The Acquisition Probability for a Minimum Distance One-Class Classifier

KEINOSUKE FUKUNAGA, Fellow, IEEE

RAYMOND R. HAYES
Purdue University

LESLIE M. NOVAK
M.I.T., Lincoln Laboratory

An approximation for the acquisition probability for a minimum distance one-class classifier is derived. An exact expression for the acquisition probability is dependent upon the operating characteristics in the distance space, the number of targets detected, and the number of other objects detected. An approximate expression replaces the operating characteristics curve by a single point. Experimental results are presented to demonstrate the validity of the approximation. Combinatorial techniques can be used when only the total number of objects detected is known. All of these results can be extended to include the multitarget, multiple-shot case.

Manuscript received January 29, 1986; revised June 24, 1986.

This work was partly supported by the National Science Foundation under Grant ECS-8300536, and by IBM under the Resident Study Program.

Authors' addresses: Fukunaga and Hayes, School of Electrical Engineering, Purdue University, Lafayette, IN 47907; Novak, Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, MA 02173.

0018-9251/87/0700-0493 $1.00

I. INTRODUCTION

In many targeting scenarios, objects from different classes are detected and classified. As long as all of the classes are well defined, standard Bayesian classification techniques work very well. However, in some cases, one class can be well defined, while the other is not. For example, when we want to distinguish tanks (targets) from all other possible objects (nontargets), the nontargets may include trucks, automobiles, and all kinds of other vehicles as well as trees and clutter discretes which are detected erroneously. Because of the wide variety, it is almost impossible to study the distributions of all possible nontargets before a classifier is designed.

One-class classification schemes have been proposed to solve this problem. Typically, they involve measuring the distance of the object from the target mean and applying a threshold to determine if it is or isn't a target [1]. This technique, however, greatly increases the classification error. The mapping from the original n-dimensional feature space to a one-dimensional distance space destroys valuable classification information which existed in the original feature space.

However, this large increase in error can be reduced if one uses ranking instead of thresholding. If many objects are detected in a field and the goal is to acquire that one object which is most target-like, rank the objects according to their distances from the target mean, and select the closest one. The acquisition probability of this procedure was derived and studied by Parenti and Tung [2] and Novak [3]. In this paper, we point out that this probability is determined by the operating characteristics in the distance space as well as the numbers of targets and nontargets detected in the field. Also, we show that, if an exact measure is not required, the probability of acquisition can be approximated from just one point of the operating characteristics.

II. COMPUTATION OF ACQUISITION PROBABILITY

Let X be an n-dimensional vector, representing an object in the feature space, and let us assume that k_1 targets $(X_1, \ldots, X_{k_1})$ and k_2 nontargets $(X_{k_1+1}, \ldots, X_{k_1+k_2})$ are detected in a field. The acquisition procedure which is studied in this paper is: 1) compute the squared distance of X_i from the target's expected vector (M_1), normalized by the target covariance matrix (Σ_1)

$$z_i = \frac{1}{n} (X_i - M_1)^{\mathrm{T}} \Sigma_1^{-1} (X_i - M_1), \qquad (i = 1, 2, \ldots, k_1 + k_2) \quad (1)$$

where T indicates the transpose of the vector (M_1 and Σ_1 are assumed to be known), and 2) rank the X_i's according to their z_i values. The X_i with the smallest z is selected as the target to be acquired.

The probability of acquiring any one of the k_1 targets in the field by this procedure (the probability of correct classification) can be expressed as [2]

$$P_a = \int_0^1 k_1(1 - u_1)^{k_1-1}(1 - u_2)^{k_2}\, du_1 \tag{2}$$

where

$$u_i(t) = \int_0^t p_i(z)dz, \qquad (i = 1, 2) \tag{3}$$

and $p_i(z)$ is the density function of z for class i. Classes 1 and 2 are assigned to the targets and the nontargets, respectively. As is seen in (3), $u_i(t)$ is the probability of a sample from class i falling in $0 \le z < t$. The variables $u_1(t)$ and $u_2(t)$ are known as the detection and false alarm probabilities in the z-space when the threshold is chosen at $z = t$. In (2), du_1, $(1-u_1)^{k_1-1}$ and $(1-u_2)^{k_2}$ represent the probability of one of the k_1 targets falling in $t \le z < t + \Delta t$, $k_1 - 1$ of the targets falling in $t + \Delta t \le z < \infty$, and all k_2 nontargets falling in $t + \Delta t \le z < \infty$. The product of these three gives the probability of the combined event. Since the acquisition of any one of the k_1 targets is a correct classification, the probability is multiplied by k_1. The integration is taken with respect to t from 0 to ∞, that is, with respect to u_1 from 0 to 1. The derivation of (2) is given in Appendix B.

Rewriting $(1-u_1)$ as v and $(1-u_2)$ as $f(v)$, the acquisition probability becomes

$$P_a = \int_0^1 k_1 v^{k_1-1} f^{k_2}(v)dv. \tag{4}$$

Equation (4) indicates that P_a is a function of k_1, k_2, and $f(v)$, and $f(v)$ is a function relating $1-u_2$ to $1-u_1$. Since u_1 and u_2 are the detection and false-alarm probabilities in the z-space, $f(v)$ represents the operating characteristics when each sample is classified in the z-space without ranking.

Fig. 1 shows typical operating characteristics from a series of experiments which are described in Section III. Also shown are plots of v^{k_1-1} for $k_1 = 5$ and 20, which were used in these experiments. The worst case in which the summation of the class 1 error and the class 2 error is always 100 percent, regardless of the operating point or the threshold value is represented by $f(v) = v$. That is, the distributions of class 1 and class 2 are identical. Therefore, if the distributions are classifiable through this ranking procedure, $f(v) > v$. Thus, v^{k_1-1} is reasonably assumed to drop to zero more quickly than $f^{k_2}(v)$, for realistic values of k_1 and k_2. This means that only the rightmost part of the operating characteristics, where v is close to 1, contributes to P_a. The other part of the operating characteristics will not affect P_a.

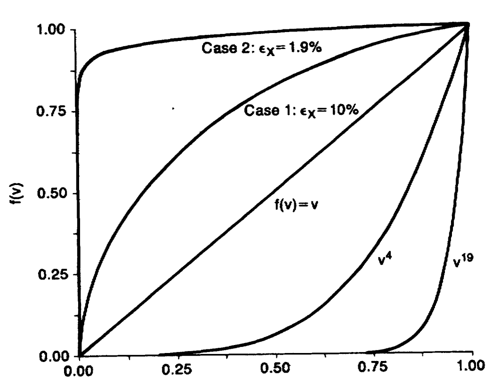


Fig. 1. Operating characteristics and v^{k_1-1}.

Although (2) is the exact expression for P_a, it is desirable to have an approximation formula through which P_a can be estimated faster and which shows the effects of k_1, k_2, and $f(v)$ more explicitly. Since only a small portion of $f(v)$ affects P_a and $f(v)$ is very flat in that portion, we tried to approximate $f(v)$ in this region with a constant, a line, and other simple constructs. We have found that a constant gives us the simplest and most robust approximation of P_a, although it is rather crude. Thus,

$$f(v) \cong 1 - \gamma \text{ for } v^{k_1-1} \not\cong 0 \tag{5}$$

$$P_a \cong \int_0^1 k_1 v^{k_1-1}(1-\gamma)^{k_2}dv = (1-\gamma)^{k_2}$$

$$\cong 1 - k_2\gamma. \tag{6}$$

We have found empirically that γ may be selected in the following manner: 1) for a given k_1, find v_0 which satisfies $v_0^{k_1-1} = 0.5$, and 2) read the operating characteristics $f(v)$ at v_0. Then, $f(v_0) = 1 - \gamma$. The experimental results of this approximation are reported in Section III.

It might seem that (6) is too sensitive to changes in the value of γ. However, a small change in γ corresponds to a significant change in the operating characteristics. So, in practice, the variation of γ stays very small and the approximation of (6) works well as reported in the next section.

III. EXPERIMENTAL RESULTS

In order to test the validity of the proposed approximation and to find a way to select the value of γ, a series of experiments were run.

For $p_i(z)$ $(i = 1, 2)$ of (1), Gamma densities were chosen as

$$p_i(z) = \frac{c_i^{b_i+1}}{\Gamma(b_i+1)} z^{b_i} e^{-c_i z} \tag{7}$$

whose expected value and variance are

$$m_i = \frac{b_i + 1}{c_i} \quad \text{and} \quad \sigma_i^2 = \frac{b_i + 1}{c_i^2}. \tag{8}$$

The reasons for this selection are as follows.

1) For class 1, if X is distributed Gaussianly with the expected vector M_1 and covariance matrix Σ_1, z of (1)

has the Gamma distribution of (7) with $m_1 = 1$ and $\sigma_1^2 = \frac{2}{n}$.

2) For class 2, even if X is distributed Gaussianly, z does not have an exact Gamma density, since the expected vector M_2 differs from M_1. However, our experiments show that the empirical distributions of z are very close visually to the Gamma distributions for a wide variety of M_2 and Σ_2 values. The empirical distribution of z was obtained from samples generated Gaussianly with given M_2 and Σ_2 in the X-space and converted to z by (1). The corresponding Gamma density function was specified by the expected value and variance computed by the following equations:

$$m_2 = \frac{1}{n}\left(\sum_{i=1}^{n} \lambda_i + \sum_{i=1}^{n} \mu_i^2\right) \tag{9}$$

$$\sigma_2^2 = \frac{1}{n^2}\left(2\sum_{i=1}^{n} \lambda_i^2 + 4\sum_{i=1}^{n} \lambda_i \mu_i^2\right) \tag{10}$$

where λ_i and μ_i are obtained as the results of simultaneous diagonalization. That is, a linear transformation $A^T X$ is applied to X such that $A^T \Sigma_1 A = I$ and $A^T \Sigma_2 A = \Lambda$. The variables λ_i and μ_i are the ith components of the diagonal matrix Λ and the transformed vector $A^T(M_2 - M_1)$, respectively. The derivations of (9) and (10) are given in the Appendix A.

In order to cover various cases for the class 2 distribution, two types of Gaussian distributions were chosen for the experiments. Note that the selection of I for Σ_1 does not hurt generality, since we can always linearly transform Σ_1 to I without changing the subsequent results. Throughout the experiments, it was assumed that a priori probabilities of classes 1 and 2 are equal. The Bayes errors in the X- and z-spaces, respectively, are indicated by ϵ_x and ϵ_z. The Bayes error is the smallest error which can be obtained by the optimal classifier (the Bayes classifier) for given distributions [1].

Case 1: $\Sigma_1 = \Sigma_2 = I$, $M_2 - M_1 = M$, $n = 20$.

The Bayes classifier in the X-space is linear in this case and ϵ_x is determined by the length of the vector M, $\|M\|$. We selected $\|M\|$'s to get 1, 5, 10, and 20 percent for ϵ_x.

Case 2: $\Sigma_1 = I$, $\Sigma_2 = \Lambda$, $M_2 - M_1 = [\mu_1, \ldots, \mu_8]^T$, $n = 8$.

The variables Λ and $M_2 - M_1$ were chosen from standard data of [4], and their components are $\lambda_1 = 8.41$, $\lambda_2 = 12.06$, $\lambda_3 = 0.12$, $\lambda_4 = 0.22$, $\lambda_5 = 1.49$, $\lambda_6 = 1.77$, $\lambda_7 = 0.35$, $\lambda_8 = 2.73$, and $\mu_1 = 3.86$, $\mu_2 = 3.10$, $\mu_3 = 0.84$, $\mu_4 = 0.84$, $\mu_5 = 1.64$, $\mu_6 = 1.08$, $\mu_7 = 0.26$, $\mu_8 = 0.01$. This data is suitable to test the case where Σ_1 and Σ_2 are significantly different, since the λ's vary from 0.12 to 12.06. The Bayes classifier is quadratic for this case and the resulting ϵ_x is 1.9 percent [4]. In order to obtain various ϵ_x's, we multiplied $M_2 - M_1$ by constants while keeping the covariances fixed.

The experiments were carried out as follows.

1) Compute m_2 and σ_2^2 of (9) and (10) from given M_2 and Σ_2.

2) We assumed that the class 2 distribution in the z-space is Gamma with m_2 and σ_2^2 computed in step 1. The class 1 distribution is Gamma with $m_1 = 1$ and $\sigma_1^2 = 2/n$. These two Gamma density functions are plotted in Fig. 2(a) for case 1 and 2(b) for case 2, respectively.

3) In Fig. 2, select the threshold t and compute $u_1(t)$ and $u_2(t)$ by (3). Changing t from 0 to ∞, plot the relationship between $1 - u_2$ and $1 - u_1$. The results are the operating characteristics as shown in Fig. 1.

4) Compute P_a of (4) and the approximated P_a of (6) for several values of γ. Table I shows the results when γ

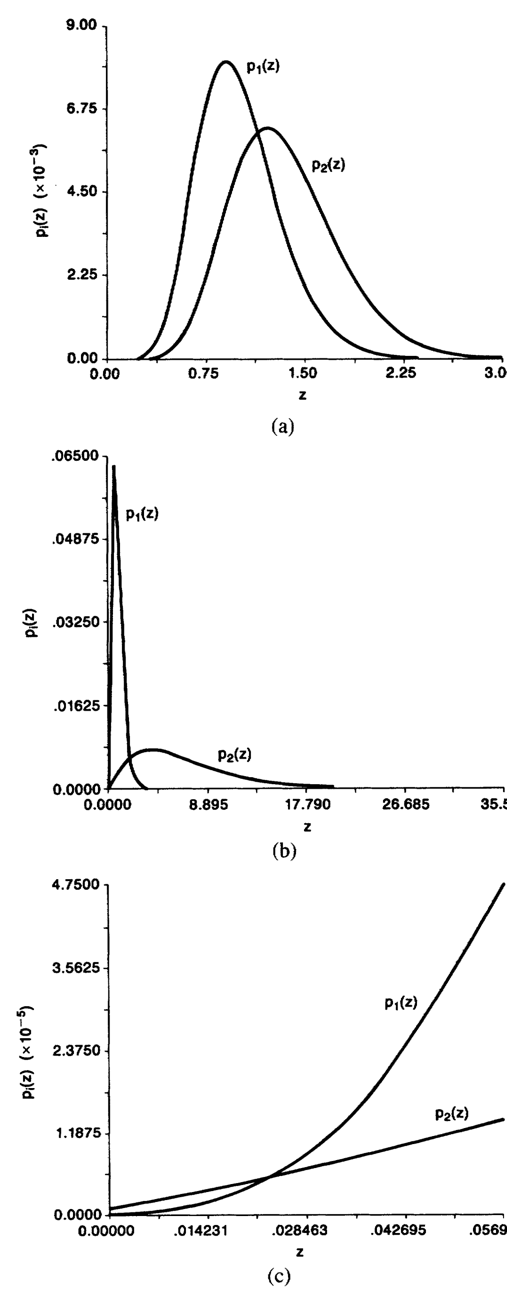


Fig. 2. Gamma densities plot (a) Case 1 with $\epsilon_x = 10$ percent, (b) Case 2 with $\epsilon_x = 1.9$ percent, (c) Blow-up of the left-most part of (b).

TABLE I
Results of P_a Approximation for cases 1 and 2.
(All numbers are percent)

	ϵ_x	ϵ_z	$k_1 = k_2 = 5$: $1-P_a$	$k_1 = k_2 = 5$: $1-(1-\gamma)^{k_2}$	$k_1 = k_2 = 5$: $k_2\gamma$	$k_1 = k_2 = 20$: $1-P_a$	$k_1 = k_2 = 20$: $1-(1-\gamma)^{k_2}$	$k_1 = k_2 = 20$: $k_2\gamma$
Case 1:	1.0	10.0	0.9	0.3	0.3	0.6	0.1	0.1
$\Sigma_1 = \Sigma_2$	5.0	24.0	8.9	6.2	6.4	4.4	2.9	3.0
$= I$	10.0	32.0	17.6	14.8	15.8	14.9	8.4	8.7
	20.0	42.0	34.2	35.5	42.0	32.0	26.9	31.1
Case 2:	1.9	12.9	4.4	3.1	3.1	6.9	4.0	4.1
$\Sigma_1 \neq \Sigma_2$	*	29.7	17.6	17.3	18.7	30.1	27.5	31.9
	*	35.8	23.1	23.4	25.9	37.0	36.3	44.6

Note: * - unknown error rates

is selected as $1 - f(v_0)$ where $v_0^{k_1 - 1} = 0.5$. Although the approximations are somewhat crude, they predict the trend of P_a reasonably well.

A counter-intuitive result was observed in the case 2 experiment. Intuitively, as k_1 and k_2 increase (i.e., as the numbers of targets and nontargets detected increase), the probability of misacquisition should decrease since there are now more targets, the acquisition of any one of which is considered correct. This is shown clearly in case 1. However, in case 2, the probability of misacquisition actually increases with an increase in k_1 and k_2. From (4), it should be apparent that an increase in k_1 makes the far rightmost position of the operating characteristics more dominant. Due to the construction of $f(v)$, the rightmost position of the operating characteristic corresponds to the integration of the leftmost portion (the section closest to zero) of the probability duration of the distances. Ordinarily, one would expect $p_1(z) > p_2(z)$ for small values of z (i.e., one would expect the probability of a target being very close to the target mean to be greater than the probability of a nontarget being very close to the target mean). However, Fig. 2 shows that, in spite of the fact that $m_2 > m_1$, $p_2(z) > p_1(z)$ for small values of z! Thus, increasing k_1 compresses the range of significant distances from the target mean and amplifies the effect of the small region in which nontargets are more likely to be closer to the target mean than the targets themselves, increasing the probability of misacquisition. This result suggests that a careful examination is needed for the starting edges of the density functions in the z-space before deriving any conclusions by intuition.

At this point, we would like to point out the purpose of the ranking procedure. As Table I shows, the transformation of (1) from n-dimensional X to one-dimensional z increases the classification error from ϵ_x to ϵ_z, if a simple threshold in the z-space is applied. Although the ranking procedure reduces ϵ_z to $1 - P_a$, this reduction barely compensates the loss from ϵ_x to ϵ_z. Therefore, there is no need to use the proposed procedure, if the class 2 distribution is unimodal Gaussian as in the experiments. The conventional Bayes classifier in the X-space gives the classification error ϵ_x. However, if the class 2 distribution consists of many Gaussians surrounding the class 1 distribution as shown in Fig. 3, we must use a one-class classifier such as $z \gtrless t$, accepting ϵ_z as the resulting error. In this case, ϵ_x merely

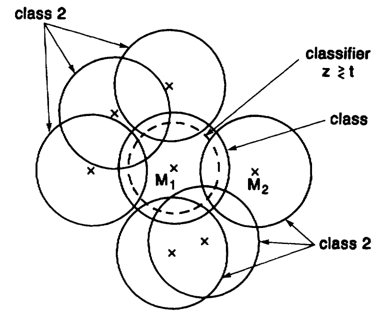


Fig. 3. A possible class 2 distribution with multi-modal Gaussian.

serves as a measure of how far the neighboring Gaussians are apart from the class 1 center. As was discussed in the introduction, in many target classification scenarios, class 2 includes various objects such as trucks, automobiles and all kinds of other vehicles as well as trees and clutter discretes, thus creating a distribution like the one in Fig. 3. Therefore we point out how much the error can be reduced (from ϵ_z to $1 - P_a$) by the ranking procedure, and discuss the effects of k_1, k_2, and the relative locations of the class 2 distributions.

IV. SUPPLEMENTARY DISCUSSIONS

Combinatorial Results

The expression and approximation derived for P_a are only good for fixed k_1 and k_2. More realistically, one is given the total number of objects detected in a field k and the a priori probability that a sample is a target, P_1. In this case, P_a can be computed by

$$P_a = \sum_{i=0}^{k} \binom{k}{i} P_1^i (1 - P_1)^{k-i} P_a(i, k-i) \qquad (11)$$

where $P_a(i, k-i)$ is the acquisition probability for $k_1 = i$ and $k_2 = k-i$.

Effect of Distance-Space Mapping

Even though the ranking procedure outperforms conventional one-class classification techniques, it is still hampered by the error introduced by the mapping from the original n-dimensional feature space to the one-dimensional distance space. In order to see how much the error is increased, the relationship between the Bayes errors in the X- and z-spaces, ϵ_x and ϵ_z, was examined for case 1 of the previous section. Results are presented in Fig. 4. Fig. 4 was obtained as follows.

1. Fix n (10, 50, 10, 150, 200).
2. Change $\|M\|$ in case 1 experiment and obtain the corresponding ϵ_x in the X-space.

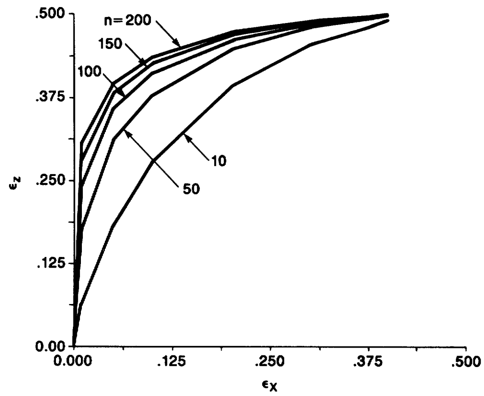


Fig. 4. ϵ_z versus ϵ_x for constant n.

3. Compute $u_1(t_0)$ and $u_2(t_0)$ of (3) by numerical integration. The variable $p_1(z)$ and $p_2(z)$ are assumed to be Gamma densities with $m_1 = 1$ and $\sigma_1^2 = 2/n$ for $p_1(z)$ and m_2 and σ_2^2 computed by (9) and (10) for $p_2(z)$. The variable t_0 is the value of z where $p_1(z)$ and $p_2(z)$ cross. When $p_1(z)$ and $p_2(z)$ cross at two values of z, as is the case for case 2 experiment, choose the larger z.

4. $\epsilon_z = \frac{1}{2}(1 - u_1(t_0)) + \frac{1}{2}u_2(t_0)$, since the a priori probabilities for classes 1 and 2 are both assumed to equal 1/2. As one would expect, ϵ_z becomes very large as n increases.

Tradeoff Between Number of Features and Original Error

Increasing the number of features n has both advantages and disadvantages in our targeting scenario. It reduces ϵ_x in general, but increases the information lost by mapping to the z-space. Thus, there should be some sort of tradeoff between the number of features which would provide a reasonable error in the X-space and limit the amount of information lost in the distance mapping. Fig. 5 shows the same experimental results as Fig. 4, but this time ϵ_x versus n is plotted for a fixed ϵ_z. Fig. 5 indicates that in order to achieve $\epsilon_z = 27$ percent for example, we may have many choices such as $\epsilon_x = 10$ percent with $n = 10$, $\epsilon_x = 3$ percent with $n = 60$, and so on. There is no reason to use an incredibly large number of features if the additional classifiability in the X-space cannot be transferred to the z-space. By keeping ϵ_z fixed, we were able to see just how much error could be introduced in the X-space while maintaining ϵ_z and reducing the number of features used.

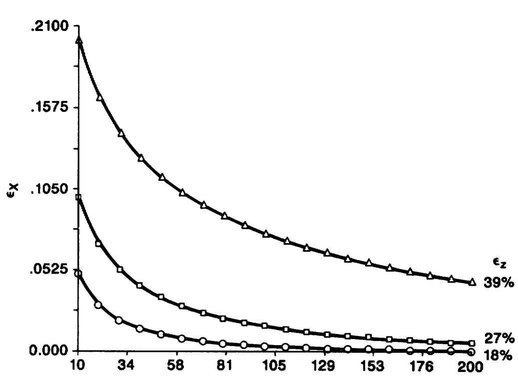


Fig. 5. ϵ_x versus n for constant ϵ_z.

Extension to Multiple-Target Multiple-Shot Case

So far, our scenarios have assumed that one and only one acquisition is attempted. However, in a different situation, where α acquisitions are attempted, we need to compute the probability that β targets are acquired in α attempts ($\alpha > \beta$). In this case, the probability of acquisition equals the probability that, if the α smallest distances are selected, β are targets and $\alpha - \beta$ are nontargets. The probability is presented here without derivation:

$$
\begin{aligned}
P_{\alpha\beta} = \int_0^1 \Bigg[& \binom{k_1}{\beta - 1} u_1^{\beta-1}(1-u_1)^{k_1-\beta+1} \\
& \times \binom{k_2}{\alpha-\beta} u_2^{\alpha-\beta} \binom{k_1-\beta+1}{1} \frac{du_1}{1-u_1} \\
& + \binom{k_1}{\beta} u_1^{\beta}(1-u_1)^{k_1-\beta} \binom{k_2}{\alpha-\beta-1} u_2^{\alpha-\beta-1} \\
& \times (1-u_2)^{k_2-\alpha+\beta+1} \binom{k_2-\alpha+\beta+1}{1} \frac{du_2}{1-u_2} \Bigg].
\end{aligned}
\tag{12}
$$

In [2], an approximation was developed which defines the multiple-target multiple-shot acquisition probability as a function of P_a, k_1, α, and β:

$$
P_{\alpha\beta} = \begin{cases} \binom{k_1}{\beta} P_a^{\beta}(1-P_a)^{k_1-\beta}, & \beta < \alpha,\ \beta \le k_1 \\ \sum_{i=\alpha}^{k_1} \binom{k_1}{i} P_a^{i}(1-P_a)^{k_1-i}, & \beta = \alpha,\ \beta \le k_1. \end{cases}
\tag{13}
$$

If our proposed approximation for P_a is used together with this expression, $P_{\alpha\beta}$ can be estimated directly from the empirical operating characteristics.

V. CONCLUSIONS

Targeting scenarios, in which one class is known and well defined and the other is unknown, point out the need for one-class classifiers. Conventional one-class classification techniques introduce a great deal of error by mapping the n-dimensional feature space into a one-dimensional distance space. An exact expression for the acquisition probability is dependent upon the empirical operating characteristics, the number of targets detected, and the number of other objects detected. An approximate expression is dependent on a single point of the operating characteristics, the number of targets detected, and the number of nontargets detected. Combinational techniques can be used when only the total number of objects detected is known. All of these results can be extended to include the multiple-target multiple-shot case.

APPENDIX A. DERIVATION OF EXPRESSIONS FOR m_2 AND σ_2^2.

Let us assume that $\Sigma_1 = I$, $\Sigma_2 = \Lambda$, $M_1 = 0$, and $M_2 = M = [\mu_1, \ldots, \mu_n]^T$. These assumptions do not hurt any generality. First, z of (1) can be modified as

$$z = \frac{1}{n} X^T X = \frac{1}{n}(X-M+M)^T(X-M+M)$$

$$= \frac{1}{n}(X-M)^T(X-M) + \frac{2}{n} M^T(X-M) + \frac{M^T M}{n}. \tag{A1}$$

The expected value of z for class 2 (ω_2) is

$$m_2 = E\{z|\omega_2\} = \frac{1}{n} E\{(X-M)^T(X-M)|\omega_2\}$$

$$+ \frac{2}{n} M^T E\{(X-M)|\omega_2\} + \frac{M^T M}{n}$$

$$= \frac{1}{n} \operatorname{tr} E\{(X-M)(X-M)^T|\omega_2\}$$

$$+ \frac{M^T M}{n} = \frac{1}{n} \operatorname{tr} \Lambda + \frac{M^T M}{n}$$

$$= \left(\sum_i \lambda_i + \sum_i \mu_i^2\right). \tag{A2}$$

Likewise, the second-order moment of z for ω_2 is

$$E\{z^2|\omega_2\} = \frac{1}{n^2} E\{(X-M)^T(X-M)(X-M)^T(X-M)|\omega_2\}$$

$$+ \frac{4}{n^2} M^T E\{(X-M)(X-M)^T|\omega_2\} M$$

$$+ \frac{1}{n^2}(M^T M)^2$$

$$+ \frac{2}{n^2} E\{(X-M)^T(X-M)|\omega_2\} M^T M$$

$$= \frac{1}{n}\left[3 \sum_i \lambda_i^2 + 2 \sum_{i>j}\sum \lambda_i \lambda_j\right]$$

$$+ \frac{4}{n^2} \sum_i \lambda_i \mu_i^2 + \frac{1}{n^2}\left(\sum_i \mu_i^2\right)^2$$

$$+ \frac{2}{n^2}\left(\sum_i \lambda_i\right)\left(\sum_i \mu_i^2\right) \tag{A3}$$

where X is assumed to be Gaussian. Thus, the variance of z for ω_2 is

$$\sigma_2^2 = E\{z^2|\omega_2\} - m_2^2$$

$$= \frac{1}{n^2}\left[2 \sum_i \lambda_i^2 + 4 \sum_i \lambda_i \mu_i^2\right] \tag{A4}$$

APPENDIX B. DERIVATION OF THE ACQUISITION PROBABILITY OF (2).

The acquisition probability of (2) is derived as follows:

$$P_a = \Pr\{\text{the smallest } z \text{ is from class } 1\}$$

$$= \sum_{i=1}^{\infty} \Pr\{A_i \text{ and } B_i \text{ and } C_i\}$$

$$= \sum_{i=1}^{\infty} \Pr\{A_i\} \Pr\{B_i|A_i\} \Pr\{C_i|A_i, B_i\} \tag{B1}$$

where $A_i = \{\text{no sample in } 0 \le z < i\Delta t\}$, $B_i = \{\text{one class 1 sample is in } i\Delta t \le z < (i+1)\Delta t\}$ and $C_i = \{k_1 - 1 \text{ class 1 samples and } k_2 \text{ class 2 samples in } (i+1)\Delta t \le z < \infty\}$. $\Pr\{A_i\}$, $\Pr\{B_i|A_i\}$, and $\Pr\{C_i|A_i, B_i\}$ may be computed as follows:

$$\Pr\{A_i\} = \binom{k_1}{0} u_1^0(i\Delta t)(1-u_1(i\Delta t))^{k_1} \binom{k_2}{0}$$

$$\times u_2^0(i\Delta t)(1-u_2(i\Delta t))^{k_2}$$

$$= (1-u_1(i\Delta t))^{k_1}(1-u_2(i\Delta t))^{k_2} \tag{B2}$$

$$\Pr\{B_i|A_i\} = \binom{k_1}{1}\left(\frac{\Delta u_1(i\Delta t)}{1-u_1(i\Delta t)}\right)^1$$

$$\times \left(1 - \frac{\Delta u_1(i\Delta t)}{1-u_1(i\Delta t)}\right)^{k_1-1}$$

$$\times \binom{k_2}{0}\left(\frac{\Delta u_2(i\Delta t)}{1-u_2(i\Delta t)}\right)^0$$

$$\times \left(1 - \frac{\Delta u_2(i\Delta t)}{1-u_2(i\Delta t)}\right)^{k_2}$$

$$\simeq k_1 \frac{\Delta u_1(i\Delta t)}{1-u_1(i\Delta t)} \tag{B3}$$

$$\Pr\{C_i|A_i, B_i\} = 1 \tag{B4}$$

where $\Delta u_j(i\Delta t)$ is the probability of a class j sample filling in $i\Delta t \le z < (i+1)\Delta t$. The approximation of (B3) is obtained by making $\Delta u_j \to 0$. Substituting (B2), (B3), and (B4) into (B1) and letting $\Delta u_j \to 0$, we can obtain

$$P_a = \int_0^1 k_1(1-u_1)^{k-1}(1-u_2)^{k_2}\, du_1. \tag{B5}$$

The summation of (B1) is taken by changing t from 0 to ∞. Since $u_i(0) = 0$ and $u_i(\infty) = 1$, and $u_i(t)$'s are the monotonic functions of t, the integration is taken with respect to u_1 from 0 to 1.

REFERENCES

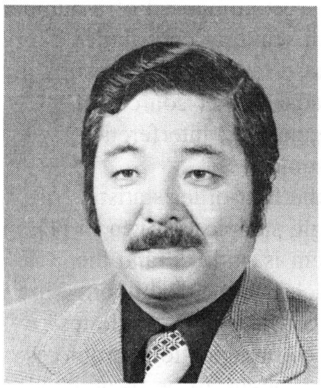
[1] Fukunaga, K. (1986)
Statistical pattern classification.

Handbook of Pattern Recognition and Image Processing
(T. Young and K.S. Fu, Ed.).
New York: Academic Press, 1986.

[2] Parenti, R.R., and Tung, E.W. (1981)
A statistical analysis of the multiple-target multiple-shot target acquisition problem.
Project Report TT-43, M.I.T., Lincoln Lab., Jan. 28, 1981.

[3] Novak, L. (1981)
Optimal target designation techniques.
IEEE Transactions on Aerospace and Electronic Systems, AES-17 (Sept. 1981), 676–684.

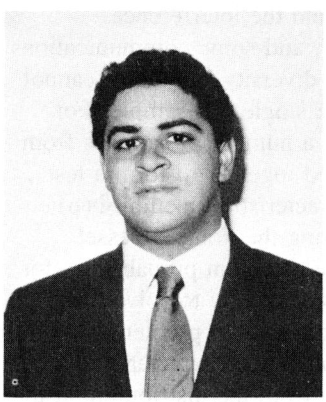
[4] Fukunaga, K. (1972)
Introduction to Statistical Pattern Recognition.
New York: Academic Press, 1972, p. 66.



Keinosuke Fukunaga (M'66—SM'74—F'79) received the B.S. degree in electrical engineering from Kyoto University, Japan, in 1953, the M.S.E.E. degree from the University of Pennsylvania, Philadelphia, in 1959, and the Ph.D. degree from Kyoto University in 1962.

From 1953 to 1966 he was with the Mitsubishi Electric Company, Japan, first with the Central Research Laboratories working on computer applications in control systems, and then with the Computer Division where he was in charge of hardware development. Since 1966 he has been with Purdue University, West Lafayette, Ind., where he is currently a Professor of Electrical Engineering. In the summers he has worked with a number of organizations. Also, he has served as a consultant to various government agencies and private companies.

Dr. Fukunaga was an Associate Editor of the *IEEE Transactions on Information Theory* for pattern recognition from 1977 to 1980. He is the author of *Introduction to Statistical Pattern Recognition*. He is a member of Eta Kappa Nu.



Raymond R. Hayes was born in Ann Arbor, Mich. on July 10, 1962. He received the B.S. degree in computer and electrical engineering from Purdue University, West Lafayette, Ind., in 1984 through the Bell Labs Engineering Scholarship Program.

He is currently working toward the Ph.D. degree, also in electrical engineering, at Purdue University under the IBM Resident Study Program. His research interests include statistical pattern recognition, artificial intelligence, and image processing.

Leslie Novak received the B.S.E.E. degree from Fairleigh Dickinson University in 1961, the M.S.E.E. degree from the University of Southern California in 1963, and the Ph.D. degree from the University of California, Los Angeles, in 1971.

Since 1977 he has been a member of the technical staff at M.I.T. Lincoln Laboratory, Lexington, Mass., where he is performing target detection and classification algorithm studies for millimeter wave radar seeker systems. He has published several papers on the sensitivity of the quadratic and linear classifiers in radar target detection. From 1972 to 1977 he was with the Raytheon Company, Bedford, Mass., where he was involved with the design of a CFAR signal processor for the Patriot system. Also, as a lead engineer he performed correlation algorithm studies and developed the adaptive threshold binary correlation algorithm utilized in the PDMM radar map-matching system. From 1968 to 1972 he was with Hughes Aircraft Company, Fullerton Calif., where he developed extended Kalman filter algorithms for the TPQ-36 and TPQ-37 artillery and mortar location radar systems. From 1961 to 1968 he was with Autonetics, Anaheim, Calif., where he performed radar system studies. While he was working toward the Ph.D. degree he received a Howard Hughes Ph.D. Fellowship. In addition, he has contributed chapters on stochastic observer theory to *Advances in Control Theory* (Academic Press, C.T. Leondes, Ed.), vol. 9, 1973, and vol. 12, 1976.