Reprinted from SPIE Vol. 98 — Assessment of Imaging Systems (Sira)

PREDICTION OF REALISTIC VISUAL TASKS FROM IMAGE QUALITY DATA

Aart van Meeteren
Institute for Perception Research TNO
Kampweg 5, Soesterberg, The Netherlands

Abstract

As a realistic visual task subjects had to recognize military vehicles out of six alternatives. Percentages of correct identifications were determined at three different observation distances in order to interpolate the distance required for 50% correct, which is taken as a measure of performance. The scenery was simulated indoors by slide-projection in favour of better control of conditions and faster procedure. The slides were photographed at the location of an earlier field trial. The performance of a portable image intensifier was measured in this way as a function of luminance and contrast in object space. It appears that recognition of the present set of objects is visually equivalent to the detection of a circular disk, and can be predicted from image quality data (displayed S/N, MTF) along that line.

Introduction

Optical and electro-optical observation devices serve to perform realistic visual tasks. Their usefulness obviously is primarily related to image quality factors like MTF and displayed S/N-ratio. Several attempts have been made to establish and to quantify this relation. Roughly, one may distinguish two approaches:

Image quality → Subjective image quality → Performance
Image quality → Elementary visual functions → Performance

Along the first line satisfactorily high correlations were found[1, 2], but no recipe was given, as far as we know, for the prediction of absolute performance. Such a recipe was formulated along the second line by Johnson[3], who concluded from a series of experiments that "complex military targets may be considered equivalent, in a visual sense, to repetitive resolution patterns of appropriate spatial frequencies". Identification of a target proved to correspond roughly with a resolution of 6.4 grating-periods per "critical" dimension. Similarly, recognition of the class of objects to which the target belongs, would correspond with 4 periods, and detection with 1 period per critical dimension. Thus, performance of such realistic visual tasks as the detection, recognition and identification of objects might be reduced to the resolution of gratings.

Contrast thresholds for sine wave gratings have been measured under a great variety of conditions in the last decade. The results are usually presented in the form of modulation sensitivity functions, which are highly comparable to MTF's. Van Meeteren[4] has shown that the modulation sensitivity function for vision through image intensifiers can be predicted straightforwardly from the MTF and the display-S/N-ratio of the device concerned. Johnson's recipe, cited above, now actually suggests that one may directly read from the modulation sensitivity function at what distance objects of a given contrast can be detected, recognized or identified, simply by substituting distance for spatial frequency. This is what we want to verify in the present paper, whereby we have to remark that Johnson's gratings were square wave gratings.

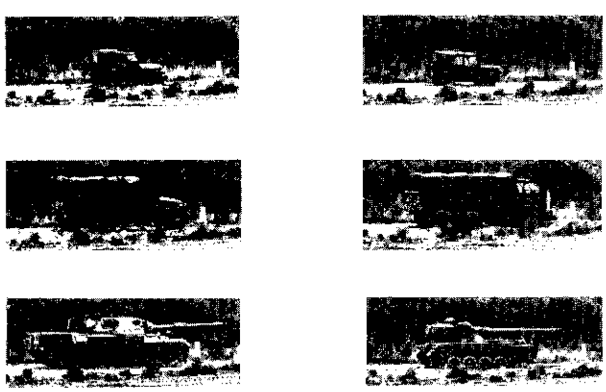

Fig. 1. Test objects. Next to identification of individual objects recognition of object classes (jeeps, trucks, tanks) can be studied with this set of test objects. 1. Munga; 2. Nekaf; 3.DAF; 4. GMC; 5. AMX; 6. Leopard.

The concepts of detection, recognition and identification need further definition. Wagenaar and Van Meeteren[5] argued that we should not use these terms with respect to individual objects in the way Johnson[3] did. Identification of say, a horse may be very difficult when the alternatives are other hoofed animals, whereas it may be easy to identify a horse among a group of reptiles. Detection, recognition and identification are only meaningful with respect to a certain set of alternative objects. As a consequence, every attempt to quantify performance must start from some realistic set of objects. Fig. 1 illustrates the standard set of objects we used. Although this is a small set, it involves a variety of discriminations with different degrees of difficulty. It also allows to define detection (is there any object?), recognition (jeep, truck, tank?) and identification (Munga, Nekaf, GMC, DAF, AMX, Leopard?) as three different visual tasks.

The general experimental procedure was to measure the percentages of correct identifications at three different observation distances, in order to interpolate the distance required for 50% correct. This distance was taken as the main measure of performance and will be called the recognition distance. Recognition distances were measured as a function of luminance for vision through an image intensifier with contrast as parameter. Next, the results could be replotted as contrast sensitivity functions with recognition distance along the horizontal axis, and reciprocal contrast along the vertical axis. This allows a direct comparison with the spatial modulation sensitivity functions mentioned above.

Experiments

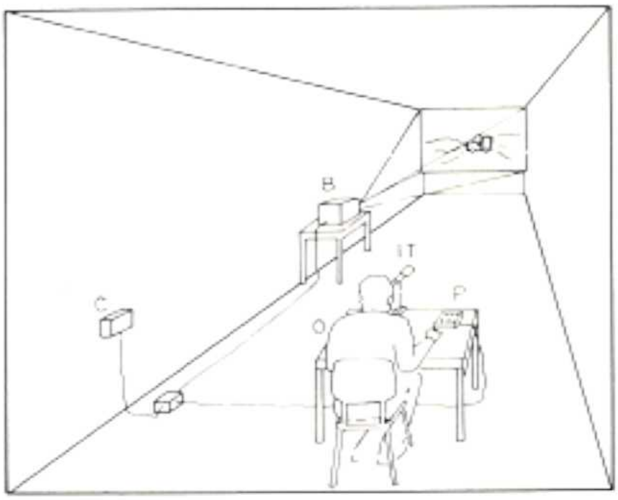

Fig. 2. Experimental set-up. O: observer, IT: image intensifier, P: response panel, B: light-tight box with projector, C: connection with digital computer.

The scenery was simulated indoors by slide-projection (Fig. 2) in order to obtain better standardization of conditions as well as faster procedure and lower costs than possibly might be achieved in field experiments. The basic photographs, however, were made at the location of a preceding field experiment, that served to verify the simulation approach, and to calibrate a possible difference in overall performance.

Visual search was not involved: the objects, if present, were always in the center of the scenery, as Fig. 3 illustrates.

Fig. 3. The scenery. The contrast between objects and background is relatively small, apart from the highlights. No search involved.

The contrast of an object upon a background is usually defined as:

$$C_o = \frac{L_o - L_b}{L_B}, \qquad (1)$$

where L_o and L_b are the luminances of object and background respectively. In our case we obviously have to compare the luminance of the dominant shining parts of the objects with the average luminance of the fir-background. It turns out then that $L_o = 3L_b$ so that the original contrast according to Eq. (1) is 2.0 in this rather arbitrary approach. This is the value which we will use throughout the paper in order to characterize the contrast of the objects. According to measurements this contrast is not substantially different in terms of the spectral sensitivity of the photocathodes that are applied in image intensifiers. The apparent contrast could be reduced by projecting a veiling luminance L_v over the scenery. It follows then, that

$$C = C_o \frac{L_b}{L_b+L_v} \tag{2}$$

The luminance of* the scenery was calibrated with a Pritchard Photometer and is expressed here in equivalent* light source A values, although the actual light was different. Luminance was controlled with density filters in front of the projector, which was placed in a light-tight box. The filters were calibrated using the image intensifier as a sensor.

The experiments were arranged in sessions during which series of 80 slides were presented all at the same luminance and with the same veiling light. These series contained 27 slides on a large scale, 26 slides on a medium scale and 27 slides on a smaller scale, so that three different observation distances, in the proportion of 1:1,35:1,82, were simulated per session. Each object was represented 4 times per distance in slightly different side views. The remaining slides represented "empty" sceneries, i.e., pictures of the background without an object in it.

The image intensifier was carefully adjusted before each session. The observer's responses were registered with the aid of a key-board, connected with a digital computer system. Each object was represented by a separate key and there was a seventh key for the empty scenery. The observer was asked to look at the scenery and to identify the object. In case of doubt between two or more objects he was asked to press all corresponding keys. A standard score was evenly distributed over all responses per presentation. The key-board was also used to call for the next slide. After a session average percentages of correct identifications were computed for the three different distances in order to determine the

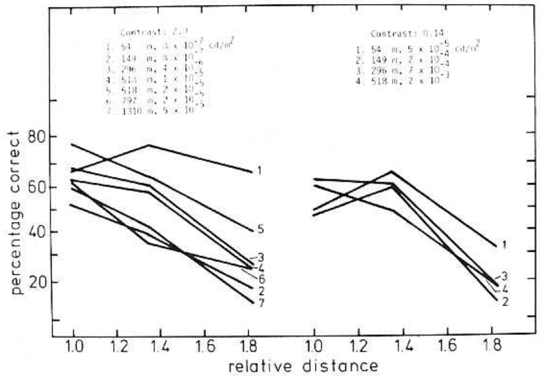

Fig. 4. Examples of raw data. Each curve represents the results of a single experimental session and allows to interpolate the 50%-recognition distance for the condition concerned.

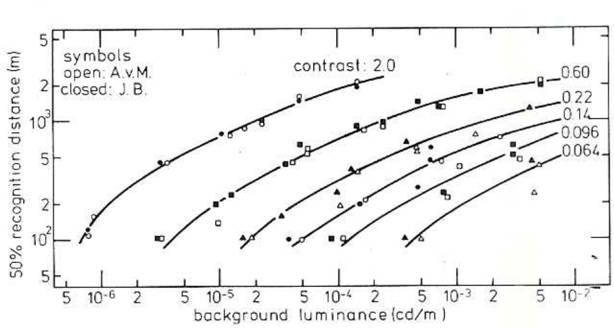

Fig. 5. Main results. Recognition distances as a function of background luminance as parameter. See text for definition of contrast.

50%-correct distance by graphical interpolation, see Fig. 4. Thus each session served to determine one recognition distance for a certain luminance and contrast condition. Frequently after a session it turned out that the conditions were such that no recognition distance could be interpolated. Of 127 sessions only 65 were successful, 32 for observer AVM. Fig. 5 illustrates these results.

* With respect to the S-25 photocathode.

Results

Recognition distances are plotted in Fig. 5 as a function of background luminance with contrast (Eq. (2)) as parameter.

The results of the two observers agree rather well. In practice one might expect families of curves shifted somewhat with respect to each other for different observers.

The recognition distance increases with luminance, roughly according to a square root relation as might be expected, considering the photon noise. Above 2000 m the recognition distance starts to level off due to the optical resolution limit of the device. Unfortunately it was not possible to simulate larger distances in our experiment.

The effect of lower contrast is to shift the curves towards higher luminances, i.e. at higher luminances the same recognition can be obtained with lower contrast. Again, contrast sensitivity proves to be proportional to the square root of luminance, following the photon noise limit.

Effect of meteorological range

Apparent luminance, contrast and distance are interrelated in the field, where atmospheric straylight is added to the scenery[6]. Field conditions are best characterized by the vertical illuminance E of the target and the meteorological range R. R is the distance over which the contrast of an object against the horizon sky is reduced to 2%. It is further required to know the luminance of the horizon sky. For the sake of simplicity we will assume that the latter is just equal to the luminance of a white object in the position of the target. This condition is roughly met when the reflectance of the ground in front of the target is in between 5-20%. In that case we find the apparent contrast C(r) and the appa-

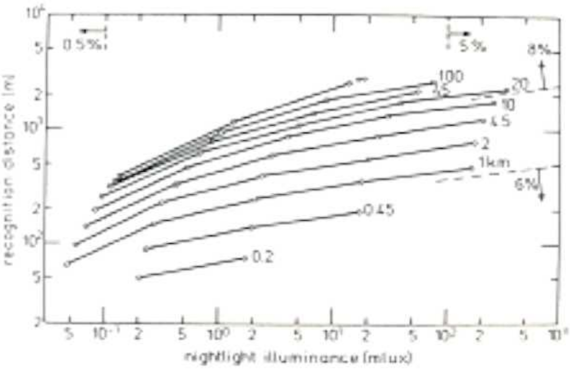

Fig. 6. Recognition distances for field conditions. Data derived from Fig. 5 accounting for atmospheric straylight, represented here by the meteorological range as parameter.

rent background luminance $L_b(r)$ from the following relations:

$$C(r) = C_o \cdot \rho_b \frac{e^{-3.912\ r/R}}{1+(\rho_b - 1)e^{-3.912\ r/R}} \qquad (3)$$

and

$$L_b(r) = \frac{E}{\pi} \left| 1+(\rho_b - 1)e^{-3.912\ r/R} \right|, \qquad (4)$$

where $\rho_b = 0.1$ is the luminance factor of the background and r is the observation distance.

Starting from Eqs. (3) and (4) we now can reconstruct recognition distances by a graphical method as a function of target illumination with meteorological range as parameter. Fig. 6 illustrates the results. An additional correction was made for the spectral difference of light source A and nightlight in the field.

Obviously the rate of increase of the recognition distance with target illuminance is lower according as atmospheric conditions are worse.

Probabilities of illuminance below 0.1 mlux and above 100 mlux, and of meteorological ranges larger than 10 km and shorter than 1 km are also indicated in Fig. 6. These probabilities were derived from Van Schie's[7] extensive statistical measurements near the Netherlands-Germany border during the nights (altitude of sun below -7^{o}) of the period between October 1967 and October 1968.

As mentioned above a field trial was arranged in order to verify the simulation approach. Performance proved to be better indoors than in the field, roughly by a factor 1.5 in terms of threshold contrast. Unfortunately we did not measure actual straylight in the field experiment, so that the above field degradation factor rests upon indirect estimation. For the meteorological rage we relied upon reports of the nearby air-base Soesterberg. Next to atmospheric straylight instrumental straylight should not be neglected in image intensifiers. Part of the field of view is covered by the relatively bright sky, scattering light over the darker parts of the scenery. Also, light from outside the field of view may be scattered into it. Instrumental straylight was measured later in conditions similar to the field experiment.

We did not allow in Fig. 6 for possible effects of instrumental straylight, nor for the field degradation just mentioned. Roughly the combined effect might be a shift towards four times higher illuminance.

Prediction

Is it possible to predict performance from image quality-data? To study this question the results of Fig. 5 were replotted by making vertical cross-sections with background luminance as parameter. In this way one can read what contrast is required as a function of

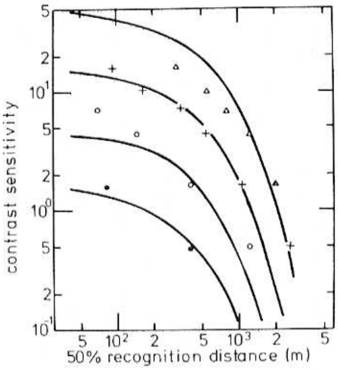

Fig. 7. Results of Fig. 5 replotted as contrast sensitivity functions with background luminance as parameter (3.2×10^{-6}, 3.2×10^{-5}, 3.2×10^{-4} and 3.2×10^{-3} cd/m^2). Curves are modulation sensitivity functions for sine wave gratings at the same luminances shifted along the axes to get the best fitting.

recognition distance. Finally the reciprocal of the threshold contrast was plotted vs recognition distance with luminance as parameter in Fig. 7. Thus performance is characterized by a kind of contrast sensitivity functions, which according to Johnsons[3] suggestion should be directly comparable to modulation sensitivity functions for sine wave gratings. These modulation sensitivity functions can be predicted straightforward for vision through image intensifiers starting from the MTF and the displayed S/N-ratio of the device[4]. This was confirmed experimentally for the device concerned. Is it possible to fit a family of modulation sensitivity functions to the performance data in Fig. 7? Fig. 7 illustrates the best fit that we could obtain by suitable shifts along the axes. A recognition distance of 100 m corresponds with a spatial frequency of 2.8 cpd in object space in that case, which would mean that recognition of the set of objects we used corresponds with the detection of 1.6 line pairs per meter, i.e. about 4 periods per critical dimension, indeed.

However, the prediction is not really satisfying. A somewhat better prediction might have been obtained by using square wave gratings, but the results themselves suggest a more basic change of the prediction rule. Actually the measuring points seem to lie more on straight lines with slope-1 which reminds of Piper's law for the detection of circular disks upon a background. Thus, instead of grating-detection it seems to be better to try disk-detection as the intermediate elementary visual function connecting image quality and recognition performance. This also is more satisfying theoretically: the relevant details of the objects are two-dimensional single patches of light contrasting against a more or less uniform background, i.e. more comparable with disks than with repetitive resolution gratings. In terms of spatial frequencies the objects must be represented by a two-dimensional spectrum* rather than just by one one-dimensional narrow band[8]. We have to do right to Johnson now immediately, because he actually assumed that resolution of gratings is equivalent to the detection of disks and finally related recognition performance to disk-detection.

In order to predict disk-detection we may follow here the succesful fluctuation theory of De Vries[9] and Rose[10], according to which vision at low luminances is limited by photon-noise. Vision is limited in general by the displayed noise and the detection of a disk upon a background is essentially a signal-noise discrimination. If we may assume that the noise is original photon-noise we may write for large disks:

$$\frac{1}{C_{th}} = f.d.L_b^{\frac{1}{2}}, \qquad (5)$$

where C_{th} is the threshold contrast, f is a constant, d is the diameter of the disk and L_b is the background luminance. With respect to d Eq. (5) represents Piper's experimental law. With respect to L_b Eq. (5) is the De Vries-Rose law. The effect of optical blurring is that the signal is spread over a large area and must be compared with the noise in that larger area. This leads to Eq. (6) for disks smaller than the blur disk d_b:

* We will leave undecided here whether or not spatial spectra would be more relevant than the original light-distribution.

$$\frac{1}{C_{th}} = f.\ \frac{d^2}{d_b}\ L_b^{\frac{1}{2}} \tag{6}$$

Eq. (6) represents Ricco's law with respect to d. When $1/C_{th}$ is plotted as a function of d on log-log graph-paper, Eqs. (5) and (6) are rendered by two straight lines with slopes -1-

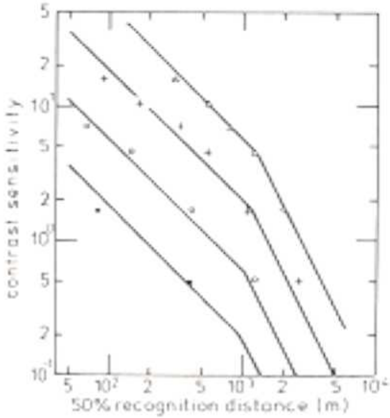

Fig. 8. See Fig. 7. Curves now refer to the detection of a circular disk of 0.70 m.

and -2 respectively. Fig. 8 illustrates these predictions and they do provide a better fit than the modulation sensitivity functions for gratings. To be short, recognition of the present set of objects corresponds to the detection of a disk with a diameter of 0.7 m, the criterion being a S/N-ratio of 4.0.

The bending point of the curves in Fig. 8 of course is related to the size of the optical blur circle and is at $d=d_b$, where d_b can be evaluated from the MTF of the device, accounting also for the transfer properties of the observer's eye. In first approximation we may characterize the modulation sensitivity functions S(w) as follows:

$$S(w) = S(o).e^{-(k_e+k_d)w}, \tag{7}$$

where w is the spatial frequency and k_e and k_d are characteristics of the human eye and the MTF of the device respectively. It can be shown than that we may write:

$$d_b = (k_e+k_d).\ \frac{\sqrt{2}}{\pi} \tag{8}$$

Unfortunately, it will be clear from Fig. 8 that our performance measurements did not extent into resolution limited conditions, so that this part of the prediction was not really verified here.

So far we discussed 50%-correct identification distances, which we called recognition distances, because 50% correct identification practically implies that the objects are correctly classified as jeeps, trucks and tanks. The experiments were especially designed to determine the 50% correct identification distance. Part of the data, however, allowed similar analyses for 20% and for 80% correct identification, which proved to correspond with disks of 0.90 m and 0.45 m respectively.

Conclusions

Recognition and identification of a set of military vehicles can be compared with the detection of disks upon a uniform background. As such, the performance of these tasks can be predicted from image quality data. It seems that it is not correct to characterize realistic objects by equivalent repetitive resolution patterns.

References

1. Scott, F., "The search for a summary measure of image quality - A progress report". Photo. Sci. Engng., Vol. 12, pp. 154-164. 1968.
2. Snyder, H.L., "Image quality and observer performance". In: Perception of displayed information, Ed. L.M. Biberman, pp. 87-118, Plenum Press 1973.
3. Johnson, J. "Image intensifier symposium", Fort Belvoir, D220160, 1958.
4. Van Meeteren, A. Visual aspects of image intensification, Report Institute for Perception TNO, Kampweg 5, Soesterberg, The Netherlands. 1973.
5. Wagenaar, W.A., and Van Meeteren, A., "Noisy line-scan pictures". Photogrammatic Eng. Vol. 35, pp. 1127-1134. 1969.
6. Middleton, W.E.K., Vision through the atmosphere. Un. of Toronto Press, 1952.
7. Van Schie, J. Noctural illumination and decrease of contrast in the atmosphere. Re-

port Ph.L. 1969-4. Physics Laboratory TNO, The Hague, The Netherlands, 1969.

8. Ginsburg, A.P. Psychological correlates of a model of the human visual system. MS Thesis SE/EE/715-2, Wright Patterson Air Force Base, Ohio: Air Force Instit. Technol., 1971.

9. De Vries, H., "The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye". Physica, Vol. 10, pp. 553-564, 1943.

10. Rose, A. "The sensitivity performance of the human eye on an absolute scale". J. Opt. Soc. Am. Vol. 38, pp. 196-208, 1948.