Computational models for search and discrimination

Anthony C. Copeland
Mohan M. Trivedi
University of California, San Diego
Computer Vision and Robotics Research (CVRR) Laboratory
Electrical and Computer Engineering Department
La Jolla, California 92093-0407
E-mail: copeland@swiftlet.ucsd.edu
E-mail: mtrivedi@ucsd.edu

Abstract. We present an experimental framework for evaluating metrics for the search and discrimination of a natural texture pattern from its background. Such metrics could help identify preattentive cues and underlying models of search and discrimination, and evaluate and design camouflage patterns and automatic target recognition systems. Human observers were asked to view image stimuli consisting of various target patterns embedded within various background patterns. These psychophysical experiments provided a quantitative basis for comparison of human judgments to the computed values of target distinctness metrics. Two different experimental methodologies were utilized. The first methodology consisted of paired comparisons of a set of stimuli containing targets in a fixed location known to the observers. The observers were asked to judge the relative target distinctness for each pair of stimuli. The second methodology involved stimuli in which the targets were placed in random locations unknown to the observer. The observers were asked to search each image scene and identify suspected target locations. Using a prototype eye tracking testbed, the integrated testbed for eye movement studies, the observers' fixation points during the experiment were recorded and analyzed. For both experiments, the level of correlation with the psychophysical data was used as the basis for evaluating target distinctness metrics. Overall, of the set of target distinctness metrics considered, a metric based on a model of image texture was the most strongly correlated with the psychophysical data. [DOI: 10.1117/1.1390297]

Subject terms: target detection; human visual search; discrimination; eye tracking; target signature metrics; image texture.

Paper ATA-14 received Nov. 22, 2000; revised manuscript received Mar. 1, 2001; accepted for publication Mar. 23, 2001.

1 Introduction

The issue of the development and assessment of useful computational models and quantitative metrics for integrated search and discrimination tasks is discussed. The approach is experimental in nature, where psychophysical data provide the guidance and support for comparative assessment of various metrics. This research is performed in the overall context of search and detection of camouflaged targets in natural scenes. Figure 1 illustrates an example of such a scene, where a human observer or a machine vision system may be required to look for and detect military targets, such as a tank. This scenario is quite general. The associated problems provide a number of interesting research issues in computational vision. For example, what is the underlying model for integrated search and discrimination? What preattentive cues affect search or discrimination? How can we evaluate the relative ease or difficulty of an observer attempting to locate a selected camouflage pattern in a natural scene? How can we design the most effective camouflage pattern for a naturally textured scene? How can we rank the capabilities of automatic target recognition systems in relative terms?

We describe our efforts directed toward the resolution of these kinds of questions. We restrict our investigation to only textured patterns and static images.

Issues related to color, range (or depth), and motion, important as they definitely are, cannot be examined in the limits of the scope of our research. We do believe that the overall experimental framework will be of utility and value for studies involving other cues.

The ultimate goal of this line of research is the development of a robust and quantitative means for characterizing the signature strength of a target in a sensed image. The signature strength measurement should be closely correlated to the ease or difficulty of a human observer attempting to detect it.[1] In this context, the signature strength of a target is equivalent to the distinctness of the image pattern representing the target from the pattern of its specific background. Metrics that are successful at measuring perceived target distinctness would be a key component of a computational model of human visual target acquisition.[2] Such a model could form the basis of an automatic target recognition system for autonomous robot sensing or military weapons applications.[3] It could also serve to improve the assessment of military camouflage patterns and the development of more effective ones.[4]

For the purpose of defining the scope of this research, we consider human target acquisition to involve target detection followed by target recognition. The detection task is that which establishes the existence and location of an ob-

Fig. 1 An illustration of camouflaged targets in a natural scene.

ject. Recognition is the task of determining the characteristics of the object that indicate its identity, such as its size, shape, etc. Further, we consider target detection to consist of the combination of the individual tasks of search and discrimination. Search is the process of locating areas of a scene in which to direct our attention. Discrimination is the process of segregating a potential object from its immediate background. This approach is very similar to the conclusion of O'Kane, Walters, and O'Agostino.[5] We are concerned with the target detection task, comprising search and discrimination, without considering recognition.

We conducted two different types of psychophysical experiments to generate quantitative measurements of perceived target distinctness for comparison to various target distinctness metrics. The first type of experiment involved paired comparisons of image stimuli that contain a target pattern embedded in a background pattern, in a constant location known to the observers. The patterns consisted of various textures extracted from images of natural scenes. For every stimulus, the target field consisted of a square shape of a constant size. We say that this experiment is a study of pure discrimination, since there is no search or recognition involved. For each pair of stimuli, the observer was required to select which of the pair possesses a target that is more distinct. By combining the decisions from a number of observers, it was possible to estimate numerical scale values for the relative levels of perceived target distinctness in the stimuli. These psychological scale values were compared to the computed values of target distinctness metrics. The second type of experiment utilized image stimuli that contain several target patterns embedded in a background scene, in random locations unknown to the observers. In this experiment, the observer needed to perform both search and discrimination. As the observer searched the scene for targets, his eye fixation points were determined by processing video of the observer's eye. The fixation point data from several observers were used to compute various statistics for each target, indicating how easily the observers located it, including the likelihood the target was fixated or identified and the time required to do so. These computed statistics served as another quantitative basis for evaluating the relative effectiveness of target distinctness metrics at representing perceived target distinctness.

2 Target Distinctness Metrics

In some previous experiments,[6–9] we observed three major perceptual cues that humans tend to utilize in judging target distinctness. These cues can roughly be called contrast, texture differences, and boundary strength. There are certainly many other possible perceptual cues, but these three seem to be the strongest. In this section, we discuss some specific metrics that attempt to measure the strengths of these three perceptual cues for a particular target and its local background.

2.1 *Measuring Contrast*

Contrast is typically measured with first-order metrics that can be computed solely from the histograms of the target and local background fields.[5] A histogram is considered a first-order probability distribution, since it can be calculated by considering the gray levels of pixels individually (one at a time). Statistics calculated from a histogram are capable of characterizing the overall brightness and variance of the patterns. Probably the earliest target distinctness metric is the area-weighted average ΔT,[10] which is simply the difference between μ_t and μ_b, the computed mean gray levels of the target and background fields:

$$\Delta T = |\mu_t - \mu_b|.$$

The Doyle ΔT[5] incorporates the computed standard deviations of the target and background fields, σ_t and σ_b:

$$\text{Doyle} = [(\mu_t - \mu_b)^2 + (\sigma_t - \sigma_b)^2]^{1/2}.$$

Effective pixels on target (Eff_POT) is computed as the number of pixels in the target pattern that have a gray level that differs from the mean gray level of the local background pattern by more than two standard deviations of the background histogram. This metric has shown promise, especially when combined with the Doyle.[5]

2.2 *Measuring Texture Differences*

The texture cue has been successfully measured with second-order metrics, ones computed from the gray level cooccurrence (GLC) probability distributions of the target and the background.[7,11,12] After Julesz made the important conjecture about the role of second-order statistics in human texture discrimination, GLC models have found many useful applications in machine vision.[13] In several studies to compare the relative power of various texture analysis techniques to perform texture discrimination, GLC matrices generally outperformed other methods.[14–16] GLCs have also been used for object detection,[17] scene analysis,[18] as well as texture synthesis.[19–21] Other studies have demonstrated the wealth of texture information contained within GLCs.[22–24]

A GLC probability distribution is calculated by considering the gray levels of pixels in pairs (two at a time), capturing information about the spatial relationships between pixels. As such, GLC probabilities are often used as a model of image texture. One second-order metric that has shown great promise is average cooccurrence error (ACE).[7] It is defined as

$$\mathrm{ACE}=\frac{1}{\tau_{\mathrm{NGLC}}}\sum_{\Delta\in D}\sum_{i=0}^{G-1}\sum_{J=0}^{G-1}|P_t(i,j|\Delta)-P_b(i,j|\Delta)|,$$

where τ_{NGLC} is the total number of displacement vectors in the set D of vectors in the texture model, G is the number of possible gray levels, $P_t(i,j|\Delta)$ is the joint probability of a pixel of gray level i and a pixel of gray level j given the displacement vector $\Delta=[\Delta_x\Delta_y]$ for the target pattern, and $P_b(i,j|\Delta)$ is the corresponding joint probability for the background pattern, For computing this metric, we normally consider all possible displacements of up to a maximum of $\tau_{NX}=\tau_{NY}=8$ pixels, yielding a total of $\tau_{\mathrm{NGLC}}=2\tau_{NX}\tau_{NY}+\tau_{NX}+\tau_{NY}=144$ displacements. If the original image is quantized to 256 gray levels, the pixel values in the target and background regions are reduced to $G=8$ possible gray levels for computation of the model. Since each of the 144 GLC matrices in the texture models is of size $G\times G$, using a full $G=256$ gray levels produces a data structure that is prohibitively large.

2.3 *Measuring Boundary Strength*

The third class of target distinctness metrics we considered consisting of metrics are those that attempt to quantify target/background boundary strength. Even if a target's texture pattern is very similar to the texture of its local background, discontinuities along the target/background boundary can still serve as a perceptual cue.[25] One way to measure this is to compute the average contrast between the pixels lying on either side of the target/background boundary. For a single point i along a boundary, the contrast is

$$c(i)=|p_t(i)-p_b(i)|,$$

where $p_t(i)$ is the gray level of the pixel just on the target side of the boundary and $p_b(i)$ is the gray level of the adjacent pixel just on the background side. For a target field that is a rectangular lattice of pixels, the lengths of the boundaries are $n_{\mathrm{top}}=n_{\mathrm{bottom}}=n_{\mathrm{horiz}}$ and $n_{\mathrm{left}}=n_{\mathrm{right}}\equiv n_{\mathrm{vert}}$. The average contrast for one boundary (such as the top boundary) is

$$C_{\mathrm{top}}=\frac{1}{n_{\mathrm{horiz}}}\sum_{i=1}^{n_{\mathrm{horiz}}}c(i),$$

where i is just a summation index for the boundary points. Then the average boundary strength (ABS) for the whole target is:

$$\mathrm{ABS}=\frac{n_{\mathrm{horiz}}(C_{\mathrm{top}}+C_{\mathrm{bottomn}})+n_{\mathrm{vert}}(C_{\mathrm{left}}+C_{\mathrm{right}})}{2n_{\mathrm{horiz}}+2n_{\mathrm{vert}}}.$$

For a target field that is a perfect square, such as in our stimulus images, we have $n_{\mathrm{horiz}}=n_{\mathrm{vert}}\equiv n_{\mathrm{bound}}$. In this case, the equation for ABS reduces to

Fig. 2 A test subject studying a displayed image scene while ITEMS tracks his eye fixation points.

$$\mathrm{ABS}=\frac{1}{4n_{\mathrm{bound}}}n_{\mathrm{bound}}(C_{\mathrm{top}}+C_{\mathrm{bottom}}+C_{\mathrm{left}}+C_{\mathrm{right}})$$

$$=\frac{1}{4}(C_{\mathrm{top}}+C_{\mathrm{bottom}}+C_{\mathrm{left}}+C_{\mathrm{right}}).$$

The ABS measure does not take into account the values of any pixels that do not lie adjacent to the target/background boundary. However, a target/background boundary that has a high value for ABS may not be very distinct if it is embedded in a region that already is characterized by a large amount of contrast. To take into account the contrast of the entire region, we use relative average boundary strength (RABS):

$$\mathrm{RABS}=\frac{\mathrm{ABS}}{\frac{1}{n_{\mathrm{region}}}\sum_{i=1}^{n_{\mathrm{region}}}c(i)},$$

where n_{region} is the number of adjacent (either vertically or horizontally) pixel pairs within the target field or in the background near the target. Essentially, RABS is the ratio of the average contrast along the target/background boundary to the average contrast between adjacent pixels in the vicinity.

3 ITEMS Testbed

3.1 *Overview and Utility of ITEMS*

This section discusses the design and implementation of the integrated testbed for eye movement studies (ITEMS). This prototype eye tracking testbed consists of an integrated system of hardware and software, which allows an experimenter to present an observer with an image displayed on a high resolution monitor and have the observer perform a visual task. Figure 2 shows a test subject studying a displayed image scene while ITEMS tracks his eye fixation points. Using ITEMS, not only can we determine whether a particular target was identified by an observer, but also whether the target was ever fixated by the observer (even if it was not identified as being a target), how long did it take before the target was first fixated, how long the target was

studied before it was identified, what search path the observer took on the way to the target, and any number of other aspects of visual search.

The hardware components of ITEMS are a Silicon Graphics Indy computer workstation with high resolution color monitor, a Sony CCD black and white video camera fitted with a 50-mm lens and 5-mm lens spacer, a Datacube MaxTD image processing system containing a MaxVideo~200 pipeline processor and MVME-167 CPU system controller, and an adaptable yet sturdy apparatus to which the camera is mounted as well as a helmet for restricting observer head movements. The software components of ITEMS include an X-Windows application to handle the image scene display and observer response registration for the Indy workstation, pupil centroid tracking and registration for the Datacube MaxTD, a utility for fixation point estimation and head movement adjustments, a utility for spatial calibration and error interpolation, and another for calculation of target fixation and identification statistics.

All required images are created by the experimenter beforehand. This includes a zero point image, several calibration images, and all desired experiment image scenes. The procedure is that the zero point image is displayed first, then all calibration images in succession, the zero point image again, and then any number of sessions of experiment images, each pair of sessions separated by another presentation of the zero point image. The number of experiment sessions and number of images in each session can vary, but it has been found that five images per session with one calibration session and two experiment sessions results in a moderate ten minutes of data collection for each observer.

The zero point image is an image with one target that is located such that it is directly ahead of the observer's left eye when displayed on the monitor. The target consists of a square region of uniform gray level against a background of a different gray level. This image is used to establish a reference point to which all eye movements can be related and also to measure periodically the change in fixation point estimates that is the result of small head movements accumulating over time. This procedure is described later in Sec. 4.5.

Each calibration image consists of a row of square targets. The targets in all the calibration images taken together constitute an array of evenly spaced points, which are used as sample points at which to measure the error in fixation point estimates due to measurement and modeling error. As these errors will vary over different spatial locations in the display image, a number of samples are taken and then adjustments are made in fixation point estimates from an interpolation of the calibration samples from the vicinity of each estimate.

3.2 *ITEMS Hardware Configuration*

Figure 3 shows the interconnectedness of the various hardware components of ITEMS. Briefly, the Silicon Graphics Indy workstation is used to load each image scene stimulus from disk and display it on its high-resolution color monitor. The Sony CCD video camera sends continuous video to the Datacube MaxTD image processor, which locates, tracks, and records the pupil centroid location of the observer's left eye. The observer's head movement is restricted using a baseball-batting helmet, which is rigidly mounted to the table, using adjustable aluminum extrusion material. This material allows the helmet to be raised or lowered to accommodate different observers and also to be locked into place when the appropriate position is found. Head movement is further restricted by a chin rest.

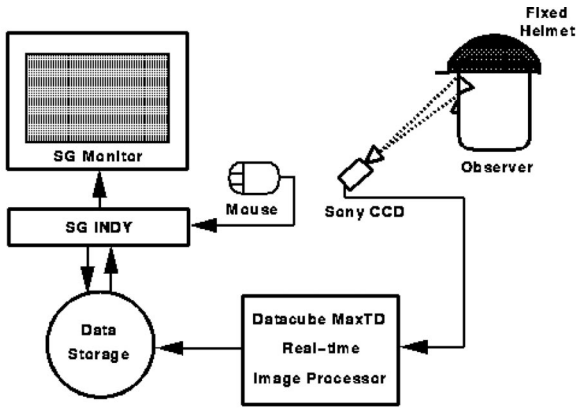


Fig. 3 The ITEMS system components.

The camera is mounted directly in front of and below the observer, just below the Silicon Graphics monitor, looking upward at the observer's left eye. This location was found to provide an adequate image of the observer's left eye and a small reference mark affixed just below the eye. This reference mark is a small, glossy black paper circle, used to distinguish eye movements from small head movements. The observer's face is illuminated with a small portable flashlight as necessary to segregate the pupil and reference mark together from the rest of the video image. The Datacube MaxTD also has a small terminal screen, which allows the experimenter to monitor the status of the image processor's eye tracking, and the video from the CCD camera is simultaneously displayed on a small monitor for the same purpose.

3.3 *Image Scene Display and Observer Response Registration*

Image scene display and observer target identification response registration for ITEMS is handled by the Silicon Graphics Indy workstation. The X-Windows application created for this purpose is called I_SPY. I_SPY is used to load each image scene stimulus from disk and display it on the high-resolution color monitor. In experiment mode, the observer uses mouse buttons to indicate when to display each image, when he wishes to identify a suspected target, and when he is finished searching a particular scene. In playback mode, I_SPY allows the experimenter to study the data by displaying the image scene stimuli with a cursor that moves about the images indicating the observer's fixation points over time.

3.4 *Pupil Centroid Tracking and Registration*

The tracking and registration of the pupil centroid position is handled by the Datacube MaxTD image processing system. The procedure is to first threshold the video frame, such that both the observer's pupil and the black paper

circle affixed just below his eye appear as black circular blobs in the image. The resulting binary image is then subjected to a connectivity analysis, which computes the number of blobs in the image and a roundness measure for each. The roundness measure is computed by finding a best-fit ellipse for each blob, and calculating the ratio of the two axes of the ellipse. The roundness measure is used to separate the pupil and reference mark blobs from various shadow artifacts, which generally do not appear as round blobs at all. The values that are stored are the centroid differences in both x and y coordinates between the upper blob (the pupil) and the lower blob (the reference mark), along with the current timestamp. Thus it is only movement of the pupil relative to the reference mark that is tracked and registered. In this way, movements of the eye can be distinguished from small head movements. That is, a small head movement will result in a change of position of both the pupil and the reference mark in the camera image. Although a helmet mounted in a fixed position and a chin rest are used to restrict observer head movement, in practice there is still a bit of a small head movement even with the most cooperative observers, due to breathing, heartbeats, etc.

3.5 *Eye Tracking Geometry and Fixation Point Estimation*

Details of the fixation point estimation process are given in Ref. 26. Briefly, the steps necessary to obtain the fixation estimate for each data sample are as follows.

1. Extract the values for the difference in x and y coordinates between the pupil centroid and the reference point centroid from the data file of the pupil centroid tracking program.
2. Compare these values to the same values from the moment the observer identified the first zero point. The change is taken to be the movement of the pupil center in the camera image from the zero state.
3. From the location of the pupil center in the camera image, find its location in world coordinates using the inverse perspective transform,[27] subject to the constraint that the point is known to lie on the front side of the sphere representing the eyeball.
4. Based on the location of the pupil center in world coordinates, find the intersection point of the line representing the visual axis and the plane representing the display image.
5. Find the fixation point estimate by converting the location of the intersection point from world coordinates to display image coordinates (x' and y').
6. Adjust the fixation point estimate for small head movements by subtracting the average of the error for the zero point at the beginning of the session and the one at the end of the session. For each zero point, the error is taken to be the change in fixation point estimate since the first zero point image at the beginning of the calibration session.

4 Studying Pure Discrimination

This section describes a psychophysical experiment designed to investigate the task of human target discrimination separate from visual search, or pure discrimination. The image stimuli used in this experiment consisted of target patterns embedded in background patterns, in a constant location known to the observers. With such stimuli, it is unreasonable to ask observers to make absolute judgments of target distinctness because of the complex nature and wide range of criteria that could be used in such a judgment. Instead, we only asked the observers to make relative judgments of target distinctness. The image stimuli were presented in pairs, and the observers were required to select which image of each pair possesses a target that is more distinct. By combining the decisions from a number of observers, it is possible to estimate numerical scale values for the relative levels of perceived target distinctness in the stimuli. These psychological scale values were used as a quantitative basis for evaluating the relative effectiveness of our target distinctness metrics to represent perceived target distinctness. The established method for accomplishing this psychological scaling is the law of comparative judgment (LCJ), introduced by Thurstone.[28,29] The LCJ is based on the postulate that if a stimulus is presented to a human subject, it excites a discriminal process, which has some value on the psychological continuum. It is also assumed that this value will not be exactly the same each time the same stimulus is presented, but rather these values will form a normal distribution along the continuum. For more information about the specific method to estimate the scale values, see Ref. 7.

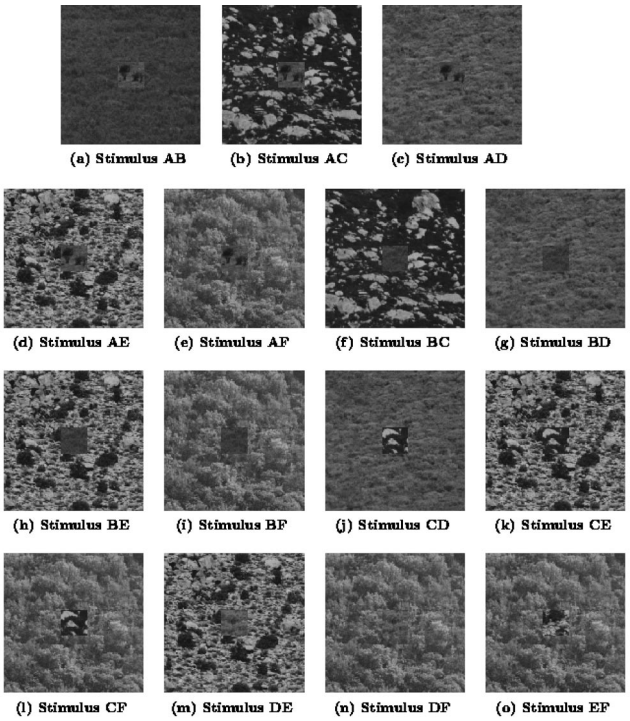


Fig. 4 The 15 256×256 test images for the discrimination experiment.

The 15 image stimuli used in the experiment are shown in Fig. 4. The computer environment that was developed to automate the sequential display of the image stimulus pairs and the registration and recording of subject responses is the X-Windows perceptual experiment tested (XPET).[6,7]

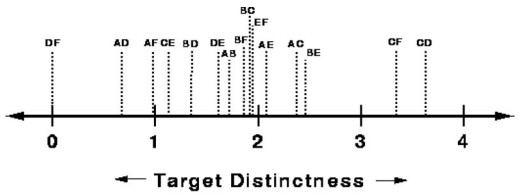


Fig. 5 The relative locations of the scale values along the perceptual continuum representing target distinctness.

XPET was used to present 20 observers with all 105 possible pairs of the 15 stimuli. The raw judgments were used to estimate an appropriate scale value for each stimulus. Figure 5 shows graphically the locations of the scale values along the perceptual continuum representing target distinctness. These scale values indicate only relative amounts of target distinctness in the stimuli as judged by the observers, and have no absolute meaning. The stimulus containing the target judged least distinct was stimulus DF. This stimulus is assigned a scale value of zero, and the scale is constructed upward from that point. The stimuli containing the most distinct targets, as judged by the observers, were stimuli CF and CD. The sample correlation coefficient was then computed between the vector of psychological scale values and the vector of each of the computed target distinctness metrics. The results are given in Table 1. Figure 6 shows the test images plotted with their LCJ scales and computed values for the ACE metric.

Table 1 The sample correlation coefficients (r) between the vector of stimulus scale values for perceived target distinctness and the vector of each of the target distinctness metrics.

Metric	r
ΔT	0.14
Doyle	0.66
Eff_POT	0.57
ACE	0.83
ABS	0.65
RABS	0.76

4.1 *Multivariable Linear Regression and Multiple Correlation*

We now compare the psychological scale values to not one, but several variables. The single variable linear regression model is of the form

$$y=\beta_0+x\beta_1+\varepsilon,$$

where y is the response (dependent) variable, x is the independent variable, β_0 and β_1 are regression parameters, and ε is the error that is presumed to be normally distributed with mean of $\mu=0$ and variance of σ^2. Previously, y represented the stimulus scale value estimated from the sychophysical data, and x represented any one of the image metrics that we were studying. With N stimuli in the experiment, we actually have N samples of both y and x, so the entire model is written

$$\mathbf{y}=\beta_0+\mathbf{x}\beta_1+\varepsilon,$$

where $\mathbf{y}'=(y_1,\ldots,y_N)$ represents the N scale values, $\mathbf{x}'=(x_1,\ldots,x_N)$ represents the N computed values of the particular image metric, and $\boldsymbol{\varepsilon}'=(\varepsilon_1,\ldots,\varepsilon_N)$ represents the error for each sample. We actually have k independent variables (image metrics) interacting simultaneously. Now, the model can be written

$$\mathbf{y}=\mathbf{X}'\boldsymbol{\beta}=\varepsilon,$$

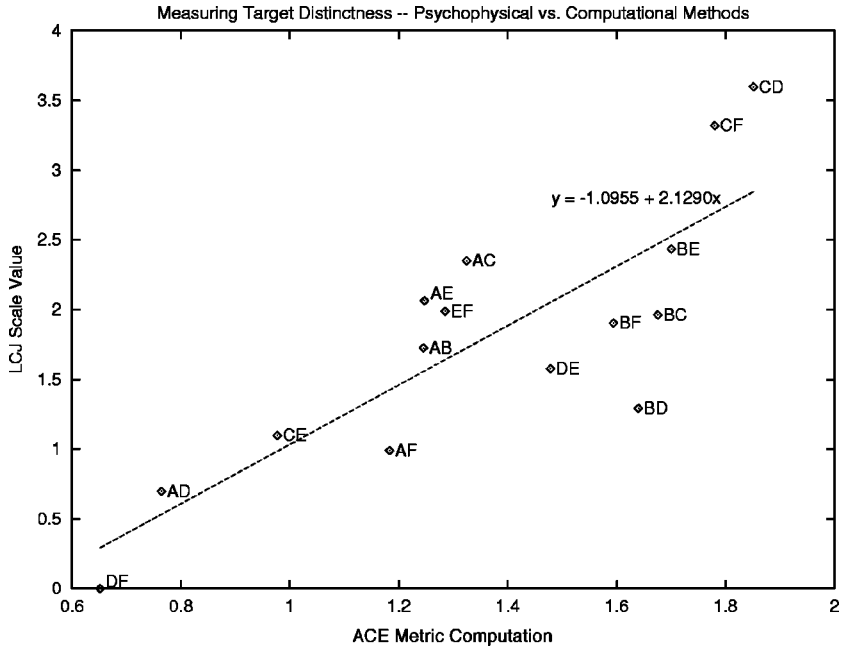


Fig. 6 The test images plotted with their LCJ scales versus computed values for the ACE metric.

Table 2 The multiple correlation coefficients for selected pairs of metrics.

	ΔT	Doyle	Eff_POT	ACE	ABS	RABS
ΔT	-	0.72	0.59	0.83	0.65	0.78
Doyle	-	-	0.90	0.83	0.75	0.89
Eff_POT	-	-	-	0.88	0.80	0.78
ACE	-	-	-	-	0.88	0.87
ABS	-	-	-	-	-	0.80
RABS	-	-	-	-	-	-

where the differences are that $\boldsymbol{\beta}$ is a $k+1$ length vector of regression parameters and $\mathbf{X}$ is a rectangular matrix of computed image metrics with $k+1$ rows and N columns. (Actually, the first row of $\mathbf{X}$ consists of all 1's, which are dummy variables so that the additive constant parameter β_0 is included.) The least squares solution for $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}}=(\mathbf{XX}')^{-1}$ (see Ref. 30).

Statistical correlation can also be extended to multiple independent variables. We previously performed a simple correlation to measure the degree of linear association between two random variables. We can now utilize multiple correlations to measure the maximum correlation between the dependent variable and a linear combination of a set of independent variables. This enables us to test the ability of various linear models for the human texture discrimination process to explain the empirical data. The multiple correlation was computed for all possible pairs of the metrics considered. This value is defined as the highest value of the correlation coefficient computed between the scale values and a linear combination of the two metrics. The results are given in Table 2.

Additionally, we can use multiple correlation to test the effectiveness of various models consisting of linear combinations of more than two metrics to predict the psychological data. For this analysis, four metrics were selected as the most promising out of the seven that were tested with pairwise correlations. These four metrics are assigned numerals 1 to 4 as follows: 1=Doyle, 2=Eff_POT, 3=ACE, and 4=RABS. The models tested are a linear combination of all four and every possible combination of three. The results of this are given in Table 3.

In the table, the second column lists the value of the maximum correlation coefficient computed between the scale values and the linear combination of metrics in the first column. The remaining columns list the values of the regression parameters for which the model yields the maximum correlation value, corresponding to the $k+1\,\beta$ parameters in $\hat{\boldsymbol{\beta}}=(\mathbf{XX}')^{-1}$. In each case, the value listed in the β_0 column is the value of the additive constant parameter in the linear model for the optimum case. The values of these regression parameters do not absolutely indicate the relative importance of each metric in the model, since they provide both weighting and normalizing of the metrics. They are included simply to illustrate that although the maximum correlations for the models are rather high, their eventual utility depends on the proper selection of values for several parameters.

When the metrics were considered two at a time, the highest correlation (0.90) was obtained for a linear combination of the Doyle metric with the Eff_POT metric. These two metrics were also found, in a previous experiment performed at the US Army Night Vision and Electronic Sensors Directorate, to be the best predictors of the probability of finding low observable military targets in simulated infrared imagery.[5]

When combinations of three or four metrics were considered, a correlation of 0.94 resulted for the combination of Doyle, Eff_POT, and RABS. The inclusion of the GLC-based ACE metric does not significantly improve this result. Thus, it seems that for the stimuli and resulting psychological scale values in this experiment, it is best to use a GLC-based error metric if a single metric is desired as a measure of target distinctness. However, if we allow the inclusion of multiple metrics in the model, it is best to discard the GLC-based metric and instead use the Doyle, Eff_POT, and RABS metrics. But before such a combination model can be used in practice, it will certainly be necessary to conduct further experimentation to either confirm the robustness of the regression parameters that were best for this experiment or determine values that are the better for the particular imagery being used.

5 Studying Integrated Visual Search and Discrimination Process

This section describes a psychophysical experiment designed to investigate the task of human target discrimination when combined with visual search. The image stimuli used in this experiment also consisted of square target patterns embedded in background patterns, but in random locations unknown to the observers. As each observer per-

Table 3 The multiple correlation coefficient and corresponding regression parameters for selected linear combinations of metrics.

		Regression Parameters				
Metrics	Max Correlation	β_0	β_1	β_2	β_3	β_4
1,2,3,4	0.94	$-1.18e+00$	$6.58e-02$	$5.48e-04$	$1.11e-01$	$3.0767e-01$
1,2,3	0.91	$-8.05e-01$	$5.34e-02$	$6.77e-04$	$8.50e-01$	-
1,2,4	0.94	$-1.16e+00$	$6.91e-02$	$5.57e-04$	-	$3.22e$-01
1,3,4	0.89	$-1.55e+00$	$4.03e-02$	-	$6.00e-01$	$4.37e-01$
2,3,4	0.89	$-1.20e+00$	-	$3.38e-04$	$1.51e+00$	$2.06e-01$

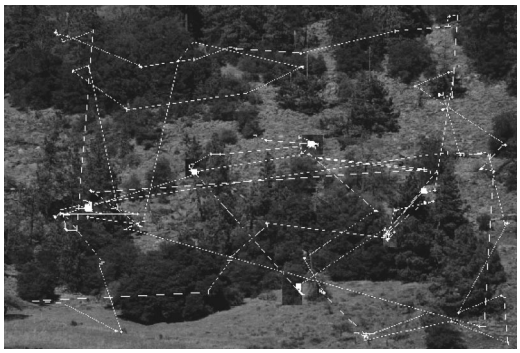


Fig. 7 The raw fixation point data from one observer for one of the stimulus images. The white streaks indicate the observer's fixation points, while suspected target locations are shown as small white square blocks.

formed a visual search of the scene for targets, his eye fixation point within the stimulus was measured by processing video of the observer's eye. By integrating search and discrimination, we can indirectly measure perceived target distinctness by measuring various statistics that indicate how easily the observers located it, including the likelihood the target was fixated or identified and the time required to do so. These computed statistics will also serve as a quantitative basis for evaluating the relative effectiveness of our target distinctness metrics at representing perceived target distinctness.

5.1 *Creation of the Image Scene Stimuli*

The images used in the visual search experiment were extracted from a set of natural scenes of various locations in southern California. All of the images were obtained using a Nikon 35-mm camera and developed as 8×10 in color enlargements. The enlargements were digitized at 120 pixels per inch using a Hewlett Packard digital scanner. The scenes include a wide variety of both terrain and vegetation conditions such as forests, mountains, fields, and deserts. Great care was taken to ensure that no man-made objects or animals appear in the scenes. The viewing perspective of each scene is such that the viewer is looking down from above, and the viewing distance varies from as close as 100 m to as far as several kilometers.

Ten 800×1200 images were selected from the database as representative of the wide variety of possible terrain and vegetation conditions. The color images were converted to gray scale by averaging the red, green, and blue channels. These ten raw images were used to create ten stimulus images according to a random scheme. For each stimulus, one of the ten raw images was designated as the background image and another of the ten was chosen as the target image. A random number of either four, five, or six was chosen for the number of targets. Every target was a square region of 48 pixels on each side. A random location was chosen for each target square, with the restrictions that no target pixels could lie within 96 pixels, or two target dimensions, of the boundaries of the image, and no target pixels could lie within 144 pixels, or three target dimensions, of another target's pixels. If a target location was chosen that did not meet these two restrictions, it was discarded and another random location was chosen. Once the number of targets and target locations for a particular stimulus were randomly chosen, the stimulus image was created by using the pixel values of the raw background image for all pixels except target pixels. The values for the target pixels were taken from the pixels in the raw target image at the corresponding locations. In this manner, we obtained a wide variety of naturally occurring target patterns against different, naturally occurring background patterns. There were a total of 52 targets in the ten image stimuli.

Table 4 The sample correlations coefficient computed between the five vectors of target fixation and identification statistics and the vector of each of the target distinctness metrics.

	P_{ID}	$\overline{T_{<ID}}$	P_{fix}	$\overline{R_{<fix}}$	$\overline{T_{fix}}$
ΔT	0.30	−0.55	0.28	−0.47	−0.32
Doyle	0.30	−0.54	0.34	−0.49	−0.29
Eff_POT	0.35	−0.43	0.24	−0.38	−0.25
ACE	0.42	−0.62	0.31	−0.56	−0.33
ABS	0.23	−0.25	0.19	−0.31	0.09
RABS	0.43	−0.43	0.35	−0.42	−0.05

5.2 *Conduct of the Experiment*

Data was collected from a total of 12 different observers. Each observer was told that each of the image scenes contained between four and six targets each, and that every target is a square region of a specified size that contains a pattern which looks as if it does not belong in its location, in that it looks ''unnatural'' or ''out of the ordinary.'' The observer was asked to identify each target as soon as he saw it, and to find as many of the targets in each image before proceeding to the next. The ten stimuli were presented to each observer in a different, randomly chosen order. Together with five calibration images and four zero point images, each observer was presented a total of 19 images in the experiment. This typically required about 10 to 15 min, during which time the observer was required to hold his head still. Figure 7 shows the raw fixation point data from one observer for one of the stimulus images. The white cross hairs show the observer's fixation points during the display of that image at the discrete sample times, with consecutive sample points connected by a straight line to indicate the eye movement. The fixation points at each of the moments when the observer pressed the middle mouse button are shown as small white square blocks. These correspond to areas suspected by the observer to be targets.

5.3 *Target Fixation and Identification Statistics*

The data provided by ITEMS for every observer consists of the fixation point coordinates in the display image and the corresponding timestamp for each sample, along with the timestamp and button identifier for every press of a mouse button during the session. Since the mouse button presses are the means by which the observer both controls the image display process and indicates he is fixating targets, and the locations of all targets in the image stimuli are known, this data is sufficient for computing various statistics describing the observer's search for and discrimination of the targets. When an observer is studying a particular target to

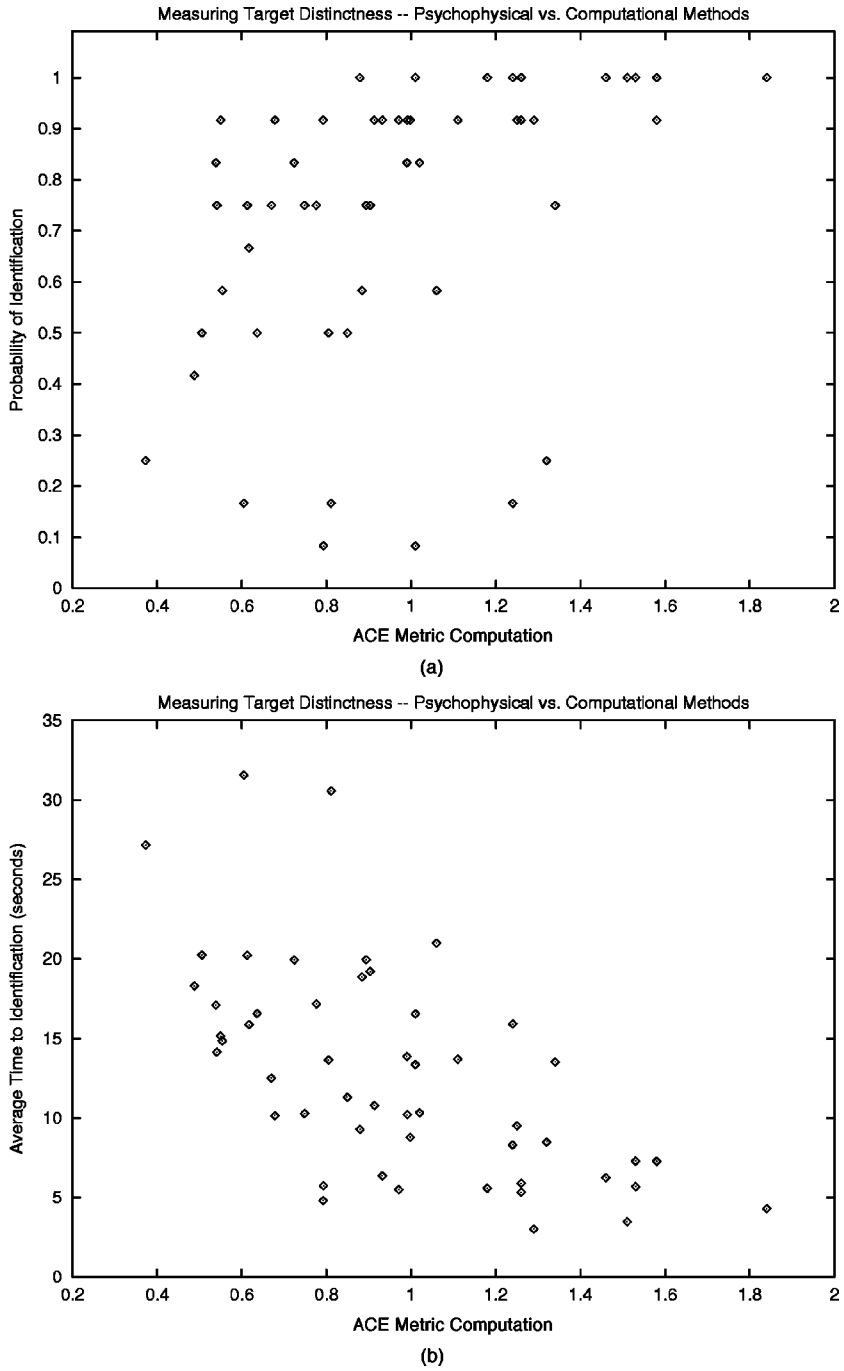


Fig. 8 The 52 targets in the search experiment plotted with their identification statistics and the computed values of the ACE metric. (a) P_{ID} versus ACE, (b) $\overline{T_{<\mathrm{ID}}}$ versus ACE.

decide whether it is indeed a target, his exact point of fixation will normally move about both within and just outside the target square, as he looks for cues to assist him in the decision. Thus, for the computation of these statistics, a fixation point was considered to be a fixation of a target if it was within the target square or within one and one-half target dimensions outside the target square. The statistics computed for the 52 targets in the experiment are identification probability (P_{ID}) average time to identification ($\overline{T_{<\mathrm{ID}}}$), fixation probability (P_{fix}) average time to first fixation ($\overline{T_{<\mathrm{fix}}}$), and average total fixation time ($\overline{T_{\mathrm{fix}}}$). The computations of P_{ID} and P_{fix} are more properly the likelihood of identification and fixation for each target, as they are simply calculated as the proportion of the 12 observers that identified and fixated the target. The statistics $\overline{T_{<\mathrm{ID}}}$ and $\overline{T_{<\mathrm{fix}}}$ are computed as the time elapsed from the moment the image was first displayed until the observer first identified or fixated the target, averaged only over those observers that did indeed identify or fixate the target. The statistic $\overline{T_{\mathrm{fix}}}$ is computed as the total time the observer spent fixating the target area, averaged over all 12 observers. The set of

target distinctness metrics were computed for all 52 targets in the experiment. For each calculation, the background was considered to consist of all pixels not in the target square but within one target dimension. Table 4 gives the sample correlation coefficient (r) computed between the five vectors of computed target fixation and identification statistics, and the vector of each of the target distinctness metrics. From Table 4, we see that for the P_{ID} and P_{fix} statistics, we have $r>0$ for all of the target distinctness metrics considered. A target that is more distinct is more likely to be fixated and/or identified. We also see that for the $\overline{T_{<\mathrm{ID}}}$ and $\overline{T_{<\mathrm{fix}}}$ statistics, we have $r<0$ for all of the metrics. A target that is more distinct will likely be fixated and/or identified in less time. The second-order ACE metric exhibited the strongest correlations for $\overline{T_{<\mathrm{ID}}}$, $\overline{T_{<\mathrm{fix}}}$, and $\overline{T_{\mathrm{fix}}}$. For P_{ID}, ACE was just behind RABS for the most strongly correlated metric.

Figure 8 shows plots of the 52 targets in the search experiment, with the horizontal axis representing the computed value of the ACE metric and the vertical axis representing the P_{ID} and $\overline{T_{<ID}}$ statistics.

5.4 *Analysis of the Results*

For this experiment, we have found that the magnitudes of the correlations between the individual target distinctness metrics and the probability of identification (P_{ID}) were as high as 0.43, and for average time to identification ($\overline{T_{<\mathrm{ID}}}$) were as high as 0.62. Although these values do indicate strong relationships, we must realize that there are many more variables contributing to whether an observer identifies a target and the time required to locate a target than just the distinctness of the target. It is also important to realize that even if there is a direct relationship between two variables, the computed value of a correlation coefficient between them may not be high if the relationship is not linear. Overall, of the set of target distinctness metrics considered, the second-order GLC-based ACE metric was the most strongly correlated with the psychophysical data. Although the observers were not instructed as to what cues they were to use in making their judgments, we can surmise that the observers probably utilized some combination of differences in brightness (contrast), texture, and abrupt discontinuities along target/background boundaries. Certainly differences in target and background first-order pixel probabilities are important, since they represent pattern contrast and variation. But second-order probabilities are important too, since they better represent the general concept of texture by taking into account the spatial relationships between pixels. A GLC model may be able to capture at least some of all of these variables. Second-order probabilities inherently contain first-order probabilities, in that a pattern's histogram can be obtained by summing over all rows or over all columns of one of its GLC matrices. Also, if two patterns have GLC models that are significantly different, it is apparent that a distinctly abrupt boundary is more likely if the two patterns are placed adjacent to each other.

6 Concluding Remarks

In our future studies, we wish to determine which cue is most important for each target and use a metric appropriate for that target, instead of trying to use the same metric for every target. Or, perhaps a proper weighting of the relative importance of the three perceptual cues could be determined for every target, and used to form a composite metric. Additionally, the variable of target size must be factored into the metrics. In our experiments, we also did not vary the size of the field of view, which most certainly has an effect on search times. We feel also that the spatial location of the target in the image (such as center or periphery) and global variables (such as scene clutter) have an effect. The model should also account for the effects of competing targets and other points of interest, as well as false alarms.[31,32]

As for the experimental methodology presented, both the pure discrimination and the search experiment allowed us to study perceived target distinctness. But the search experiment provided us with data that can be used to develop or test models describing various aspects of the search and discrimination processes, rather than only the final result. And not only do we have fixation data that include two-dimensional image coordinates, but also a third dimension of time, which will allow us to include this dimension in the model.

Besides target search and discrimination, it is apparent any study of human visual perception can benefit from measuring the eye fixations of observers. Although we can always have an observer report his judgments of a visual stimulus, knowledge of the eye fixation points provides us with invaluable insights into the process through which the observer reached his decisions. We plan to expand the scope of our studies to include other applications that depend on human visual perception, such as advanced human-computer interfaces, adaptive videoconferencing systems, and assessment of digital display quality and television advertising effectiveness.

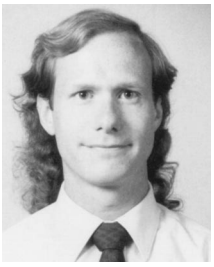
Acknowledgments

The original version of this material was first published by the Research and Technology Organization, North Atlantic Treaty Organization (RTO/NATO) in MP-45 (Search and Target Acquisition) in March 2000. This proceedings is available at http://www.rta.nato.int.

References

1. M. M. Trivedi and M. V. Shirvaikar, "Quantitative characterization of image clutter: Problems, progress, and promises," *Proc. SPIE* **1967**, 288–299 (1993).
2. J. D'Agostino, W. Lawson, and D. Wilson, "Concepts for search and detection model improvements," *Proc. SPIE* **3063**, 14–22 (1997).
3. R. Hecht-Nielsen and Yi-Tong Zhou, "Vartac: A foveal active vision ATR system," *Neural Networks* **8**(7/8), 1309–1321 (1995).
4. G. W. Walker and J. R. McManamey, "Characterization of natural background clutter for design of camouflage," *Proc. SPIE* **1687**, 254–264 (1992).
5. B. L. O'Kane, C. P. Walters, and J. D'Agostino, "Report on perception experiments in support of low observables thermal performance models," Technical report, U.S. Army, Night Vision and Electronic Sensors Directorate, Fort Belvoir, VA (Feb. 1993).
6. A. C. Copeland, M. M. Trivedi, and J. R. McManamey, "Evaluation of image metrics for target discrimination using psychophysical experiments," *Opt. Eng.* **35**(6), 1714–1722 (June 1996).
7. A. C. Copeland and M. M. Trivedi, "Texture perception in humans and computers: Models and psychophysical experiments," *Proc. SPIE* **2742**, 436–446 (1996).
8. C. Copeland and M. M. Trivedi, "Integrated framework for developing search and discrimination metrics," *Proc. SPIE* **3062**, 53–58 (1997).
9. C. Copeland and M. M. Trivedi, "Models and metrics for signature strength evaluation of camouflaged targets," *Proc. SPIE* **3070**, 194–199 (1997).

10. J. A. Ratches, ''Static performance model for thermal imaging systems,'' *Opt. Eng.* **15**(6), 525–530 (1976).
11. M. V. Shirvaikar and M. M. Trivedi, ''Developing texture-based image clutter measures for object detection,'' *Opt. Eng.* **31**, 2628–2639 (Dec. 1992).
12. S. R. Rotman, A. Cohen, D. Shamay, D. Hsu, and M. L. Kowalczyk, ''Textural metrics for clutter affecting human target acquisition,'' *Proc. SPIE* **2743**, 99–112 (1996).
13. B. Julesz, ''Visual pattern discrimination,'' *IRE Trans. Inf. Theory* **8**(2), 84–92 (Feb. 1962).
14. J. S. Weszka, C. R. Dyer, and A. Rosenfeld, ''A comparative study of texture measures for terrain classification,'' *IEEE Trans. Syst. Man Cybern.* **6**, 269–285 (Apr. 1976).
15. R. W. Conners and C. A. Harlow, ''A theoretical comparison of texture algorithms,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-2**(3), 204–222 (May 1980).
16. P. P. Ohanian and R. C. Dubes, ''Performance evaluation for four classes of textural features,'' *Pattern Recogn.* **25**(8), 819–833 (1992).
17. M. M. Trivedi, C. A. Harlow, R. W. Conners, and S. Goh, ''Object detection based on gray level cooccurrence,'' *Comput. Vis. Graph. Image Process.* **28**, 199–219 (Nov. 1984).
18. C. A. Harlow, M. M. Trivedi, R. W. Conners, and D. Phillips, ''Scene analysis of high resolution aerial scenes,'' *Opt. Eng.* **25**(3), 347–355 (Mar. 1986).
19. A. Gagalowicz and S. De Ma, ''Sequential synthesis of natural textures,'' *Comput. Vis. Graph. Image Process.* **30**, 289–315 (1985).
20. G. Lohmann, ''Co-occurrence-based analysis and synthesis of textures,'' in *12th IAPR Intl. Conf. Patt. Recog. (ICPR)*, vol. 1, pp. 449–453, Jerusalem, Israel (Oct. 1994).
21. G. Ravichandran, E. J. King, and M. M. Trivedi, ''Texture synthesis: A multiresolution approach,'' in *Proc. Ground Target Modeling Validation Conf.*, Keweenaw Research Center, Houghton, MI (1994).
22. A. R. Figueiras-Vidal, J. M. Paez-Borrallo, and R. Garcia-Gomez, ''On using cooccurrence matrices to detect periodicities,'' *IEEE Trans. Acoust., Speech, Signal Process.* **35**(1), 114–116 (Jan. 1987).
23. J. Parkinnen, K. Selkainaho, and E. Oja, ''Detecting texture periodicity from the cooccurrence matrix,'' *Pattern Recogn. Lett.* **11**, 43–50 (Jan. 1990).
24. C. C. Gotlieb and H. E. Kreyszig, ''Texture descriptors based on co-occurrence matrices,'' *Comput. Vis. Graph. Image Process.* **51**, 70–86 (1990).
25. M. J. Muller, ''Texture boundaries: Important cues for human texture discrimination,'' in *IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 464–468, Miami Beach, FL (June 1986).
26. A. C. Copeland, ''Image metrics for human search and discrimination of textured targets and backgrounds,'' PhD thesis, Univ. of Tennessee, Knoxville (1996).
27. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, MA (1992).
28. L. L. Thurstone, ''A law of comparative judgment,'' *Psychol. Rev.* **34**, 273–286 (1927).
29. W. S. Torgerson, *Theory and Methods of Scaling*, John Wiley and Sons, New York (1958).
30. M. S. Srivastava and E. M. Carter, *An Introduction to Applied Multivariate Statistics*, North-Holland, New York (1983).
31. S. Grossman, Y. Hadar, A. Rehavi, and S. R. Rotman, ''Target acquisition and false alarms in clutter,'' *Opt. Eng.* **34**(8), 2487–2495 (Aug. 1995).
32. T. J. Doll and D. E. Schmieder, ''Observer false alarm effects on detection in clutter,'' *Opt. Eng.* **32**(7), 1675–1684 (July 1993).



Anthony C. Copeland received the BSEE degree in 1988 from the U.S. Military Academy at West Point, New York, and the MSEE and PhD degrees in 1993 and 1996 from the University of Tennessee, Knoxville. He was a signal corps officer in the U.S. Army from 1988 to 1991, which included service as a communications node platoon leader for the 24th Infantry Division (Mechanized) during Operation Desert Shield/Storm. He was employed as a research engineer during 1996-97 in the Electrical and Computer Engineering Department at the University of California, San Diego, performing research in image understanding, texture analysis, and human visual perception. Since 1997, he has developed real-time signal processing systems for hyperspectral sensors at PAR Government Systems Corp. in San Diego.

Mohan M. Trivedi is a professor in the Electrical and Computer Engineering Department of the University of California, San Diego, where he serves as the Director of the Computer Vision and Robotics Research Laboratory (http://cvrr.ucsd.edu). He and his team are engaged in a broad range of research studies in active perception and novel machine vision systems, intelligent ("smart") environments, distributed video networks, and intelligent systems. At UCSD, Trivedi also serves on the Executive Committee of the California Institute for Telecommunication and Information Technologies, Cal (IT) 2, leading the team involved in the Intelligent Transportation and Telematics projects. Trivedi serves as the Editor-in-Chief of *Machine Vision and Applications*, the official journal of the International Association of Pattern Recognition. He is a frequent consultant to various national and international industry and government agencies. Trivedi is a recipient of the Pioneer Award (Technical Activities) and the Meritorious Service Award of the IEEE Computer Society and the Distinguished Alumnus Award from the Utah State University. He is a Fellow of the International Society for Optical Engineering (SPIE).