TEXAS INSTRUMENTS INCORPORATED

POB 655936 • MS: 134 • Dallas Texas 75265
Central Research Laboratories

Hebbian Learning, Principal Components and Automatic Target Recognition

Alan J. Katz

TECHNICAL REPORT

TR 08-91-14 May 31, 1991

J. Anthony	147	J. Florence	134	J. B. Sampsell	134
T. Barlow	8441	G. Frazier	154	P. Saunier	134
B. Barton	3669	M. Gately	238	H. Schaake	150
R. T. Bate	134	B. Gnade	147	A. Seabaugh	134
B. Bayraktaroglu	134	L. Hornbeck	134	T. Shaffner	147
W. Bean	31	M. Johnson	145	H.-D. Shih	154
J. D. Beck	150	J. Keenan	147	D. W. Shaw	147
S. Borrello	150	W. F. Keenan	150	A. Simmons	150
B. Boyd	8464	M. Kinch	150	G. C. Smith	134
K. Carson	150	R. Koestner	147	R. Stratton	136
W. W. Chan	150	C. Lindahl	8513	C. Tew	150
D. Chandra	154	R. Logan	3131	P. Thrift	134
C. Chapoton	8518	J. Luscombe	154	J. Tregilgas	154
P. Chatterjee	944	T. Moore	147	H.-Q. Tserng	134
M.-C. Chen	150	F. Morris	134	C. Turner	8707
D. R. Collins	134	E. Nelson	134	M. Villalba	238
L. Columbo	154	B. Newell	134	R. Wiggins	238
P. Congdon	8513	C. Penn	238	W. R. Wisseman	134
M. Cowens	147	A. Penz	134	J. Younse	134
M. Douglas	147	R. Pickens	8518	H.T. Yuan	134
W. Duncan	147	A. Ploysongsang	8518	TRS (2)	8433
G. Feather	145	A. Purdes	147		
J. Fish	3986	J. Randall	134		
B. Flinchbaugh	238	C. G. Roberts	150		

TI Overview: A key TI problem domain is automatic target recognition (ATR). ATR requires determining image features to distinguish targets from background clutter. Determining a set of robust image features that work well under a variety of conditions (time-of-day, season, geographic location, etc.) has proven extremely challenging. Even the simpler problem of finding a suitable set of image features for a fixed set of conditions is often difficult and requires considerable development time.

In this paper, we use a neural network learning technique that is closely related to a well-known statistical method to automatically generate image features for distinguishing targets from clutter. Extraction of these features from the images takes only local operations that can be parallelized to yield fast implementation times. We test our approach on two DSEG data sets: in the first, the targets are vehicles of military interest and, in the second, the targets are camouflage nets. We obtain robust image features that separate targets from clutter under a variety of conditions.

HEBBIAN LEARNING, PRINCIPAL COMPONENTS AND AUTOMATIC TARGET RECOGNITION

A. J. Katz
Central Research Laboratories
Texas Instruments Incorporated
Dallas, Texas 75265

Abstract

We apply a neural network generalized Hebbian learning algorithm, which is closely related to statistically-based principal component analysis, to the problem of automatic target recognition. Although our focus is on automatic target recognition, the described learning and classification methods extend to a variety of pattern recognition domains. We use the learning algorithm to generate features specific to a given target class and present results for two image sets. The resulting features provide more robust classification of target objects.

Introduction

Since the groundbreaking work of Hubel and Wiesel[1] and Marr[2], much attention has been focused on what features in the visual field are perceived by the visual system and how the receptive fields, neurons sensitive to particular features, are organized. Not only does this work provide important clues for understanding biological vision systems, it lays groundwork for artificial vision systems in such areas as robotics and automatic target recognition.

Recent work[3,4,5,6] in artificial neural networks suggests mechanisms and optimization strategies that explain the formation of receptive fields and their organization in mammalian vision systems. Linsker[5] has demonstrated how Hebbian learning algorithms, which change synaptic connections according to the degree of correlation between neuronal inputs and outputs, give rise to layers of center-surround and orientation-selective cells, even if the input to the initial layer is random white Gaussian noise.

Kammen and Yuille[6] show that orientation-selective receptive fields can also develop from a symmetry-breaking mechanism. Under certain conditions, the receptive fields perform a principal component analysis of the input data, as was first shown by Oja[7]. Since Hebbian learning occurs

in nature[8], the intimate connection between certain forms of Hebbian learning and the well-known statistical technique of principal components is compelling.

In this paper, we apply a generalized Hebbian learning algorithm (GHA) due to Sanger[9] to extract features for automatic target recognition from long-wavelength infrared (IR) and TV images. Sanger has proven that GHA determines the principal components of the data set in order of decreasing eigenvalue. Principal components or receptive features generated with GHA are very similar in appearance to those found in Linsker's work[5]. The novelty of our work lies in the use of GHA to generate a set of distinguishing target characteristics that separates targets from background clutter, where clutter is defined by a conventional screener algorithm to be non-target regions in the image that have target-like characteristics. We use only target statistics to generate the distinguishing features to enhance the robustness of the feature set: the background clutter can change significantly from one scene to the next, so using any particular set of clutter statistics to do feature generation can introduce unwanted biases[10]. We compute a signal-to-noise ratio to select which of the learned receptive fields to use for target recognition. When further discrimination is necessary, we generate a hierarchy of receptive fields over the spatial length scales covering the target and examine relationships between receptive fields at different length scales. We perform classification in terms of a binary tree structure, where the decision criteria are parameters defined by the eigenvalues associated with the principal components. We provide results from two data sets, which include examples of partially occluded targets, targets along tree lines, and targets very similar in appearance to background clutter.

Algorithm

We use the generalized Hebbian learning algorithm to train a one-layer neural network, where the input nodes define arrays of pixel intensity values from image data and the output nodes index the principal components. The form of the algorithm is[9]:

$$c_{ij}(\mathrm{T}+1) = c_{ij}(\mathrm{T}) + \gamma(\mathrm{T})[y_i(\mathrm{T})x_j(\mathrm{T}) - y_i(\mathrm{T})\sum_{k\le i} c_{kj}(\mathrm{T})y_k(\mathrm{T})] \qquad (1)$$

where c_{ij} is the weight or connection strength between the j^{th} input neuron and the i^{th} output neuron (c_{ij} is initially assigned random weights), x_j is the j^{th} component of the input vector, y_i is the i^{th} component of the output vector, and $\gamma(T)$ is a learning parameter that decreases with time such that

$$\lim_{T\to\infty} \gamma(T) = 0 \quad \textit{and} \quad \sum_{T=0}^{T=\infty} \gamma(T) = \infty$$

The second term on the rhs of equation 1 is the Hebbian term and the third term ensures that the algorithm learns successive eigenvectors (which are the principal components) of the covariance matrix of the input vectors ordered by decreasing eigenvalue. This decomposition of the covariance matrix in terms of eigenvectors is the well-known Karhunen-Loeve transform. Sanger[9] shows how equation 1 can be effected using only local operations; such a local implementation distinguishes equation 1 from other algorithms for computing the Karhunen-Loeve transform and underscores the importance of equation 1 for training neural networks. Sanger[9] applies equation 1 to image coding, texture segmentation, and the generation of receptive fields; other authors[11] have also used principal components to characterize image texture.

Experiments

We apply equation 1 to the development of receptive fields for identifying a specific target object. The extracted characteristics of the target object are embedded in its covariance statistics. Inputs to the network are $r \times s$ arrays of pixel values, which are rastered into $r \times s$ component vectors, from image subregions that contain the target of interest. The resulting principal components (extracted from the weight matrix c_{ij}) are directions in the $r \times s$ dimensional input space with maximum variance: these directions are the most informative ones in the sense that projections of input vectors along the principal component directions are maximally distinguishable. Eigenvalues corresponding to the principal components determined from equation 1 provide a measure of the variance in the principal component directions. Since vectors in the input space are made up of pixel intensities, the principal components correspond to prominent intensity patterns or features in the object of interest. We train on several examples of the target object to smooth out noise present in individual examples and to generate principal components that signify features common to different occurrences of the object.

The generated principal components are arrayed in $r \times s$ matrices to produce receptive fields or filters that are convolved with the original image data during classification. We convolve these filters in a way that preserves the spatial sampling of pixel intensities used to construct the input vectors. We multiply every $r \times s$ array of pixels contained in the image subregion of interest by the generated filters and then compute the variances of the resulting convolutions. Variances (these are related to the eigenvalues of the principal components) or ratios of the variances (these provide a measure of the relative content of two patterns) compose the parameter sets used for classification.

In our approach, we assume that range information is available, so that as we scan the image, we can properly adjust the size of the box or window circumscribing the subregion of interest to reflect the target size.

An important parameter in the generation of the receptive fields is the spatial sampling density that enters into the construction of the input vectors. This parameter corresponds to the synaptic connection density in Linsker's Hebbian algorithm[5], which only yields principal components if the connection density is held fixed. In our approach, the spatial sampling density determines the scale of the feature. Since the target object occurs at different ranges in the image data (note that range information is provided in our data sets), the spatial sampling density must be appropriately scaled to ensure that the same feature scale is measured in all cases. As we reduce the spatial sampling density for targets at nearer ranges, we average over the shorter length scales to avoid aliasing effects, though we found averaging had little effect on results from test cases we examined.

To illustrate how scaling of the spatial sampling density is done, we assume that the target at the longest range fits into a $u \times v$ pixel box. Input vectors for this case are formed from intensity values of $r \times s$ blocks of pixels (where these blocks are smaller than the box size) extracted from the box circumscribing the target object. For targets at half the initial range, we compose input vectors from $2r \times 2s$ blocks of pixels, where we extract the intensity value from every second pixel. We continue in this fashion as we move to closer ranges.

We also scale the spatial sampling density for a given target sample to generate a hierarchy of receptive fields at different scales. The relevant scales are set by the smallest scale detectable (effectively the resolution) for the target seen at longest range and the size of the target object. This hierarchy characterizes the target object in terms of what features become relevant at different length scales. For self-similar objects, we expect to find an invariant feature set as a function of scale. There are similarities here to renormalization group analysis[12] where system behavior is governed by how the physical operators scale.

Data and Results

The data sets studied are characterized in Table 1; the targets were objects of military interest. Images in data set I were TV whereas those in data set II were long-wavelength IR. Target objects for both data sets were of several different types[10], so it was important to find receptive fields common to all types. Different orientations of the target objects in these data sets did not generate problems (though, in general, they could): targets from data set I were positioned in the field-of-

view at long enough range that there was little sensitivity to orientation, and those from data set II were sufficiently spherically symmetric to ignore orientation effects. No preprocessing of the images was done except to normalize linearly the pixel intensities so they fell in the range from 0 to 255.

Clutter objects for both data sets were defined in terms of a conventional screener algorithm: any region in the image passed by the screener and not a target fell into the clutter class. With this definition of clutter, we measured the capability of the principal component features to distinguish targets from objects similar in appearance.

For data set I, we generated 3×3 pixel filters using five targets contained in a single image and tested filter performance on 9 additional images; the targets in the training image represented different types of the target class (See Figure 1). Target heights in data set I ranged from 4 pixels to 32 pixels . The first three learned filters, which are displayed in Figure 2a, are intuitively plausible. The first filter emphasizes regions that have strong grayshade contrast with the background environment: most of the targets show strong contrast with background. The second and third features highlight regions with strong horizontal and vertical grayshade gradients, respectively; target regions all have sharp transitions in grayshade, from pixels within the targets to pixels outside the targets.

Figure 2b shows the separability of target and clutter objects based on variance values derived from the third filter in Figure 2a. Figure 2b indicates that most of the background clutter has much smaller vertical gradient content than the targets: tree lines and roads and horizons in the image set extend mostly parallel to the horizontal edges of the image and provide little vertical gradient content. We also examined the same images with a standard in-house feature set, which is used for automatic target recognition. We found that between 8 and 16 of these features were required to achieve the same level of separability found in Figure 2b and that the discriminant surface, which divided targets and clutter, was highly nonlinear.

We generated eight 5×5 pixel filters (shown in Figure 3) for target set II from four examples in a single image and tested filter performance on 23 images. Target heights ranged from 15 pixels in the far range to 175 pixels in the near range. Data set II was in several ways more challenging than the first data set because the noise level was higher and the target texture was very similar to the texture of the clutter. We ranked the filters by a signal-to-noise ratio S/N (see Figure 3), where S is the mean of the variances from the four examples in the training set and N is the variance of the variances from the same training examples. This criterion is a natural one, since we seek target

characteristics that are both prominent and invariant over the data set, and that, therefore, lead to large values of S/N. Filter 5 had the largest S/N ratio among the eight generated and alone provides considerable separation of targets and clutter (see Figure 4). In Figure 4, we show a lower cutoff for the target region: the training examples only establish a lower bound on the target region in the feature space since these examples were particularly noisy ones and gave smaller variance values than more prototypical examples. Choosing noisy examples (and, therefore, examples more easily confused with clutter) for training is important for purposes of classification and generalization, because we can better estimate the true position (as determined from an infinitely large data set) of the classification discriminant surface from examples that lie close to the boundary between target and clutter than from examples that lie further away[13].

To achieve further separability of targets from clutter, we also examined ratios of the variance values from the filters: the ratio of the variance outputs from filters 4 and 5 in Figure 3 further distinguish targets and clutter (as seen from Figure 5a). In addition, we used the training image to generate a second set of eight 5×5 pixel filters at a larger length scale by halving the spatial sampling density of the input vectors to the neural network. The filter with the second largest S/N ratio was the same as filter 5 in Figure 3; this filter further reduces the number of clutter points mistakenly identified as targets (Figure 5b). This points to some self-similarity in the target characteristics (though a factor of two in length scales is not sufficient to establish any sort of fractal[4] characteristics) and indicates the persistence of a particular feature over a factor of two in length scales. Together the three filters and their corresponding variance outputs establish a binary classification tree for distinguishing targets from background clutter. The final probability of detection (ratio of the number of targets detected to total number of targets) was 0.89 whereas the false alarm rate (ratio of the number of clutter points mistakenly classified as targets to total number of clutter points) was 0.07.

Conclusions

Connections between biologically motivated learning algorithms and statistical classification techniques provide important clues for generating features for artificial vision applications and for a variety of pattern recognition applications where feature extraction plays an important role. Tailoring the features (in the form of principal component filters) to reflect the characteristics of a specific target or target class leads to separability of the targets based on a relatively small number of features. Reducing the size of the feature space reduces the size of the training set required for adequate generalization[15]. Using only target training examples for feature generation removes biases that arise from nonrepresentative clutter training sets. Signal-to-noise ratios aid in

identifying features that are more robust over the training data. Hierarchies of feature filters that cover the relevant length scales in the image set can provide further discrimination of the object classes and indicate any scale invariant properties of the objects. Finally, carrying out filter convolutions is a local operation and can be parallelized over the image to yield fast implementation times.

Acknowledgement

We thank Bill Boyd, Jerry Herstein, and Rod Pickens for providing the image data.

References

1. Hubel, D.J. and Wiesel, T.N. *J. Physiol.* **160**, 106-154 (1962).
2. Marr, D. *Vision* (W.H. Freeman and Co. San Francisco, 1982).
3. Miller, K.D., Keller, J.B. and Stryker, M.P. *Science* **245**, 605-615 (1989).
4. Durbin, R. and Mitchison, G. *Nature* **343**, 644-647 (1990).
5. Linsker, R. *Computer* **21**, 105-117 (1988).
6. Kammen, D.M. and Yuille, A.L. *Biol. Cybern.* **59**, 23-31 (1988).
7. Oja, E. J. *Math. and Biol.* **15**, 267-273 (1982).
8. Brown, T.H., Kairiss, E.W. and Keenan, C.L. *Annu. Rev. Neurosci.* **13**, 475-511 (1990).
9. Sanger, T.D. *Neural Networks* **2**, 459-473 (1989).
10. Katz, A.J., Gately, M.T., and Collins, D.R. *Neural Computation* **2**, 472-479 (1990).
11. Ade, F. *Signal Processing* **5**, 451-457 (1983).
12. Wilson, K.G. *Rev. Mod. Phys.* **47**, 773-840 (1975).
13. Katz, A.J., Collins, D.R., and Lugowski, M. *Complex Systems* **3**, 185-207 (1989).
14. Mandelbrot, B.B. *The Fractal Geometry of Nature* (W.H. Freeman and Co. San Francisco, 1982).
15. Denker, J. et al. *Complex Systems* **1**, 877-922 (1987).

Figure Captions

Table 1. Description of two data sets discussed in text.

Figure 1. TV image from data set I. This image contains the five targets (marked with an "x") that were used used for training.

Figure 2a. 3 x 3 pixel filters determined by learning with the generalized Hebbian algorithm. These filters are the first three principal components for the training target regions from data set I. Dark gray shade corresponds to positive numbers, white corresponds to negative numbers, and lighter gray shade corresponds to numbers near zero.

Figure 2b. Abscissa values V_3 are the variances derived from Filter 3 in Figure 2a. The left-hand curve is the probability for a clutter region to have a variance V_c larger than variance V_o; the right-hand curve is the probability for a target region to have a variance V_t smaller than variance V_o. The dotted line indicates the lower-bound on V_t determined from the training data. Observe that the dotted line lies close to the point where the two probability curves cross (which marks the demarcation point in a Bayes classification approach).

Figure 3. 5 x 5 pixel filters derived from training data for data set II. Filters represent top eight principal components and are ordered by decreasing eigenvalue. Only filters 4 and 5 were used for classification. S/N denotes the signal-to-noise ratio defined in the text. Dark gray shades correspond to positive numbers, white corresponds to numbers near zero, and lighter gray shades correspond to negative numbers.

Figure 4. V_5 denotes variances derived from Filter 5 in Figure 3. The curves are defined in the figure caption for Figure 2b. The dotted line is the lower bound on V_t determined from the training data. With this lower bound, the probability of detection, P_d, is 0.89 and the false alarm rate, P_{fa}, is 0.27.

Figure 5a. V_4/V_5 is the ratio of variances derived from Filters 4 and 5 in Figure 3. The curves are defined in the figure caption for Figure 2b, except V_c, V_t, and V_o here indicate the ratio of variances from Filters 4 and 5. The dotted line is determined from training data. The clutter curve uses only false alarms from the results in Figure 4 whereas

the target curve uses all of the target points. Using the dotted line to discriminate targets and clutter (together with results from Figure 4), we obtain a detection probability of 0.89 and a false alarm rate of 0.10.

Figure 5b. V_4 is the variance derived from Filter 4 in a second set of 5 x 5 pixel filters (not shown but described in text). The clutter curve uses only false alarms from the results in Figure 5a; all target points, however, are used. The dotted line is the discriminator between targets and clutter set by the training data. If we use results from Figures 4, 5a, and 5b in sequence, we obtain a detection probability of 0.89 and a false alarm rate of 0.07. The three discriminators can be viewed as forming a three-level binary classification tree.

Image Set	Number of Images	Location	Scene Description
1	4	G	a
2	2	G	b
3	1	S	c
4	1	S	d
5	2	N	e

a. Fields and Trees; Urban Clutter; Open and Partially Occluded Targets
b. Fields; Urban Clutter; Targets in Open
c. Fields; Urban Clutter; Targets in Open
d. Heavy Forests; Partially Occluded Targets
e. Fields and Heavy Forests; Partially Occluded Targets

Data Set I

Image Set	Number of Images	Location	Scene Description
1	21	S	Fields With Clusters of Trees; Urban Clutter; Open and Partially Occluded Targets.
2	3	N	Heavy Vegetation; Rural Clutter; Open targets.

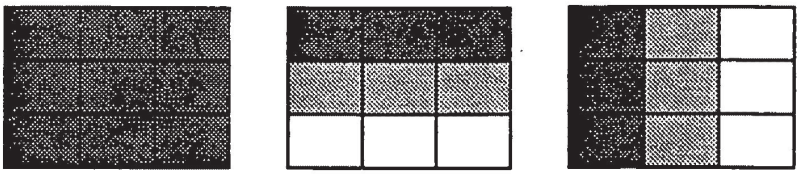
Data Set II

04650

Table I

Figure 1

04656



Filter 1 Filter 2 Filter 3

Figure 2a

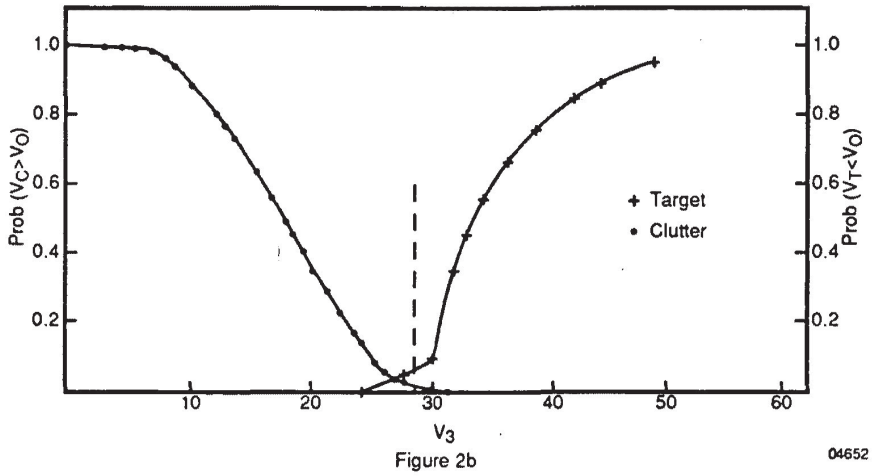
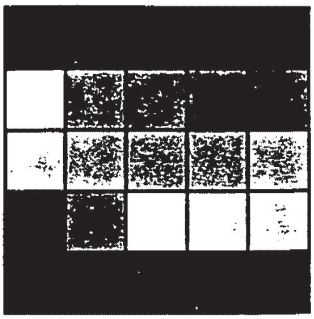


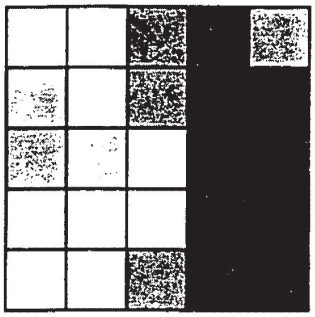
Figure 2b

Filter 1 s/n = 1

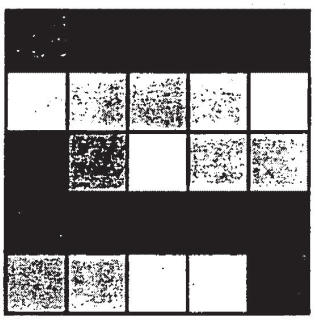
Filter 2 s/n = 7



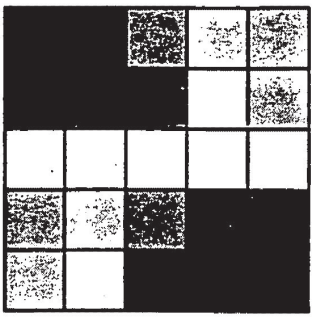
Filter 3 s/n = 7



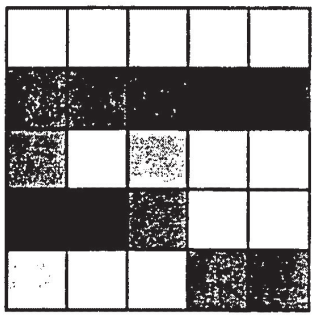
Filter 4 s/n = 5



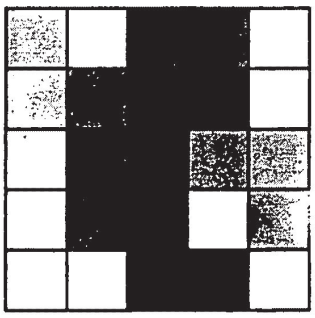
Filter 5 s/n = 23



Filter 6 s/n = 8



Filter 7 s/n = 13



Filter 8 s/n = 9

Figure 3

04651

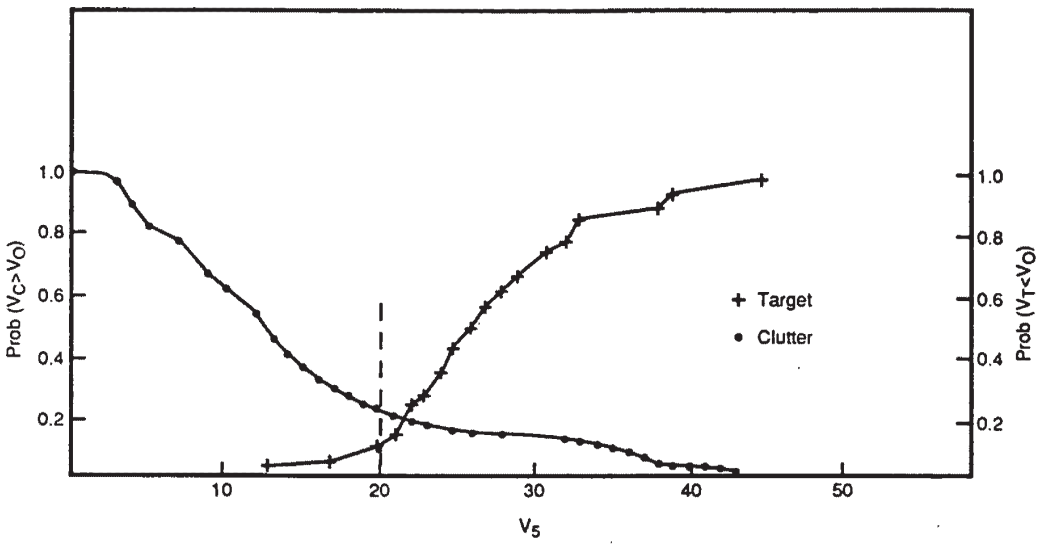


Figure 4

04653

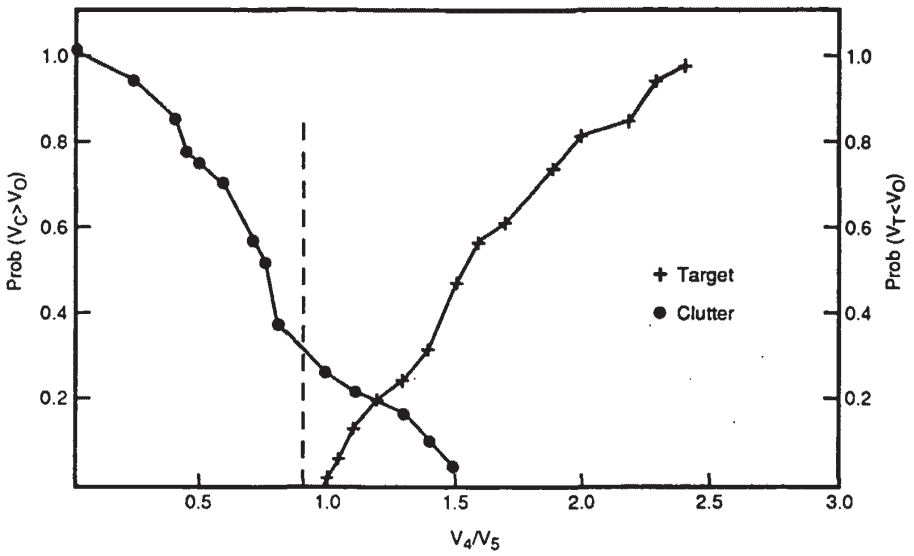


Figure 5a

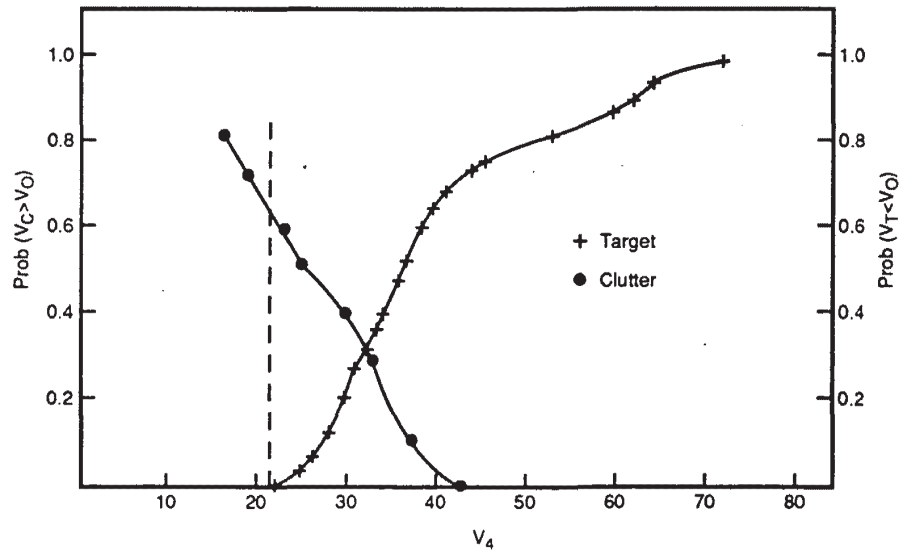


Figure 5b

04655