

Image dataset for testing search and detection models

Alexander Toet, MEMBER SPIE

Piet Bijl

J. Mathieu Valetton, MEMBER SPIE

TNO Human Factors

Kampweg 5

3769 DE Soesterberg

The Netherlands

E-mail: toet@tm.tno.nl

Abstract. The TNO Human Factors Search_2 image 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. The images in the Search_2 dataset are a subset of a larger set that has been used in a visual search and detection experiment. Each scene (image) contains a single military vehicle that serves as a search target. The image dataset, an Excel file with the ground truth and observer data, and an extensive report describing the dataset are available on CD-ROM (requests by email to the first author). The dataset can be used to develop and validate digital metrics that compute the visual distinctness (contrast, conspicuity, or saliency) of targets in complex scenes, and models of human visual search and detection. 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. In addition to this work, eight other articles in this special section address the Search_2 dataset. © 2001 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.1388608]

Subject terms: Search_2 dataset; visual search; conspicuity; psychophysics.

Paper ATA-01 received Oct. 15, 2000; revised manuscript received Feb. 3, 2001; accepted for publication Mar. 23, 2001.

1 Introduction

The TNO Human Factors Search_2 image dataset has been compiled to provide a set of data that can be used to evaluate and validate digital metrics and early vision models that compute the visual distinctness (conspicuity, saliency) of targets in complex natural scenes, and models of the human visual search and detection capability. We give a concise overview of the contents of this dataset. An extensive description is given elsewhere.¹

1.1 Measuring Target Distinctness

Human observer experiments designed to quantify visual target distinctness usually involve search and detection tasks or contrast detection tasks.

Targets that are highly distinct are usually noticed quickly. In contrast, targets that are indistinct (hard to detect, highly similar to their surroundings) usually yield large search times. Visual search and detection tasks yield temporal measures that characterize visual target distinctness, like the cumulative detection probability (i.e., the fraction of all observers that detect the target after a given amount of search time) and the mean search time.

The concept of visual lobe or conspicuity area is an overall measure of target distinctness that captures all factors contributing to the visual contrast of a target and its surroundings. It can operationally be defined as the peripheral area around the central fixation point from which specific target information can be extracted in a single glimpse.²⁻⁴

The size and shape of the conspicuity area have been measured for a range of static targets in static scenes.²⁻⁸ It is found that the conspicuity area is small if the target is embedded in a complex background (a surrounding with high feature variability) or if the target is surrounded by irregularly positioned nontargets of high similarity (a surrounding with high spatial variability). The conspicuity area is large if the target stands out clearly from a homogeneous background.

TNO Human Factors recently developed a simple and efficient psychophysical procedure to quantify the visual conspicuity of a target in a complex (natural) scene,^{9,10} which has been successfully applied to a range of practical problems involving the optimization of target visibility.¹¹ With this procedure visual conspicuity can quickly and easily be determined. Only a few (typically two or three) observers are needed to obtain sufficient accuracy. It has been shown that conspicuity determines human target acquisition performance in realistic and military relevant complex scenarios.⁹ Also, conspicuity measured on photographic slides agrees with conspicuity measured in the field.^{9,10} This implies that conspicuity can be used in combination with photosimulation studies to obtain estimates of human observer performance in the field.

1.2 Computing Target Distinctness

A range of different metrics have been proposed to compute the visual distinctness of image subregions (target areas)

from digital imagery. Visual distinctness metrics can be used to compare and rank target detectability, and to quantify background or scene complexity.

Some metrics are based directly on the sampled luminance values, others involve some nonlinear or noninvertible transformation on the pixel values. They may be computed locally over the target area and its immediate surround, semilocally on distinct locations in the scene, or globally over the entire scene.

Local metrics or signal-to-noise ratios quantify the distinctness of a target in its immediate surroundings. The general idea is that a target that is highly similar to its local background will be hard to see. Semilocal metrics are based on the calculation of likely fixation points for a human observer searching the scene for a target. The general idea is that a target will be hard to find (will be inconspicuous) if the inspection of the scene requires a large number of fixations. Fixation points are assumed to correspond to local extrema of variance of the graylevel distribution;¹² busyness (e.g., Ref. 13); curvature of the edge map;^{14,15} or symmetry of the graylevel distribution;^{16,17} or saliency, which may be any combination of the output of generalized difference operators operating on length, orientation, contrast, contour curvature, size, perimeter, and average graylevel.^{18–20}

Global conspicuity metrics or signal-to-clutter ratios take into account the overall structural composition of the scene. The general idea is that a target situated in a busy scene (a complex scene with a large amount of detail similar to the target, or a scene with a high structural variability) will be less conspicuous than the same target situated in a relatively empty scene (a scene with low variability).

Current target acquisition models typically use first- or second-order statistical metrics to describe the scene information content. First-order metrics are only a function of pixel intensities and contain no information about relative pixel locations (spatial image structure). Second-order metrics do contain some spatial information, but it is very difficult to determine how much and whether it is relevant to human spatial vision.

Computational models of early human vision typically process an input image through various spatial and temporal bandpass filters and compute first-order statistical properties of the filtered images to compute a target distinctness metric.^{21–24}

Models of the human visual search and detection capability that predict the detection probability for targets in complex (natural) scenes as a function of time (e.g., Visdet,²⁵ Oracle,²⁶ and GTV²⁷) typically require a large number of input parameters. The most important apparent scene parameters are: the mean target luminance, mean target background luminance, overall luminance level of the scene, local luminance contrast, angular size of the target, and the amount of clutter in the scene.

In addition, some models require the display gamma function²¹ or the RGB-XYZ color coordinate transform.²³ Most of the previously mentioned parameters are supplied for the images in the Search_2 dataset.¹

2 Search_2 Image Dataset

2.1 General Information

The TNO-TM Search_2 dataset contains a well documented collection of high resolution digital images. Each image represents a military vehicle in a complex rural background (Fig. 1(a)). The exact position in the scene, the viewing distance, and the orientation of each vehicle are given in a separate datasheet. Target mask (binary) images are included (Fig. 1(c)), so that each target can easily be located (extracted), and computations can be restricted to the visible parts of the target support (Fig. 1(b)). Close-up views of all targets are also provided (Fig. 2). A digitized Kodak color calibration slide allows the user to compute the gamma correction and the XYZ-RGB transform. Finally, human visual search time statistics and visual lobe measurements of the target are provided for each scene.

2.2 Image Registration

The original images were taken during the distributed interactive simulation, search, and target acquisition fidelity (DISSTAF) field test, which was designed and organized by Night Vision and Electronic Sensors Directorate (NVESD), Fort Belvoir, Virginia, and which was held in May and June 1995 in Fort Hunter Liggett, California. The scenes were captured on Kodak 5045 EB 100 36×24-mm color slides using a Canon T70 camera equipped with a 300-mm lens. This corresponds to a field of view of 6.9 deg horizontal by 4.6 deg vertical.

A Kodak PCD Imaging Workstation 4220, equipped with a Kodak Professional PCD Film Scanner 4045, is used to digitize the slides in 64-Base Kodak PCD format onto a Kodak Digital Science Pro Photo CD Master Disc. This scanner digitizes 6144×4096 pixels over a 35-mm frame of film, whose dimensions are 36×24 mm (1.42×0.94 in). This translates to a scanning resolution of approximately 4400 pixels/inch in both dimensions. The angular size of one (square) pixel is therefore approximately 4 sec of arc. The scene balance algorithm of the Kodak PCD Imaging Workstation 4220 was turned off.

2.3 Image Data

The nine military vehicles that serve as visual search targets are listed in Table 1 and shown in Fig. 2 (an extensive description and detailed images of these vehicles are given elsewhere²⁸).

The ground truth and observer data for each of the 44 images in the Search_2 dataset includes the values of the following parameters:

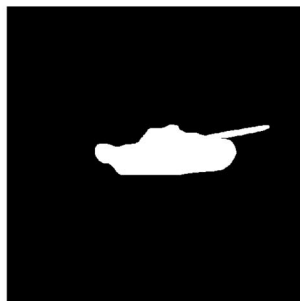
- the distance (in meters) between the target and the location of the camera
- the aspect angle of the vehicle (in degrees)
- the coordinates of the target center (in pixels) in the 6144×4096 images
- the horizontal and vertical size of the visible (i.e., not occluded) parts of the target (both in pixels and in meters)
- the target area, defined as the number of pixels that represent the visible part of the target



(a)



(b)



(c)

Fig. 1 (a) A full size (6144×4096) target scene, (b) a 512×512 subregion containing the target, as indicated by the white outlines in (a), and (c) the corresponding subregion of the binary mask representing the visible parts of the target support.

Table 1 Targets occurring in the Search_2 images.

Type	Description	Length (m)	Width (m)	Height (m)
BMP-1	Russian armed personnel carrier (APC)	6.74	2.94	1.94
BTR-70	Russian APC	7.54	2.80	2.13
HMMVV-Scout	Jeep (general purpose vehicle)	4.72	2.18	1.83
HMMVV-Tow	Jeep with antitank weapon	4.72	2.18	1.83
M1A1	American main battle tank (MBT)	7.92	3.65	2.96
M3-Bradley	American APC	6.45	3.30	2.97
M60	American MBT	6.95	3.63	3.29
M113	American APC	4.86	2.69	2.20
T72	Russian MBT	6.91	3.46	2.19

- the mean luminance (in cd/m^2) of, respectively, the scene, the target, and the surround of the target. Some targets have clearly distinct bright and dark parts. For these targets the mean luminance of the bright and dark parts are given separately. Some targets are partly surrounded by grass and partly by trees. For these targets the mean luminance of the grass and the mean luminance of the trees are given separately. The illuminance of the scene is also given (in lux).
- the number of correct detections, the number of false responses, and the number of missed detections, for a total of 62 observers



Fig. 2 The nine military vehicles used as search targets (for an explanation of the abbreviations see Table 1).

- the mean, geomean, and median search time, and the size of the visual lobe for detection and identification, respectively.

The visibility of the targets varies largely throughout the entire stimulus set. This is mainly due to variations in the structure of the local background, the viewing distance, the luminance distribution over the target support (shadows), the orientation of the targets, and the degree of occlusion of the targets by vegetation.

2.4 Target Close-Ups

Search and detection models that use learning methods, e.g., neural nets, to derive characteristic target classification features need to be trained on high resolute target representations. The same holds for methods that depend on template matching and correlational approaches. The dataset therefore includes three different close-up views of each target vehicle: a front view, an oblique front view, and an oblique rear view. These images can also be used to familiarize observers performing search and detection tasks with the scenes.

2.5 Target Masks

Some computational methods require a precise definition of the target support. Other methods need a definition of the local background²⁴ or the definition of a zone within which background elements are able to interfere with target details.²⁹ The image set therefore includes masks of the visible parts of the target support. A local background support can easily be constructed by dilating these masks with an appropriate structuring element, e.g., by using the maximum filter in Photoshop with a square structuring element³⁰ and subtracting the original mask from the dilated mask. The size of the structuring element determines the size of area around the target support, which is considered as the local background.

2.6 Calibration Image

The information that is needed to compute the conversion of pixel values to display luminance and chromaticity values is also supplied with the dataset.

An opaque slide was digitized to quantify the effect of the dark current of the Kodak PCD Film Scanner 2000. A blank slide was also digitized to test the homogeneity of the digitizer and its maximum digital output value.

A Kodak calibration slide is included to enable the estimation of the display gamma function (from the grayscale patches) and the RGB-XYZ transformation matrix (from the color patches). This slide was projected on the same screen that was used in the observer experiments. The (x,y) chromaticity coordinates of the Kodak Color Control Patches and the luminance coordinate (Y ; cd/m^2) of the Kodak grayscale patches, as measured from the image projected onto the screen, are included in the dataset.

3 Search Experiment

The 44 digitized target scenes in the TNO Human Factors Search_2 dataset are part of a larger set of images that have been used as stimuli in a visual search and detection experiment. As a result, there are a number of measures available that characterize observer performance for each of

these images. The following sections briefly describe the experiment and its results. A detailed description is given elsewhere.^{9,31}

3.1 Stimuli

The 256 images used in this experiment represent 44 different rural scenes and are captured at Fort Hunter Liggett, California. This set includes 144 images that contain exactly one military vehicle that serves as a search target and 112 images without a target.

3.2 Apparatus

A Kodak Ektapro 7000 carousel slide projector, equipped with a 90-mm lens, was used to project the slides onto a white screen. A second projector was used to create a bright boundary around the projected scene. A PC was used to control the stimulus presentation and record the observer responses.

3.3 Procedure

This section presents a brief description of the experimental procedure. A detailed description of the experimental design is given elsewhere.^{9,31}

3.3.1 General procedure

First the visual acuity of the subject was tested. Then the subject was shown three close-ups of each of the nine targets (27 slides in total): one front view and two side views. The presentation of the close-ups of the vehicles merely served to familiarize the subject with the outline of the search targets. The subject was instructed not to memorize the targets or their characteristic details, since there is no need to classify the targets in the actual search experiment. The intention of this instruction is to remind the subject that the targets are man-made objects with straight and clear-cut edges and sharp corners, in contrast to natural objects (rocks, bushes) that generally have jagged and fuzzy edges.

After the presentation of the close-up views of the targets, a test run consisting of ten trials was performed to familiarize the subject with the visual search procedure. The observer was free to choose a search strategy, and was not requested to scan the image in a certain order or to start the search at a certain location.

A computer was used to control the order and duration of the presentation of the stimuli and to register the response times and the estimated target locations.

A search trial started with the presentation of a new scene. The subject's task was to search for a military vehicle in the image and to press the space bar of the computer keyboard immediately on detection of the target. The temporal interval that elapsed between the onset of the displayed scene and the moment the subject indicated that the target had been found (by pressing the space bar) was registered and adopted as the search time. The displayed scene disappeared immediately after the subject responded, and a slide showing a 10×10 rectangular grid with cells numbered from 0 to 99 was displayed instead. The subject was then requested to indicate the perceived location of the target by entering the number of the grid cell that covered the perceived target location. The total duration of the stimulus presentation was limited to 60 s. An auditory warning sig-

nal was presented 50 s after the onset of the presentation if no response has been registered during the preceding presentation interval. In the remaining 10 s of the presentation interval, the subject either selected the most likely target from a number of perceived candidate targets (i.e., anything in the scene that to some degree resembles a target), or entered a “don’t know” response code. This procedure serves to prevent the occurrence of extremely long search times, to reduce the number of “don’t know” responses, and to increase the number of false alarms.

The analysis of false alarms is very important, since this may indicate what features are used by human observers to detect the presence of a target.

The 256 slides were distributed over four carousels of 64 slides each. A Latin square design³² was used to control the order in which the slides within each carousel and the carousels themselves were presented to the subjects. This is done to minimize learning effects in the course of the experiment. The duration of a complete run was about two hours.

3.3.2 Scene familiarity

Familiarity with a scene may affect search performance, since an observer may notice a change in the scene without actually distinguishing a target. To analyze this effect the observers were divided into two groups. One group was only presented with the target images. For the other group, the presentation of a target image was preceded by the presentation of the corresponding empty scene. The empty scene was presented during 15 s. The time delay between the presentation of the empty scene and the presentation of the target image was 3 s. For 8 of the 44 scenes, no empty scene was available. In this case a text slide containing the message “No empty scene” was presented for three seconds prior to the target image presentation.

3.3.3 Edge effect

Observers tend to fixate on the center of an image.³³ As a result search times for targets near the edges of an image can be up to 30% longer than search times for the same targets when they are close to the center of the image, independent of the target distinctness.³⁴ The image set used in this experiment includes each scene either two or four times (28 scenes were captured with the target at four different locations in the camera field of view, and 16 scenes with the target at two different locations, amounting to a total of 144 images). The different images of the same scene correspond to different orientations of the optical axis of the camera, so that the target is at different locations in the field of view. Estimates of the previously mentioned edge effect can therefore be obtained from the analysis of the corresponding observer data.

3.4 Subjects

A total of 64 civilian observers, aged between 18 and 45 years, participated in the experiment. Approximately half of these were women, the other half were men. All subjects have (corrected to) normal vision, with an acuity better than 1.25 arcmin⁻¹.

3.5 Viewing Conditions

Viewing was binocular. The experiments were performed in a dimly lit room. The images are captured on 36×24-mm slides using a 300-mm lens. This corresponds to a field of view of 6.9 deg horizontal by 4.6 deg vertical. The slides were projected onto a screen area with a width of 1.65 m and a height of 1.11 m. The distance between the lens of the projector and the screen was 4.13 m. The observer was seated at a distance of 1.3 m from the screen. At this viewing distance the projected image subtends 65 deg horizontal by 46 deg vertical. This corresponds to an enlargement of the original scene by a factor of about 10.

A second projector was used to present a continuously visible 13-deg-wide bright border around the projection of the search scenes. This bright border serves to reduce the variations in the adaptation level of the observers resulting from large overall variations in brightness that occur between successive slide presentations.

3.6 Results and Discussion

The basic data collected in this experiment are the detection time and the correctness of the response (whether the true or false target was detected). The correctness of the response was determined by comparing the reported location of the target with its actual location (stored in file for each target image). These data are collected for each of the 144 target images, and for 64 observers, making a total of 9216 responses. For each target image the mean, median, and geometric mean of the search time, are computed for all correct detections. In addition, the number of correct detections, the number of false detections, and the number of missed targets are given.

The time needed to detect a target depends on the position of the target in the scene, the starting position of the eye, and the strategy that is used. As a result, temporal measures of visual search performance are highly variable. A large number of repetitions is therefore required to obtain a reliable estimate of a search time measure, e.g., the mean search time. Each observer should view each scene only once (the search is obviously over once the target is found). This means that a large number of observers are required to perform a search experiment.

Search performance is usually expressed as the cumulative detection probability as a function of time, and approximated by³⁵⁻⁴¹

$$P_d(t) = \begin{cases} 0 & : t < t_0 \\ P_\infty [1 - \exp(-t - t_0)/\tau] & : t \geq t_0 \end{cases} \quad (1)$$

where $P_d(t)$ is the fraction correct detections at time t , P_∞ is the probability of a correct response after an infinite amount of search time, t_0 is the minimum time required to respond, and τ is a time constant.

As Eq. (1) clearly shows, search times are not normally distributed. Therefore, simple search time statistics may not correctly characterize observer performance. In most situations, search time is restricted and $P_d(t)$ need not approach P_∞ for targets that are hard to find. As a result, the mean of the observed search times may underestimate the true mean, i.e., the mean over the observation period may be less than the mean that would result if the search went on

for an infinite amount of time. Other measures may therefore be more suitable to characterize observer search performance, such as the median, geometric mean, and upper and lower quartiles.

3.6.1 Scene familiarity

The analysis of the results shows no effect of scene familiarity on search performance. Therefore, in all further analyses, no distinction is made between target image presentations with and without a preceding empty scene, and the data for all remaining 62 observers are combined.

3.6.2 Edge effect

The results show a strong effect of target location on average search time. Search times are lowest for target locations near the center of the image, and up to 50% longer for targets that are positioned near the edges of the image. This result agrees with the finding that observers preferably perform searches near the center of the display.³³ However, for the present experiment this hypothesis can not be verified, since no eye movements were recorded. For the stimulus set used in this experiment, the target positions are carefully balanced over the display area (over different angular distances from the center of the screen and over different radial orientations). As a result, the edge effect only contributes about 3% to the total standard error in the search time data. Since the total standard error in the search time data for each individual target image is about 14%, the contribution of the edge effect is relatively small.

The 44 target images selected for the Search_2 dataset correspond to the target locations that are closest to the center of the image. This is done to ensure that the target images are surrounded on all sides by a large fraction of the background, which is a prerequisite for some computational algorithms.

4 Visual Lobe Measurements

4.1 Concept

Target conspicuity is operationally defined as the maximum distance between target and foveation, measured in the fronto-parallel plane through the target (i.e., in the plane that is parallel to the image plane and at the same distance from the observer as the target), at which the target can still be visually resolved from its surroundings. This conspicuity measure has been shown to be⁹ independent of viewing distance, consistent among observers, and meaningful in the sense that it correlates with search time. The conspicuity distance is easy and quick to measure with only a few observers, and can be used on familiar scenes.¹¹

The conspicuity distance is closely related to the concept of conspicuity area, which is operationally defined as the peripheral area around the central fixation point from which specific target information can be extracted in a single glimpse.²⁻⁴ The conspicuity area is small if the target is embedded in a complex background (a surround with high feature variability) or if the target is surrounded by irregularly positioned nontargets of high similarity (a surround with high spatial variability). The conspicuity area is large if the target stands out clearly from a homogeneous background.

4.2 Procedure

The conspicuity distance measurement procedure is as follows. The slide representing the scene and the target is projected continuously. First a moveable fixation dot is superimposed on the projected image by means of a laser pointer. This fixation dot is initially positioned at a large angular distance from the target location. Subjects are then instructed to move the pointer slowly in the direction of the target while fixating the laser dot projected on the screen, and to stop moving the pointer when the target first becomes noticeable. The image is then replaced by the projection of a reference grid with numbered cells, and the position of the fixation dot relative to this grid is recorded. Since the position of the target is known, the distance from the target at which its visibility is first reported can then be computed. This distance is adopted as the characteristic spatial extent of the conspicuity area of the target.

A subject can use two different criteria to determine whether the target is visible or not. The first criterion is whether there is anything at the location of the target that contrasts with the local background in some way (color, shape, texture, luminance, etc.). This criterion yields a visual lobe for the detection of the target. The second criterion that can be used is whether the spatial structure at the location of the target really originates from the target (can be discriminated as being the target). This criterion yields a visual lobe for the identification of the target.

Each lobe measurement is repeated at least three times. Subjects are usually able to make a setting within one minute.

4.3 Results

The detection lobes range between 0.73 and 3.58 deg. The identification lobes range between 0.06 and 2.66 deg.

5 Concluding Remarks

The Search_2 image dataset has already been used in the literature to establish the relation between visual conspicuity and search time,⁹⁻¹¹ develop and validate computational visual target distinctness metrics,⁴²⁻⁴⁷ validate visual search and detection models and metrics,^{48,49} and develop computational models of human visual attention.⁵⁰⁻⁵³ In the current special section, eight articles address the Search_2 dataset.⁵⁴⁻⁶¹

The authors are currently involved in experiments in which the eye fixations of human observers searching the scenes will be recorded. These data will be made available in future updates of the dataset.

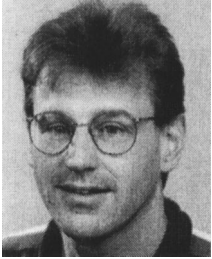
References

1. A. Toet, P. Bijl, F. L. Kooi, and J. M. Valeton, "A high-resolution image data set for testing search and detection models," Report TNO-TM 1998 A020, TNO Human Factors, Soesterberg, The Netherlands (1998).
2. F. L. Engel, "Visual conspicuity. Directed attention and retinal locus," *Vision Res.* **11**, 563-575 (1971).
3. F. L. Engel, "Visual conspicuity and selective background interference in eccentric vision," *Vision Res.* **14**, 459-471 (1974).
4. F. L. Engel, "Visual conspicuity, visual search and fixation tendencies of the eye," *Vision Res.* **17**, 95-100 (1977).
5. J. R. Bloomfield, "Visual search in complex fields: size differences between target disc and surrounding discs," *Hum. Factors* **14**, 139-148 (1972).
6. Y. M. Bowler, "Towards a simplified model of visual search," in *Visual Search*, D. Brogan, Ed., pp. 303-309, Taylor & Francis, London (1990).

7. B. L. Cole and S. E. Jenkins, "The effect of variability of background elements on the conspicuity of objects," *Vision Res.* **24**, 261–270 (1984).
8. S. E. Jenkins and B. L. Cole, "The effect of the density of background elements on the conspicuity of objects," *Vision Res.* **22**, 1241–1252 (1982).
9. A. Toet, F. L. Kooi, P. Bijl, and J. M. Valeton, "Visual conspicuity determines human target acquisition performance," *Opt. Eng.* **37**(7), 1969–1975 (1998).
10. A. Toet, P. Bijl, F. L. Kooi, and J. M. Valeton, "Quantifying target distinctness through visual conspicuity," in *Targets and Backgrounds: Characterization and Representation IV*, W. R. Watkins and D. Clement, Eds., *Proc. SPIE* **3375**, 152–163 (1998).
11. A. Toet and F. L. Kooi, "Conspicuity: an efficient alternative for search time," in *Vision in Vehicles*, A. G. Gale, I. D. Brown, C. M. Haslegrave, and S. P. Taylor, Eds., pp. 451–462, Elsevier Science Ltd., Oxford (1999).
12. H. P. Moravec, "Towards automatic visual obstacle avoidance," in *Proc. 5th Int. Joint Conf. Artificial Intell.*, pp. 584–590, Cambridge, MA (1977).
13. P. Burt, "Smart sensing with a pyramid vision machine," *Proc. IEEE* **76**, 1006–1015 (1988).
14. Y. Lamdan, J. T. Schwartz, and H. Wolfson, "On recognition of 3-d objects from 2-d images," in *Proc. IEEE Int. Conf. Robotics Automation*, pp. 1407–1413, Philadelphia, PA (1988).
15. Y. Yeshurun and E. L. Schwartz, "Shape description with a space-variant sensor: algorithm for scan-path, fusion, and convergence over multiple scans," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(11), 1217–1222 (1989).
16. D. Reisfeld, H. Wolfson, and Y. Yeshurun, "Detection of interest points using symmetry," *Proc. Third. Int. Conf. Computer Vision*, pp. 62–65, IEEE Computer Society Press, Washington (1990).
17. D. Reisfeld, H. Wolfson, and Y. Yeshurun, "Context-free attentional operators: the generalized symmetry transform," *Int. J. Comput. Vis.* **14**, 119–130 (1995).
18. R. Milanese, "Detecting salient regions in an image: from biology to implementation," PhD Thesis, Computer Science Department, University of Genova, Switzerland (1993). Available via ftp at <http://cui.unige.ch/as/pub/milanese/thesis>.
19. R. Milanese, J.-M. Bost, and T. Pun, "Bottom-up attention system for active vision," in *Proc. 10th European Conf. Artificial Intell.*, pp. 808–810, John Wiley, New York (1992).
20. R. Milanese, H. Wechsler, S. Gil, J.-M. Bost, and T. Pun, "Integration of bottom-up and top-down cues for visual attention using non-linear relaxation," in *Proc. IEEE Conf. Computer Vision Patt. Recog.*, pp. 781–785 (1994).
21. A. J. Ahumada, A. M. Rohaly, and A. B. Watson, "Image discrimination models predict object detection in natural backgrounds," supplement to *Invest. Ophthalmol. Visual Sci.* **36**(4) abstract 2013, S439 (1995). Available at <http://vision.arc.nasa.gov>.
22. T. J. Doll, S. W. McWhorter, and D. E. Schmieder, "Target and background characterization based on a simulation of human pattern perception," *Proc. SPIE* **1967**, 432–454 (1993).
23. G. Gerhart, T. Meitzler, E. Sohn, G. Witus, G. Lindquist, and J. R. Freeling, "Early vision model for target detection," *Proc. SPIE* **2470**, 12–23 (1995).
24. R. Hecker, "Camaleon-Camouflage assessment by evaluation of local energy, spatial frequency and orientation," *Proc. SPIE* **1687**, 342–349 (1992).
25. G. Waldman, J. Wootton, and G. Hobson, "Visual detection with search: an empirical model," *IEEE Trans. Syst. Man Cybern. SMC-21(3), 596–606 (1991).*
26. L. Overington, "Towards a complete model of photopic visual threshold performance," *Opt. Eng.* **21**, 2–13 (1982).
27. T. J. Doll, S. W. McWhorter, D. E. Schmieder, and A. A. Wasilewski, "Simulation of selective attention and training effects in visual search and detection," in *Vision Models for Target Detection and Recognition*, E. Peli, Ed., pp. 396–418, World Scientific, Singapore (1995).
28. *Jane's Military Vehicles and Logistics*, 17th ed., C. F. Foss and T. J. Gander, Eds., Butler & Tanner Ltd., London (1997).
29. G. Tidhar, G. Reiter, Z. Avital, Y. Hadar, S. R. Rotman, V. George, and M. L. Kowalczyk, "Modeling human search and target acquisition performance: IV, detection probability in the cluttered environment," *Opt. Eng.* **33**, 801–808 (1994).
30. J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, New York (1982).
31. P. Bijl, F. K. Kooi, and M. van Dorresteijn "Visual search performance for realistic target imagery from the DISSTAF field trials," Report TNO-TM 1997 A055, TNO Human Factors, Soesterberg, The Netherlands (1997).
32. W. A. Wagenaar, "Note on the construction of diagram-balanced Latin squares," *Psychol. Bull.* **72**, 384–386 (1969).
33. S. K. Mannan, K. H. Ruddock, and D. S. Wooding, "The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images," *Spat. Vis.* **10**(3), 165–188 (1996).
34. T. H. Monk, "The interaction between the edge effect and target conspicuity in visual search," *Hum. Factors* **23**(5), 615–625 (1981).
35. E. S. Krendel and J. Wodinsky, "Visual search in an unstructured visual field," *J. Opt. Soc. Am.* **50**, 562–568 (1960).
36. L. G. Williams, "Target conspicuity and visual search," *Hum. Factors* **8**, 80–92 (1966).
37. C. P. Greening, "Mathematical modeling of air-to-ground target acquisition," *Hum. Factors* **18**(2), 111–148 (1976).
38. A. Akerman and R. E. Kinzly, "Predicting aircraft detectability," *Hum. Factors* **21**(3), 277–291 (1979).
39. J. A. Ratches, W. R. Lawson, F. J. Shields, C. W. Hoover, L. P. Obert, S. P. Rodak, and M. C. Sola, *Status of Sensor Performance Modelling at NV&EOL*, Night Vision and Electro-Optics Laboratory, Fort Belvoir, VA (1981).
40. S. R. Rotman, E. S. Gordon, and M. L. Kowalczyk, "Modeling human search and target acquisition performance: I. First detection probability in a realistic multitarget scenario," *Opt. Eng.* **28**(11), 1216–1222 (1989).
41. G. Waldman, J. Wootton, and G. Hobson, "Visual detection with search: an empirical model," *IEEE Trans. Syst. Man Cybern. SMC-21(3), 596–606 (1991).*
42. J. M. Baena, A. Toet, X. R. Fdez-Vidal, A. Garrida, and R. Rodriguez-Sanchez, "A computational visual distinctness metric," *Opt. Eng.* **37**(7), 1995–2005 (1998).
43. X. R. Fdez-Vidal, A. Toet, J. A. Garcia, and J. Fdez-Valdivia, "Computing visual target distinctness through selective filtering, statistical features, and visual patterns," *Opt. Eng.* **39**(1), 267–281 (2000).
44. J. A. Garcia, J. Fdez-Valdivia, X. R. Fdez-Vidal, and R. Rodriguez-Sanchez, "Information theoretic measure for visual target distinctness," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(3) (in press).
45. J. A. Garcia, J. Fdez-Valdivia, X. R. Fdez-Vidal, and R. Rodriguez-Sanchez, *Computational Models for Predicting Visual Target Distinctness*, SPIE Press, Bellingham, WA (2001).
46. X. R. Fdez-Vidal, J. A. Garcia, J. Fdez-Valdivia, and R. Rodriguez-Sanchez, "Defining the notion of visual pattern for predicting visual target distinctness in a complex rural background," *Opt. Eng.* **39**(2), 415–429 (2000).
47. J. A. Garcia, J. Fdez-Valdivia, X. R. Fdez-Vidal, and R. Rodriguez-Sanchez, *Computational Models for Predicting Visual Target Distinctness*, SPIE Press, Bellingham, WA (2001).
48. G. Witus and D. Ellis, "Perception testing for development of computer models of ground vehicle visual discrimination performance," in *Targets and Backgrounds VI: Characterization, Visualization, and the Detection Process*, W. R. Watkins, D. Clement, and W. R. Reynolds, Eds., *Proc. SPIE* **4029**, 2–9 (2000).
49. A. Toet, P. Bijl, and J. M. Valeton, "Test of three visual search and detection models," *Opt. Eng.* **39**(5), 1344–1353 (2000).
50. L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Res.* **40**, 1489–1506 (2000).
51. L. Itti and O. Koch, "Computational modeling of visual attention," *Nature Reviews Neuroscience* (in press).
52. E. Niebur, L. Itti, and C. Koch, "Controlling the focus of visual selective attention," in *Models of Neural Networks 4*, Van Hemmen, Ed., Springer Verlag, Berlin (in press).
53. L. Itti, "Models of bottom-up and top-down visual attention," PhD Thesis, California Institute of Technology (2000).
54. G. Aviram and S. R. Rotman, "Analyzing the effect of imagery wavelength on the agreement between various image metrics and human detection performance of targets embedded in natural images," *Opt. Eng.* **40**(9) (2001).
55. C. M. Birkemark, "CAMEVA, a methodology for estimation of target detectability," *Opt. Eng.* **40**(9) (2001).
56. J. A. Garcia, J. Fdez-Valdivia, X. R. Fdez-Vidal, R. Rodriguez-Sanchez, and J. M. Fuertes, "Minimum error gain for predicting visual target distinctness," *Opt. Eng.* **40**(9) (2001).
57. L. Itti, C. Gold, and C. Koch, "Visual attention and target detection in cluttered natural scenes," *Opt. Eng.* **40**(9) (2001).
58. T. Nilsson, "Evaluation of target acquisition difficulty using recognition distance to measure required retinal area," *Opt. Eng.* **40**(9) (2001).
59. S. Nyberg and L. Bohman, "Assessing camouflage methods using textural features," *Opt. Eng.* **40**(9) (2001).
60. D. L. Wilson, "Image based contrast-to-clutter modeling of detection," *Opt. Eng.* **40**(9) (2001).
61. G. Witus, G. R. Gerhart, and R. D. Ellis, "A contrast model for 3-D vehicles in natural lighting and search performance analysis," *Opt. Eng.* **40**(9) (2001).



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.



Piet Bijl received his PhD in physics from the University of Utrecht in the Netherlands in 1991. He has worked on visual contrast detection. He is involved in the development of procedures to quantify electro-optical system performance, in particular for staring array systems, the development of test equipment for these cameras, and the improvement of target acquisition models. His current research interests include visual information processing, electro-optical system performance, target acquisition, and fashion models.



J. Mathieu Valeton received his PhD in physics from the University of Utrecht in the Netherlands in 1983. He has worked on the electrophysiology of the monkey retina, on the detection of signals in band-limited static noise, on spatial scaling in the human visual system, and on human visual search performance in realistic scenarios. His current research interests include visual information processing, electro-optical system performance, search and target acquisition modeling, image quality, simulators, display systems, and work schedules.