SPECIAL SECTION

Computational Modeling of Foveal Target Detection

Gary Witus, Turing Associates, Ann Arbor, Michigan, and **R. Darin Ellis,** Wayne State University, Detroit, Michigan

This paper presents the VDM2000, a computational model of target detection designed for use in military developmental test and evaluation settings. The model integrates research results from the fields of early vision, object recognition, and psychophysics. The VDM2000 is image based and provides a criterion-independent measure of target conspicuity, referred to as the *vehicle metric* (VM). A large data set of human responses to photographs of military vehicles in a field setting was used to validate the model. The VM adjusted by a single calibration parameter accounts for approximately 80% of the variance in the validation data. The primary application of this model is to predict detection of military targets in daylight with the unaided eye. The model also has application to target detection prediction using infrared night vision systems. The model has potential as a tool to evaluate the visual properties of more general task settings.

INTRODUCTION

Need and Scope

Visual detection by human observers to is considered to be the major operational threat to individual vehicle and unit operation security. The U.S. Army, as well as the NATO countries, have recognized the need for a greater understanding of both the visual signature of their vehicles and the process of detection by human observers using unaided eyes and direct view optics. To this end, the military community implemented initiatives such as the U.S. Army Target Acquisition Model Improvement Program in the early 1990s and the NATO Research and Technology Organization working group (SCI-12) on camouflage evaluation methods and models. This project was motivated by the need for tools for developmental test and evaluation of military vehicles employing designs and technologies for detection avoidance (i.e., "signature management"). This project was specifically motivated by the need for models of visual detection.

The performance specification for developmental test and evaluation is expressed in terms of detection, given that the observer is looking at, or in the direction of, the vehicle (i.e., foveal target detection). It is not expressed in terms of search time or probability of detection during wide-field-of-view search. Search performance is of interest in operational test and evaluation of large-scale combat, but it is not used in performance requirements for materiel development because search outcome and performance is influenced by many factors over which the materiel designer has no control (e.g., the tactics and operational employment of friendly and threat forces; the large-scale terrain properties). The materiel designer has control over only the vehicle signature and, to some degree, how it interacts with its local surrounding. The performance specifications for signature management are expressed in terms of detection range and probability of detection. In the context of military target acquisition, *detection* means determining that an object is a potential target, specifically a military vehicle. Detection avoidance

Address correspondence to R. Darin Ellis, Wayne State University, 87 E. Ferry St., Detroit, MI 48187. HUMAN FACTORS, Vol. 45, No. 1, Spring 2003, pp. 47–60.

requirements for ground vehicles invariably address detection avoidance for a stationary vehicle. Moving target detection is considered in combat models but is not currently a major component in developmental test and evaluation.

Objectives

The goal of the project was to develop a robust and accurate analytic model to predict human observer performance in visual vehicle discrimination at the "detection" level for stationary targets, given that the observer was looking at, or in the direction of, the target. The result of this work was a model of target detection called VDM2000.

VDM2000 THEORY

Comparison with Recent Approaches

The visual search paradigm. An extensive body of research in experimental psychology has been developed based on the visual search paradigm. This paradigm, although useful in studies of basic visual attention and perception, does not apply to search for vehicles in natural scenes. The visual search paradigm enables experimental psychologists to control the visual content to which the participants respond and thereby to isolate specific aspects of vision. The visual search paradigm, as described by Wolfe (1998), is characterized by (a) discrete target and distractor figures, (b) a well-defined specific target description, (c) well-defined distractors, (d) well-defined visual attributes with distinctive and discrete values, (e) randomized placement, and (f) a noninterfering and noninformative background.

The standard experimental psychology visual search paradigm eliminates uncertainty on several dimensions important to understanding search in natural scenes. First, there is no uncertainty as to the appearance of the target or the question of what the objects are. Second, this paradigm eliminates the contribution of local and global context in search. These properties make the standard visual search paradigm useful for basic vision and attention research, but they also make it inapplicable to search for vehicles in natural terrain.

Salience theory of search and computer vision models. The salience theory of search holds that the strength with which locations draw visual attention is proportional to the magnitude of resemblance between the target and scene locations (Itti & Koch, 2000; Wolfe, 1994). Computer-vision-based models of search and performance based on salience theory have the challenge of developing a computational metric of the extent to which locations resemble a target. The computational salience models attempt to find targets in an image using cues and criteria that correspond, in theory, to those used by people. Computer vision systems have been effective in structured environments and tasks, but they have not proven effective in unconstrained environments or for ill-structured tasks.

Even when computer vision systems are effective, they do not use the same methods that human vision uses. Itti and Koch (2000) concluded that the performance of their algorithm is uncorrelated with human performance. Computational salience modeling is a special case of the more general problem of automatic target recognition (ATR) algorithms. ATR algorithms focus on target detection as the end goal, rather than on increasing understanding of human vision. Over the past 20 years, the U.S. Department of Defense has provided significant funding for ATR development. The U.S. Army, Navy, Air Force, and Defense Advanced Research Projects Agency all have active ATR programs. To date, no robust and effective computational method has been demonstrated to find locations in images that resemble ground vehicles. No ATR algorithms have been successfully placed in the field, even with the more limited goal of cuing human observers.

Operational effectiveness models. The U.S. Army Night Vision Laboratory's target acquisition models (O'Kane, 1995; Wilson, 2001) are the preeminent examples of the "operational effectiveness" approach. These models have shown some limited success in explaining performance for search and detection with low-resolution night vision devices. These models were intended to predict average performance over a set of similar images, not to predict detection performance for specific images (as the current modeling effort does). The Night Vision Lab models have been less accurate when applied to high-resolution visual search or to evaluate specific images. In 1992 the U.S. Army initiated a 3-year Target Acquisition Model Improvement Program, which ultimately failed to produce an improved model (Mazz, Kistner, Bushra, & Pibil, 1997).

Development and Historical Antecedents

The VDM2000 is a cascading sequence of equations representing front-end vision, perceptual organization of the vehicle, local contrast and clutter, evidence accumulation, and psychophysical response. The richness of the model comes primarily from the number of different factors and stages of processing that are represented. It is a low-threshold model, consistent with basic vision research results for search and cued detection.

The VDM2000 makes numerous contributions to the science and practice of modeling human observers in areas such as contrast mechanisms and measurement, the ability to account for masking attributable to local clutter, and representation of internal target structure. Table 1 presents a list of VDM2000 contributions vis-à-vis classical models (Matchko & Gerhart, 2001; O'Kane, 1995; Wilson, 2001). A process flow of the VDM2000 is show in Figure 1.

Low-Level Vision Module: Achromatic and Color Vision

The model's front end (the side of the model that interacts with the inputs) represents bottom-up visual processing, including pupil reflex; cone saturation and spectral response; spatial filtering and sampling resulting in the retinal output response to the image formed on the retina; and finally achromatic and color-opponent response (Kaiser & Boynton, 1996). See the box labeled "low-level vision module" in Figure 1 for a depiction of this module relative to the overall VDM2000 architecture.

The red-green-blue (RGB) image is converted to the standard Commission Internationale de l'Éclairage (CIE) tri-stimulus XYZ coordinates (CIE, 1932). The XYZ image is converted to an image of the long-, medium-, and short- (LMS) wavelength cone responses. The perceived color

TABLE 1: VDM2000 Contributions with Respect to Classical Models

Model Feature	Classical Models[a]	VDM2000
Basis for contrast measurement	Based on the input to the visual system	Based on the output of the receptors (cones) after luminance adaptation and cone nonlinearities (Boynton & Whitten, 1970; Kaiser & Boynton, 1996)
Contrast channels accounted for	Evaluate only luminance contrast	Includes both luminance and chromatic contrast (Brainard, 1996; Wandell, 1995)
Ability to account for effects of object's internal structure	Compute aggregate statistics treating the entire vehicle as homogeneous	Uses a simple model of the structure and appearance of 3-D objects under natural illumination (Moore & Cavanagh, 1998; Witus, Gerhart, & Ellis, 2001)
Generality of contrast model	Use an equation to compute contrast that is well defined only for uniform targets against uniform backgrounds	Uses a band-limited, adaptive function to compute local contrast in inhomogeneous surroundings (Ahumada & Beard, 1998; Peli, 1990, 1997)
Ability to account for local clutter	Do not include a measure of local clutter or its effect on detection	Computes local clutter and represents its masking effect on the efficiency of contrast for detection (DeValois & DeValois, 1990)
Model output	Predict detection independent of the response biases and false alarm context	Generates a receiver operating characteristic curve, expressing the probability of positive response as a function of the false alarm rate and the vehicle detection metric (Palmer et al., 2000)

[a] Matchko & Gerhart, 2001; O'Kane, 1995; Wilson, 2001.

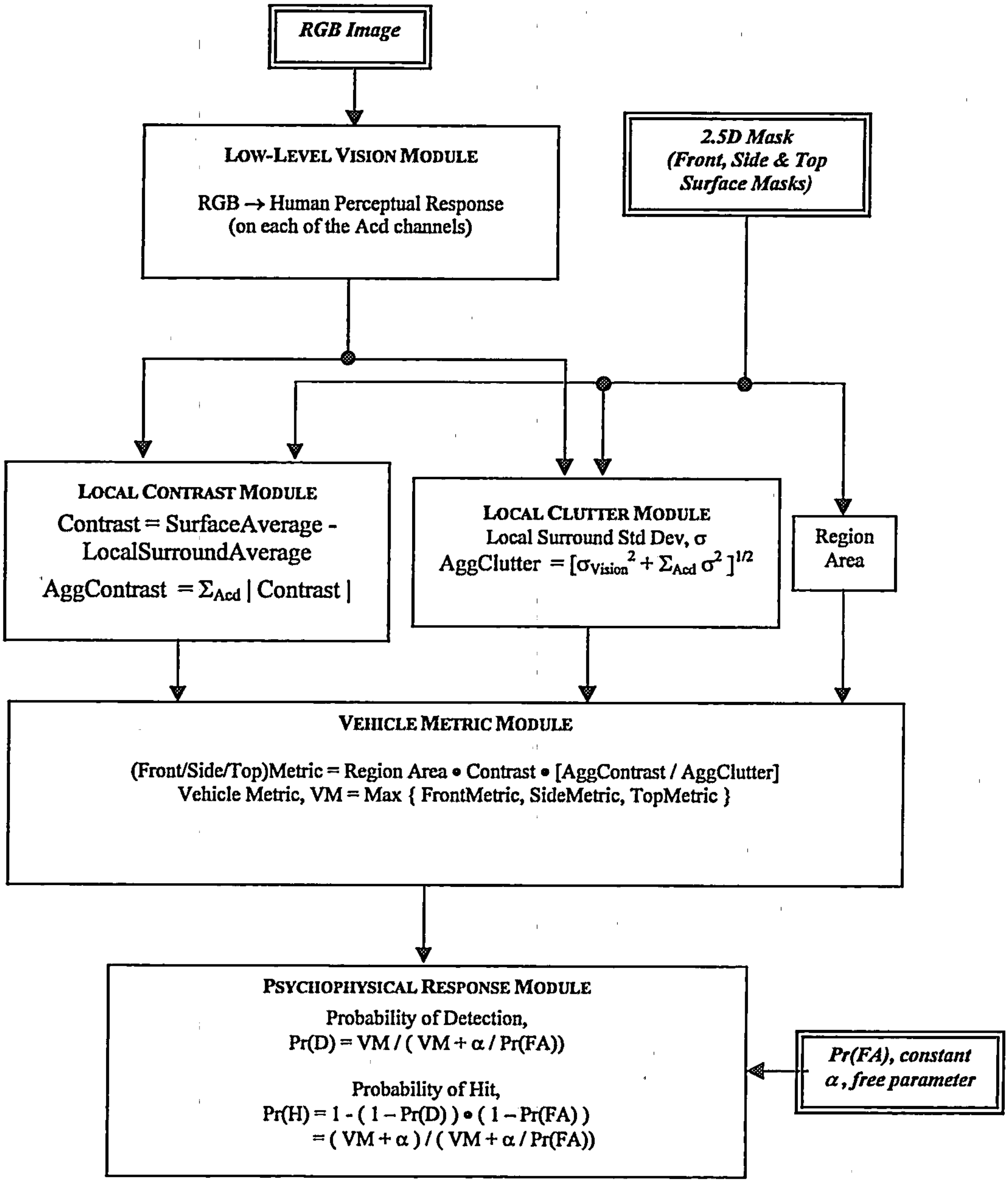

Figure 1. VDM2000 processing flow.

image (Acd; achromatic, red-green, and yellow-blue color-opponent channels) is modeled as a linear transformation of the LMS cone response.

Local Contrast and Local Clutter

Local contrast and clutter are calculated based on a combination of the bottom-up information provided from the front-end vision module and some top-down assumptions regarding the cognitive processes involved in the perceptual organization of the target. These modules are implemented as independent modules in the model architecture and are depicted in Figure 1.

Target organization: Shape from shading. With simple two-dimensional targets there is typically no need to consider the organization of target perception. Ground vehicles are three-dimensional (3-D) structures that rarely present a uniform appearance. The simple physics of lighting and geometry of configuration dictate that some surfaces of a 3-D object will receive more illumination than others because of shading from surface orientation and shadowing from occlusion. Some surfaces are bright, some are dark. These systematic variations provide the pop-out cues that the human visual system uses to segment objects from their surroundings (Sun & Perona, 1996). Tarr, Kersten, and Buelthoff (1998) found that the human visual system encodes the direction of illumination and its related effects (e.g., shading and shadows) and that these function to reveal a 3-D object's shape. Moore and Cavanagh (1998) found that two-tone images of 3-D objects were highly effective at inducing 3-D object perception despite their lack of shading, hue, or texture. They concluded that the mechanisms of 3-D object perception in natural scenes include a mechanism for processing scene illumination with respect to an internal memory representation of a 3-D object's shaded appearance. Witus and Gerhart (2000) found that aggregate differences among the front, side, and top regions accounted for more than 65% of the luminance variance over the entire vehicle for a sample of 44 images of military vehicles in natural settings.

VDM2000 does not attempt to emulate the process by which the visual system recognizes objects. Rather, the model uses *as an input* a representation of the results of the top-down processing to measure the amount of signal available for an observer to detect the object. In VDM2000, the generic representation of a vehicle is the projection onto the image plane of the orthogonal front (or rear), side, and top vehicle surfaces – a 2.5-D representation. These vehicle surfaces have different orientations with respect to the illumination source and observer. Consequently, there tends to be low variation within a region and high contrast between regions. These properties are characteristic of 3-D objects and are natural for a simplified representation of a military vehicle.

Clutter and contrast: Defining the local surround. For each of the three characteristic target regions, and across the achromatic and two color-opponent channels of the Acd images in perceptual coordinates (the outputs of the model's front end), the VDM2000 computes local contrast and local clutter. The VDM2000 local-contrast and local-clutter measures are single-channel band-pass metrics. The upper limit on the spatial band-pass is the Nyquist limit of the input image (45 cycles/°). The lower limit is a function of the size and shape of the target region and its local surround. The local surrounds are computed with the same algorithm; however, the local surround for calculating contrast is narrower than the local surround for calculating local clutter. Aggregate values for each region are then calculated for determination of the vehicle metric in the next step. Aggregate contrast is the sum of the contrast magnitude on the individual achromatic and color-opponent channels. Aggregate clutter is the root-sum-square of the clutter (σ) on the individual achromatic and color-opponent channels plus a component representing the internal noise of the visual system (σ_{VISION}). See Figure 1 for a depiction of the contrast and clutter modules as well as their relation to the overall model architecture.

Evidence Accumulation: The Vehicle Metric Module

The visual evidence is based on the spatial signal from the bottom-up path (contrast and clutter metrics), organized by the region of interest from the top-down path (2.5-D mask). The contrast, clutter, and area are all combined into a detectability metric for each surface region (see Figure 1, vehicle metric module). The region metric is equal to region area times aggregate contrast times contrast efficiency. The contrast efficiency is the ratio of aggregate contrast to aggregate clutter. Thus the region metric is equal to area times contrast squared, divided by clutter.

The vehicle metric is the maximum of the three surface region metrics. This is consistent with the observation that detection is a function of the dominant cue: Suppressing secondary cues has no effect if the dominant cue is not treated, but suppressing the dominant cue reduces detectability until it is reduced to the point where it is no longer dominant.

Psychophysical Response Module

The vehicle metric is a criterion-independent measure of target signal. Observer detection response also depends on a number of other factors unrelated to the image (e.g., perceived cost of a missed detection vs. perceived cost of a false alarm; expectations regarding the frequency or density of targets; nontarget objects that resemble targets). The effect of these factors is measured by the probability of false alarm (i.e., false alarm rate). The psychophysical function expresses the probability of hit (i.e., positive response in the presence of, but not necessarily in response to, a target) as a function of the criteria-independent measure of perceived signal and the probability of false alarm in the form of a receiver operating characteristic curve. Several different psychophysical models have been used successfully in visual search and detection (Palmer, Verghese, & Pavel 2000). VDM2000 uses a two-stage psychophysical model. Pr(D), report of "detection" in response to a target, is computed as a function of the vehicle metric (VM), the probability of false alarm, Pr(FA), and the calibration parameter, α. The value of α is scaled to the false alarm rate, and one value of α is used for all response levels. The false alarm rate is used as a measure of bias, or the willingness of an observer to call some signal a target.

At any given response level characterized by Pr(FA), Pr(D) is computed as the vehicle metric divided by the vehicle metric plus a constant. The constant is divided by the false alarm rate.

$$Pr(D) = VM/[VM + \alpha/Pr(FA)]. \quad (1)$$

The equation for Pr(H) (probability of a "hit") as a function of Pr(D) and Pr(FA) is based on a simple theoretical model. The model assumes that positive response to the target signal and positive response to other signals in the image are processes that are parallel and independent. A "detection" is reported if there is either a positive perception of the target (at probability Pr[D]) or positive misperception of nontarget stimuli (at probability Pr[FA]). The standard equation for the probability of the union of two events is used:

$$Pr(H) = 1 - [1 - Pr(D)][1 - Pr(FA)] \quad (2)$$

– that is, no detection is reported only when there is no correct detection and no false alarm. These two equations can be combined into a single psychometric equation for Pr(H) as a function of the vehicle metric, Pr(FA), and the calibration parameter α:

$$Pr(H) = (VM + \alpha)/[VM + \alpha/Pr(FA)]. \quad (3)$$

VDM2000 IMPLEMENTATION AND OPERATION

VDM2000 takes two images as inputs: an image of the vehicle in the scene and a mask image designating the front (or rear), side, and top surfaces of the vehicle. The specification of the projection of the front (rear), side, and top surfaces onto the image plane is based on the geometry of illumination, shadowing, and reflection. The mask image is a three-color image in which the projections of the front, side, and top are arbitrarily color coded red (255, 0, 0), green (0, 255, 0), and blue (0, 0, 255). The nontarget area is black (0, 0, 0). An example target image and accompanying mask is shown in Figure 2. Finally, the VDM2000 predicts Pr(H) and Pr(D) for several values of Pr(FA). The values of Pr(FA) for which the analyst wants to predict Pr(H) and Pr(D) are entered as well.

Figure 2. Original image and "2.5-D" mask of the three characteristic surfaces.

A fully detailed discussion of the technical requirements for images (e.g., resolution, blur), calibration requirements (e.g. computing the RGB-to-XYZ transform), construction of masks, and other aspects of operating the model is beyond the scope of this paper but appears in Witus (2001).

VALIDATION DATA: METHODS, PROCEDURE, AND RESULTS

Participants

The model validation experiments were conducted using paid participants recruited from the general population. Participants (18–45 years of age) had vision corrected to 20/20 and normal color vision (screened with a Bausch and Lomb Orthorater). Of the 20 participants, there were 19 men and 11 women.

Stimulus Material

The image set consisted of 1150 distinct images, 800 images with vehicles in the scene and 350 without. The images were created from 44 color slide images taken during the 1995 Distributed Interactive Simulation, Search, and Target Acquisition Fidelity field test held by the U.S. Army Communications-Electronics Command Night Vision and Electronic Sensors Directorate (Toet, Bijl, Kooi, & Valeton, 1998). The color slides were digitized at high resolution to produce a digital image of 6144 × 4096 pixels. The original 44 images contained nine different types of tracked and wheeled U.S. and foreign military vehicles in a variety of locations, aspects, postures (including partially obscured), and lighting conditions. Ranges were from 500 to 5000 m. The original 44 images did not produce wide variation in probability of detection in search and detection test (Toet et al., 1998). The 1150 images used for model validation in this project were down sampled with low-pass filtering, cropped to 1080 × 720 pixels, and then digitally manipulated to produce a wider range of expected observer response. Details of the stimulus image manipulation process are given in Witus (2001). Image manipulations were chosen so that the image set would vary with respect to factors that are known to influence visual perception, such as brightness, contrast, chromaticity, and spatial scale. Table 2 briefly describes the image manipulations.

Apparatus

Stimuli were presented and responses collected via custom-made visual search experimentation software written in Visual Basic 6. All responses were collected through a two-button mouse. The experimental software ran on a Gateway 2000 Pentium II PC using a 17-inch (43-cm) EV700 monitor with an ATI Rage Pro 128 graphics card set to 24-bit color at 1280 × 1024 pixel resolution.

TABLE 2: Important Dimensions of Variation in the Image Set

Dimension	Source of Variation
Vehicle type	9 types of military vehicles
Viewing azimuth	360°
Viewing elevation	0° to 45°
Range	13× variation in linear scale, resulting from a combination of original image target size and the down-sampling scale factor (2:1 and 3:1)
Lighting	direct, diffuse, or in shadow
Lighting	from front, back, side, or top
Overall illumination	light or dark
Contrast attenuation	clear or haze
Color	full color or gray scale
Vehicle exposure	fully exposed, partially exposed, or foreground foliage
Clutter	in the open or amid clutter
Signature	baseline, reduced contrast, or suppressed cue features
Vehicle shadows	present or absent
Camouflage nets	present or absent
Camouflage paint patterns	present or absent

The monitor was calibrated prior to testing by displaying red, green, and blue images at a staircase of intensities and measuring the CIE xyY coordinates with a Minolta CS-100 colorimeter. We also measured the display xyY with only the low level of ambient illumination maintained for the test. These data were used to estimate the parameters of the RGB-to-XYZ transfer function in the computer model (the monitor bias, the gamma exponent, and a linear matrix). Ambient illumination from the wall behind the monitor varied somewhat but was approximately 30 cd/m^2. Peak monitor output was 110 cd/m^2. Reflection from the dark screen was 5 cd/m^2. The participants were seated 60 inches (152 cm) from the monitor and behind a table to prevent them from moving forward. At this distance, the angular resolution of the monitor was 110 pixels/°, roughly equal to the limit of visual acuity for high-contrast signals at the central fovea.

Experimental Procedure

Instructions and pretest training. Observers were individually tested in a self-paced manner. Prior to the experiment, the observers were presented with a set of 27 closeup images of the various vehicles in the natural terrain. The closeup images were presented in a brief training session to familiarize the participants with the procedure. The results from the familiarization trials were not included in the experiment results, and the training images were not used in the experiment.

Block procedure. The test was organized into four experimental blocks based on systematic differences in the overall scene appearance: baseline images, darkened images, lightened images with contrast attenuated, and gray-scale images. Within each block, image order was randomized without replacement across trials.

Trial procedure. The testing was self-paced. Before a scene was presented, a random spatial noise pattern was displayed for 750 to 2000 ms. Target location was cued with a red circle 300 pixels (approximately 3°) in diameter, centered on the target location. When a trial used an image that did not contain a target, the cuing circle was centered at a location where it was physically possible for a vehicle to be. This cue oriented the participant to the vehicle location without distracting from or interfering with vehicle perception. The participant would click the mouse to display the image and then click again when he or she had decided whether or not a vehicle was in the scene. For trials in which the participants identified a vehicle (or thought that they did), they were instructed to click on the vehicle itself to indicate its location in the image. The scene was then masked with the noise image again, and a response menu appeared. The observer selected from one of the following four responses: (a) "definitely no vehicle was present," (b) "unsure whether or not a vehicle was present," (c) "confident that a vehicle was present," and (d) "certain that a vehicle was present."

In addition to the menu choice, the response time between image display and mouse click was recorded. There was a maximum response time of 60 s, at which time the four-choice response menu was displayed. Following menu selection, a dialog box appeared that provided feedback on target presence, response time, and the number of trials remaining in the block. Clicking "OK" on this dialog box started the next experimental trial.

Data Treatment

For the specific purposes of model validation, each trial response provided a single data point: the rating of target-present confidence. The first step in data reduction and analysis was coding the observers' target-present response level (on a 1–4 scale) into correct and incorrect decisions. Three different response levels were used: liberal, moderate, and conservative. For target-present images, hit rate (HR) at the liberal level was the proportion of responses that had a rating of 2 (*maybe*) or higher. HR at the moderate and conservative levels were the proportion of responses that had a rating of 3 (*probably*) or higher or 4 (*definitely*), respectively. The false alarm rate (FAR) at each response level was calculated from the 350 images without targets. Thus, for each of the images, we obtained three estimates of HR as a function of FAR.

Aggregate Results

Empirical estimates of Pr(FA) were calculated at each response level by aggregating over all participants and over all images without targets. Empirical estimates of Pr(H) were calculated at each response level for each image with a vehicle present by aggregating over all participants. Figure 3 shows a plot of Pr(H) versus Pr(FA) at each of the three response levels. Pr(H) is aggregated over all vehicles, and Pr(FA) is aggregated over all vehicle-absent scenes. We computed Pr(H) for an image by simply pooling the participants' responses and dividing the number of positive responses by the total number of responses. Table 3 shows the mean Pr(H) for different partitions of the image set.

Table 4 shows the proportion of variance in Pr(H) explained by each of the major factors over all images with vehicles, calculated as η^2. The base scene accounts for 49% to 66% of the variance in Pr(H). Interaction effects account for 25% to 39% of the variance in Pr(H). Individually, the variations in vehicle signature modification, scene modification, and scale modification had only small effects. Large vehicles in the open were still large vehicles in the open. In combination, however, and in combination with the variation in the base scene, these factors had significant effects:

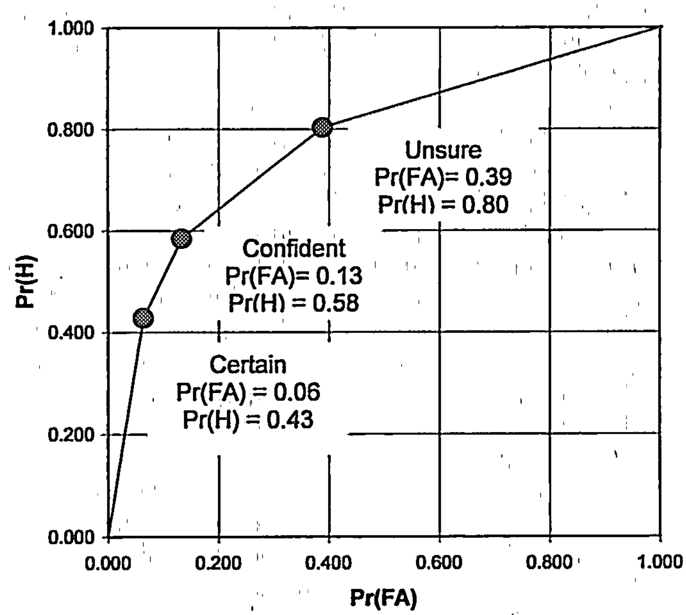

Figure 3. Aggregate Pr(H) versus Pr(FA).

TABLE 3: Mean Value of Pr(H) by Level of Factors in the Experimental Design

Factor	Image Set Partition	Certain	Confident	Unsure
Vehicle	Baseline vehicles	.51	.67	.86
	Low-contrast vehicles	.34	.50	.74
	Special vehicle variations	.39	.54	.76
Scene	Unmodified scene	.46	.61	.81
	3:1 Scale factor	.39	.53	.77
	2:1 Scale factor	.46	.62	.83
	Darkened scene	.42	.58	.80
	Hazy scene	.38	.52	.76
	Gray-scale scene	.43	.59	.81

TABLE 4: Pr(H) η^2 Results for All Images with Vehicles

	Proportion of Variance Explained, η^2		
Factor	Certain	Confident	Unsure
Base scene	.66	.60	.49
Vehicle modification	.07	.07	.08
Scene modification	.01	.01	.01
Scale modification	.01	.02	.03
Total main effects	.75	.70	.61

Table 5 shows the average Pr(H) for each of the partitions of the major factors in the experimental design. Comparing the unmodified vehicle aggregates with the reduced contrast vehicle aggregates shows that the vehicle contrast reduction lowered Pr(H) by .17 at the "certain" criterion, .13 at the "confident" criterion, and .09 at the "unsure" criterion. Comparing the unmodified scene aggregates with the darkened scene aggregates shows that darkening the scene lowered Pr(H) by .04 at the "certain" criterion, .04 at the "confident" criterion, and .01 at the "unsure" criterion. Comparing the unmodified scene aggregates with the reduced contrast scene aggregates shows that reducing contrast over the entire scene lowered Pr(H) by .08 at the "certain" criterion, .09 at the "confident" criterion, and .06 at the "unsure" criterion. Comparing the unmodified scene aggregates with the gray-scale scene aggregates shows that removing color from the scene lowered Pr(H) by .04 at the "certain" criterion, .02 at the "confident" criterion, and .00 at the "unsure" criterion. These are only aggregate effects. For specific scenes, the effects will be more or less depending on the interactions in the specific scene.

MODEL EVALUATION

Our evaluation goal was to determine if the model is accurate in aggregate and if the model has systematic biases with respect to identifiable properties or characteristics of the input image. The factors and levels of the experiment were designed to enable us to assess biases with respect to a variety of characteristics known to influence target detection (e.g., size, contrast, luminance).

Outliers. The validation data set was examined to determine whether or not all the images were appropriate for applying the model. Of the 800 images with vehicles present in the validation data set used, 118 were not appropriate for applying the model. Of these 118 inappropriate images, 95 were derived from 4 of the 44 base scenes. In these scenes, one edge of the vehicle is aligned with a linear terrain feature (such as a ridge), and on the remaining three sides the vehicle had low contrast with its surroundings. When this combination of conditions occurred, the observers tended to interpret the contrast at the target edge as a continuation of the terrain feature. The model, which does

TABLE 5: Mean Pr(H) by Data Partition over All Scenes with Vehicles

Data Partition	Certain	Confident	Unsure
BAll images with vehicles	.43	.58	.80
Unmodified vehicle	.52	.68	.87
Reduced contrast vehicle	.35	.51	.75
Special variation vehicle	.40	.55	.78
Unmodified scene	.47	.62	.82
Darkened scene	.43	.58	.81
Reduced contrast scene	.39	.53	.76
Gray-scale scene	.43	.60	.82
3:1 scale factor	.40	.54	.77
2:1 scale factor	.46	.63	.83

not analyze terrain features, interpreted the contrast across the edge as visual evidence for detection. The other 23 inappropriate images were derived from 1 of the 44 base scenes: In this scene there was unusual lighting and shadowing such that a patch of glint from an elevation-angle view of the side was all that was visible. Observers could see the patch of glint but did not interpret it as evidence of a vehicle, whereas the model did. In both of these situations, the images violate the assumption of the nature of the top-down processing involved in detecting a 3-D object in a natural scene. The model validation proceeded with 682 images and their associated observer responses.

Accuracy and Explanatory Power

The direct measure of accuracy is the root-mean-square (RMS) prediction error. Over a set of N images, the RMS prediction error, ε, is

$$\varepsilon = \{\Sigma_j[Pr(H)_{\text{Predicted}}(j) - Pr(H)_{\text{Observed}}(j)]^2/N\}^{1/2}. \quad (4)$$

A related measure of performance is the explanatory power of the model. It is equal to the fraction of the variance in experimental Pr(H) explained by the model. The fraction of variance explained by the model, ϕ, is 1 minus the ratio of the RMS prediction error squared to the variance in $Pr(H)_{\text{Observed}}$ over the image set.

$$\phi = 1 - \{\varepsilon^2/\text{variance}[Pr(H)_{\text{Observed}}]\}. \quad (5)$$

The term *explanatory power* is used rather than r^2 to maintain the distinction that linear regression allows two free parameters, whereas VDM2000 has only one free parameter. The measures computed in Equations 4 and 5 understate the performance of the model because they include sampling error in empirical Pr(H) as part of the prediction error.

Figures 4 to 6 show scatter plots of observed versus predicted Pr(H) at the three response levels. The scatter between predicted Pr(H) and observed Pr(H) includes sampling error in observed Pr(H) in addition to error between predicted Pr(H) and the underlying true Pr(H). Table 6 summarizes the accuracy and explanatory power of the model and the mean error in observed Pr(H).

The VDM2000 accounts for more than 80% of the variance in Pr(H). The model remains accurate, although with somewhat reduced explanatory power at the lower confidence response levels. The accuracy of the model is actually higher at the lower confidence response levels (i.e., the RMS prediction error is lower; see Table 6). The RMS error is lower because the increased false alarm rate compresses the range of responses. The explanatory power of the model – that is, the fraction of variance in

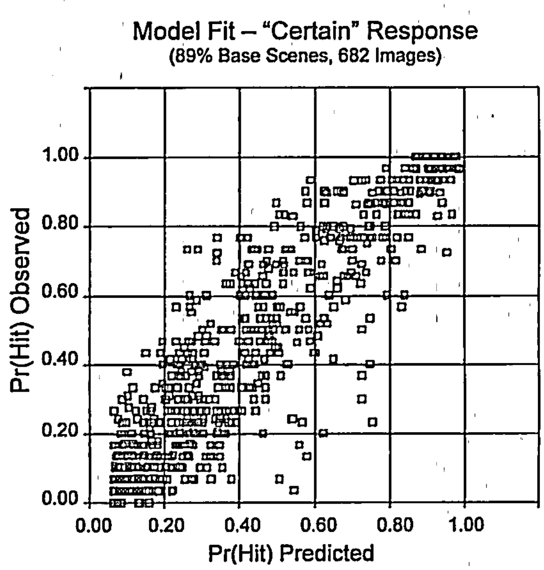

Figure 4. Predicted versus observed Pr(H) at the "certain" response level.

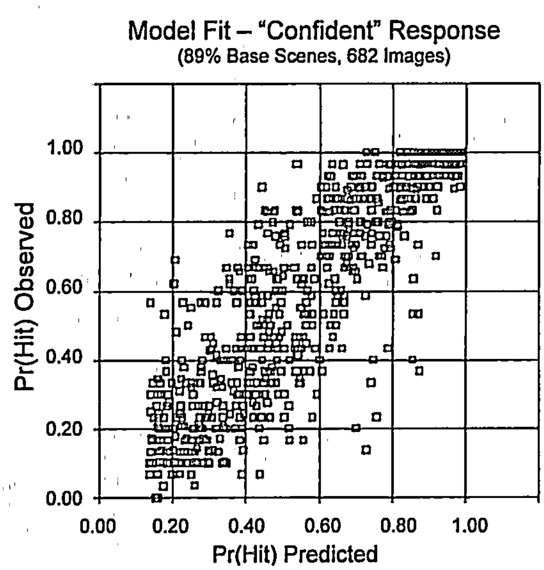

Figure 5. Predicted versus observed Pr(H) at the "confident" response level.

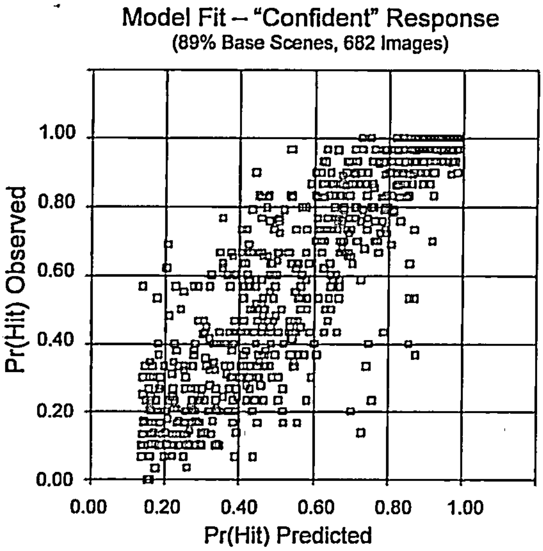

Figure 6. Predicted versus observed Pr(H) at the "unsure" response level.

Pr(H) that is explained – is a better measure of performance because it normalizes to the variance in the phenomena to be explained. The explanatory power of the model is lower at the lower confidence response levels because of the contribution of false alarms to Pr(H). The average rate of false alarms is an input to the model, but false positive responses to images add variance to Pr(H).

Bias

The term *model bias* refers to a situation in which the model's prediction errors are not zero-mean normally distributed (i.e., the model underpredicts or overpredicts actual behavior in a reliable manner). Bias in a model can be evidence that a model is improperly or incompletely specified. The prediction bias with respect to subset S, $B(S)$, of the images is

$$B(S) = \Sigma_S[Pr(H)_{Pred}(j) - Pr(H)_{Obs}(j)] - \Sigma_{All}[Pr(H)_{Pred}(j) - Pr(H)_{Obs}(j)]. \quad (6)$$

The bias of the model is shown in Table 7. The net bias is the average prediction error for the partition (i.e., factor of interest) minus the average prediction error over all the cases. A negative bias means that the model's prediction of observer hit rate was, on average, less than the empirical Pr(H) – that is, that the model underestimated Pr(H). A positive bias means that the model overestimated Pr(H). For comparison, the expected error magnitude – the sampling error in empirical Pr(H) (at the "certain" response level) divided by the square root of the number of cases – is also shown.

There are several important observations to make with regard to the data in Table 7. There are no large biases. Except for the special variations at the "unsure" response level, all of the biases are less than 2%. Pr(H) is systematically underestimated for the baseline vehicles and overestimated for the reduced signature variants, but the bias is not large compared with the adjusted sampling error.

CONCLUSIONS

The vehicle metric as computed by the VDM-2000 does a good job of accounting for variance in foveal detection performance. The VDM2000 represents a substantial contribution to both the state of the art in vision modeling and the state of the art in developmental test and evaluation tools. As previously noted, in some visual target acquisition situations, the model is not completely applicable. These are situations in which observers are deceived or misinterpret the visual signals. VDM2000 implicitly assumes that all of the perceived elements of vehicle

TABLE 6: VDM2000 Performance in Terms of Accuracy (*RMS* Prediction Error), Explanatory Power (Fraction of Variance Explained), and Mean Error

Rating	*RMS* Prediction Error	Explanatory Power	Mean Error
Certain	.136	.803	.017
Confident	.153	.740	.028
Unsure	.119	.603	.008

TABLE 7: Net Prediction Bias by Factor and Level in the Experimental Design

Data Partition	Bias at "Certain"	Bias at "Confident"	Bias at "Unsure"	Sampling Error/$N^{1/2}$
High resolution (2:1)	.0042	–.0005	.0035	.0037
Reduced resolution (3:1)	–.0028	.0005	–.0032	.0034
Darkened scene	.0045	.0030	–.0031	.0051
Lightened "haze" scene	–.0136	–.0003	.0020	.0051
Original scene	.0044	.0019	.0069	.0051
Baseline vehicle signature	–.0173	–.0192	–.0179	.0039
Reduced contrast signature	.0153	.0153	.0100	.0039
Special variation signature	.0051	.0117	.0228	.0061

signature constitute evidence for vehicle detection. More development effort could be placed in a cognition/decision-making module to help disentangle some of these effects as well as in extending the model to applications requiring higher levels of target discrimination (e.g., recognition and identification). It is possible that a richer vehicle template may be needed (e.g., distinguishing turret and chassis regions or other characteristic features, such as a gun tube). This is a trivial extension to the software; however, there are no available data with which to test and evaluate the model.

The model could potentially be adapted to target detection tasks, such as viewing thermal images. However, the characteristic vehicle regions for visual image detection will probably not be the same for thermal image detection. The characteristic visual regions (i.e., the image organization that accounts for variance over the vehicle image) are created by the vehicle surface geometry and angles with respect to illumination and observer, in concert with the observer's innate understanding of 3-D objects in a 3-D world. Thermal signatures are created by temperature gradients over the vehicle. The thermal regions (i.e., those characteristic regions that account for temperature variation) are created by various heat-generation processes and lags associated with different thermal mass. In thermal imaging applications, the vehicle image regions should reflect areas with different thermal mass (because they will heat and cool at different rates), regions corresponding to different heat sources (e.g., engine compartment, exhaust, and tracks), and surface geometry (reflection and shadowing of infrared illumination). It may be the case that the top-down processes involved in interpreting a thermal image of a scene are very different from those involved in the task of interpreting a naturally lit scene.

Computer models of vehicle detection are needed for early evaluation of design alternatives under a wide variety of terrain, lighting, and weather conditions. Field exercises have a number of key limitations that make the VDM2000 an efficient alternative: They are expensive, they require physical prototypes for each design alternative, they typically have a limited variety of terrain conditions, and their field atmospheric and environmental conditions are not controlled. The field of computer graphics is advancing rapidly in its ability to capture, represent, and render highly realistic imagery. The combination of these techniques with models of human performance such as the VDM2000 could conceivably give guidance to designers for any possible combination of design alternatives and operational conditions.

ACKNOWLEDGMENTS

This research was funded under Department of Defense Small Business Innovation Research Program Contract DAAE07-C-97-X101 with the U.S. Army Tank-Automotive and Armaments Command. The views and opinions expressed in this paper are those of the authors and do not represent a position of the sponsoring agency.

REFERENCES

Ahumada, A. J., Jr., & Beard, B. L. (1998). A simple vision model for inhomogeneous image quality assessment. In J. Morreale (Ed.), *Society for Information Display international symposium digest of technical papers* (Vol. 29, paper 40.1). Santa Ana, CA: Society for Information Display.

Boynton, R. M., & Whitten, D. N. (1970). Selective chromatic adaptation in primate photoreceptors. *Vision Research, 12,* 855–874.

Brainard, D. H. (1996). Appendix IV: Cone contrast and opponent modulation color spaces. In P. K. Kaiser & R. M. Boynton (Eds.), *Human color vision* (pp. 563–579). Washington, DC: Optical Society of America.
Commission Internationale de l'Éclairage. (1932). *Commission Internationale de l'Éclairage Proceedings, 1931*. Cambridge, UK: Cambridge University Press.
DeValois, R. L., & DeValois, K. K. (1990). *Spatial vision*. New York: Oxford University Press.
Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.
Kaiser, P. K., & Boynton, R. M. (1996). *Human color vision*. Washington, DC: Optical Society of America.
Matchko, R. M., & Gerhart, G. (2001). ABCs of foveal vision. *Optical Engineering, 40*, 2735–2745.
Mazz, J., Kistner, R., Bushra, A., & Pibil, W. (1997). *Search and target acquisition model comparison: Unaided eye analysis* (Division Note No. DN-CI-11). Aberdeen Proving Ground, MD: U.S. Army Material Systems Analysis Activity.
Moore, C., & Cavanagh, P. (1998). Recovery of 3D volume from 2-tone images of novel objects. *Cognition, 67*, 45–71.
O'Kane, B. L. (1995). Validation of prediction models for target acquisition with electro-optical sensors. In E. Peli (Ed.), *Vision models for target detection and recognition* (pp. 192–218). River Edge, NJ: World Scientific.
Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40*, 1227–1268.
Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America A, 7*, 2030–2040.
Peli, E. (1997). In search of a contrast metric: Matching the perceived contrast of Gabor patches at different phases and bandwidths. *Vision Research, 37*, 3217–3224.
Sun, J., & Perona, P. (1996). Preattentive perception of elementary three-dimensional shapes. *Vision Research, 36*, 2515–2529.
Tarr, M. J., Kersten, D., & Buelthoff, H. H. (1998). Why the visual recognition system might encode the effects of illumination. *Vision Research, 38*, 2259–2275.
Toet, A., Bijl, P., Kooi, F. L., & Valeton, M. (1998). *A high-resolution image data set for testing search and detection models* (TNO Report TM-98-A020). Soesterberg, Netherlands: TNO Human Factors.
Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer Associates.
Wilson, D. (2001). Image-based contrast-to-clutter modeling of detection. *Optical Engineering, 40*, 1852–1857.
Witus, G. (2001). *Vehicle discrimination model, VDM2000, version 2.001, final results and documentation* (Final Report to U.S. Army Tank-Automotive and Armaments Command, Contract Number: DAAE07-97-C-X101, P00005). Ann Arbor, MI: Turing Associates.
Witus, G., & Gerhart, G. (2000). A contrast metric for 3-D vehicles in natural lighting. In *RTO Meeting Proceedings 45, search and target acquisition* (pp. 12.1–12.10). Neuilly-sur-Seine, France: North Atlantic Treaty Organization, Research and Technology Organization.
Witus, G., Gerhart, G. R., & Ellis, R. D. (2001). Contrast model for three dimensional vehicles in natural lighting and search performance analysis. *Optical Engineering, 40*, 1858–1868.
Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review, 1*, 202–238.
Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Research, 9*(1), 33–39.

Gary Witus is president of Turing Associates, Ann Arbor, Michigan. He received his Ph.D. in industrial engineering from Wayne State University in 2002.

R. Darin Ellis is an associate professor at Wayne State University, where he is on the faculty of the Institute of Gerontology, Department of Industrial and Manufacturing Engineering, and Department of Biomedical Engineering. He received his Ph.D. in industrial engineering from Pennsylvania State University in 1994.

Date received: October 17, 2001
Date accepted: January 15, 2003