Theoretical Performance of a Multivalued Recognition System

Deba Prasad Mandal, C. A. Murthy, and Sankar K. Pal, *Fellow, IEEE*

Abstract—A multivalued recognition system was formulated by the authors which has the ability of discriminating the nonoverlapping, and overlapping and no-class (i.e., ambiguous/doubtful) regions and of analyzing the associated uncertainties by providing output decisions in four states, namely, *single*, *first-second*, *combined*, and *null choices*. The *single choices* correspond to the nonoverlapping regions, whereas the overlapping regions are reflected by the *first-second* and *combined choices*. The *null choices* reflect the portions outside the pattern classes and/or the portions of the pattern classes uncovered by the training samples. A theoretical analysis of these characteristics and of the performance of the recognition system has been provided in the present article. It has been shown theoretically that with the increase in the size of the training samples, the estimates of the overlapping, nonoverlapping, and no-class regions tend to their actual sizes. All analytical findings have been substantiated with experimental results various situations in one- and two-dimensional feature spaces. Bayes decision boundaries are always found to lie within the *combined choice* region as provided by the multivalued recognition system. The present investigation, in turn, establishes analytically the justification of providing multivalued output decisions in four states for managing uncertainties arising from ambiguous regions.

I. Introduction

THE multivalued recognition system, described by the authors in [1], decomposes the feature space based on geometric structure and the relative position of the pattern classes, and provides output as multiple choice for classes with their preferences. It classifies a sample either as a *single choice* (possibility to be only in one class), or as a *combined choice* (possibility of belonging to more than one class with same preference), or as a *first-second choice* (possibility of belonging to more than one class with different preferences), or as a *null choice* (possibility of not belonging to any of the classes). The *single choices* correspond to the nonoverlapping regions whereas the overlapping regions are reflected by the *first-second* and *combined choices*. The *null choices* reflect the regions outside the pattern classes and/or the regions of the pattern classes uncovered by the training samples (even with its extended portions). Thus, the system has the ability of discriminating the overlapping and no-class (i.e., ambiguous/doubtful) regions and of analyzing the assoicated uncertainties by providing output in four states.

In this paper, a theoretical analysis of the aforementioned characteristics and the performance of the recognition system

is provided. Initially, the regions corresponding to the four output forms are determined analytically and the estimates of the nonoverlapping, overlapping, and no-class regions in the feature space are then determined. Various possible situations in one- and two-dimensional feature spaces with two classes are considered here. It has been shown that with the increase in the size of the training samples, the estimates of the overlapping, nonoverlapping, and no-class (null choice) regions tend to their actual sizes.

All the analytical findings have been substantiated with experimental results. For a comparative study of the recognition system with a conventional one, the Bayes classifier [2] is implemented on the considered various data sets. The Bayes thresholds between classes are always found to lie within the *combined choice region* of the multivalued recognition system. The present investigation, in turn, analytically establishes the justification of considering output decision in four states for managing uncertainties arising from ambiguous regions.

A brief description of the recognition system [1] is furnished in Section II. Section III provides some theorems which are used in the subsequent sections. The theoretical analysis of the system in one- and two-dimensional feature spaces are dealt with in Sections IV and V, respectively. Section VI contians the conclusions.

II. Multivalued Recognition System [1]

The multivalued recognition system developed by Mandal, Murthy, and Pal is described here in brief. The system has the capability of handling various input patterns and it provides multiple choice for classes as the output decision. For describing the system, let us consider an M class $C_1, C_2, \cdots, C_j, \cdots C_M$ and N feature $(F_1, F_2, \cdots, F_i, \cdots, F_N)$ problem. The block diagram of the recognition system is shown in Fig. 1. It consists of two parts, namely, *learning* and *fuzzy processor*. The learning section basically decomposes the entire feature space into some *space subdomains* and finds a relational matrix. The fuzzy processor uses the relational matrix in the modified compositional rule of inference [3] to decide about the class or classes in which a pattern X may belong. The operations of various blocks in Fig. 1 are discussed below.

A. Learning

The learning section has two blocks, namely, *preprocessing* and *relational matrix estimator*. The *space subdomains* in the feature space are obtained in the preprocessing block and the block relational matrix estimator finds a relational matrix R.
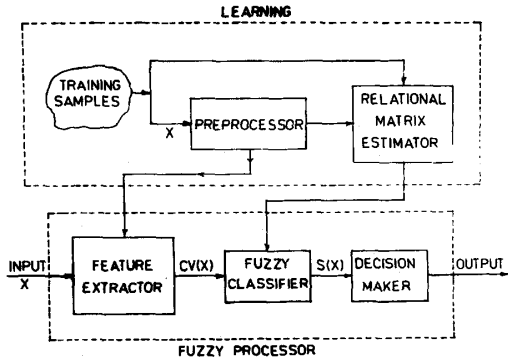
Fig. 1. Block diagram of the multivalued recognition system [1].

Fig. 2. Piecewise linear triangular function [1].

Initially depending on the geometric structure [4]–[6] and the relative positions of the pattern classes in the feature space, the training sample set of each pattern class (say C_j) is decomposed into a few sample groups (say, m_j). That is, the training sample sets of all the M pattern classes are decomposed into $\Sigma_{j=1}^{M} m_j (= \hat{M}$, say) groups. Accordingly, each individual feature axis (say, F_i) is divided into a number (say, n_i) of subdomains (referred to as *feature subdomains*) to highlight the sample goups. Each of the *feature subdomains* is extended to an extent to incorporate the portion (of the pattern classes) possibly uncovered by the training samples. Thus the whole feature space is decomposed into some, say $\hat{N} (= \Pi_{i=1}^{N} n_i)$ overlapping *space subdomains*.

The feature *subdomains* are then characterized by the piecewise linear triangular functions. That is, the gth *feature subdomain* along the ith feature space is characterized by $T(x, \alpha_{ig}, \beta_{l_ig}, \beta_{u_ig}, \Gamma_{l_ig}, \Gamma_{u_ig})$ in which α_{ig} is the central part where the membership value is 1.0; β_{l_ig} and β_{u_ig} are the lower and upper most ambiguous (cross-over) points where the membership values are 0.5; Γ_{l_ig} and Γ_{u_ig} are the lower and upper end points beyond which the membership values are zero. Here

$$\Gamma_{l_ig} = \beta_{l_ig} - \varepsilon_{ig}$$

and

$$\Gamma_{u_ig} = \beta_{u_ig} + \varepsilon_{ig}$$

where

$$\varepsilon_{ig} = \delta_t (\beta_{u_ig} - \beta_{l_ig}) \qquad (1)$$

is the extension factor for the gth domain along ith feature axis and δ_t is the accuracy factor. The value of δ_t is decided as [4]–[6]

$$\frac{1}{t^{1/N}} \leq \delta_t \leq \frac{1}{t^{1/(N+1)}} \qquad (2)$$

so that as the number of training samples $(t) \to \infty, \delta_t \to 0$, and $t\delta_t^N \to \infty$.

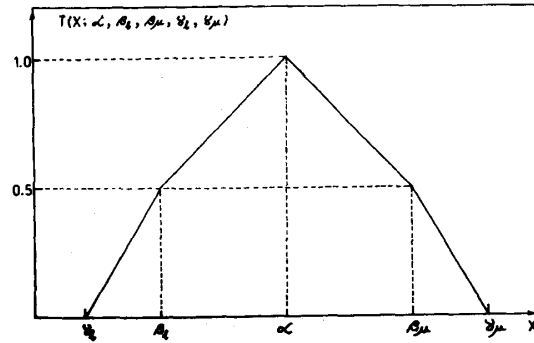The functional form of the piecewise linear triangular function (T) is stated below.

$$T(x; \alpha, \beta_l, \beta_u, \Gamma_l, \Gamma_u) = \begin{cases} \dfrac{1}{2}\left(\dfrac{x - \Gamma_l}{\beta_t - \Gamma_l}\right) \\ \qquad \text{if } \Gamma_l \leq x \leq \beta_l \\ \dfrac{1}{2} + \dfrac{1}{2}\left(\dfrac{x - \beta_l}{\alpha - \beta_l}\right) \\ \qquad \text{if } \beta_l \leq x \leq \alpha \\ \dfrac{1}{2} + \dfrac{1}{2}\left(\dfrac{x - \beta_u}{\alpha - \beta_u}\right) \\ \qquad \text{if } \alpha \leq x \leq \beta_u \\ \dfrac{1}{2}\left(\dfrac{x - \Gamma_u}{\beta_u - \Gamma_u}\right) \\ \qquad \text{if } \beta_u \leq x \leq \Gamma_u \\ 0 \qquad \text{otherwise} \end{cases} \qquad (3)$$

Such a piecewise linear triangular function is graphically shown in Fig. 2. Note that a π function [1] (which is a quadratic function) may also be used to characterize the *feature subdomains*.

The relational matrix R denotes the compatibility of the pattern classes corresponding to the space subdomains. The order of R is $\hat{N} \times M$. The matrix R is estimated from the training samples in the relational matrix estimator block. Let r_{hj} denote the (h, j)th element of R, i.e., the element corresponding to the hth space subdomain and jth pattern class. The value of r_{hj} is decided as

$$r_{hj} = \begin{cases} 0 & \text{if } h\text{th space subdomain does not} \\ & \text{highlight } j\text{th pattern class;} \\ 1 & \text{if } h\text{th space subdomain highlights} \\ & \text{only } j\text{th pattern class;} \\ (0.8)^{\frac{NS_h}{NG_h NC_j^h}} & \text{if } h\text{th space subdomain highlights} \\ & j\text{th pattern class along with} \\ & \text{some other classes.} \end{cases}$$

$$\qquad (4)$$

Here NG_h is the number of training sample groups highlighted by the space subdomain h; NC_j^h is the number of training samples from the jth class (C_j) is the hth space subdomain and NS_h is the total number of training samples in the hth space subdomain i.e., $NS_h = \Sigma_{j=1}^{M} NC_j^h$.

So the block relational matrix estimator provides R which is utilized in the fuzzy classifier block to find the final output of the recognition system.

B. Fuzzy Processor

This section consists of three parts, namely, *feature extractor*, *fuzzy classifier*, and *decision maker*. It uses the relational matrix in the modified compositional rule of inference [1], [3] to decide about the class or classes in which a pattern X may belong.

The feature extractor block determines the membership values (degree of belonging) of an unknown pattern $X(= [F_1, F_2, \cdots, F_N])$ to the space subdomains. Initially, each individual feature information is considered separately to find the membership values (f) of X to the corresponding feature subdomains.

For an unknown pattern of X, a characteristic vector $CV(X)$ was then defined as

$$CV(X) = (cv_1(X), cv_2(X), \cdots, cv_{\hat{N}}(X)) \qquad (5)$$

where the hth element $cv_h(X)$ denotes the degree of its belonging to the hth space subdomain. Let the hth $(h = 1, 2, \cdots, \hat{N})$ space subdomain consist of the feature subdomains $g_1^h, g_2^h, \cdots, g_i^h, \cdots, g_N^h$. Suppose $f_{g_i^h}(X)$ represents the membership of X to belong in the g_i^hth feature subdomain. Then the hth element of $CV(X)$, i.e., the membership value of X corresponding to hth space subdomain, was defined [1] as

$$cv_h(X) = \begin{cases} \dfrac{1}{N}\sum_1^N f_{g_i^h}(X) & \text{if } f_{g_i^h}(X) > 0 \\ & \quad \text{for all } i = 1, 2, \cdots, N \\ 0 & \text{otherwise} \end{cases}$$
$$h = 1, 2, \cdots, \hat{N}. \qquad (6)$$

So the block feature extractor finds a characteristic vector with $\hat{N}$ elements (for $\hat{N}$ space subdomains) corresponding to each input pattern X.

Suppose the membership $(cv_h(X))$ values of a feature point are positive for two or more neighboring subdomains and one particular pattern class is highlighted by the said subdomains. Then it indicates that the said feature point lies in two or more training sample groups of a pattern class. This in turn increases the possibility for the said feature point to belong in the actual pattern class. We call this notion as the neighboring effect and to incorporate this effect in the characteristic vectors $CV(X)$, M neighboring characteristic vectors corresponding to M pattern classes are defined as

$$CNV_j(X) = (cnv_{j1}(X), cnv_{j2}(X), \cdots, cnv_{j\hat{N}}(X))$$
$$j = 1, 2, \cdots, M \qquad (7)$$

where the hth element if $cnv_{jg}(X)$ denotes the degree of belonging to the jth class based on hth subdomain and the effect of neighboring subdomains of h on the jth class. It is

defined as

$$cnv_{jh}(X) = \begin{cases} \min\{1, cv_h(X) + \dfrac{1}{2N}cv_{\hat{h}}(X)\} \\ \quad \text{when } r_{hj} > 0 \text{ and } r_{\hat{h}j} > 0 \\ \quad \text{and } cv_{\hat{h}}(X) \text{ is the maximum} \\ \quad \text{among all the elements in the} \\ \quad \text{neighbor of } h \text{ which} \\ \quad \text{highlights the class } C_j \\ cv_h(X) \quad \text{otherwise.} \end{cases} \qquad (8)$$

These $CNV_j(X)$'s along with the relational matrix R are utilized in the fuzzy classifier to find the degree of similarity of the input X to the various pattern classes.

A *class similarity vector* $S(X)$ was then defined as

$$S(X) = (s_1(X), s_2(X), \cdots, s_M(X)) \qquad (9)$$

where the jth element $s_j(X)$ denotes the degree of similarity of a pattern X to the jth class. The $s_j(X)$ is determined as

$$s_j(X) = \begin{cases} \max\limits_{h=1,2,\cdots,\hat{N}} \left\{\tfrac{1}{2}(cnv_{jh}(X) + r_{hj})\right\} \\ \quad \text{if } cnv_{jh}(X) > 0 \text{ and } r_{hj} > 0 \\ 0 \quad \text{otherwise} \end{cases} \qquad (10)$$

where $cnv_{jh}(X)$ is the hth element of $CNV_j(X)$; r_{hj} is the (h, j)th entry of R, and $\hat{N}$ is the number of space subdomains. Therefore, the block fuzzy classifier finds a class similarity vector $S(X)$ corresponding to an unknown input X.

The similarity vector $S(X)$ is then analyzed in the decision marker block. The system provides output decision in one of the following forms.

- *Single Choice:* If the entry in $S(X)$ corresponding to only one class, say C_j, is positive then the class C_j is considered as the output with single choice.
- *Combined Choice:* If the entries in $S(X)$ corresponding to more than one class are positive and if they are nearly same (difference ≤ 0.05) then the said classes are considered as output with combined choice.
- *First-Second Choice:* If the entries in $S(X)$ corresponding to at least two classes are positive and the said entries do not satisfy the criteria for combined choice then first-second choice is considered. The highest two entries in $S(X)$ are taken as the first and second choices, respectively.
- *Null Choice:* If all the entries in $S(X)$ are zero then the system refuses to assign the unknown sample to any class, i.e., null choice is given.

It needs to be mentioned here that the single choices correspond to the nonoverlapping regions whereas, the overlapping regions are reflected by the first-second and combined choices. The null choices reflect the regions outside the pattern classes and/or the regions of the pattern classes uncovered by the training samples.

In order to give the final decision in linguistic form regarding the class or classes to which the unknown input pattern X may belong, the confidence factor (CF), as defined in [1], is calculated and accordingly the final output decision in linguistic form is provided.

III. SOME THEOREMS

In order to provide the theoretical analysis of the multi-valued recognition system [1], we have presented here some theorems.

Definition 1: A set $A \subseteq \mathbb{R}^1$ is said to be a pattern class [7] if A is a compact interval. ∎

Definition 1.1: A set $\mathcal{A} \subseteq \mathbb{R}^N$ is said to be a pattern class [7] if

1) $\mathcal{A}$ is path connected compact,
2) $cl(Int(\mathcal{A})) = \mathcal{A}$, [$cl$ means closure, Int means interior]
3) $Int(\mathcal{A})$ is path connected, and
4) $\lambda(\delta\mathcal{A}) = 0$ where $\delta\mathcal{A} = \mathcal{A} \cap cl(\mathcal{A}^c)$ and λ is the Lebesgue measure on $\mathbb{R}^N$.

Let $\mathbb{B} = \{\mathcal{A} : \mathcal{A} \subseteq \mathbb{R}^N, \mathcal{A}$ is a pattern class$\}$. ∎

Definition 2: $\mathcal{P}$ is said to be probability measure on $\mathcal{A}, \mathcal{A} \in \mathbb{B}$ if $\mathcal{P}$ satisfies the following properties [7]–[9].

1) $\mathcal{P} \ll \lambda$ [λ is the Lebesgue measure on $\mathbb{R}^N$].
2) Let $f = d\mathcal{P}/d\lambda$ be the density on $\mathcal{A}$. Then $f(x) > 0 \quad \forall x \in Int(\mathcal{A})$. ∎

Note that Definition 2 restricts the probability measures under consideration on $\mathcal{A}$.

Result: Let $X_1, X_2, \cdots, X_t$ be independent and identically distributed random variables with density f. If $\mathcal{A}$ is a class and $\mathcal{P}$ is a probability measure on $\mathcal{A}$ then

$$\forall \mathcal{B} \subseteq \mathcal{A}, \quad \mathcal{B}$$

is open,

$$\Rightarrow \quad \mathcal{P}(X_1 \notin \mathcal{B}, X_2 \notin \mathcal{B}, \cdots, X_t \notin \mathcal{B}) \rightarrow 0 \text{ as } t \rightarrow \infty.$$

Proof: Proof is obvious. ∎

Theorem 1: Let $X_1, X_2, \cdots, X_t$ be independent and identically distributed random variables taking values in $[a, b]$ with continuous density f.

Let $X_{(1)} = \min\{X_1, \cdots, X_t\}$ and $X_{(t)} = \max\{X_1, \cdots, X_t\}$. Then $X_{(1)}$ goes to a in probability as $t \rightarrow \infty$ and $X_{(t)}$ goes to b in probability as $t \rightarrow \infty$.

Proof: Let F be the density function, i.e., $F(x) = \int_a^x f(y)\,dy, a \le x \le b$. Then $F(a) = 0$ and $F(b) = 1$.

$$\mathcal{P}(X_{(1)} > y) = \mathcal{P}(X_1 > y, X_2 > y, \cdots, X_t > y)$$
$$= [\mathcal{P}(X_1 > y)]^t$$
$$= [1 - \mathcal{P}(X_1 \le y)]^t$$
$$= [1 - F(y)]^t \quad a \le y \le b \quad (11)$$

Let $\varepsilon > 0$.

$$\mathcal{P}(|X_{(1)} - a| > \varepsilon) \rightarrow 0 \quad \text{for any} \quad \varepsilon > 0.$$
$$\Leftrightarrow \quad \mathcal{P}(X_{(1)} - a > \varepsilon) \rightarrow 0$$
$$\Leftrightarrow \quad \mathcal{P}(X_{(1)} > a + \varepsilon) \rightarrow 0$$
$$\Leftrightarrow \quad [1 - F(a + \varepsilon)]^t \rightarrow 0 \quad [\text{by } (11)]$$

which is true as $t \rightarrow \infty$.

So $X_{(1)}$ goes to a in probability.

$$\mathcal{P}(X_{(t)} \le y) = \mathcal{P}(X_1 \le y, X_2 \le y, \cdots, X_t \le y)$$
$$= [F(y)]^t \quad a \le y \le b \quad (12)$$

Let $\varepsilon > 0$.

$$\mathcal{P}(|X_{(t)} - b| \ge \varepsilon) \rightarrow 0 \quad \text{for any} \quad \varepsilon > 0.$$
$$\Leftrightarrow \quad \mathcal{P}(b - X_{(t)} \ge \varepsilon) \rightarrow 0.$$
$$\Leftrightarrow \quad \mathcal{P}(X_{(t)} \le b - \varepsilon) \rightarrow 0.$$
$$\Leftrightarrow \quad [F(b - \varepsilon)]^t \rightarrow 0 \quad [\text{by } (12)]$$

which is true as $t \rightarrow \infty$.

So $X_{(t)}$ goes to b in probability as $t \rightarrow \infty$. Hence the theorem. ∎

Theorem 2: Let $X_1, X_2, \cdots, X_t$ be independent and identically distributed random variables with continuous density f in $[a, b]$.

Let $\mathcal{A} = [a, b]$ and $\delta_t \rightarrow 0$ as $t \rightarrow \infty$. Let $\mathcal{B}_t = [X_{(1)} - \delta_t, X_{(t)} + \delta_t]$. Then $\lambda(\mathcal{B}_t \Delta \mathcal{A}) \rightarrow 0$ in probability as $t \rightarrow \infty$, where λ is the lebesgue measure on $\mathbb{R}$.

Proof: $a \le X_{(1)} \le X_{(t)} \le b \quad \forall t$.

It is sufficient to show that

$$|a - X_{(1)} + \delta_t| \rightarrow 0 \text{ in probability}$$

and

$$|b - X_{(t)} - \delta_t| \rightarrow 0 \text{ in probability}.$$

Here we shall show that $|b - X_{(t)} - \delta_t| \rightarrow 0$ in probability. The proof for the other statement is similar.

Claim: $\mathcal{P}[|b - X_{(t)} - \delta_t| \le \varepsilon] \rightarrow 1$ as $t \rightarrow \infty$ for every $\varepsilon > 0$.

Proof: Let $\varepsilon > 0$.

Note that $|b - X_{(t)} - \delta_t| \le |b - X_{(t)}| + \delta_t = b - X_{(t)} + \delta_t \,\exists t_1$ such that $t > t_1 \Rightarrow \delta_{<\varepsilon/2}$ since $\delta_t \rightarrow 0$. Let $t_2 = [\frac{2(b-a)}{\varepsilon}] + 1$ ([τ] means the largest integer $\le \tau$). Let $\mathcal{C} = [b - \frac{b-a}{N_2}, b]$.

$\Rightarrow \quad \mathcal{P}(X_1 \notin \mathcal{C}, X_2 \notin \mathcal{C}, \cdots, X_t \notin \mathcal{C}) \rightarrow 0$ as $t \rightarrow \infty$. Hence $\mathcal{P}(\exists t$ such that $X_t \in \mathcal{C}) \rightarrow 1$ as $t \rightarrow \infty$. Then $\mathcal{P}(\exists t$ such that $b - X_t \le \frac{\varepsilon}{2}) \rightarrow 1$ as $t \rightarrow \infty$. or $\mathcal{P}(\exists \cdot t$ such that $b - X_{(t)} \le \frac{\varepsilon}{2}) \rightarrow 1$ as $t \rightarrow \infty$.

Hence the theorem. ∎

Theorems 1 and 2 are used in the subsequent sections to verify the following proposition.

Proposition 1: As the sample size increases, the estimated overlapping, nonoverlapping, and no-class regions tend to their actual size, respectively. ∎

Here, we have considered overlapping and nonoverlapping classes in $\mathbb{R}^1$ and $\mathbb{R}^2$. The classes in $\mathbb{R}^2$ are considered to be regular shaped (rectangular and circular). Piecewise linear triangular functions (3) has been considered as membership functions.

For the nonoverlapping pattern classes, the training samples of the clases are obviously nonoverlapping. But for the overlapping pattern classes, the training samples are either overlapping or nonoverlapping. Note that if the training samples are nonoverlapping then the membership functions of the feature *feature subdomains* may be overlapping. All these situations have been considered in the proposed analysis.

IV. ANALYSIS IN 1-D FEATURE SPACE

To provide the theoretical analysis of the multivalued recognition system in one-dimensional feature (F) space, a two
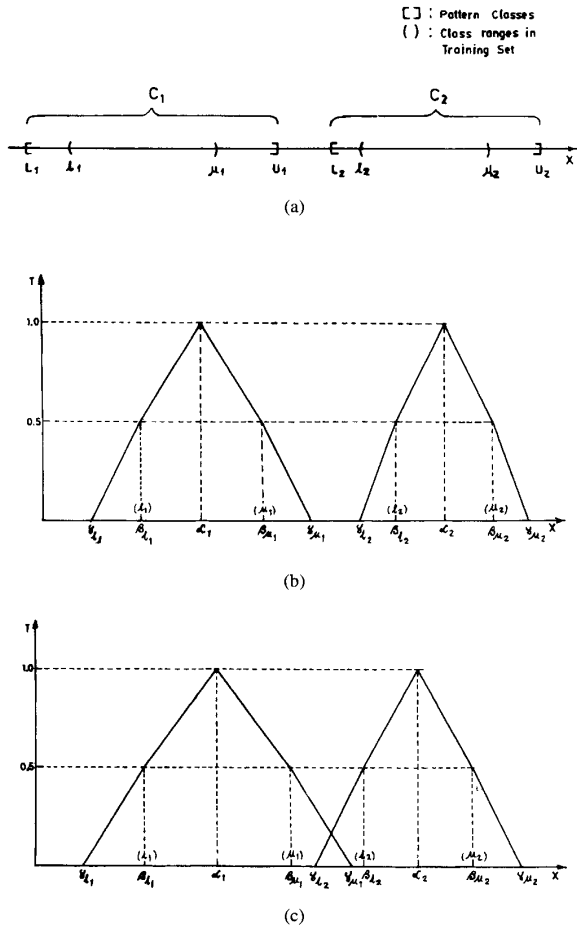
(a)

(b)

(c)

Fig. 3. (a) Two nonoverlapping classes (1-D) with a training set. (b) Subdomains (disjoint) with their membership functions for the pattern classes in (a). (c) Subdomains (overlapping) with their membership functions for the pattern classes in (a).

class (C_1 and C_2) problem is considered here. Since there is only one feature, the space subdomains are same as the feature subdomains. Suppose $[L_1, U_1]$ and $[L_2, U_2]$ [Fig. 3(a)] denote the actual class ranges corresponding to the classes C_1 and C_2, respectively. Suppose also, $[l_1, u_1]$ and $[l_2, u_2]$ [Fig. 3(a)] are the ranges of training samples corresponding to C_1 and C_2 for a particular training set. To find the regions corresponding to *single, first-second, combined,* and *null choices,* and to verify Proposition 1, different possible cases are considered.

A. Nonoverlapping Pattern Classes

This case is shown in Fig. 3(a). Initially, the system finds two distinct feature subdomains $[l_1, u_1]$ and $[l_2, u_2]$ corresponding to the classes C_1 and C_2. The feature subdomains are denoted by D_1 and D_2. Now these feature subdomains are extended to some extent using piecewise linear triangular functions (3) T_1 and T_2, respectively. The feature subdomains with their characterizing membership functions are shown in

Fig. 3(b). Note that

$$\beta_{l_g} = l_g; \qquad \beta_{u_g} = u_g;$$

$$\Gamma_{l_g} = l_g - \varepsilon_g \quad \text{and} \quad \beta_{u_g} = u_g + \varepsilon_g$$

for $g = 1, 2$.

Here ε_g (1) is the extended portion for the gth feature subdomain D_g ($g = 1, 2$).

The relational matrix R in this case will be

$$R = ((r_{gj}))_{g=1,2; j=1,2}$$

where

$$r_{gj} = \begin{cases} 1.0 & \text{for } g = j \\ 0 & \text{for } g \neq j. \end{cases}$$

Case 1: Feature subdomains are disjoint.

This case is shown in Fig. 3(b) where the feature subdomains D_1 and D_2 are disjoint. Here, it is easy to show that the single choice region for C_1 is $(\Gamma_{l_1}, \Gamma_{u_1})$; the singel choice region for C_2 is $(\Gamma_{l2}, \Gamma_{u_2})$ and the remaining points in the feature space represent the null choice region. ∎

Case 2: Feature subdomains are overlapping.

This case is demonstrated in Fig. 3(c) where the feature subdomains D_1 and D_2 are overlapping in their extended portions. It is clear from Fig. 3(c) that $(\Gamma_{l_1}, \Gamma_{l_2}]$ and $[\Gamma_{u_1}, \Gamma_{u2})$ represent the single choice regions for C_1 and C_2, respectively. For $X \in (\Gamma_{l_2}, \Gamma_{u_1})$, the elements in the similarity vector $S(X)$ are all positive, i.e., $s_1(X) > 0$ and $s_2(X) > 0$. This implies that $(\Gamma_{l_2}, \Gamma_{u_1})$ is the overlapping region and we will concentrate now only on this region.

Here two points, say τ_1 and τ_2, can be found such that $\Gamma_{l_2} < \tau_1 < \tau_2 < \Gamma_{u_1}$, and

$$\text{for } X \in (\Gamma_{l_2}, \tau_1), \qquad [s_1(X) - s_2(X)] > 0.05;$$
$$\text{for } X \in (\tau_2, \Gamma_{u_1}), \qquad [s_2(X) - s_1(X)] > 0.05$$

and

$$\text{for } X \in [\tau_1, \tau_2], \qquad |s_1(X) - s_2(X)| \leq 0.05.$$

The values of τ_1 and τ_2 are given by

$$\tau_1 = \frac{u_1\varepsilon_2 + l_2\varepsilon_1 - 1.1\varepsilon_1\varepsilon_2}{\varepsilon_1 + \varepsilon_2}$$

and

$$\tau_2 = \frac{u_1\varepsilon_2 + l_2\varepsilon_1 - 0.9\varepsilon_1\varepsilon_2}{\varepsilon_1 + \varepsilon_2} \tag{13}$$

where ε_g (1) is the extended portion of the feature subdomain D_g ($g = 1, 2$), and l_j and u_j are the lowest and highest values, respectively, among the training samples from the class C_j ($j = 1, 2$).

So, as a summary, we can state the following: The single choice for C_1 is $(\Gamma_{l_1}, \Gamma_{l_2}]$; the first-second choice region with first choice as C_1 is (Γ_{l_2}, τ_1); the combined choice region is $[\tau_1, \tau_2]$; the first-second choice region with first choice as C_2 is (τ_2, Γ_{u_1}); the single choice for C_2 is $[\Gamma_{u_1}, \Gamma_{u_2})$, and the remaining points in the feature space correspond to the no-class (null choice) region. However, if $T_1(\Gamma_{l_2}) \leq 0.05$, i.e.,

$l_2 - u_1 \geq \varepsilon_2 + 0.9\varepsilon_1$, then (Γ_{l_2}, τ_1) becomes a combined choice region instead of a first-second choice region. Again, if $T_1(\Gamma_{l_2}) \leq 0.05$, i.e., if $l_2 - u_1 \geq \varepsilon_1 + 0.9\varepsilon_2$, then (τ_2, Γ_{u_1}) will be a combined choice region instead of a first-second choice region. If both the conditions are satisfied, i.e., if $l_2 - u_1 \geq \varepsilon_2 + 0.9\varepsilon_1$ and $l_2 - u_1 \geq \varepsilon_1 + 0.9\varepsilon_2$, then the complete set $(\Gamma_{l_2}, \Gamma_{u_1})$ will represent a combined choice region.   ∎

Proof of Proposition 1: By Theorem 1, it can be stated that with the increase of training sample size,

$$l_j \to L_j \quad \text{and} \quad u_j \to U_j$$
$$\text{in probability for } j = 1, 2. \tag{14}$$

With the increase of training sample size, the sample sizes in the feature subdomains also increase. Thus, by Theorem 2,

$$\Gamma_{l_g} \to \beta_{l_g} \quad \text{and} \quad \Gamma_{u_g} \to \beta_{u_g}$$
$$\text{in probability for } g = 1, 2. \tag{15}$$

This implies that the uncovered portions of the feature space (by the training sample set) decrease in probability with the increase in sample size.

In case 1, there are no overlapping region, and the single choice region for C_j ($j = 1, 2$) is $(\Gamma_{l_j}, \Gamma_{u_j})$ which tends to the set (L_j, U_j) (the actual nonoverlapping region) in probability.

In case 2, the estimated overlapping region is $(\Gamma_{l_2}, \Gamma_{u_1})$. By (14) and (15), this overlapping region tends to (L_2, U_1) in probability. But $U_1 < L_2$. This implies that the overlapping region decreases in probability.

The single choice regions for C_1 and C_2 are $(\Gamma_{l_1}, \Gamma_{l_2}]$ and $[\Gamma_{u_1}, \Gamma_{u_2})$, respectively. By (14) and (15) these regions converge to (L_1, L_2) and (U_1, U_2) in probability. As $U_1 < L_2$, the regions under single choices for C_1 and C_2 tend to (L_1, U_1) and (L_2, U_2) (actual nonoverlapping regions) is probability.

Hence the proof.   ∎

Experimental Results: To verify the aforementioned analytical results, a two class problem with class ranges [2, 6] and [7, 11] is considered. It may be noted that the pattern classes are nonoverlapping. To implement the recognition system, five training sample sets with 50, 100, 150, 200, and 250 samples from each of the classes are chosen randomly.

The ranges of the training sample sets for the classes C_1 and C_2, and for the feature subdomains D_1 and D_2 corresponding to the five sample sets are shown in Table I(a). It is to be noticed that for first two sample sets (with 50 and 100 samples from each class), the feature subdomains obtained are overlapping whereas, for the remaining three sets (with 150, 200, and 250 samples from each class), the feature subdomains obtained are nonoverlapping. So the first two sample sets represent the case 2, i.e., the feature subdomains are overlapping when the pattern classes are nonoverlapping. The remaining three sample sets represent the case 1, i.e., the feature subdomains are nonoverlapping when the pattern classes are nonoverlapping. The various regions obtained for the first two sample sets are shown in Table I(b). The regions obtained for the remaining three sample sets are shown in Table I(c).

These results verify that with the increase in sample size, the training classes (i.e., portions covered by the training samples)

TABLE I(a)
RANGES OF TRAINING SETS AND FEATURE SUBDOMAINS
FOR THE PATTERN CLASSES [2, 6] AND [7, 11].

	SAMPLE SIZES IN EACH CLASS				
	50	100	150	200	250
Class C_1	2.09–5.92	12.07–5.94	12.06–5.97	12.02–5.98	12.01–5.99
Class C_2	7.16–10.89	7.05–10.92	7.04–10.93	7.02–10.95	7.01–10.98
Subdomain D_1	1.29–6.74	1.37–6.59	1.52–6.44	1.67–6.32	1.78–6.19
Subdomain D_2	6.49–11.65	6.42–11.62	6.56–11.47	6.71–11.33	6.82–11.22

TABLE I(b)
VARIOUS REGIONS AND BAYES THRESHOLD POINTS FOR THE FIRST
TWO TRAINING SETS OF THE PATTERN CLASSES [2, 6] AND [7, 11].

VARIOUS GROUP OF CHOICES	SAMPLE SIZES IN EACH CLASS	
	50	100
Null choice	$-\infty$–1.2908	$-\infty$–1.3727
Single choice (C_1)	1.2909–6.4947	1.3728–6.4263
First-second choice (C_1)	6.4948–6.5402	6.4264–6.4494
Combined choice	6.5403–6.6966	6.4495–6.5748
First-second choice (C_2)	6.6967–6.7494	6.5749–6.5984
Single choice (C_2)	6.7495–11.6507	6.5985–11.6236
Null choice	11.6508–$+\infty$	11.6237–$+\infty$
Bayes threshold point	6.4859152	6.4435577

TABLE I(c)
VARIOUS REGIONS AND BAYES THRESHOLD POINTS FOR THE REMAINING
THREE TRAINING SETS OF THE PATTERN CLASSES [2, 6] AND [7, 11].

Various Group of Choices	Sample Sizes in Each Class		
	150	200	250
Null choice	$-\infty$–1.5250	$-\infty$–1.6690	$-\infty$–1.7779
Single choice (C_1)	1.5251–6.4461	1.6691–6.3288	1.7480–6.2519
Null choice	6.4462–6.5638	6.3289–6.6969	6.2520–6.7619
Single choice (C_2)	6.5639–11.4736	6.6970–11.3288	6.7620–11.2496
Null choice	11.4737–$+\infty$	11.3289–$+\infty$	11.2197–$+\infty$
Bayes threshold point	6.3621473	6.4193249	6.3900852

tend to the actual classes (Theorem 1) and the feature subdomains with the extended portions tend to their actual sizes (Theorem 2). It is to be noticed that the overlapping regions (i.e., corresponding to combined and first-second choices), the nonoverlapping regions (i.e., corresponding to single choices) and no-class regions (i.e., corresponding to null choices) are tending to their actual sizes with the increase in sample sizes. Thus Proposition 1 has been verified experimentally for the case of nonoverlapping pattern classes in one-dimensional feature space.

To show the recognition performance of the system experimentally, a test sample set with 1000 samples from each of the considered pattern classes is generated. With all the aforementioned five training sets, the system recognized the total test sample set correctly under single choices.

For comparison, the Bayes classifier is also applied on the same pattern classes (assuming normal distributions). The threshold points found between the classes C_1 and C_2 are shown in Table I(b) and (c) corresponding to the first two and the remaining three training sets, respectively. These thresholds are seen to lie in the combined choice region of our recognition system. The Bayes classifier also recognized

$$s_1(X) = \tfrac{1}{2}[T_1(X) + \tfrac{1}{2}T_2(X) + 1.0] \quad \text{as } r_{11} = 1.0$$

and

$$s_2(X) = \tfrac{1}{2}[T_2(X) + r_{22}].$$
$$\Rightarrow \quad s_1(X) - s_2(X) = \tfrac{1}{2}[T_1(X) - \tfrac{1}{2}T_2(X)$$
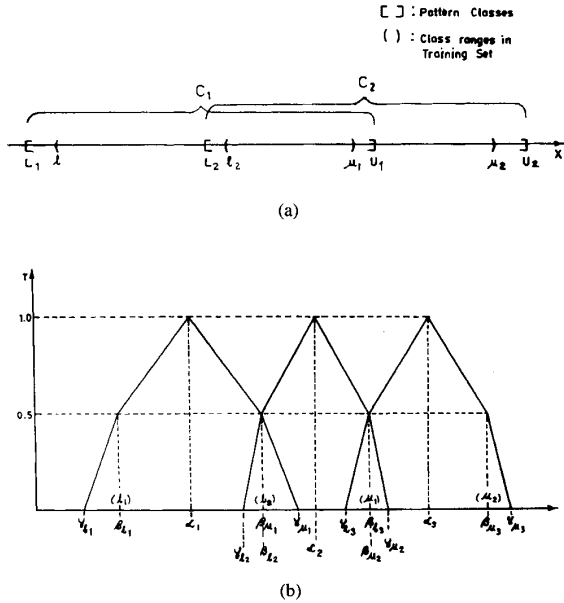$$+ 1.0 - r_{22}].$$

(a)

(b)

Fig. 4. Two overlapping classes (1-D) with a training set. (b) Subdomains with their membership functions for the pattern classes in (a).

all the test samples correctly.

B. Overlapping Pattern Classes

This case is shown in Fig. 4(a). Here the actual overlapping region between the classes C_1 and C_2 is (L_2, U_1). In the training samples, the overlapping region is (l_2, u_1).

Initially the algorithm decomposes the feature space into three feature subdomains as $[l_1, l_2), [l_2, u_1], (u_1, u_2]$ denoted by $D_1, D_2,$ and D_3, respectively. Here the feature subdomain D_1 reflects only C_1, the feature subdomain D_3 reflects only C_2 and the feature subdomain D_2 is overlapping by reflecting both C_1 and C_2. These feature subdomains are extended to some extent and are characterized by piecewise linear triangular functions (3) $T_1, T_2,$ and T_3. The feature subdomains with their membership functions are shown in Fig. 4(b).

The relational matrix R in this case will be

$$R = ((r_{gj}))_{g=1,2,3; j=1,2}$$

where

$$r_{11} = 1.0; \qquad r_{12} = 0;$$
$$r_{21} > 0; \qquad r_{22} > 0;$$
$$r_{31} = 0; \qquad \text{and} \quad r_{32} = 1.0.$$

In the feature region $(\Gamma_{l_1}, \Gamma_{l_2}], T_1(X) > 0$ and $T_2(X) = T_3(X) = 0$. This implies that $s_1(X) > 0$ and $s_2(X) = 0$. So $(\Gamma_{l_1}, \Gamma_{l_2}]$ is the single choice region for C_1.

Similarly, $[\Gamma_{u_2}, \Gamma_{u_3})$ is the single choice region for C_2.

Let us now consider the feature region $(\Gamma_{l_2}, \beta_{l_2}]$. Here, $T_1(X) \geq 0.5 \geq T_2(X) > 0 = T_3(X)$. Hence (see top of page)

As here $T_1(X) \geq 0.5 \geq T_2(X) > 0$ and $r_{22} < 1.0, [s_1(X) - s_2(X)] > 0.05$. Thus, $(\Gamma_{l_2}, \beta_{l_2}]$ comes under a first-second choice region with first choice as C_1.

Similarly, $[\beta_{l_3}, \Gamma_{u_2})$ comes under a first-second choice region with first choice as C_2.

Now consider the region $[\Gamma_{u_1}, \Gamma_{l_3}]$. Here $T_1(X) \geq 0.5$ and $T_2(X) = T_3(X) = 0$. Hence,

$$s_1(X) = \tfrac{1}{2}[T_2(X) + r_{21}]$$

and

$$s_2(X) = \tfrac{1}{2}[T_2(x) + r_{22}].$$

Now

$$|s_1(X) - s_2(X)| \leq 0.05 \Rightarrow |r_{21} - r_{22}| \leq 0.1.$$

So, if $|r_{21} - r_{22}| \leq 0.1$ then $(\Gamma_{u_1}, \Gamma_{l_3}]$ will be a combined choice region. If $r_{21} > r_{22} + 0.1$, then $(\Gamma_{u_1}, \Gamma_{l_3}]$ will be a first-second choice region with first choice as C_1. Otherwise (i.e., if $r_{22} > r_{21} + 0.1$), it becomes a first-second choice region with first choice as C_2.

Now in the region $(\beta_{u_1}, \Gamma_{u_1}), T_2(X) \geq 0.5 \geq T_1(X) \geq 0 = T_3(X)$. Hence

$$s_1(X) = \tfrac{1}{2}[T_2(X) + \tfrac{1}{2}T_1(X) + r_{21}]$$

and

$$s_2(X) = \tfrac{1}{2}[T_2(X) + r_{22}]$$
$$\Rightarrow \quad s_1(X) - s_2(X) = \tfrac{1}{2}[\tfrac{1}{2}T_1(X) + r_{21} - r_{22}].$$

Here a point τ_1 may be found such that $[s_1(\tau_1) - s_2(\tau_1)] = 0.05$ where

$$\tau_1 = l_2 + \varepsilon_1[4(r_{21} - r_{22}) + 0.6] \tag{16}$$

in which ε_1 (1) is the extended portion of the *feature subdomain* D_1 and l_2 is the lowest among the training samples from C_2.

Thus, (β_{u_1}, τ_1) is a first-second choice region with first choice as C_1 and $[\tau_1, \Gamma_{u_1})$ is a combined choice region. If $r_{22} \geq r_{21} + 0.15$, then the complete set $(\beta_{u_1}, \Gamma_{u_1})$ will be a combined choice region. Again, if $r_{21} \geq r_{22} + 0.1$, then $(\beta_{u_1}, \Gamma_{u_1})$ will be a first-second choice region with first choice as C_1.

Similarly, in the region $(\Gamma_{l_3}, \beta_{l_3})$, a point τ_2 may be found such that $[s_2(\tau_2) - s_1(\tau_2)] = 0.05$ where

$$\tau_2 = u_1 - \varepsilon_3[4(r_{22} - r_{21}) + 0.6] \qquad (17)$$

in which ε_3 is the extended portion of the feature subdomain D_3 and u_1 is the highest among the training samples from C_1.

So, (τ_2, β_{l_3}) is a first-second choice region with first choice as C_2 and $(\Gamma_{l_3}, \tau_2]$ is a combined choice region. If $r_{21} \geq r_{22} + 0.15$, then the complete set $(\Gamma_{l_3}, \beta_{l_3})$ will be a combined choice region. Again, if $r_{22} \geq r_{21} + 0.1$, then $(\Gamma_{l_3}, \beta_{l_3})$ will be a first-second choice region with first choice as C_2.

The conclusions, after combining all the previous results, are given below.

The single choice region for C_1 is $(\Gamma_{l_1}, \Gamma_{l_2}]$; the single choice region for C_2 is $[\Gamma_{u_2}, \Gamma_{u_3})$; the overlapping region (i.e., corresponding to combined and first-second choices) is $[\Gamma_{l_2}, \Gamma_{u_2}]$ and the remaining portion in the feature space represent the no-class region.

Note that in the overlapping region, $(\Gamma_{l_2}, \beta_{l_2}]$ is a first-second choice region with first choice as C_1 and $[\beta_{u_2}, \Gamma_{u_2}]$ is a first-second choice region with first choice as C_2.

In the region $(\Gamma_{u_1}, \Gamma_{l_3}]$, if $|r_{21} - r_{22}| \leq 0.1$, then it becomes a combined choice region. If $r_{21} > r_{22} + 0.1$, then $(\Gamma_{u_1}, \Gamma_{l_3}]$ becomes a first-second choice region with first choice as C_1, and if $r_{22} > r_{21} + 0.1$, then $(\Gamma_{u_1}, \Gamma_{l_3}]$ becomes a first-second choice region with first choice as C_2.

In the region $(\beta_{u_1}, \Gamma_{u_1})$, if $r_{22} \geq r_{21} + 0.15$, then it becomes a combined choice region, and if $r_{21} \geq r_{22} + 0.1$, then it comes under a first-second choice region with first choice as C_1. Otherwise (i.e., if $r_{21} - 0.1 < r_{22} < r_{21} + 0.15$), a point τ_1 (16) is found such that (β_{u_1}, τ_1) becomes a first-second choice region with first choice as C_1 and $[\tau_1, \Gamma_{u_1})$ becomes a combined choice region.

In the region $(\Gamma_{l_3}, \beta_{l_3})$, if $r_{21} \geq r_{22} + 0.15$, then it becomes a combined choice region. If $r_{22} \geq r_{21} + 0.1$, then $(\Gamma_{l_3}, \beta_{l_3})$ will be a first-second choice region with first choice as C_2. Otherwise (i.e., if $r_{22} - 0.1 < r_{21} < r_{22} + 0.15$), a point τ_2 (17) is found such that (τ_2, β_{l_3}) will be a first-second choice region with first choice as C_2 and $(\Gamma_{l_3}, \tau_2]$ will be a combined choice region.

Proof of Proposition 1: Here, the actual overlapping region is $[L_2, U_2]$ and the nonoverlapping regions for C_1 and C_2 are (L_1, L_2) and (U_1, U_2), respectively. In the training set, the region $[\Gamma_{l_2}, \Gamma_{u_1}]$ is overlapping between C_1 and C_2, and the regions $(\Gamma_{l_1}, \Gamma_{l_2})$ and $(\Gamma_{u_1}, \Gamma_{u_2})$ are nonoverlapping for C_1 and C_2, respectively.

TABLE II(a)
RANGES OF TRAINING SETS AND FEATURE SUBDOMAINS
FOR THE PATTERN CLASSES [2, 7] AND [5, 10].

	Sample Sizes in Each Class				
	50	100	150	200	250
Class C_1	2.11–6.88	2.08–6.93	2.06–5.93	2.04–5.96	2.02–5.99
Class C_2	5.13–9.89	5.12–9.92	5.07–9.94	5.03–9.95	5.01–9.99
Subdomain D_1	1.34–5.64	1.43–5.58	1.56–5.44	1.69–5.31	1.77–5.21
Subdomain D_2	4.57–7.35	4.68–7.32	4.76–7.24	4.83–7.16	4.86–7.13
Subdomain D_3	6.26–10.63	6.40–10.54	6.55–10.40	6.68–10.29	6.79–10.22

By Theorem 1, it can be stated that as the training sample size increases

$$l_1 = \beta_{l_1} \to L_1 \qquad \text{in probability}$$
$$l_2 = \beta_{u_1} = \beta_{l_2} \to L_2 \quad \text{in probability}$$
$$u_1 = \beta_{l_3} = \beta_{u_2} \to U_1 \quad \text{in probability}$$

and

$$u_2 = \beta_{u_3} \to U_1 \qquad \text{in probability.}$$

With the increase in the size of training samples, the number of samples in the feature subdomains also increases. Then, by Theorem 2,

$$\Gamma_{l_g} \to \beta_{l_g} \quad \text{and} \quad \Gamma_{u_g} \to B_{u_g} \quad \text{in probability}$$
$$\text{for } g = 1, 2, 3$$
$$\Rightarrow \quad \Gamma_{l_2} \to \beta_{l_2} - \beta_{u_1} = l_2 \to L_2 \qquad \text{in probability}$$

and

$$\Gamma_{u_2} \to \beta_{u_2} = \beta_{l_3} = u_1 \to U_1 \qquad \text{in probability.}$$

Therefore, with the increase of the training sample size, the estimated overlapping region $[\Gamma_{l_2}, \Gamma_{u_1}]$ tends toward the actual overlapping region $[L_2, U_2]$ in probability. At the same time, the estimated nonoverlapping (single choice) regions $(\Gamma_{l_1}, \Gamma_{l_2})$ and $(\Gamma_{u_1}, \Gamma_{u_2})$ for the classes C_1 and C_2 tend toward their actual sizes (L_1, L_2) and (U_1, U_2), respectively, in probability. The no-class region also tends to its actual size.

Hence the proof.   ■

V. EXPERIMENTAL RESULTS:

To substantiate the analytical results, a two class problem with classes [2, 7] and [5, 10] is considered. Here, the actual nonoverlapping regions for the classes C_1 and C_2 are [2, 5] and [7, 10], respectively, and the actual overlapping region between the classes is [5, 7]. To implement the recognition system, five training sample sets with 50, 100, 150, 200, and 250 samples from each of the pattern classes are chosen randomly.

The ranges of the training samples of the classes C_1 and C_2, and the ranges of the obtained three feature subdomains D_1, D_2, and D_3 for the considered five sample sets are shown in Table II(a). The regions obtained corresponding to various output choices are shown in Tables II(b) and (c) for the first two (with 50 and 100 samples from each class) and the remaining three (with 150, 200, and 250 samples from each class) sample sets, respectively.

TABLE II(b)
VARIOUS REGIONS AND BAYES THRESHOLD POINTS FOR THE FIRST
TWO TRAINING SETS OF THE PATTERN CLASSES [2, 7] AND [5, 10].

Various Group of Choices	Sample Sizes in Each Class	
	50	100
Null choice	$-\infty$–1.3402	$-\infty$–1.4274
Single choice (C_1)	1.3403–4.5741	1.4275–4.6752
First-second choice (C_1)	4.5742–5.3090	4.6753–5.3175
Combined choice	5.3091–6.4549	5.3176–6.6060
First-second choice (C_2)	6.4550–7.3457	6.6061–7.3173
Single choice (C_2)	7.3458–10.6274	7.3174–10.5438
Null choice	10.6275–$+\infty$	10.5439–$+\infty$
Bayes threshold point	5.7445097	5.8655739

TABLE II(c)
VARIOUS REGIONS AND BAYES THRESHOLD POINTS FOR THE REMAINING
THREE TRAINING SETS OF THE PATTERN CLASSES [2, 7] AND [5, 10].

Various Group of Choices	Sample Size in Each Class		
	150	200	250
Null choice	$-\infty$–1.5625	$-\infty$–1.6921	$-\infty$–1.7684
Single choice (C_1)	1.5626–4.7565	1.6922–4.8287	1.7685–4.8549
First-second choice (C_1)	4.7566–5.2758	4.8288–5.1997	4.8550–5.1318
Combined choice	5.2759–6.7329	5.1998–6.8013	5.1319–6.8505
First-second choice (C_2)	6.7330–7.2361	6.8014–7.1621	6.8506–7.1279
Single choice (C_2)	7.2362–10.3957	7.1622–10.2906	7.1280–10.2228
Null choice	210.3958–$+\infty$	10.2907–$+\infty$	110.2229–$+\infty$
Bayes threshold point	5.9670115	5.9635067	5.9716177

TABLE II(d)
RECOGNITION SCORE FOR THE PATTERN CLASSES [2, 7] AND [5, 10].

Various Group of Choices	% Recognition Score				
	Sample Sizes in Each Class				
	50	100	150	200	250
Single Correct Choice	55.40	58.30	60.15	61.60	63.05
First Correct Choice	16.85	15.50	13.15	9.60	7.25
Combined Correct Choice	18.60	18.55	21.10	24.10	26.50
Second Correct Choice	9.15	7.65	5.60	4.70	3.20
Fully Wrong Choice	0.00	0.00	0.00	0.00	0.00
Bayes Classifier — *Correct*	79.00	79.35	79.30	79.30	79.25
Bayes Classifier — *Wrong*	21.00	20.65	20.70	20.70	20.75

These results substantiate that with the increase in sample size, the training classes (i.e., the portions covered by the training samples) tend to their actual classes (Theorem 1) and the feature subdomains with the extended portions tend to their actual sizes (Theorem 2). It is to be noticed that the overlapping and the nonoverlapping regions are tending to their actual sizes with the increase in sample sizes. Hence Proposition 1 is verified experimentally for the case of overlapping pattern classes in one-dimensional feature space.

For analyzing the recognition performance of the system, 1000 test samples from each of the pattern classes are generated. The recognition scores corresponding to the considered five training sets are shown in Table II(d). Note that the recognition scores are grouped into five categories, namely, single correct choice, first-correct choice, combined correct choice, second correct choice, and fully wrong choice.

For comparison, the Bayes classifier is implemented on the same pattern classes (assuming normal distributions). The threshold points found between the classes C_1 and C_2 are

shown in Tables II(b) and (c) corresponding to the considered five training sets. The recognition scores of the Bayes classifier are included in Table II(d).

Note from Tables II(b) and (c) that the Bayes threshold points lie in the combined choice region of our recognition system. By adding up the scores corresponding to single correct, first-second correct, and half of the combined correct choices in Table II(d), the recognition score becomes higher than the corresponding correct choices of Bayes classifier. Again, the wrong choices of Bayes classifier are found to be distributed in the combined correct and second correct choices of the proposed system. Therefore, the proposed recognition system has a provision of improving its efficiency significantly by incorporating combined and second choices under the control of a supervisory scheme.

VI. ANALYSIS IN 2-D FEATURE SPACE

Let us consider a two-class (C_1 and C_2) problem to analyze the performance of the multivalued recognition system in a two-dimensional feature space ($F_1 \times F_2$). For the sake of convenience, the classes are initially assumed to be of rectangular shape. Then the results are extended to circular pattern classes. These results can easily be generalized to the pattern classes of any shape.

A. Rectangular Classes

Suppose $[L_{11}, U_{11}] \times [L_{21}, U_{21}]$ and $[L_{12}, U_{12}] \times [L_{22}, U_{22}]$ denote the classes C_1 and C_2, respectively. Here $[L_{12}, U_{11}] \times [L_{22}, U_{21}]$ is the overlapping portion between the classes. Suppose also that $[l_{11}, u_{11}] \times [l_{21}, u_{21}]$ and $[l_{21}, u_{12}] \times [l_{22}, u_{22}]$ denote the training sets corresponding to C_1 and C_2, respectively, where $[l_{12}, u_{11}] \times [l_{22}, u_{21}]$ is the overlapping portion in the training set. Such pattern classes with the span of their training sets are shown in Fig. 5(a).

Initially, based on the training set, each individual feature domain is partitioned into three feature subdomains [Fig. 5(b)]. Recall that the gth ($g = 1, 2, 3$) feature subdomain in the ith ($i = 1, 2$) feature axis is denoted by D_{ig}. As a result, the total feature space is decomposed into nine space subdomains, which are dentoed by $SD_1, SD_2, \cdots, SD_9$, respectively [Fig. 5(b)]. Here SD_1, SD_2, and SD_4 uniquely correspond to the class $C_1; SD_6, SD_8$, and SD_9 uniquely correspond to $C_2; SD_5$ is overlapping by reflecting both C_1 and C_2, and SD_3 and SD_5 are the no-class regions, i.e., they reflect neither C_1 nor C_2.

Thus, the feature subdomains along F_1 and F_2 are extended to some extent to highlight the portions possible uncovered by the training samples. These feature subdomains are characterized by different piecewise linear triangular functions of the form $T_{ig}(x_i; \alpha_{ig}, \beta_{l_{ig}}, \beta_{u_{ig}}, \Gamma_{l_{ig}}, \Gamma_{u_{ig}})$ ($i = 1, 2; g = 1, 2, 3$) (3). The feature subdomains and the space subdomains are shown in Fig. 5(b). Note that

$$\beta_{l_{i1}} = l_{i1}; \qquad \beta_{u_{i3}} = u_{i2};$$
$$\beta_{l_{i2}} = \beta_{u_{i1}} = l_{i2}; \qquad \beta_{l_{i3}} = \beta_{u_{i2}} = u_{i1};$$
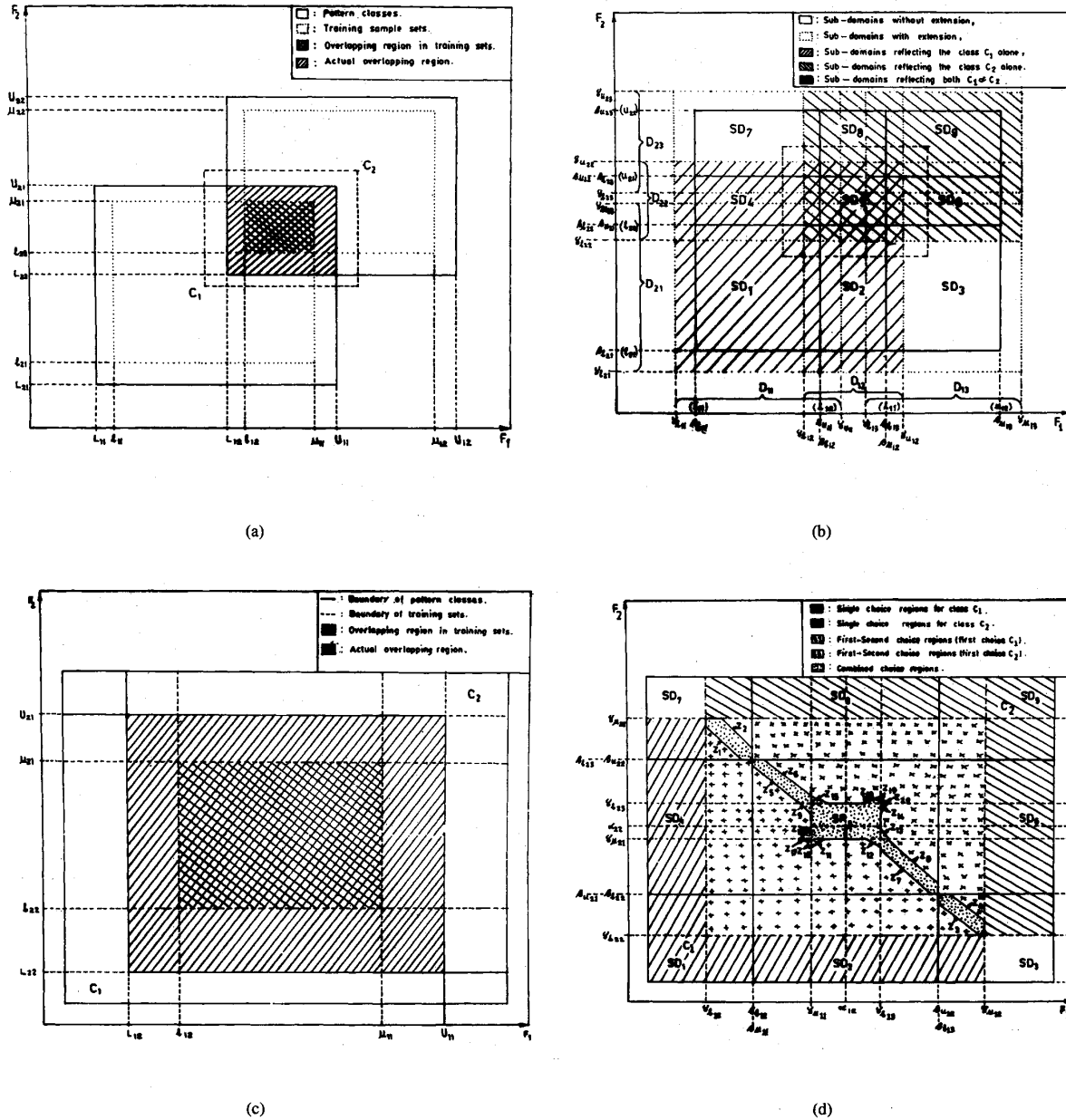
(a)

(b)

(c)

(d)

Fig. 5. (a) Two rectangular classes with a training set. (b) Subdomains and choice regions for the pattern classes in (a). (c) Enlarged version of the rectangular portion (enclosed by dotted lines) in (a). (d) Enlarged version of the rectangular portion (enclosed by dotted lines) in (b).

and

$$\varepsilon_{ig} = (\beta_{u_{ig}} - \beta_{l_{ig}})\delta_{ig}$$

where ε_{ig} (1) is the extended portion for the gth ($g = 1, 2, 3$) feature subdomain in the ith ($i = 1, 2$) feature axis. It is obvious from Fig. 5(b) that $[\Gamma_{l_{11}}, \Gamma_{l_{12}}] \times [\Gamma_{l_{21}}, \Gamma_{u_{22}}]$ and $[\Gamma_{l_{12}}, \Gamma_{u_{12}}] \times [\Gamma_{l_{21}}, \Gamma_{l_{22}}]$ are the single choice regions for C_1 and $[\Gamma_{l_{12}}, \Gamma_{l_{13}}] \times [\Gamma_{u_{22}}, \Gamma_{u_{23}}]$ and $[\Gamma_{u_{12}}, \Gamma_{u_{13}}] \times [\Gamma_{l_{22}}, \Gamma_{u_{22}}]$ are ths single choice regions for C_2.

If $X \in [\Gamma_{l_{12}}, \beta_{l_{12}}] \times [\Gamma_{l_{22}}, \beta_{u_{22}}]$ or $X \in [\Gamma_{l_{12}}, \beta_{u_{12}}] \times [\Gamma_{l_{22}}, \beta_{l_{22}}]$, the elements in $S(X)$ are positive for both the classes, i.e., $s_1(X) > 0$ and $s_2(X) > 0$. It can be easily shown that $[s_1(X) - s_2(X)] > 0.05$ for X's lying in these regions. This implies that $[\Gamma_{l_{12}}, \beta_{l_{12}}] \times [\Gamma_{l_{22}}, \beta_{u_{22}}]$ and $[\Gamma_{l_{12}}, \beta_{u_{12}}] \times [\Gamma_{l_{22}}, \beta_{l_{22}}]$ are the first-second choice regions with first choice as C_1. Similarly, $[\Gamma_{l_{12}}, \beta_{l_{12}}] \times [\Gamma_{l_{22}}, \beta_{u_{22}}]$ and $[\Gamma_{l_{12}}, \beta_{u_{12}}] \times [\Gamma_{l_{22}}, \beta_{l_{22}}]$ are the first-second choice regions with first choice as C_2.

Now consider the feature region $[\Gamma_{l_{12}}, \beta_{l_{12}}] \times [\beta_{u_{22}}, \Gamma_{u_{22}}]$ which is overlapping. That is, $s_1(X) > 0$ and $s_2(X) > 0$. In this region, two lines Z_1 and Z_2 may be found so that

for X lying below the line Z_1,
$$[s_1(X) - s_2(X)] > 0.05,$$
for X lying above the line Z_2,
$$[s_2(X) - s_1(X)] > 0.05,$$

and

for X lying between Z_1 and Z_2,
$$|s_1(X) - s_2(X)| \leq 0.05.$$

The equations of Z_1 and Z_2 are

$$A_1 x_1 + B_1 x_2 - \eta_1 = 0$$

and

$$A_2 x_1 + B_2 x_2 - \eta_2 = 0 \tag{18}$$

where

$$A_1 = A_2 = k_2 k_3 k_4 - k_1 k_3 k_4;$$
$$B_1 = B_2 = k_1 k_2 k_4 - k_1 k_2 k_3$$
$$\eta_1 = \beta_{u_{11}} k_2 k_3 k_4 - \Gamma_{l_{12}} k_1 k_3 k_4$$
$$\qquad + \Gamma_{u_{22}} k_1 k_2 k_4$$
$$\qquad - \beta_{l_{23}} k_1 k_2 k_3 + 0.4 k_1 k_2 k_3 k_4;$$
$$\eta_2 = \beta_{u_{11}} k_2 k_3 k_4 - \Gamma_{l_{12}} k_1 k_3 k_4$$
$$\qquad + \Gamma_{u_{22}} k_1 k_2 k_4$$
$$\qquad - \beta_{l_{23}} k_1 k_2 k_3 + 0.4 k_1 k_2 k_3 k_4$$

with

$$k_1 = \alpha_{11} - \beta_{u_{11}}; \quad k_2 = \beta_{l_{12}} - \Gamma_{l_{12}};$$
$$k_3 = \beta_{u_{22}} - \Gamma_{u_{22}} \text{ and } k_4 = \alpha_{23} - \beta_{l_{23}}.$$

Similarly, $[\beta_{u_{12}}, \Gamma_{u_{12}}] \times [\Gamma_{l_{22}}, \beta_{l_{22}}]$ is a overlapping region. Here also, two lines Z_3 and Z_4 may be found such that

for X lying below the line Z_3,
$$[s_1(X) - s_2(X)] > 0.05,$$
for X lying above the line Z_4,
$$[s_2(X) - s_1(X)] > 0.05,$$

and

for X lying between Z_3 and Z_4,
$$|s_1(X) - s_2(X)| \leq 0.05.$$

The equations of Z_3 and Z_4 are

$$A_3 x_1 + B_3 x_2 - \eta_3 = 0$$

and

$$A_4 x_1 + B_4 x_2 - \eta_4 = 0 \tag{19}$$

where

$$A_3 = A_4 = k_2 k_3 k_4 - k_1 k_3 k_4;$$
$$B_3 = B_4 = k_1 k_2 k_4 - k_1 k_2 k_3;$$
$$\eta_3 = \beta_{u_{12}} k_2 k_3 k_4 - \Gamma_{l_{13}} k_1 k_3 k_4$$
$$\qquad + \beta_{u_{21}} k_1 k_2 k_4$$
$$\qquad - \Gamma_{l_{22}} k_1 k_2 k_3 + 0.4 k_1 k_2 k_3 k_4;$$
$$\eta_4 = \beta_{u_{12}} k_2 k_3 k_4 - \Gamma_{l_{13}} k_1 k_3 k_4$$
$$\qquad + \beta_{u_{21}} k_1 k_2 k_4$$
$$\qquad - \Gamma_{l_{22}} k_1 k_2 k_3 + 0.4 k_1 k_2 k_3 k_4$$

with

$$k_1 = \beta_{u_{12}} - \Gamma_{u_{12}}; \quad k_2 = \alpha_{13} - \beta_{l_{13}};$$
$$k_3 = \alpha_{21} - \beta_{u_{21}} \text{ and } k_4 = \beta_{l_{22}} - \Gamma_{l_{22}}.$$

Let us now consider the region $[\beta_{u_{11}}, \Gamma_{u_{11}}] \times [\Gamma_{l_{23}}, \beta_{l_{23}}]$ which is overlapping. In this region, two lines Z_5 and Z_6 may be found so that

for X lying below the line Z_5,
$$[s_1(X) - s_2(X)] > 0.05,$$
for X lying above the line Z_6,
$$[s_2(X) - s_1(X)] > 0.05,$$

and

for X lying between Z_5 and Z_6,
$$|s_1(X) - s_2(X)| \leq 0.05.$$

The equations of Z_5 and Z_6 are

$$A_5 x_1 + B_5 x_2 - \eta_5 = 0$$

and

$$A_6 x_1 + B_6 x_2 - \eta_6 = 0 \tag{20}$$

where

$$A_5 = A_6 = k_2 k_3 k_4 - k_1 k_3 k_4;$$
$$B_5 = B_6 = k_1 k_2 k_4 - k_1 k_2 k_3;$$
$$\eta_5 = \Gamma_{u_{11}} k_2 k_3 k_4 - \beta_{l_{12}} k_1 k_3 k_4$$
$$\qquad + \beta_{u_{22}} k_1 k_2 k_4$$
$$\qquad - \Gamma_{l_{23}} k_1 k_2 k_3 + [1.6 - 16(r_{51} - r_{52})] k_1 k_2 k_3 k_4;$$
$$\eta_6 = \Gamma_{u_{11}} k_2 k_3 k_4 - \beta_{l_{12}} k_1 k_3 k_4$$
$$\qquad + \beta_{u_{22}} k_1 k_2 k_4$$
$$\qquad - \Gamma_{l_{23}} k_1 k_2 k_3 - [1.6 + 16(r_{51} - r_{52})] k_1 k_2 k_3 k_4$$

with

$$k_1 = \beta_{u_{11}} - \Gamma_{u_{11}}; \quad k_2 = \alpha_{12} - \beta_{l_{12}};$$
$$k_3 = \alpha_{22} - \beta_{l_{22}} \text{ and } k_4 = \beta_{l_{23}} - \Gamma_{l_{23}}.$$

Similarly, in the overlapping region $[\Gamma_{l_{13}}, \beta_{l_{13}}] \times [\beta_{u_{21}}, \Gamma_{u_{21}}]$, two lines Z_7 and Z_8 may be found so that

for X lying below the line Z_7,
$$[s_1(X) - s_2(X)] > 0.05,$$
for X lying above the line Z_8,
$$[s_2(X) - s_1(X)] > 0.05,$$

and

for X lying between Z_7 and Z_8,
$$|s_1(X) - s_2(X)| \leq 0.05.$$

The equations of Z_7 and Z_8 are

$$A_7 x_1 + B_7 x_2 - \eta_7 = 0$$

and

$$A_8 x_1 + B_8 x_2 - \eta_8 = 0 \qquad (21)$$

where

$$A_7 = A_8 = k_2 k_3 k_4 - k_1 k_3 k_4;$$
$$B_7 = B_8 = k_1 k_2 k_4 - k_1 k_2 k_3;$$
$$\eta_7 = \beta_{u_{12}} k_2 k_3 k_4 - \Gamma_{l_{13}} k_1 k_3 k_4$$
$$\qquad + \Gamma_{u_{21}} k_1 k_2 k_4$$
$$\qquad - \beta_{l_{22}} k_1 k_2 k_3 + [1.6 - 16(r_{51} - r_{52})] k_1 k_2 k_3 k_4;$$
$$\eta_8 = \beta_{u_{12}} k_2 k_3 k_4 - \Gamma_{l_{13}} k_1 k_3 k_4$$
$$\qquad + \Gamma_{u_{21}} k_1 k_2 k_4$$
$$\qquad - \beta_{l_{22}} k_1 k_2 k_3 - [1.6 + 16(r_{51} - r_{52})] k_1 k_2 k_3 k_4$$

with

$$k_1 = \alpha_{12} - \beta_{u_{12}}; \quad k_2 = \beta_{l_{13}} - \Gamma_{l_{13}};$$
$$k_3 = \beta_{u_{21}} - \Gamma_{u_{21}} \text{ and } k_4 = \alpha_{22} - \beta_{l_{22}}.$$

Now consider the region $[\beta_{u_{11}}, \Gamma_{u_{11}}] \times [\alpha_{22}, \Gamma_{l_{23}}]$. Here, a line Z_9 may be found so that

for X lying below the line Z_9;
$$[s_1(X) - s_2(X)] > 0.05$$

and

for X lying above the line Z_9;
$$|s_1(X) - s_2(X)| > 0.05.$$

The equation of Z_9 is

$$A_9 x_1 + B_9 x_2 - \eta_9 = 0 \qquad (22)$$

where

$$A_9 = k_2; \qquad B_9 = k_1;$$
$$\eta_9 = \Gamma_{u_{11}} k_2 + \beta_{u_{22}} k_1$$
$$\qquad + [0.6 - 16(r_{51} - r_{52})] k_1 k_2$$

with

$$k_1 = \beta_{u_{11}} - \Gamma_{u_{11}} \text{ and }$$
$$k_2 = \alpha_{22} - \beta_{u_{22}}.$$

Similarly, in the region $[\beta_{u_{11}}, \Gamma_{u_{11}}] \times [\Gamma_{u_{21}}, \alpha_{22}]$, a line Z_{10} may be found such that

for X lying above the line Z_{10};
$$[s_1(X) - s_2(X)] > 0.05$$

and

for X lying below the line Z_{10};
$$|s_1(X) - s_2(X)| > 0.05.$$

The equation of Z_{10} is

$$A_{10} x_1 + B_{10} x_2 - \eta_{10} = 0 \qquad (23)$$

where

$$A_{10} = k_2; \qquad B_{10} = k_1;$$
$$\eta_{10} = \Gamma_{u_{11}} k_2 + \beta_{l_{22}} k_1$$
$$\qquad + [0.6 - 16(r_{51} - r_{52})] k_1 k_2$$

with

$$k_1 = \beta_{u_{11}} - \Gamma_{u_{11}} \text{ and }$$
$$k_2 = \alpha_{22} - \beta_{u_{22}}.$$

In the regions $[\beta_{u_{11}}, \alpha_{12}] \times [\beta_{u_{21}}, \Gamma_{u_{21}}]$ and $[\alpha_{12}, \Gamma_{l_{13}}] \times [\beta_{u_{21}}, \Gamma_{u_{21}}]$, two lines Z_{11} and Z_{12} are found such that

for X lying below both Z_{11} and Z_{12};
$$[s_1(X) - s_2(X)] > 0.05$$

and

for X lying above Z_{11} or Z_{12};
$$|s_1(X) - s_2(X)| > 0.05.$$

The equations of Z_{11} and Z_{12} are

$$A_{11} x_1 + B_{11} x_2 - \eta_{11} = 0$$

and

$$A_{12} x_1 + B_{12} x_2 - \eta_{12} = 0 \qquad (24)$$

where

$$A_{11} = A_{12} = k_2;$$
$$B_{11} = k_1 \text{ and } B_{12} = \hat{k}_1;$$
$$\eta_{11} = \beta_{l_{12}} k_2 + \Gamma_{u_{21}} k_1 + [0.6 - 16(r_{51} - r_{52})] k_1 k_2;$$
$$\eta_{12} = \beta_{u_{12}} k_2 + \Gamma_{u_{21}} \hat{k}_1 + [0.6 - 16(r_{51} - r_{52})] \hat{k}_1 k_2$$

with

$$k_1 = \alpha_{12} - \beta_{l_{12}}; k_2 = \beta_{u_{21}} - \Gamma_{u_{21}}$$

and

$$\hat{k}_1 = \alpha_{12} - \beta_{u_{12}}.$$

In the regions $[\Gamma_{l_{13}}, \beta_{l_{13}}] \times [\Gamma_{u_{21}}, \alpha_{22}]$ and $[\Gamma_{l_{13}}, \beta_{l_{13}}] \times [\alpha_{22}, \Gamma_{l_{23}}]$, two lines Z_{13} and Z_{14} may be found in their respective regions such that

for X lying right to both Z_{13} and Z_{14};
$$[s_2(X) - s_1(X)] > 0.05$$

and

for X lying left to Z_{13} or Z_{14};
$$|s_1(X) - s_2(X)| > 0.05.$$

The equations of Z_{13} and Z_{14} are

$$A_{13} x_1 + B_{13} x_2 - \eta_{13} = 0$$

and

$$A_{14} x_1 + B_{14} x_2 - \eta_{14} = 0 \qquad (25)$$

where

$$A_{13} = k_2; A_{14} = \hat{k}_2;$$
$$B_{13} = B_{14} = k_1;$$
$$\eta_{13} = \Gamma_{l_{13}} k_2 + \beta_{l_{22}} k_1 + [0.6 - 16(r_{52} - r_{51})] k_1 k_2;$$
$$\eta_{14} = \Gamma_{l_{13}} \hat{k}_2 + \beta_{u_{22}} k_1 + [0.6 - 16(r_{52} - r_{51})] k_1 \hat{k}_2$$

with

$$k_1 = \beta_{l_{13}} - \Gamma_{l_{13}}; k_2 = \alpha_{22} - \beta_{l_{22}}$$

and

$$\hat{k}_2 = \alpha_{22} - \beta_{u_{22}}.$$

Similarly, in the regions $[\Gamma_{u_{11}}, \alpha_{12}] \times [\Gamma_{l_{23}}, \beta_{l_{23}}]$ and $[\alpha_{12}, \Gamma_{l_{13}}] \times [\Gamma_{l_{23}}, \beta_{l_{23}}]$, two lines Z_{15} and Z_{16} may be found in their respective regions such that

for X lying above both Z_{15} and Z_{16};
$$[s_2(X) - s_1(X)] > 0.05$$

and

for X lying below Z_{15} or Z_{16};
$$|s_1(X) - s_2(X)| > 0.05.$$

The equations of Z_{15} and Z_{16} are

$$A_{15} x_1 + B_{15} x_2 - \eta_{15} = 0$$

and

$$A_{16} x_1 + B_{16} x_2 - \eta_{16} = 0 \qquad (26)$$

where

$$A_{15} = A_{16} = k_2;$$
$$B_{15} = k_1; B_{16} = \hat{k}_1;$$
$$\eta_{15} = \beta_{l_{12}} k_2 + \Gamma_{l_{23}} k_1 + [0.6 - 16(r_{52} - r_{51})] k_1 k_2;$$
$$\eta_{16} = \beta_{u_{12}} k_2 + \Gamma_{l_{23}} \hat{k}_1 + [0.6 - 16(r_{52} - r_{51})] \hat{k}_1 k_2$$

with

$$k_1 = \alpha_{12} - \beta_{l_{12}}; k_2 = \beta_{l_{23}} - \Gamma_{l_{23}}$$

and

$$\hat{k}_1 = \alpha_{12} - \beta_{u_{12}}.$$

Let us now consider the region $[\beta_{u_{11}}, \Gamma_{u_{11}}] \times [\beta_{u_{21}}, \Gamma_{u_{21}}]$. Here two lines Z_{17} and Z_{18} may be found such that

for X lying below both Z_{17} and Z_{18};
$$[s_1(X) - s_2(X)] > 0.05$$

and

for X lying above Z_{17} or Z_{18};
$$|s_1(X) - s_2(X)| > 0.05.$$

The equations of Z_{17} and Z_{18} are

$$A_{17} x_1 + B_{17} x_2 - \eta_{17} = 0$$

and

$$A_{18} x_1 + B_{18} x_2 - \eta_{18} = 0 \qquad (27)$$

where

$$A_{17} = k_2; A_{18} = \hat{k}_2;$$
$$B_{17} = k_1; B_{18} = \hat{k}_1;$$
$$\eta_{17} = \Gamma_{u_{11}} k_2 + \beta_{l_{22}} k_1 + [0.6 - 16(r_{51} - r_{52})] k_1 k_2;$$
$$\eta_{18} = \beta_{l_{12}} \hat{k}_2 + \Gamma_{u_{21}} \hat{k}_1 + [0.6 - 16(r_{51} - r_{52})] \hat{k}_1 \hat{k}_2$$

with

$$k_1 = \beta_{u_{11}} - \Gamma_{l_{11}}; k_2 = \alpha_{22} - \beta_{l_{22}};$$
$$\hat{k}_1 = \alpha_{12} - \beta_{u_{12}} \text{ and}$$
$$\hat{k}_2 = \beta_{u_{21}} - \Gamma_{u_{21}}.$$

Again, in the region $[\Gamma_{l_{13}}, \beta_{l_{13}}] \times [\Gamma_{l_{23}}, \beta_{l_{23}}]$, two lines Z_{19} and Z_{20} may be found such that

for X lying above both Z_{19} and Z_{20};
$$[s_2(X) - s_1(X)] > 0.05$$

and

for X lying below Z_{19} or Z_{20};
$$|s_1(X) - s_2(X)| \le 0.05.$$

The equations of Z_{19} and Z_{20} are

$$A_{19} x_1 + B_{19} x_2 - \eta_{19} = 0$$

and

$$A_{20} x_1 + B_{20} x_2 - \eta_{20} = 0 \qquad (28)$$

where

$$A_{19} = k_2; A_{20} = \hat{k}_2;$$
$$B_{19} = k_1; B_{20} = \hat{k}_1;$$
$$\eta_{19} = \Gamma_{l_{13}} k_2 + \beta_{u_{22}} k_1 + [0.6 - 16(r_{52} - r_{51})] k_1 k_2;$$
$$\eta_{20} = \beta_{u_{12}} \hat{k}_2 + \Gamma_{u_{23}} \hat{k}_1 + [0.6 - 16(r_{52} - r_{51})] \hat{k}_1 \hat{k}_2$$

with

$$k_1 = \beta_{l_{13}} - \Gamma_{l_{13}}; k_2 = \alpha_{22} - \beta_{u_{22}}$$

and

$$\hat{k}_1 = \alpha_{12} - \beta_{u_{12}} \text{ and}$$
$$\hat{k}_2 = \beta_{l_{23}} - \Gamma_{l_{23}}.$$

At last, let us consider the region $[\Gamma_{u_{11}}, \Gamma_{l_{13}}] \times [\Gamma_{u_{21}}, \beta_{l_{23}}]$. Here, $s_1(x) > 0$ and $s_2(X) > 0$ and there does not exist any effect from the neighboring feature subdomains. In this region, if $r_{51} > r_{52} + 0.1$ then $[s_1(X) - s_2(X)] > 0.05$, i.e., the region will be a first-second choice region with first choice as C_1. If $r_{52} > r_{51} + 0.1$ then $[s_2(X) - s_1(X)] > 0.05$, i.e., the region willl be a first-second choice region with first choice as C_2. Otherwise (i.e., if $|r_{51} - r_{52}| \le 0.1$ then $|s_1(X) - s_2(X)| \le 0.05$), the region will become a combined choice region.

The conclusions, after combining all the previous cases, are given below.

The single choice regions for C_1 are $[\Gamma_{l_{11}}, \Gamma_{l_{12}}] \times [\Gamma_{l_{21}}, \Gamma_{u_{22}}]$ and $[\Gamma_{l_{12}}, \Gamma_{u_{12}}] \times [\Gamma_{l_{21}}, \Gamma_{l_{22}}]$. The single choice regions for C_2 are $[\Gamma_{l_{12}}, \Gamma_{l_{13}}] \times [\Gamma_{u_{22}}, \Gamma_{u_{23}}]$ and $[\Gamma_{u_{12}}, \Gamma_{u_{13}}] \times [\Gamma_{l_{22}}, \Gamma_{u_{22}}]$. The overlapping (corresponding to combined and first-second choices) regions are $[\Gamma_{l_{12}}, \beta_{l_{12}}] \times [\Gamma_{l_{22}}, \beta_{u_{22}}]$ and $[\Gamma_{l_{12}}, \beta_{u_{12}}] \times [\Gamma_{l_{22}}, \beta_{l_{22}}]$. The remaining portion in the feature space is termed as the no-class (null choice) region.

In the overlapping region, $[\Gamma_{l_{12}}, \beta_{l_{12}}] \times [\Gamma_{l_{22}}, \beta_{u_{22}}]$ and $[\Gamma_{l_{12}}, \beta_{u_{12}}] \times [\Gamma_{l_{22}}, \beta_{l_{22}}]$ correspond to the first-second choice regions with first choice as C_1. Again, $[\Gamma_{l_{12}}, \beta_{l_{12}}] \times [\Gamma_{l_{22}}, \beta_{u_{22}}]$ and $[\Gamma_{l_{12}}, \beta_{u_{12}}] \times [\Gamma_{l_{22}}, \beta_{l_{22}}]$ correspond to the first-second choice regions with first choice as C_2. The region surrounded by the lines $Z_1, Z_2, \cdots$ and Z_{20} [Fig. 5(d)] is the combined choice region. The portions in the overlapping region, not falling in the combined choice region, are the first-second choice regions.

The pattern classes C_1 and C_2 are shown in Fig. 5(a). A typical training set is also shown in this figure. The overlapping portions in the pattern classes and also in the training set are marked here. Fig. 5(b) shows the typical feature subdomains and space subdomains corresponding to the training set [Fig. 5(a)]. The regions corresponding to single, first-second, combined, and null choices are also shown in Fig. 5(b). Note that the regions are drawn based on the previous analytical findings.

To show the overlapping regions more prominently, the rectangular portion [shown by dotted lines in Fig. 5(a)] which includes the overlapping regions, are enlarged in Fig. 5(c). The corresponding rectangular portion in Fig. 5(b) is also enlarged in Fig. 5(d). Figs. 5(b) and (d) show the various choice regions corresponding to the pattern classes in Figs. 5(a) and (c), respectively. The lines $Z_1, Z_2, \cdots, Z_{20}$ corresponding to (18)–(28) are shown in Fig. 5(d).   ∎

Proof of Proposition 1: In the present case, the actual overlapping region between the classes C_1 and C_2 is $[L_{12}, U_{11}] \times [L_{22}, U_{21}]$. The nonoverlapping regions for C_1 are $[L_{11}, U_{11}] \times [L_{21}, L_{22}]$ and $[L_{11}, L_{12}] \times [L_{22}, U_{21}]$. The nonoverlapping regions for C_2 are $[L_{12}, U_{12}] \times [U_{21}, U_{22}]$ and $[U_{11}, U_{12}] \times [L_{22}, U_{21}]$. In the training samples, the feature region $[l_{12}, u_{11}] \times [l_{22}, u_{21}]$ is overlapping. Based on the training set, $[\Gamma_{l_{12}}, \Gamma_{u_{11}}] \times [\Gamma_{l_{22}}, \Gamma_{u_{21}}]$ is the estimated overlapping region. The estimated nonoverlapping regions for C_1 are $[\Gamma_{l_{11}}, \Gamma_{u_{12}}] \times [\Gamma_{l_{21}}, \Gamma_{u_{22}}]$ and $[\Gamma_{l_{11}}, \Gamma_{u_{12}}] \times [\Gamma_{l_{22}}, \Gamma_{u_{22}}]$. On the other hand, the estimated nonoverlapping regions for C_2 are $[\Gamma_{l_{12}}, \Gamma_{u_{13}}] \times [\Gamma_{u_{22}}, \Gamma_{u_{23}}]$ and $[\Gamma_{u_{12}}, \Gamma_{u_{13}}] \times [\Gamma_{l_{22}}, \Gamma_{u_{22}}]$.

By Theorem 1, it can be stated that as the size of the training samples increases

$$l_{ij} \to L_{ij} \quad \text{in probability}$$

and

$$u_{ij} \to U_{ij} \quad \text{in probability} \quad (i = 1, 2; j = 1, 2). \quad (29)$$

With the increase of the size of the training samples, the samples in the feature subdomain also increase. Thus, by Theorem 2,

$$\Gamma_{l_{ig}} \to \beta_{l_{ig}} \quad \text{in probability}$$

and

$$\Gamma_{u_{ig}} \to \beta_{u_{ig}} \quad \text{in probability}$$
$$(i = 1, 2; g = 1, 2, 3). \quad (30)$$

From Fig. 5(c), we have

$$\beta_{l_{i1}} = l_{i1}; \qquad \beta_{u_{i1}} = \beta_{l_{i2}} = l_{i2};$$
$$\beta_{l_{i3}} = \beta_{u_{i2}} = u_{i1} \quad \text{and} \quad \beta_{u_{i3}} = u_{i2} \quad (i = 1, 2). \quad (31)$$

Combing (29), (30), and (31), one gets

$$\Gamma_{l_{i1}} \to \beta_{l_{i1}} = l_{i1} \to L_{i1} \quad \text{in probability}$$
$$\Gamma_{u_{i1}} \to \beta_{u_{i1}} = l_{i2} \to L_{i2} \quad \text{in probability}$$
$$\Gamma_{l_{i2}} \to \beta_{l_{i2}} = l_{i2} \to L_{i2} \quad \text{in probability}$$
$$\Gamma_{u_{i2}} \to \beta_{u_{i2}} = u_{i1} \to U_{i1} \quad \text{in probability}$$
$$\Gamma_{l_{i3}} \to \beta_{l_{i3}} = u_{i1} \to U_{i1} \quad \text{in probability}$$
$$\Gamma_{u_{i3}} \to \beta_{u_{i3}} = u_{i2} \to U_{i2} \quad \text{in probability}$$
$$(i = 1, 2).$$

Therefore, as the size of the training sample increases, the estimated overlapping region $[\Gamma_{l_{12}}, \Gamma_{u_{11}}] \times [\Gamma_{l_{22}}, \Gamma_{u_{21}}]$ tends to the actual overlapping regin $[L_{12}, U_{11}] \times [L_{22}, U_{21}]$ in probability. At the same time, the estimated nonoverlapping regions for C_1 and C_2 go to their actual sizes, respectively, in probability. Using the previous results, it can also be concluded that the no-class (null choice) region also tends to their actual size in probability.

Hence the proposition.   ∎

Experimental Results: To substantiate the analytical results, a two-class problem with classes [2, 7] × [2, 7] and [5, 10] × [5, 10] is considered [Fig. 6(a)]. Here, the nonoverlapping regions for C_1 are [2, 7] × [2, 5] and [2, 5] × [5, 7], and the nonoverlapping regions for C_2 are [5, 10] × [7, 10] and [7, 10] × [5, 7]. The overlapping region is [5, 7] × [5, 7].

To implement the recognition system, five traning sample sets with 50, 100, 150, 200, and 250 samples from each of the classes are chosen randomly. Figs. 6(b)–(f) show the regions corresponding to various output choices for the five sample sets. Here the character "A" represents the single choice for C_1; the character "B" represents the single choice for C_2; the character "a" represents the first-second choice with first choice as C_1; the character "b" represents the first-second choice with first choice as C_2; the character "C" represents the combined choice reflecting both C_1 and C_2, and the blank character "\thinspace " represents the null choice.

The results in Figs. 6(b)–(f) verify that with the increase of sample sizes, the training classes tend to the actual classes (Theorem 1) and the feature subdomains with the extended portions tend to their actual sizes (Theorem 2). The overlapping (combined and first-second choices), nonoverlapping (single choice) and no-class (null choice) regions are seen to tend to their actual sizes with the increase in the size of training
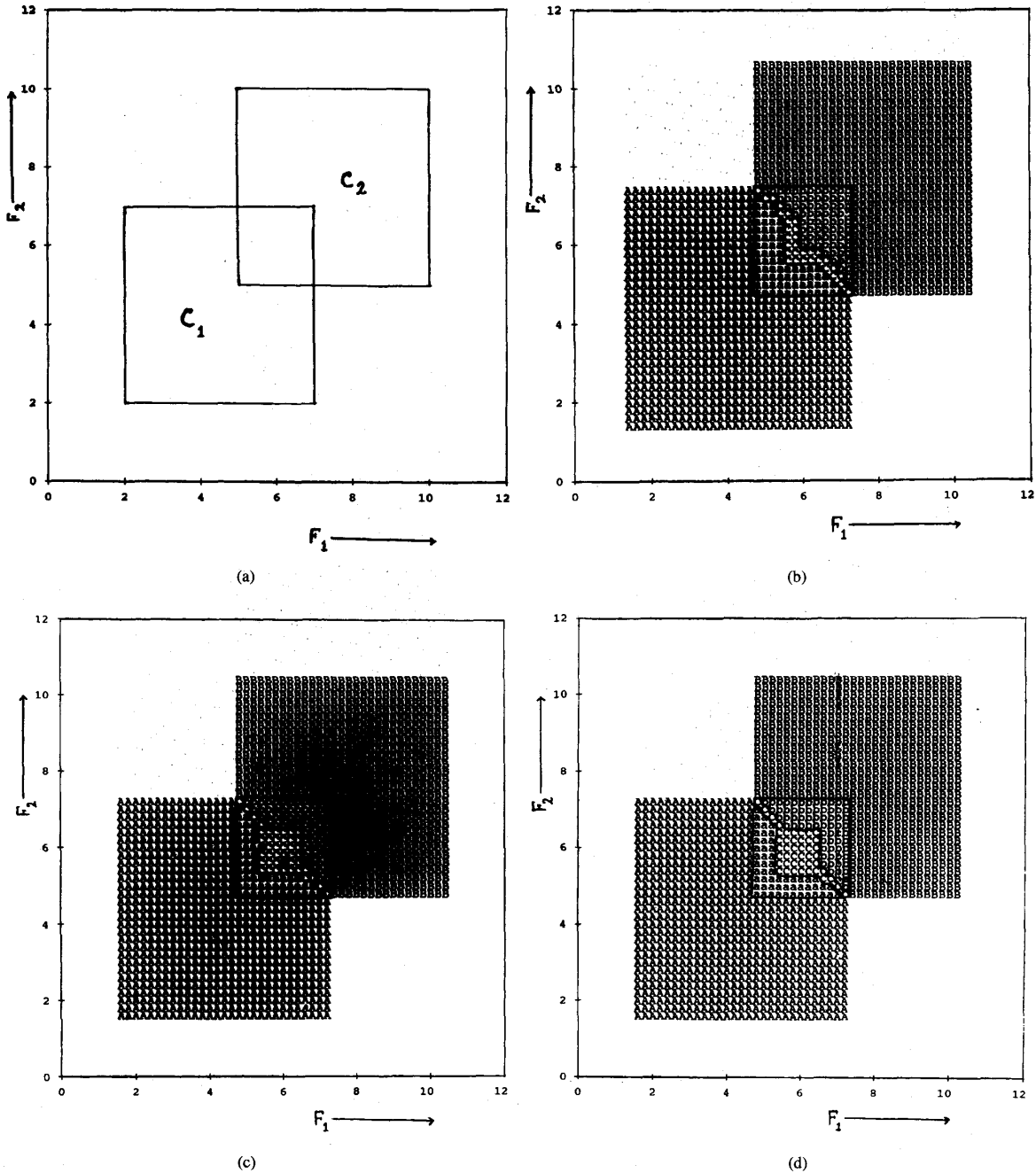
(a)

(b)

(c)

(d)

Fig. 6. (a) Two rectangular classes. (b)–(f) Various regions corresponding to five training sets (with 50, 100, 150, 200, and 250 samples from each class) for the pattern classes in (a). (g) Various regions of the Bayes classifier corresponding to the fifth training set (with 250 samples from each class) for the pattern classes in (a).

samples. Hence Proposition 1 is verified experimentally for the case of rectangular pattern classes. The experimental results in Figs. 6(b)–(f) also support the analytical findings and diagrams in Figs. 5(b) and (d).

For analyzing the performance of the proposed recognition system, a test set with 1000 samples from each of the classes is generated. Table III provides the recognition scores under

various choice groups corresponding to the aforesaid five training sample sets.

The Bayes classifier is also applied on the same pattern classes assuming rectangular distributions of the classes. For a comparative study, the hard regions for the classes C_1 and C_2, and the no-class region obtained with the Bayes classifier corresponding to the fifth sample set (with 250
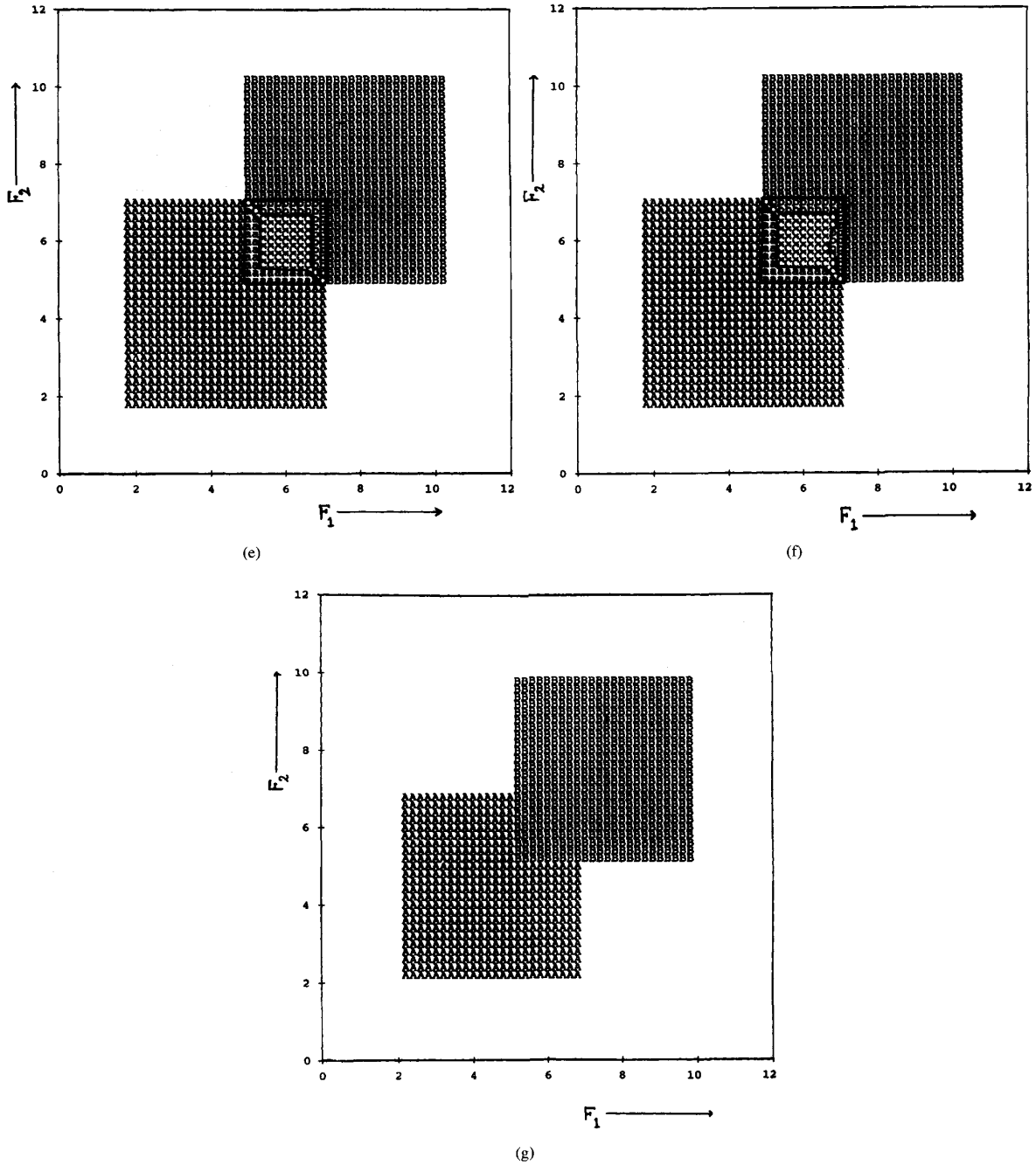
(e)

(f)

(g)

Fig. 6.   (continued)

samples from each class) are shown in Fig. 6(g). The feature points represented by the characters "A" and "B" in Fig. 6(g) indicate their belonging to the hard regions for C_1 and C_2, respectively.

Note from Fig. 6(g) that Bayes classifier (with rectangular distributions) assigned, unlike the proposed system [Fig. 6(f)], the while overlapping region (based on the training samples)

to one class, i.e., C_2. Again, corresponding to a feature point represented by the blank characters " " in Fig. 6(g), the values of the discriminating function are zero for both the classes. These feature points may be considered as belonging to the no-class region (although this no-class region, in fact, includes some portions of the actual pattern classes). These findings are due to the inherent properties of the rectangular distribution.

TABLE III
RECOGNITION SCORE FOR THE PATTERN CLASSES IN FIG. 6(a)

Various Group of Choices		% Recognition Score				
		Sample Sizes in Each Class				
		50	100	150	200	250
Single Correct Choice		79.60	80.35	80.60	81.60	82.45
First Correct Choice		11.05	8.75	8.35	6.45	4.80
Combined Correct Choice		3.70	6.05	6.60	8.15	9.95
Second Correct Choice		5.65	4.85	4.45	3.80	2.80
Fully Wrong Choice		0.00	0.00	0.00	0.00	0.00
Bayes	Correct	84.25	88.00	88.35	88.55	88.90
Classifier	Wrong	15.75	12.00	11.65	11.45	11.10

TABLE IV
RECOGNITION SCORE FOR THE PATTERN CLASSES IN FIG. 8(a)

Various Group of Choices		% Recognition Score				
		Sample Sizes in Each Class				
		50	100	150	200	250
Single Correct Choice		73.95	80.70	84.40	88.70	89.90
First Correct Choice		20.80	13.85	10.20	6.45	4.80
Combined Correct Choice		3.10	3.85	4.00	2.85	3.35
Second Correct Choice		2.15	1.60	1.60	1.95	1.95
Fully Wrong Choice		0.00	0.00	0.00	0.00	0.00
Bayes	Correct	93.00	94.10	94.25	94.50	94.45
Classifier	Wrong	7.00	5.90	5.75	5.50	5.55

Note that the assumption of any other distribution is not valid here. On the other hand, the output decisions, as shown in Figs. 6(b)–(f), of our multivalued recognition system are seen to be very appropriate.

The recognition scores of the Bayes classifier (with rectangular distributions) corresponding to the five training sets are included in Table III. Note that the score obtained by adding single correct, first correct, and half of the combined correct choices becomes much higher than the correct recognition score of the Bayes classifier. Again, a significant portion of the wrong choices of Bayes classifier is seen to be corrected by the proposed system and the remaining portion is distributed among the combined and second correct choices.

B. Circular Classes

The analytical results of rectangular classes obtained in the previous section are extended here to circular pattern classes [Fig. 7(a)]. A typical training sample set is assumed for carrying out the theoretical analysis of the proposed multivalued recognition system on the circular classes in Fig. 7(a). The actual overlapping portions and the overlapping portions from the training set are distinctly marked in Fig. 7(a).

Some typical feature subdomains and space subdomains corresponding to the training samples [Fig. 7(a)] are drawn in Fig. 7(b). Based on these feature subdomains and space subdomains, and using the results obtained for the rectangular pattern classes [Figs. 5(a)–(d)], the regions corresponding to the single, first-second, combined, and null choices are shown in Fig. 7(b). To show the overlapping regions more prominently, the rectangular portion (enclosed by dotted lines) in Fig. 7(a) which includes the overlapping region, is enlarged in Fig. 7(c). The corresponding rectangular portion (enclosed

by thick lines) in Fig. 7(b) is also enlaged in Fig. 7(d). Figs. 7(b) and (d) show the regions corresponding to various output choices for the pattern classes in Figs. 7(a) and (c), respectively.

When the size of the training samples increases, the values of the accuracy factor (2) decrease, and correspondingly the number of feature subdomains and space subdomains increases. Therefore, the sizes of the feature subdomains and space subdomains decrease with the increae in the size of the training samples. Hence, it can be concluded that the increase in sample sizes results in increase in the accuracy of the choice regions with respect to their actual sizes. Therefore, Proposition 1 is claimed for the circular shaped pattern classes. ∎

Experimental Results: To substantiate the analytical findings, a two-clas problem with circular classes is considered. The centers of the classes C_1 and C_2 are taken to be (5, 5) and (8.5, 8.5), respectively, and their radii are considered to be 3.5 and 3, respectively [Fig. 8(a)].

To implement the recognition system, five training sample sets with 50, 100, 150, 200, and 250 samples from each of the pattern classes are chosen randomly. Figs. 8(b)–(f) show the regions corresponding to various output choices for the five sample sets. Here the character "A" represents the single choice for C_1; the character "B" represents the single choice for C_2; the character "a" represents the first-second choice with first choice as C_1; the character "b" represents the first-second choice with first choice as C_2; the character "C" represents the combined choice for both C_1 and C_2, and the blank character " " represents the null choice. These results demonstrate that with the increase of sample sizes, the estimated classes tend to the actual classes (Theorem 1) and the feature subdomains (with the extended portions) tend to their actual sizes (Theorem 2). The estimated nonoverlapping, overlapping and no-class regions are seen to tend to their actual sizes with the increase in the size of training samples. Hence the claim of Proposition 1 is justified experimentally. The experimental results in figs. 8(b)–(f) also support the analytical findings in Figs. 7(b) and (d).

The distribution functions of the aforesaid pattern classes are assumed to be Gaussian for applying the Bayes classifier. The hard regions of the Bayes classifier for the classes C_1 and C_2 corresponding to the fifth sample set (with 250 samples from each class) are shown in Fig. 8(g). The feature points represented by the characters "A" and "B" in Fig. 8(g) indicate their belonging to the hard regions for C_1 and C_2, respectively. As expected, the Bayes decision boundary is seen to pass through the combined choice region of the proposed system. Note that in the proposed system, the combined choice region is considered as the boundary between overlapping classes.

As before, 1000 test samples from each of the classes are generated artificially for analyzing the performance of the proposed system. The recognition scores of our system are shown in Table IV corresponding to the five training sets. The recognition scores of the Bayes classifier with the same tranining sets are also included in Table IV. The correct decision (with single correct, first correct, and half of the
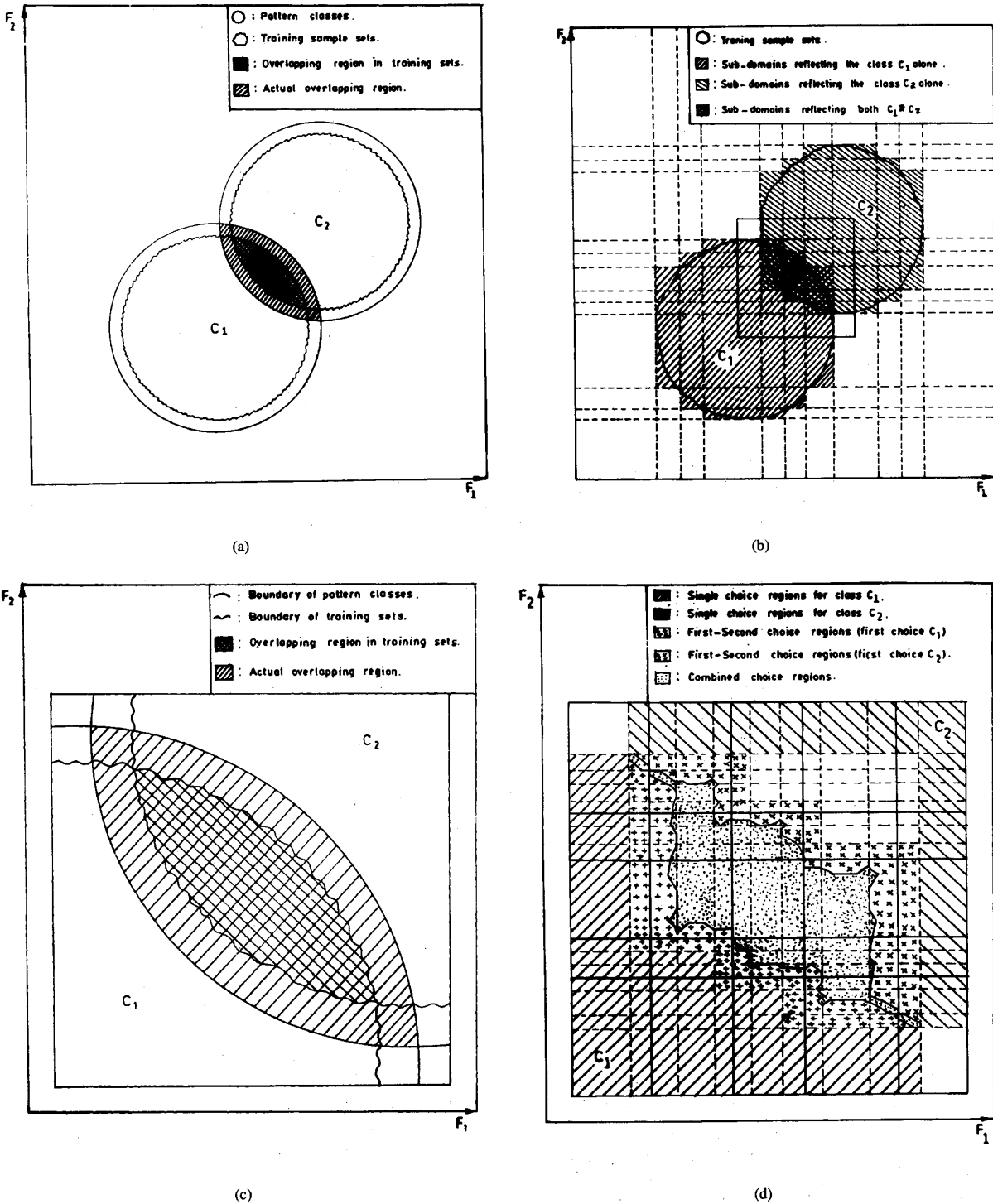
(a)

(b)

(c)

(d)

Fig. 7.   (a) Two circular classes with a training set. (b) Subdomains and choice regions for the pattern classes in (a). (c) Enlarged version of the rectangular portion (enclosed by dotted lines) in (a). (d) Enlarged version of the rectangular portion (enclosed by solid lines) in (b).

combined correct choices) of the proposed system is seen to be much higher than that of the Bayes classifier. In other words, among the samples which were misclassified by Bayes some are found to be corrected by our system and the remaining samples are distributed among the proposed combined and second correct choices.
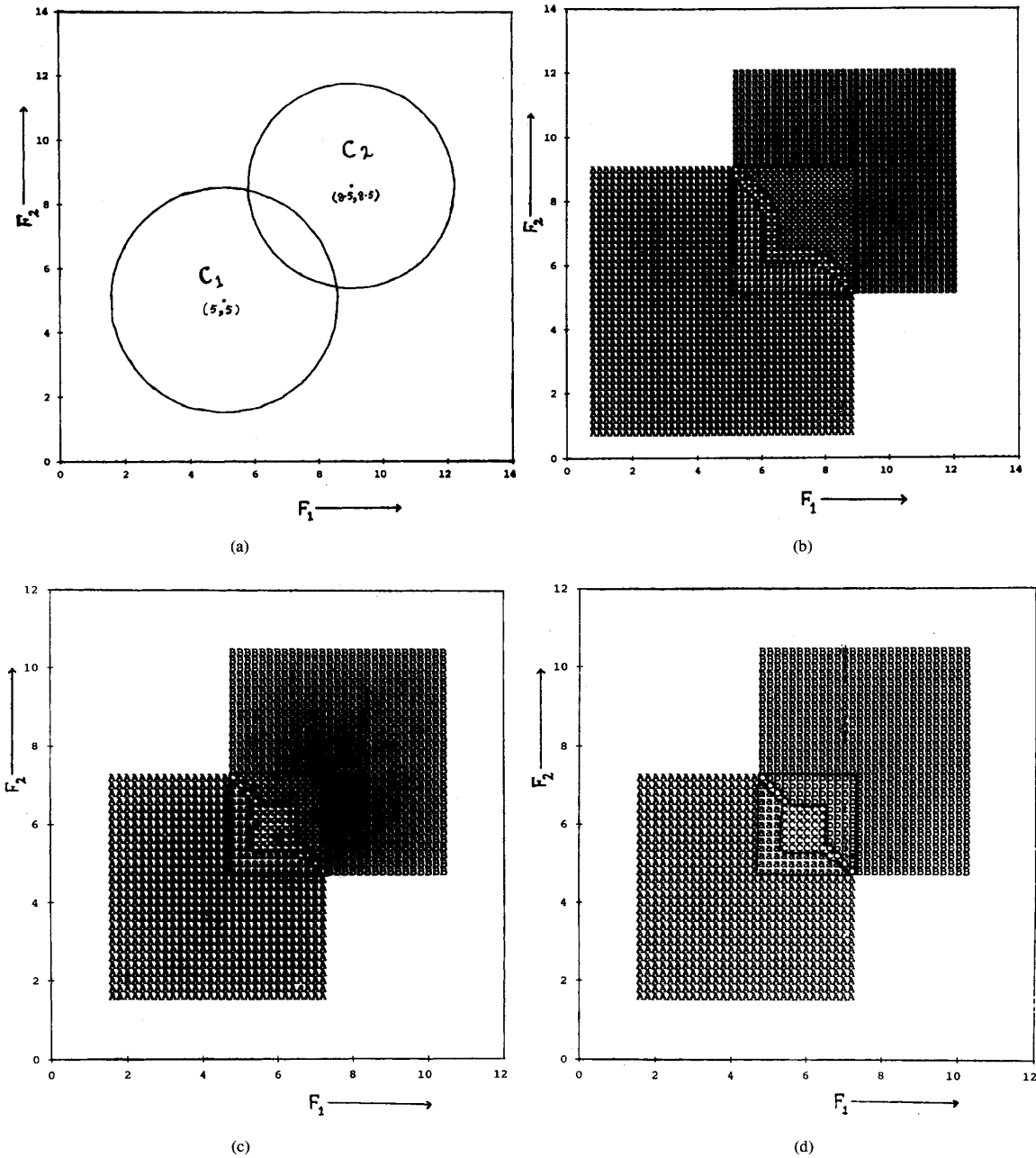
(a)

(b)

(c)

(d)

Fig. 8.   (a) Two circular classes with centers at (5.5, 5.5) and (8.5, 8.5), and with radii 3.5 and 3, respectively. (b)–(f) Various regions corresponding to five training sets (with 50, 100, 150, 200, and 250 samples from each class) for the pattern classes in (a). (g) Hard regions of the Bayes classifier corresponding to the fifth training set (with 250 samples from each class) for the pattern classes in (a).

VII. CONCLUSIONS AND DISCUSSIONS

The four state decisions of the multivalued recognition system, as described by the authors in [1], reflect the characteristics of the overlapping, and nonoverlapping and no-class regions (i.e., ambiguous/doubtful) of the feature space by providing output decision in four states, namely, single, first-second, combined, and null choices. In the present work, the regions corresponding to aforementioned four output forms are calculated analytically under two-class problem in one and two dimensional feature spaces. The similar findings can be obtained for a general M class N feature problem with classes of any shape. It has also been verified here that with the increase in the size of the training samples, the estimated versions of these regions tend to their actual sizes (Proposition 1).
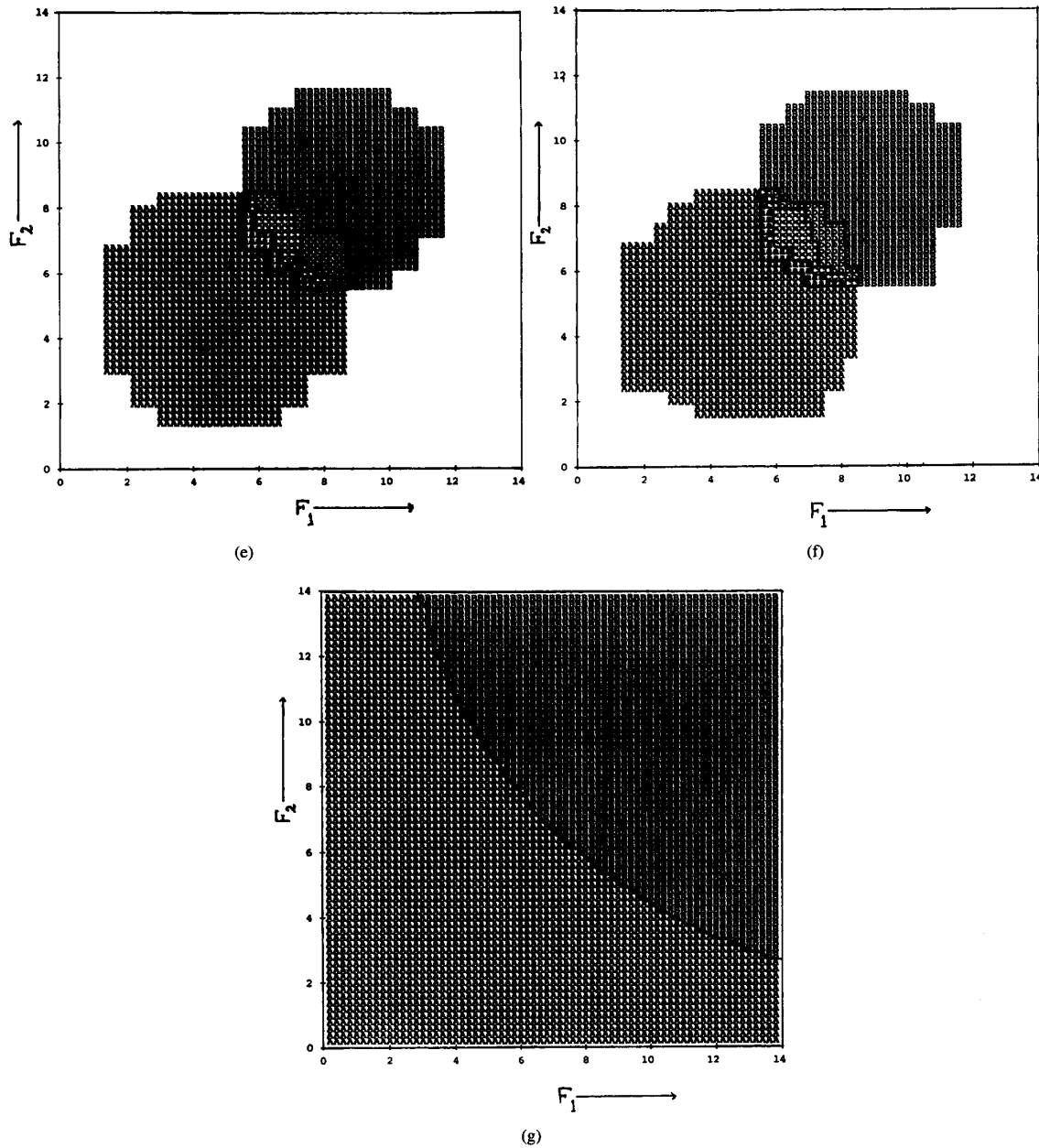
(e)


(f)

(g)

Fig. 8. (continued)

All the analytical findings have been substantiated with experimental results. It has been observed that Bayes decision boundaries always lie within the combined choice regions provided by the multivalued recognition system. Among the wrong decisions made by the Bayes classifier, a part is seen to be corrected by the single and first choices of the proposed system. The remaining portion is found to be distributed among the combined correct and second correct choices. The present theoretical analysis, therefore, establishes that the multilayered system has a provision of improving its effi-

ciency significantly by incorporating the second and combined choices under the control of a supervisory scheme. The present investigation, in turn, analytically justifies the consideration of output decisions in fours states for managing uncertainties arising from ambiguous regions.

REFERENCES

[1] D. P. Mandal, C. A. Murthy, and S. K. Pal, "Formulation of a multivalued recognition system," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-22, no. 4, pp. 307–320, 1992.

[2] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1958.

[3] L. A. Zadeh, "Fuzzy logic and approximate reasoning," *Synthese*, vol. 30, pp. 407–428, 1977.

[4] D. P. Mandal, C. A. Murthy, and S. K. Pal, "Determining the shape of a pattern class from sampled points in $\mathbb{R}^2$," *Int. J. General Syst.*, vol. 20, no. 2, pp. 307–339, 1992.

[5] D. P. Mandal, C. A. Murthy, and S. K. Pal, "Determining the shape of a pattern class from sampled points: Extension to $\mathbb{R}^N$," *Int. J. General Syst.*, submitted.

[6] D. P. Mandal and C. A. Murthy, "Determining the multivalued shape of a pattern class," in *Fuzzy Reasoning in Information, Decision and Control*, S. Tzafestas, Ed. Boston, MA: Kluwer, 1993, to be published.

[7] C. A. Murthy, "On consistent estimation of classes in $\mathbb{R}^2$ in the context of cluster analysis," Ph.D. dissertation, Indian Statistical Inst., Calcutta, India, 1988.

[8] P. Billingsley, *Probability and Measure*. New York: Wiley, 1979.

[9] R. B. Ash, *Real Analysis and Probability*. New York: Academic, 1972.
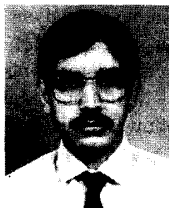
Deba Prasad Mandal received the B.Sc. (honors) degree in statistics from Kalyani University, West Bengal, India, in 1984, and the Master of Computer Applications (MCA) degree from Jawaharlal Nehru University, New Delhi, India, in 1988, and submitted his Ph.D. thesis to Indian Statistical Institute, Calcutta, in December 1992.

He is a Senior Research Fellow in the Machine Intelligence Unit at the Indian Statistical Institute. His research interests include pattern recognition, image processing, remote sensing, and fuzzy sets and systems.

Mr. Mandal received the Young Scientist Award in Computer Sciences from the Indian Science Congress Association in 1992. He is a member of the Indian Society for Fuzzy Mathematics and Information Processing (ISFUMIP) and the Indian Unit for Pattern Recognition and Artificial Intelligence (IUPRAI), and a life member of the Indian Science Congress Association (ISCA).

C. A. Murthy was born in Ongole, India, in 1958. He received the B.Stat. (honors), M.Stat., and Ph.D. degrees from Indian Statistical Institute, Calcutta, in 1979, 1980, and 1989, respectively.

He visited the Department of Computer Science of Michigan State University, East Lansing, as UNDP fellow in 1991–1992. His research interests include pattern recognition, image analysis, fuzzy sets, cluster analysis, fractals, and neural networks. He is currently working as a lecturer in the Machine Intelligence Unit at the Indian Statistical Institute.

Dr. Murthy is a member of the Indian Society for Fuzzy Mathematics and Information Processing (ISFUMIP) and the Indian Unit for Pattern Recognition and Artificial Intelligence (IUPRAI).

Sankar K. Pal (M'81–SM'84–F'93) received the B.Sc. (Hons.) degree in physics and the B.Tech., M.Tech., and Ph.D. degrees in radiophysics and electronics from the University of Calcutta, India, in 1969, 1972, 1974, and 1979, respectively. He received another Ph.D. degree in electrical engineering along with DIC from Imperial College, University of London in 1982.

He is a Professor and Head of the Machine Intelligence Unit at the Indian Statistical Institute, Calcutta. He served as a Professor-in-Charge of the Physical and Earth Sciences Division, Indian Statistical Institute, during 1988–1990. He was also a Guest Lecturer (1983–1986) in Computer Science at Calcutta University. His research interests include pattern recognition, image processing, artificial intelligence, neural nets, genetic algorithms, and fuzzy sets and systems. He is a coauthor of the book *Fuzzy Mathematical Approach to Pattern Recognition* (New York: Wiley, 1986) which received the Best Production Award in the 7th World Book Fair, New Delhi, and a Co-editor of the book *Fuzzy Models for Pattern Recognition* (New York: IEEE Press, 1992). He is listed in *Reference Asia*, Asia's *Who's Who of Men and Women of Achievements*, and *Biography International*.

Dr. Pal is a Fellow of the IEEE and the IETE, a member of the Executive Advisory Editorial Board of IEEE *Transactions on Fuzzy Systems* and the *International Journal of Approximate Reasoning* (North Holland), an Associate Editor of the *Far-East Journal of Mathematical Sciences*, a member of the Editorial Reviewing Board of IEEE *Computer* magazine, *International Journal of Applied Intelligence*, and *Mathematical Reviews*, and an Executive Committee Member of the ISFUMIP and IUPRAI. He is also a Permanent Member of the INDO–US Forum for Cooperative Research and Technology Transfer (IFCRTT). In 1986 he was awarded a Fulbright Postdoctoral Visiting Fellowship to work at the University of California, Berkeley, and the University of Maryland, College Park. In 1989 he received an NRC–NASA Senior Research Award to work at the NASA Johnson Space Center, Houston, TX. He received the 1990 Shanti Swarup Bhatnagar Prize in Engineering Sciences (which is the most coveted and highest scientific award in India) for this contribution to pattern recognition. In 1993 he was awarded the Jawaharlal Nehru Fellowship.