

## Advances in Target Acquisition Modeling II

**Thomas Meitzler**

US Army TACOM  
AMSTA-TR-R MS 263  
Warren, MI 48397-5000

**Alexander Toet**

TNO Human Factors  
Kampweg 5  
3769 DE Soesterberg  
The Netherlands

The problem of modeling electro-optical (EO) systems for the purpose of ground vehicle countermeasure development and system performance evaluation has been around for many years. This special section is devoted to recent advances in (1) computational techniques and testing procedures to predict the detectability of man-made objects in the field and in (2) methods to validate and calibrate these techniques and procedures.

Most metrics that are currently used to quantify visual target distinctness and to predict the probability of detection of a target in clutter do not relate to properties of the human visual system. As a result, their predictions do not correlate with the results of human observer tests. A well-known example is the mean square error (MSE) in intensity. Although this metric has a good physical and theoretical basis, it correlates poorly with observer performance. This is due to the fact that the human visual system does not analyze an image in a simple point-by-point manner. Bottom-up grouping mechanisms appear to drive the formation of emergent perceptual units from preattentively extracted stimulus features (e.g., edges or texture elements). When searching for known targets, top-down priming signals may influence the organization of search regions. Salient areas may then be selected for further inspection.

Only recently has there been a paradigm shift within the modeling community to transform the methods and results of recent research in the area of neurophysiology and human vision research into target acquisition modeling. However, there are still no standard and validated computational perceptual difference metrics available. Because of their computational simplicity, MSE-based measures are still widely used. Attempts to tune these metrics to the properties of the human visual system are only partly successful. These considerations have recently led to the development of visual difference metrics that are firmly based on principles of the initial stages of the human visual system.

Presently, target-acquisition-model strategies can be divided into three broad classes:

1. Variants of the “classical” approach to modeling target acquisition performance that assume a simplified target and background. Target size and average contrast are taken as the most important signature parameters for predicting target detectability. This type of model has IR and visual versions. Signal detection theory (SDT) is also used with this approach.
2. Models that use the multi-channel and multi-resolution idea adopted from human vision research together with classical psychophysics, i.e., SDT. It is assumed that the eye and the visual cortex transform the input scene into a mental image from which the observer detects a target. This is the so-called “bottom-up” approach based on first-principles of human vision and psychophysics. This is a visual model to start with but can be applied to IR scenes as well since it is in both cases the eye that looks at a displayed image on a monitor.
3. Models that use neural networks and/or fuzzy logic to predict target detectability based on the input of a data set of “feature vectors.” This type of model can be used for both IR and visual images, as well as images from radar and acoustics.

Carefully designed and performed psychophysical experiments are essential to provide data for the quantitative comparison and tuning of a model’s outcome to the judgment of observers performing visual discrimination tasks. A validated perceptual difference metric or acquisition model eliminates the need for time-consuming visual evaluation and optimization procedures involving human observers.

The NATO RTO Workshop on Search & Target Acquisition, which was held in Utrecht, The Netherlands, June 1999, was initiated by the Systems Concepts and Integration Panel SCI-12 (the former RSG-2), on “Camouflage, Concealment and Deception Evaluation Techniques.” The goal of this workshop was to provide a state-of-the-art

review of computational and psychophysical evaluation of visual target distinctness. Several of the papers in this special section were presented in an earlier form at this workshop.

Toet, Bijl, and Valetton present the TNO Human Factors SEARCH\_2 image dataset. This dataset consists of a set of 44 high-resolution digital color images of different complex natural scenes, the ground truth corresponding to each of these scenes, and the results of psychophysical experiments on each of these images. Although the dataset is small and rather limited it should be regarded as a first attempt to create a freely available database of natural imagery with corresponding human search and detection performance results that can be used to develop and validate target acquisition models and target distinctness metrics. The dataset has already been used in more than ten different studies in the literature, ranging from studies evaluating target detectability metrics to eye movement studies and attempts to model the human visual system. The following eight papers in this special section address the SEARCH\_2 dataset.

O'Kane, Bonzo, and Hoffman discuss the challenges involved in perception studies conducted to gain insight into surveillance and target acquisition by military users of thermal imagery. The goal is to emulate as accurately as possible what a military observer will actually see and how he will use the sensor to detect and identify targets. The issues include prior training, panning effects on eye movements, and contrast and brightness controls. The latest advances in these areas and some remaining challenges are discussed.

Doll and Home argue that the scope of most current human visual search and target acquisition (STA) models is restricted because only a limited part of the visual system is taken into account. He emphasizes the importance of complex pattern perception, visual attention, learning, and cognition for STA performance and suggests approaches for modeling them. He also provides guidelines for testing and validating STA models. Finally, he presents and compares alternative approaches to field testing for the purpose of model validation.

Itti, Gold, and Koch present a bottom-up model of visual attention based on the architecture of the primate visual system. The model is based on the assumption that preattentive target selection is stimulus driven. Orientation, color, and intensity information is combined into a single 2-D map that encodes the visual saliency of objects in the visual field. Competition among neurons in this map gives rise to a single winning location that corresponds to the most salient object, which constitutes the next target. If this location is subsequently inhibited, the system automatically shifts to the next most salient location, endowing the search process with internal dynamics. Application of the model to the SEARCH\_2 image set shows that the model finds the targets faster than human observers in 75% of the studied cases. It is argued that this may be a result of the lack of top-down flow of information that may bias attentional shifts in human observers.

Garcia et al. present a new computational method to quantify the visual distinctness of a target relative to its background. First they compute the optimal interest points

in a target scene. These points are defined as the spatial locations of partially invariant features that minimize the detection error probability between the scene with and without the target. Then they compute the visual target distinctness as a generalization of the Kullback-Leibler joint information gain over the optimal interest points of the target image. The method is applied to quantify the visual distinctness of targets in the SEARCH\_2 image set. The results show that the computed target distinctness correlates strongly with visual target distinctness as estimated by human observers.

Krebs, Scribner, and McCarley compare and contrast behavioral and matched filter ROC plots to determine whether matched filtering is a good predictor of human performance in search and detection tasks on short-, mid-, and long-wave infrared and gray-level and color fused imagery. They conclude that a matched filter can predict human visual sensitivity for different sensor types by target characteristics. Matched filtering may therefore be used for rapid system prototyping, for the optimization of image enhancement methods, and for the development of multispectral image fusion algorithms.

Nilsson introduces a distinctness measure based on the relative number of neural pathways required to process a target. Using images from the SEARCH\_2 dataset, he determined recognition distances for target vehicles. The rationale for this approach is that informative (highly visible, clearly delineated) targets require only few neural pathways at recognition threshold, corresponding to a small retinal projection area or, equivalently, a large recognition distance, whereas less informative (less visible, obscured, or camouflaged) targets require more processing power, and therefore a larger retinal projection or a smaller recognition distance. The results are compared with the search times provided with the SEARCH\_2 images. This comparison indicates that recognition distance thresholds effectively quantify target distinctness, in a way that is complementary to search time. Recognition distance thresholds correspond to the number of neural pathways required for recognition (retinal projection area). Search time corresponds to the duration required for recognition. Together, recognition distance thresholds and search time describe the total amount of information required for recognition.

Birkemark presents the CAMEVA model, which is a methodology developed at the Danish Defence Research Establishment (DDRE) for computational CAMouflage EVALuation and for estimation of target detectability. CAMEVA computes the dissimilarity between the statistical distributions of a set of features on a target and a corresponding set on its local background, using digitized imagery as input. The selected features depend on the detection system that is modeled. In the case of the unaided human eye typical features are contrast, texture, shape, and edge content. CAMEVA predicts the target detectability as a function of range from the dissimilarity measure and the limitations of the sensor (visual) system. CAMEVA is a man-in-the-loop model, since it requires human operator interaction to delineate the target and its local background. This paper presents validation experiments and the results of the application of the model to the SEARCH\_2 dataset.

Meitzler, Sohn, Singh, and Elgarhi discuss their research in the modeling area of predicting the probability of detection. Their approach is to use the SEARCH\_2 dataset to build and test a prediction model based on the fuzzy logic approach. The authors have achieved a 0.9 correlation to experimental results by using half the data set for training the model and half the data set for testing.

Wilson combines contrast, size, and clutter metrics to predict human observer performance on the SEARCH\_2 dataset. To calculate the contrast metric, a new image is generated from a gray-scale version of the original image by replacing the target with an "expected background" using the local background surrounding the target. The contrast metric is then obtained from the difference of this new image and the original image. The ratio of the contrast and clutter metrics is shown to correlate with human performance.

Witus, Gerhart, and Ellis introduce a contrast metric that accounts for the 3-D structure of target vehicle. First it computes the contrast for the front (or rear), side, and top surfaces. Then it computes the overall target contrast as a weighted sum of the contrasts of the component surfaces. The metric is applied to the ground target vehicles in the SEARCH\_2 dataset. The metric values are compared to experimental observer results. When the effects of false alarms are discounted, the metric accounts for 89% of the variance in the probability of detection and 95% of the variance in search time.

Nyberg and Bohman applied a number of texture descriptors and similarity metrics to quantify the distinctness of the targets in the SEARCH\_2 images relative to their local background. Using only one or two texture features they achieved a high correlation with human observer performance. The best results were obtained with edge concentration and shape of the local Wiener spectrum as texture descriptors, in combination with mean and variance based distance measures.

Aviram and Rotman address the effects of imagery wavelength on the agreement level between various image metrics and human detection performance for targets embedded in natural scenes. The metrics studied were designed to agree with human perceptual cues. The metrics were applied to natural scenes registered in the 3–5  $\mu$ , the 8–12  $\mu$ , and the visual bands of the spectrum. The results were correlated with human performance measures. It is found that scene complexity dominates human detection performance for longer wavelengths, and local target distinctness correlates with performance for short wavelengths. A statistical texture metric is shown to correlate strongly with human performance, independent of wavelength.

Copeland and Trivedi performed two psychophysical experiments to test human search and discrimination performance for natural texture patterns in natural backgrounds. In the first experiments the subjects judged the relative visual target distinctness in a paired comparison paradigm. In the second experiment the observers searched a natural scene for suspected target locations, while their eye movements were recorded. Of all the metrics considered, a metric based on a model of image texture correlated most strongly with human performance data.

Moorhead et al. present a synthetic scene simulation system (CAMEO-SIM) that generates high-fidelity imagery within the 0.4–14  $\mu$ m spectral band. The system consists of a scene design tool, an image generator, which incorporates both radiosity and ray-tracing processes, and an experimental trials tool. The scene design tool allows the user to develop a three-dimensional representation of the scenario of interest from a fixed viewpoint. Targets of interest can be placed anywhere within this 3-D representation and may be either static or moving. Different illumination conditions and effects of the atmosphere can be modeled together with directional reflectance effects. The user has complete control over the level of fidelity of the final image. The output from the rendering tool is a sequence of radiance maps that may be used by sensor models or for experimental trials in which observers carry out target acquisition tasks. A range of verification and validation tests is also discussed.

Krapels et al. argue that the performance of infrared target acquisition systems is limited by atmospheric turbulence for long-range imaging paths. The effects of atmospheric turbulence blur should therefore be represented in target acquisition models. They show that the effects of turbulence blur on detection and recognition tasks can in good approximation be modeled as a linear shift invariant process.

Watkins et al. report the results of visual search and target detection experiments for binocular viewing of single line of sight images versus stereoscopic display of wide baseline stereo images. The results indicate that stereo vision effectively reduces false alarm detection by a factor of two. Guidelines for optimum stereo display are obtained that can be used to improve target detection.

We are pleased with the manuscripts submitted for this special section and the interest in the subject of target acquisition modeling. We would like to thank all the authors and the reviewers for their contributions. We hope you will enjoy these papers and find them useful in your studies related to target acquisition.

#### *Acknowledgments*

We dedicate this special section to Mr. Randall Williams, the originator, chairman, and driving force behind NATO RTO SCI-12, who brought us all together.

We gratefully acknowledge the financial support from the following sponsors, who made the Workshop on Search and Target Acquisition possible: Turing Associates, Inc. (USA), United States Air Force European Office of Aerospace Research and Development, United States Army Research Development and Standardization Group (UK), United States Office of Naval Research, Europe, and THALES Ground Based (NL).



**Thomas J. Meitzler** received his BS and MS in physics from Eastern Michigan University and a PhD in electrical engineering from Wayne State University in Detroit, Michigan. His PhD thesis was "Modern methods of computing the probability of detection in cluttered scenes." He has coauthored several papers in the fields of target detection, infrared and visual system simulation, fuzzy logic, and wavelets. He has held adjunct teaching

positions at the University of Michigan-Dearborn in both the physical science and electrical and computer engineering departments. From 1988 to the present he has been a research scientist at the US Army TACOM Research and Engineering Center (TARDEC), Survivability Technology Center. He has been involved with the validation verification, and development of electro-optical, visual detection models, and atmospheric propagation computer programs. He directs the operation of a visual perception laboratory at TARDEC for the testing of civilian and military ground systems. He was a recipient of the 1995 Army Research and Development Award for the Computational Vision Model Development for Dual-Use Applications.



**Alexander Toet** received his PhD in physics from the University of Utrecht in the Netherlands in 1987. He has worked on visual spatial localization (hyperacuity) and image processing. His current research interests are image fusion, image quality, computational models of human visual search and detection, and the quantification of visual target distinctness.