-hulls corresponding to a point set in $\mathbb{R}^2$. The selection criterion is tested with several examples and obtained good results. The results are discussed in Section 4.

4. EXPERIMENTAL RESULTS AND CONVERGENCE

4.1. Experimental results

To show the effectiveness of the proposed selection procedure of α for α-hulls, different possible pattern classes were generated artificially and the α-hulls were determined for each of the classes where the value of αs were decided by the proposed selection procedure. The obtained shapes were found to be quite satisfactory in all the cases. Figure 1(a) and (b) show two typical pattern classes in the plane. Note that the pattern class shown in Fig. 1(b) has two holes (one circular and one rectangular), whereas the other [Fig. 1(a)] does not have any hole. Training samples of size 200 [as shown in Fig. 1(a) and (b)] are chosen randomly from each of the two classes.

Initially in all the cases, the MST of the training sample sets are obtained [Fig. 2(a) and (b)] and correspondingly the values of α are decided as -3.0527 and -2.7381 corresponding to the training point sets in Fig. 1(a) and (b), respectively. Accordingly, the α-hulls are determined and are shown in Fig. 3(a) and (b), are respectively. It is to be observed here that the estimated classes (α-hulls) are good representations of their original pattern classes.

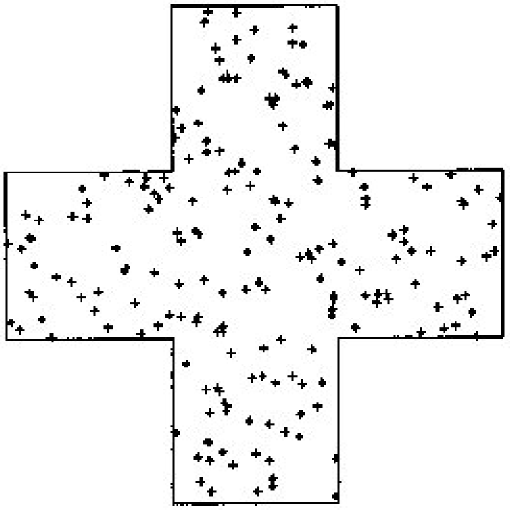
4.2. Convergence with sample size

Convergence of the α-hulls of the training sample sets to the original pattern class is shown here experimentally as well as analytically. For any shape determining approach based on sampled points, the accuracy of the estimated shapes, in general, should improve with the increase of the size (number) of the sampled points. It will be shown in this section that with the αs selected from the training sets by the proposed procedure, the α-hulls also have this property.

An artificially generated pattern class (circular ring shaped) with center at (2,2) (Fig. 4) has been considered in this section to demonstrate the said convergence property of the α-hulls using the proposed selection of α. Four different sets of data are chosen randomly from it with sizes 200, 300, 400, and 500. The proposed procedure initially finds the values of α corresponding to the aforesaid four data sets and these are shown in Table 1. The corresponding α-hulls are shown in Fig. 5(a)–(d).

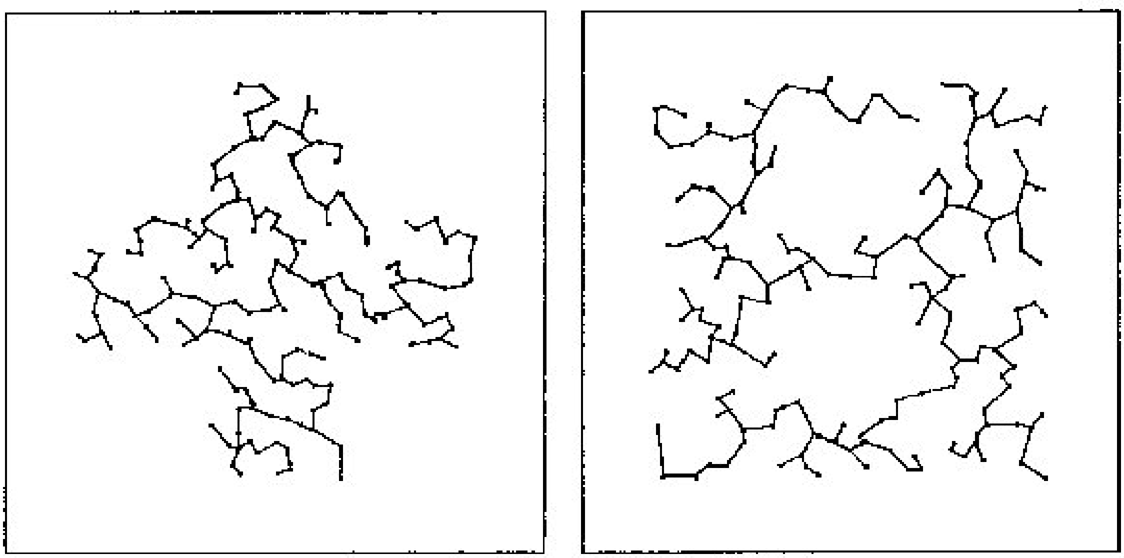
It can be seen from these results that as the sample size (n) increases, the estimated classes (α-hulls) gradually converge to the original pattern classes.

To visualize the convergence property more prominently, we have found the symmetric differences of the estimated classes or sets (α-hulls) (Fig. 5(a)–(d)) with the original pattern class (Fig. 4) and shown them in Fig. 6(a)–(d). The areas under the symmetric differences are also calculated (Table 1) and it is found that the area



(a) (b)

Fig. 1. Two typical pattern classes.



(a) (b)

Fig. 2. (a) and (b): MST corresponding to pattern classes in Fig. 1(a) and (b).

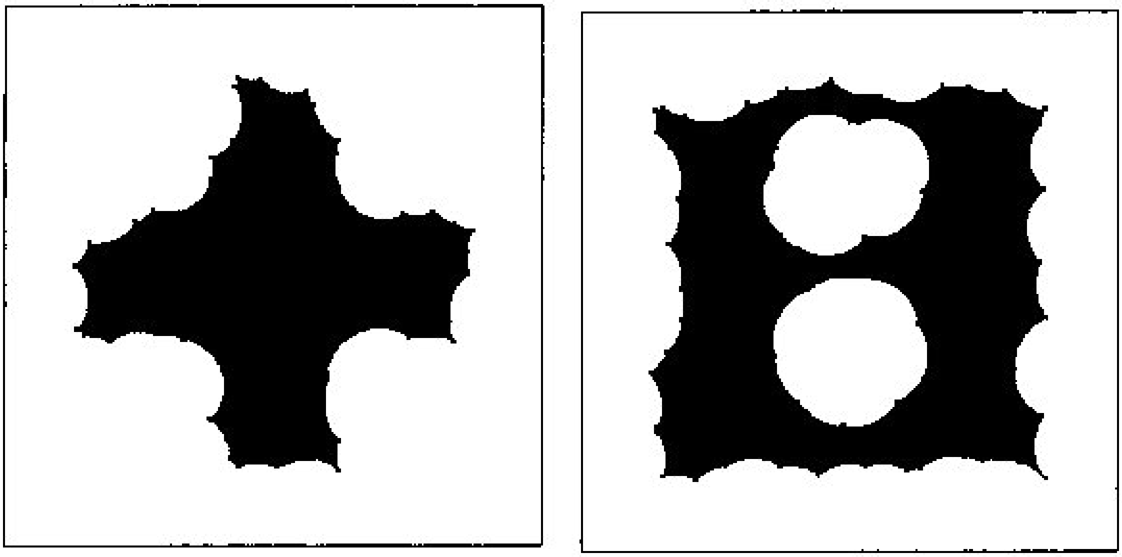
Table 1. The goodness of fit between the α-hulls, based on the proposed selection criterion of α, and the original class in Fig. 4

Estimated class (α-hull)	Sample size	Value of α	Symmetric difference	Hausdorff metric	Similarity metric
Fig. 5(a)	200	3.2239	1.1732	0.3719	0.5473
Fig. 5(b)	300	3.5745	0.9422	0.3264	0.4591
Fig. 5(c)	400	3.8463	0.7447	0.2511	0.3537
Fig. 5(d)	500	4.0624	0.6997	0.2158	0.2954

decreases with an increase in the size of the sampled points.

The aforementioned convergence property is also verified analytically using two distance measures (metrics). One of them is the Hausdorff metric[13] and the other is called the similarity metric (*Sim*) that has been defined recently by Mandal *et al.*[14,15] It has been shown that the values of both the metrics tend toward zero as $n \to \infty$.

Definition 8 (Kuratowski[13]). Let $\mathscr{A}$ and $\mathscr{B}$ be two finite sets in $\mathbb{R}^2$. Then the Hausdorff distance (metric)



(a) (b)

Fig. 3. (a) and (b): α-hulls corresponding to pattern classes in Fig. 1(a) and (b).

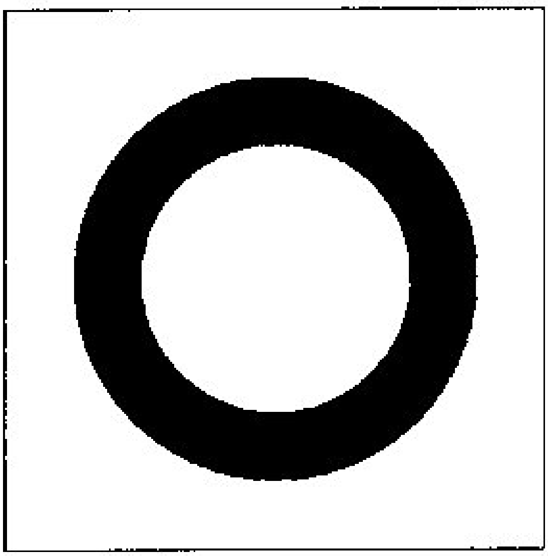


Fig. 4. A circular ring-shaped pattern class.

between $\mathscr{A}$ and $\mathscr{B}$, denoted by Dist($\mathscr{A}$,$\mathscr{B}$), is defined as

$$\mathrm{Dist}(\mathscr{A},\mathscr{B}) = \max\left\{\max_{x\in\mathscr{A}} \delta(x,\mathscr{B}),\ \max_{y\in\mathscr{B}} \delta(y,\mathscr{A})\right\}, \quad (7)$$

where

$$\delta(x,\mathscr{B}) = \min_{y\in\mathscr{B}} d(x,y), \quad \delta(y,\mathscr{A}) = \min_{x\in\mathscr{A}} d(x,y).$$

and $d(x,y)$ is the Euclidean distance between x and y.

This distance measure Dist($\mathscr{A}$,$\mathscr{B}$) is used here to check the convergence property analytically, where $\mathscr{A}$ is considered as the boundary of the estimated set or class (α-hull) and $\mathscr{B}$ is considered as the boundary of the original class. This distance measure is applied on the four estimated sets or classes (α-hulls) (Fig. 5(a)–(d)) with the original set (Fig. 4) in the following way.

The boundary of the ring (Fig. 4) is approximated by 1246 equally spaced points. This set of 1246 points is considered here as the set $\mathscr{B}$. Similarly, the boundary of each of the estimated classes is approximated by a few equally spaced points to represent the set $\mathscr{B}$. Thus, the values of the Dist measure between each of the estimated classes (Fig. 5(a)–(d)) and the original classes (Fig. 4) are found and shown in Table 1. These results confirm the convergence property of the α-hulls based on the proposed selection criterion of α.

Note that the Hausdorff metric reflects the overall similarity between two closed sets. In order to incorporate the similarity of each of the elements of the sets, Mandal *et al.*[14,15] recently proposed a new metric, named the similarity metric.

Definition 9. Let $\mathscr{A}$ and $\mathscr{B}$ be two finite sets with $t_{\mathscr{A}}$ and $t_{\mathscr{B}}$ elements, respectively. Then a similarity measure between $\mathscr{A}$ and $\mathscr{B}$, denoted by $Sim(\mathscr{A},\mathscr{B})$, is defined as

$$\mathrm{Sim}(\mathscr{A},\mathscr{B}) = \frac{1}{t_{\mathscr{A}}}\sum_{x\in\mathscr{A}}\delta(x,\mathscr{B}) + \frac{1}{t_{\mathscr{B}}}\sum_{y\in\mathscr{B}}\delta(y,\mathscr{A}) - \mathrm{Dist}(\mathscr{A},\mathscr{B}). \quad (8)$$

This Sim metric has also been applied in the same way as the previous case between each of the four estimated sets [Fig. 5(a)–(d)] and the original set (Fig. 4). The values of the Sim measure are shown in Table 1 where the convergence property is seen to be verified.

Hence, the convergence property of the α-hulls using the proposed selection procedure for α is established both experimentally and analytically.

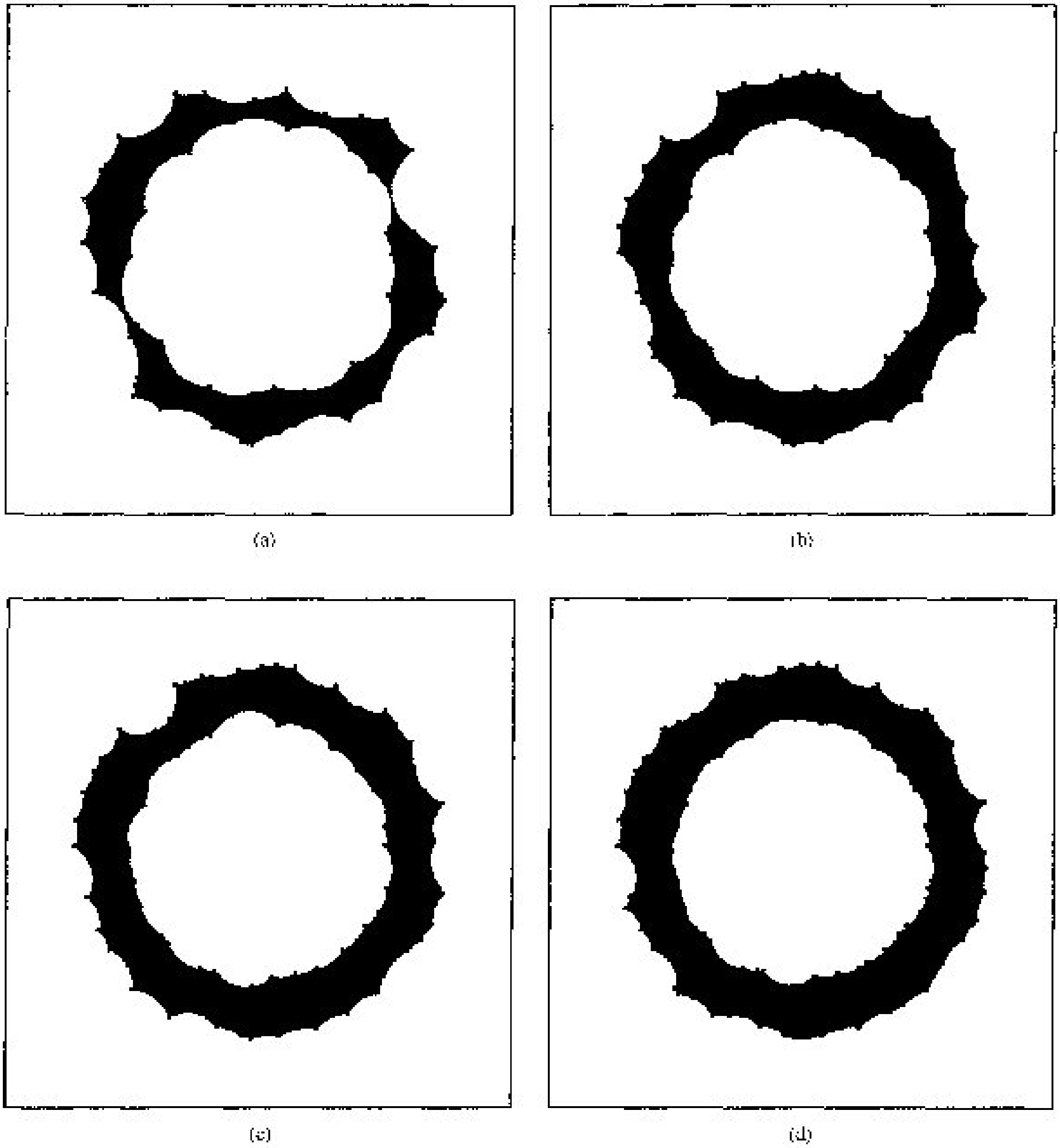


Fig. 5. (a)–(d): Estimated versions (α-hulls) of the class in Fig. 4 based on 200, 300, 400, and 500 samples, respectively.

5. CONCLUSIONS AND DISCUSSION

A decade has passed since Edelsbrunner *et al.*[6] proposed the useful concept of the α-hull to find the shape (external) of a planar set. But in the literature, there exists no work which deals the criterion for the proper selection of the value of α. This paper finds the said criterion based on the concept of the minimum spanning tree. The intuition behind the proposed criterion is justified with the existing mathematical results. Experimental results (alongwith the convergence property) on some artificially generated pattern classes are also provided to demonstrate the usefulness of the proposed procedure in selecting the value of α.

Similar to the α-hulls in the plane, Edelsbrunner and Mucke[16] recently proposed α-hulls for higher dimensions. Probably, a similar selection of α (as proposed here) would be valid for higher dimensions.

We have considered an MST-based criterion for choosing the value of α. There may exist other functions based on inter-point distances which may provide good results.

One of the inherent observations about the α-hulls is that the boundary of the class is restricted by the sampled points. The resulting boundary may omit certain regions of the pattern class which are not confined in it by the sample points. So, it may be necessary to extend the boundaries to some extent to handle the possible

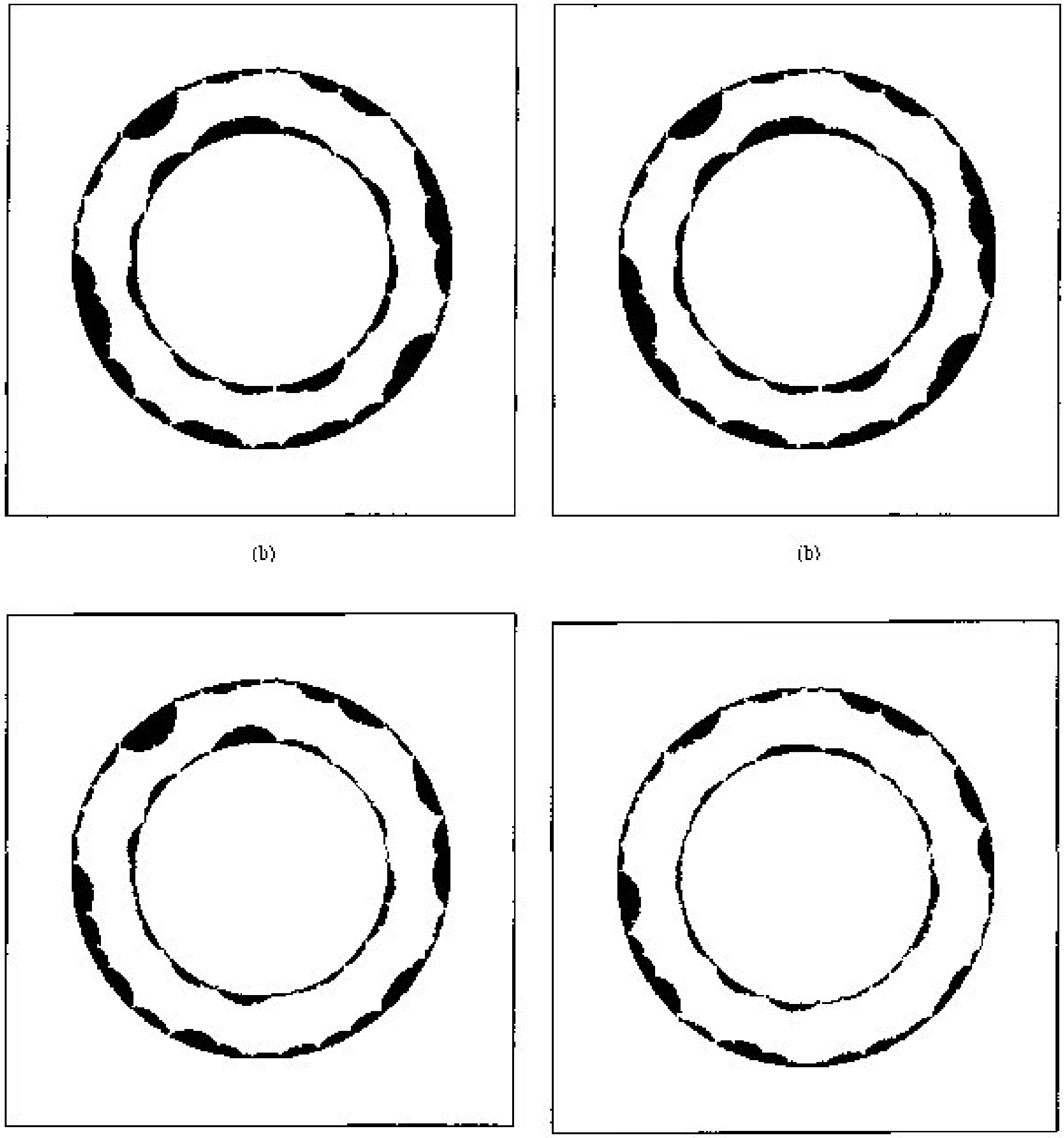


Fig. 6. (a)–(d): Symmetric difference between the estimated classes (α-hulls) in Fig. 5(a)–(d) and the original class in Fig. 4.

uncovered portions of the pattern class by the sampled points. Thus there is the scope of incorporating this concept in α-hulls to define the multivalued shape of a pattern class. For this purpose, the theory of fuzzy sets[17] may be useful. It may be noted that some literature already exists[14,15] for determining the multivalued shape of a pattern class.

REFERENCES

1. R. A. Jarvis, Computing the shape hull of points in the plane, in *Proc. IEEE Comp. Soc. Conf. Pattern Recognition and Image Process.*, pp. 231–241 (1977).
2. S. K. Akl and G. T. Toussaint, Efficient convex hull algorithm for pattern recognition applications, in *Proc. Fourth Int. Jt. Conf. Pattern Recognition*, Kyoto, pp. 483–487 (1978).
3. J. Fairfield, Contoured shape generation forms that people see in dot patterns, in *Proc. IEEE Conf. on Cybernet. and Soc.*, pp. 60–64 (1979).
4. G. T. Toussaint, The relative neighborhood graph of a finite planar set, *Pattern Recognition* **12**, 261–268 (1980).
5. G. T. Toussaint, Pattern recognition and geometrical complexity, in *Proc. Fifth Int. Conf. Pattern Recognition*, Florida, pp. 1324–1347 (1980).
6. H. Edelsbrunner, D. G. Kirkpatrick and R. Seidel, On the shape of a set of points in a plane, *IEEE Trans. Inform. Theory* **IT-29**, 551–559 (1983).

7. M. Worring and A. W. M. Smeulders, Multi-scale analysis of discrete point sets, in *Proc. 11th Int. Conf. Pattern Recognition*, The Hague, pp. 145–148 (1992).
8. C. A. Murthy, On consistent estimation of classes in R^2 in the context of cluster analysis, Ph.D. Thesis, Indian Statistical Institute, Calcutta, India (1988).
9. J. B. Kruskal, On the shortest spanning subtree of a graph and travelling salesman problem, *Proc. Am. Math. Soc.* **7**, 48–50 (1956).
10. C. T. Zahn, Graph theoretic methods for detecting and describing gestalt clusters, *IEEE Trans. Comput.* **C-20**, 68–86 (1971).
11. U. Grenander, *Abstract Inference*. Wiley, New York (1981).
12. C. A. Murthy and D. Dutta Majumder, A method for consistent estimation of compact regions for cluster analysis, in *Proc. 10th Int. Conf. Pattern Recognition*, New Jersey (1990).
13. K. Kuratowski, *Topology*, Vol. I. Academic Press, New York (1966).
14. D. P. Mandal, A multivalued approach for uncertainty management in pattern recognition problems using fuzzy sets, Ph.D. Thesis, Indian Statistical Institute, Calcutta, India (1992).
15. D. P. Mandal, C. A. Murthy and S. K. Pal, Determining the shape of a pattern class from sampled points in R^2, *Int. J. General Systems* **20**, 307–339 (1992).
16. H. Edelsbrunner and E. P. Mucke, Three-dimensional alpha shapes, in *Proc. Workshop Volume Visualization*, pp. 75–82 (1992).
17. L. A. Zadeh, Fuzzy sets, *Inform. Control* **8**(3), 338–353 (1965).

About the Author — DEBA PRASAD MANDAL was born in 1963 at Bagula, West Bengal, India. He obtained the B.Sc. (Honors) degree in Statistics from the Kalyani University, West Bengal, India, in 1984; the Master of Computer Applications (MCA) degree from the Jawaharlal Nehru University, New Delhi, India, in 1988; the Ph.D. degree in Computer Science from the Indian Statistical Institute, Calcutta, India, in 1994. Between February 1994 and March 1996 he was with the Department of Industrial Engineering, University of Osaka Prefecture, Japan under a Japanese Government Postdoctoral Fellowship. At present, he is a lecturer in the Machine Intelligence Unit of the Indian Statistical Institute, Calcutta. His research interests mainly include pattern recognition, image processing, fuzzy sets and systems, remote sensing, and neural networks. Dr Mandal received the Young Scientist Award in Computer Sciences from the Indian Science Congress Association in 1992. He is listed in the World's Who's Who of Men and Women of Distinction and International Directory of Distinguished Leadership. He is a life member of the Indian Science Congress Association (ISCA), and Indian Statistical Institute (ISI); a member of Indian Society for Fuzzy Mathematics and Information Processing (ISFUMIP), and Indian Unit for Pattern Recognition and Artificial Intelligence (IUPRAI).

About the Author — C. A. MURTHY was born in 1958 in Ongole, India. He received the B.Stat. (Honors) degree in 1979, the M.Stat. degree in 1980, and the Ph.D. degree in 1989 from Indian Statistical Institute, Calcutta, India. He visited the Computer Science Department of Michigan State University, East Lansing, as UNDP fellow in 1991–1992. His research interests mainly include pattern recognition, image analysis, fuzzy sets, fractals, neural networks, and genetic algorithms. He is currently working as an Associate Professor in the Machine Intelligence Unit of the Indian Statistical Institute, Calcutta. He is a member of Indian Society for Fuzzy Mathematics and Information Processing (ISFUMIP) and Indian Unit for Pattern Recognition and Artificial Intelligence (IUPRAI).