TOD PREDICTS TARGET ACQUISITION PERFORMANCE FOR STARING AND SCANNING THERMAL IMAGERS

Piet Bijl[*a], J. Mathieu Valeton[a], Arie N. de Jong[b]
[a] TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
[b] TNO Physics and Electronics Laboratory, P.O. Box 96864, 2509 JG, The Hague, The Netherlands

ABSTRACT

Identification and recognition performance for four staring and two scanning thermal imagers, were measured in an observer experiment using images that were collected during a NATO field trial in Nettuno, Italy, in 1998. The dataset allows validation of the MRTD and alternative sensor performance measures such as the TOD[1,2] (Triangle Orientation Discrimination threshold). The 75% correct Target Acquisition (TA) ranges were compared with the TOD sensor acuity and the MRTD spatial frequency at $\Delta T = 2$ K. The results show that the ratio between the 75% correct TA range and TOD sensor acuity (which is the equivalent to the cycle criterion in MRTD-based models), is *independent* of the sensor type used, which means that the TOD is a good predictor of range performance. A systematic relationship between TA range and MRTD spatial frequency, which is the primary assumption for many MRTD-based TA models, was *not* found. Not only do sampling and staring systems have different ratios, it also occurs that TA performance for staring systems depends on the internal noise of the sensor, while range predictions based on the MRTD are determined solely by the resolution of the sampling array. In conclusion, a single (dimensionless) factor suffices to predict TA field performance from TOD sensor acuity for both staring and scanning systems, while MRTD-based models fail to predict TA ranges, in particular for staring systems. This result in fact further validates the TOD as a tool for electro-optical imager performance characterization.

Keywords: Electro-Optical system performance testing, TOD methodology, standard measurement procedure, MRTD, Target Acquisition, field performance, validation

1. INTRODUCTION

It is a well-known fact that the traditional end-to-end thermal imager performance measure, the MRTD (Minimum Resolvable Temperature Difference), which has been developed for the first generation or scanning imagers, is unable to adequately predict to Target Acquisition (TA) performance for second generation or staring imagers [1,2,3]. The reason is, that the image of a staring sensor is usually under-sampled which causes artifacts that are essentially different for the periodic four-bar pattern that is used in the MRTD procedure than for real targets. Theoretically, the maximum allowed MRTD spatial frequency for staring sensors is half the sample frequency of the Focal Plane Array (FPA). As a consequence, there is no basis to assume a one to one relationship between MRTD and acquisition of real targets. Experimental data also show that, when the translation from MRTD to field performance for scanning systems (known as the 'cycle criteria' or 'Johnson criteria'), is applied to staring systems, this generally leads to far too pessimistic range predictions.

Recently, three promising alternative sensor performance measures have been proposed: the TOD[1,2,4,5] (Triangle Orientation Discrimination threshold), the MTDP[6] (Minimum Temperature Difference Perceived), and the MTF squeeze model[7]. The TOD method is suitable for all types of imaging systems, is representative for a real TA task, and is characterized by a solid measurement procedure and an easy observer task. Experimental assessment of the TOD is much easier than measuring an MRTD, and the results are more accurate. In a first validation study, it was shown that the shape of the TOD for a CCD camera matches the relationship between contrast and TA range for ship targets very well, while the MRC, which is the visual equivalent of MRTD, appeared to be much too steep.

In the present study, a further validation of the TOD will be performed. It will be studied how well the relationship between TA performance and TOD is preserved when applied to thermal imagers of different types: scanning and staring cameras, cooled and uncooled, in the MW (mid wave or 3-5 μm) and LW (long wave or 8-12 μm). For this study, an observer performance experiment was carried out using imagery that was recorded from a number of thermal imagers during the Nettuno field trials organized by NATO RTO AC/323SET TG.12.

In section 2, the approach will be described. The TA performance experiment is described in sections 3 and 4, the TOD and MRTD measurements in section 5, and the comparison between field and lab performance in section 6.

---

* Further information: bijl@tm.tno.nl; tel. +31 346 356277; fax +31 346 353977

In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XI*, Gerald C. Holst, Editor, Proceedings of SPIE Vol. 4030 (2000) • 0277-786X/00/$15.00

2. APPROACH

In ACQUIRE[8] or other MRTD-based TA models, it is assumed that identification or recognition of a target with a certain probability correct, corresponds to a fixed number of resolvable cycles from the MRTD over the target size (the so-called *cycle criteria N*). In recent versions of the model, target size is defined as target angular square-root area, and the 2D-MRTD is used. The recommended cycle criteria for a 50% correct identification and recognition (N_{50}) are 6 and 3, respectively, and for 75% (N_{75}) they are 8 and 4. Note that the criteria apply to a *set* of targets rather than an individual target, and that the exact values depend on the experimental conditions.
ACQUIRE can easily be modified into a TOD target acquisition model by replacing the MRTD with the TOD curve, and the cycle criteria by the *TOD criteria M* that represent the relationship between target angular square-root area and triangle square-root area.

The assumption in ACQUIRE means that, if there are no atmospheric transmission losses, the ratio between target angular size at 75% correct identification and recognition on the one hand, and MRTD cycle width (=1/spatial frequency) at target thermal contrast ΔT on the other, should equal N_{75}, independent of target contrast and sensor used. The same applies to the TOD target acquisition model, except that N_{75} is replaced by M_{75}. For high target contrasts, the MRTD is relatively insensitive to contrast variation and the relationship holds in a good approximation even if target contrast is slightly reduced by the atmosphere[6].

In the present study, both the TOD and MRTD based model are validated as follows. First, a TA field performance experiment is carried out with a number of thermal imagers in parallel. For these sensors, the TOD and MRTD are determined at approximately the same contrast as the targets in the experiment.
Then, for the MRTD validation, the number of cycles over the target (i.e. the ratio between the 75% correct target angular square-root area and the corresponding MRTD cycle width) is calculated. If the model is correct, this ratio is close to the recommended N_{75}, and the values are independent of the sensor used. If the values depend on the type of sensor, the MRTD is not a good predictor of field performance.
For the TOD validation, the ratio between the 75% correct target angular square-root area and the corresponding triangle square-root area is calculated (dimensionless). If the TOD is a good predictor of TA performance, then the ratio is independent of sensor type. The average ratio gives an estimate of the TOD criteria for identification and recognition.

3. FIELD PERFORMANCE EXPERIMENT

3.1 The Nettuno field trials

The field trials were held in Nettuno, Italy, from 1-15 July 1998. A detailed description of the trials is given by de Jong and Winkel[9] and by Williams et al.[10]. In two (of seven) experiments, imagery was collected that will be used in the observer performance experiments:

1) Experiment A, carried out by five of the participating countries.
   In this experiment, images of single stationary target vehicles were recorded at 11 distances ranging from 400 to 2477 m and at a close-up range of 125 m. Six target types were recorded at four different aspect angles, including front view. The position of the sensors was fixed, which means that the background was the same for each target range. 11 Sensors were used in the experiment.
2) Experiment B, carried out by the Netherlands.
   In this experiment, images of single stationary target vehicles (front view) were recorded at 19 distances ranging from 150 to 800 m. The sensors were positioned in a truck and the target position was fixed. Eight target types (including the six targets used in experiment A) and 3 sensors were used in the experiment.

For most sensors, imagery was recorded on video (S-VHS, U-Matic or Betamax). The bandwidths of the video systems were higher than those of the sensors. For some sensors, single or multiple frames were recorded digitally (with 10 or 12-bit luminance resolution).

3.2 Image set

Six sensors were selected for the observer experiment: three from Nettuno Experiment A and three from Experiment B. These are listed in Table 1.

Six targets from three classes were selected: two Tanks, two APC's (Armed Personnel Carriers), and two Trucks. Only front views are used. All targets had approximately the same area in front view, and average target square-root area was 2.39 m. Target thermal contrasts were medium to high.

All 11 ranges from Nettuno experiment A and all 19 ranges from experiment B were used (if available).

All imagery was converted to 8-bit static images and stored on a computer hard disk. If 12-bits images were available, these images were converted to 8-bit with optimal level and gain settings for target identification and recognition. Otherwise, the images were grabbed from video footage. In order to avoid picture recognition, characteristic details in the imagery, such as the presence of other targets, people, the moon, or dead camera pixels, were removed using an image editing program.

Table 1: The six sensors used in the observer experiment

Sensor name	Sensor type
FPA 1	Focal Plane Array, uncooled, MW (3-5 μm)
Scanner 1	Scanner, cooled, MW
FPA 2	Focal Plane Array, cooled, LW (8-12 μm)
FPA 3	Focal Plane Array, cooled, MW
FPA 4	Focal Plane Array, cooled, LW
Scanner 2	Scanner, cooled, LW

3.3 Experimental setup

The experiment was carried out on a PC with a 17" monitor (1024 x 768 pixels, dot pitch 0.25 mm). A software program was developed to present the images in the desired order (random or sequential). Soft response keys were presented on the display, and could be selected by the observers with a mouse click. Before each experimental session, the contrast and brightness controls were set by the experimenter. The observers were not allowed to touch the controls.

The observers were placed in a dimly lit room. Care was taken that no stray light fell on the monitor screen. The observers were allowed to choose their own optimal viewing distance and to scrutinize the display if they wished to do so.

3.4 Measurement procedure

Each image was presented for 7 seconds. During or after a presentation, the observers had to name the target. If they were not sure, they had to guess. Next, they had to indicate the confidence in their response by classifying it as an Identification (I), Recognition (R) or Detection only (D). Identification is correct if an (I) is given and the target type is named correctly. Recognition is correct if an (I) or (R) is given and if the chosen target belongs to the correct class. During the experiment, no feedback was given. The procedure is described in detail by Valeton & Bijl[11].

The images from Experiment A (one run per target) were presented twice to each observer; images from Experiment B (two runs per target) were presented once. 28 Images were missing. The total number of presentations to each observer was 1052.

In order to minimize the possibility of picture recognition (see 2.2), all images from the two experiments were taken together and presented in random order with respect to sensor type, target type, and target range.

The experiment was subdivided into four sessions of approximately 30 minutes. To minimize systematic effects of learning, the imagery in these sessions was balanced with respect to sensor type, target type and target range. Further, for each of the four observers (see 3.6) the sessions were carried out in a different order using a 4 by 4 digram-balanced Latin square design[12].

3.5 Observer training

A short training session, containing 72 images of the targets at the two shortest available ranges, preceded the experiment in order to make the observers familiar with their task and to check whether the observers were able to identify the targets at short range.

During the experiment the observers were allowed to use high quality close-up views of the six targets, that were recorded during the Nettuno trials at 150 m from a staring MW imager and a scanning LW imager that were not used in the observer experiments.

3.6 Observers

Two experienced military observers (MK, OB) and two civilian observers (SK, PP) participated in the experiment. All observers had better than normal visual acuity.

3.7 Analysis

For each sensor, the 75% correct identification and recognition ranges and the corresponding 75% correct target angular square-root areas were calculated as follows. First, for each sensor and each observer separately, the identification and recognition score (see 2.4) were averaged over the six targets and plotted as a function of target angular square-root area (in mrad). The relationship between target angular square-root area S_{ang} (in mrad), target square-root area S (in m, see 3.2), and target range r (in km) is given by:

$$S_{ang}(mrad) = \frac{S(m)}{r(km)} \tag{1}$$

Then, a Weibull function (psychometric or s-shaped function) was fitted to the data using the maximum-likelihood procedure described by Bijl & Valeton[5]. In general, the Weibull function is given by**:

$$P_{\alpha\beta\gamma\delta}(x) = (1-\delta) - (1-\gamma-\delta)\cdot 2^{-(x/\alpha)^{\beta}} \tag{2}$$

where x = target angular size (in mrad), α is a threshold angular size (see below), β determines the steepness of the function, γ is the guess rate, and δ is called 'finger', i.e. the probability that the observer erroneously pushes a wrong button or misses a presentation. For $x = 0$ (target at infinity), $P = \gamma$; for $x \gg \beta$ (target at close range), $P = 1 - \delta$. In the present experiment, $\gamma = \delta = 0$. This leaves two free parameters for the fit: α and β.

Next, the 75% correct theshold angular size α_{75} was calculated using the following equation**:

$$\frac{\alpha_{75}}{\alpha} = \left[-{}^{2}\log\left(\frac{0.25-\delta}{(1-\gamma-\delta)}\right)\right]^{1/\beta} \tag{3}$$

Finally, $S_{ang,75}$ is defined as the average of α_{75} (with standard error in the mean σ_m) over the observers. The 75% correct acquisition range can be calculated from $S_{ang,75}$ using equation (1).

4. RESULTS OF THE FIELD PERFORMANCE EXPERIMENT

In Fig. 1 A-F, identification and recognition scores, averaged over targets and observers, is plotted as a function of target range for the six sensors (see Table 1). Open circles represent target recognition, filled circles represent identification. For each sensor, acquisition performance decreases gradually with target range. The solid lines are the maximum likelihood fits of a Weibull function to the data (see 3.7). Note that the range scales of the upper plots (from Nettuno experiment B) and the lower plots (experiment A) are different.

From Fig. 1, 75% correct ranges can be determined. For example, the 75% correct recognition range for FPA 1 (Fig. 1A) is 623 m. The corresponding target angular square-root area $S_{ang,75}$ = 2.39 (m) / 0.623 (km) = 3.84 mrad (equation 1). In the analysis (see 3.7), calculation of the 75% target angular square-root areas is performed for each observer separately, and averaged over the observers. The data for different observers are very similar: on average, the standard error in the mean due to observer differences and statistical fluctuations, is only 7% for identification and 5% for recognition.

** Equations (2) and (3) in Bijl & Valeton[5] were incorrect; Here the correct equations are given.

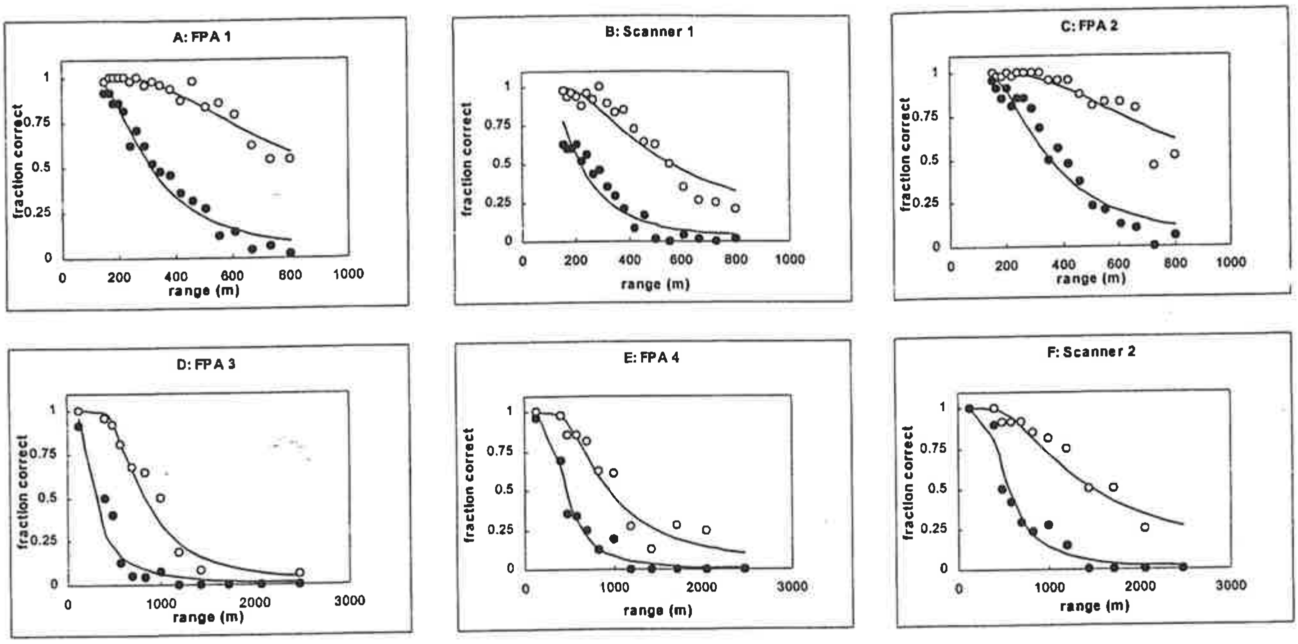

Fig. 1 A-F. Acquisition *vs.* range performance for the six sensors (see 2.2), averaged over all targets and all observers. Open circles: Recognition. Filled circles: Identification. Solid lines: best fits. See text for details.

5. TOD AND MRTD MEASUREMENTS

5.1 TOD measurements

The TOD measurement procedure is defined in detail by Bijl & Valeton[5]. For the present experiment, only the 75% correct triangle angular square-root area S_{75} at $\Delta T = 2K$ is measured. The reciprocal of this value is defined as TOD sensor acuity *SA* (in $mrad^{-1}$), which is a measure of the smallest detail that can be resolved with a sensor at high contrast (comparable to resolution limit). Measurement and analysis of the TOD are largely similar to that for real targets.

For each sensor, S_{75} is estimated for two observers and the average is taken. For each threshold estimate, the number of triangle sizes is $I = 5$, and the number of presentations per triangle size is $N = 32$. Orientation of the test patterns is random (Up, Down Left or Right), and the observer has to indicate which orientation he or she perceives. The fraction correct is plotted as a function of triangle angular square-root area. Analysis is performed in exactly the same way as described in 3.7, except that the guess rate $\gamma = 0.25$. In Fig. 2, the results for sensor FPA 1 for 1 observer are shown as an example. The results for different observers agree to within 5%. At present, not all TOD measurements have been collected. For two sensors, FPA 3 and FPA 4 (cooled, low noise FPA's), accurate estimates of S_{75} were made on the basis of their pixel sizes. No TOD data are available for Scanner 2 yet.

5.2 MRTD measurements

For the four FPA's, the 2D-MRTD spatial frequency at $\Delta T = 2K$ is simply determined by half the sampling frequency. Horizontal and vertical sampling frequency are equal for these sensors. The horizontal MRTD of Scanner 1 was measured by de Jong[9]. The 2D-MRTD for this sensor has to be determined but the 2D spatial frequency at $\Delta T = 2K$ is estimated to be 10% below the horizontal MRTD. For Scanner 2, the 2D-MRTD was taken from Wittenstein[13].

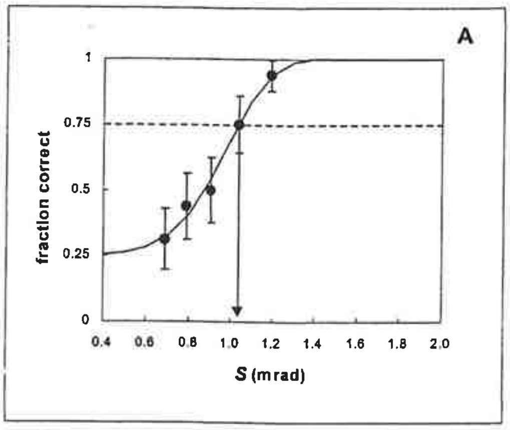

Fig. 2: Relation between triangle angular square-root area S (in mrad) and fraction correct for sensor FPA 1at $\Delta T = 2K$ for one observer. Filled circles: observer score. Solid line: best fit of the Weibull function (see 3.7). The arrow indicates the maximum likelihood threshold size S_{75}. In this case, $S_{75} = 1.04$ mrad.

6. COMPARISON OF FIELD AND LAB PERFORMANCE

6.1 TOD validation

For each sensor, the ratio between the 75% correct identification and recognition target angular square-root area $S_{ang,75}$ (from the field performance experiment, chapter 4) and the 75% correct triangle angular square-root area S_{75} (from the TOD measurements, section 5.1) was calculated. Fig 3A shows the results for identification. For recognition (not shown), similar results are found. The dark bars in Fig. 3A show the ratio for each sensor separately. The dotted line represents the average value over all sensors, and the error bar on the right-hand side of the figure shows the average +/- 1 standard deviation.
The figure shows two important results:

1. The ratio is *independent* of the sensor type used. The standard deviation is only 10% for identification and 15% for recognition, and this is within the experimental error in the data.
2. The average ratio is approximately 9 for identification. For recognition, the value is 4. These values are estimates of the TOD criteria M_{75}, which are equivalent to the cycle criteria in the ACQUIRE model (see chapter 2).

In conclusion, the TOD accounts for field performance differences between (well-sampled) scanners and (under-sampled) FPA's, and between uncooled (FPA 1) and cooled (FPA 2, 3 and 4) sensors. The TOD criteria for 75% correct identification and recognition of military targets (M_{75}) are 9 and 4, respectively.

6.2 MRTD validation

For each sensor, the required number of cycles over the target for a 75% correct identification and recognition were calculated (this is the product of the target angular square-root area $S_{ang,75}$ from the field performance experiment in chapter 4 and the 2D-spatial frequency at $\Delta T = 2K$). Fig 3B shows the results for identification. For recognition (not shown), similar results are found. The dark bars in Fig. 3B show the ratio for each sensor separately. The dotted line represents the average value over all sensors, and the error bar shows the average +/- 1 standard deviation.
Fig. 3B shows that:

1. The required number of cycles over the target is *largely dependent* on the sensor type used. The standard deviation is 33% for both identification and recognition. In more detail:
   a) The number of cycles is highest for the two scanners. The values are possibly different because these MRTD's were determined in different laboratories.
   b) The values are lowest for the three cooled FPA's (FPA 2, 3 and 4). This indicates that field performance for these sensors is highly *underestimated* compared to scanners when the MRTD is used. For a fair comparison with scanners, range predictions based on the MRTD should be increased by about 70%!
   c) The number of cycles for the uncooled FPA (FPA 1) is 33% higher for identification and 7% for recognition, when compared to the values for the cooled FPA's. This means, that identification performance for the uncooled sensor is poorer than for a cooled sensor (due to system noise), but this is not predicted on the basis of the MRTD cut-off frequency.

2. The average number of cycles is approximately 11 for identification. For recognition, the value is 5. These values are slightly higher than the recommended cycle criteria N_{75} for 75% correct identification and recognition in the ACQUIRE model (8 and 4, respectively, see chapter 2).

In conclusion: The results confirm that the MRTD is not suited to predict range performance for staring or undersampled cameras. Not only does the MRTD make an unfair comparison between scanning and staring systems, but also between cooled and uncooled cameras.

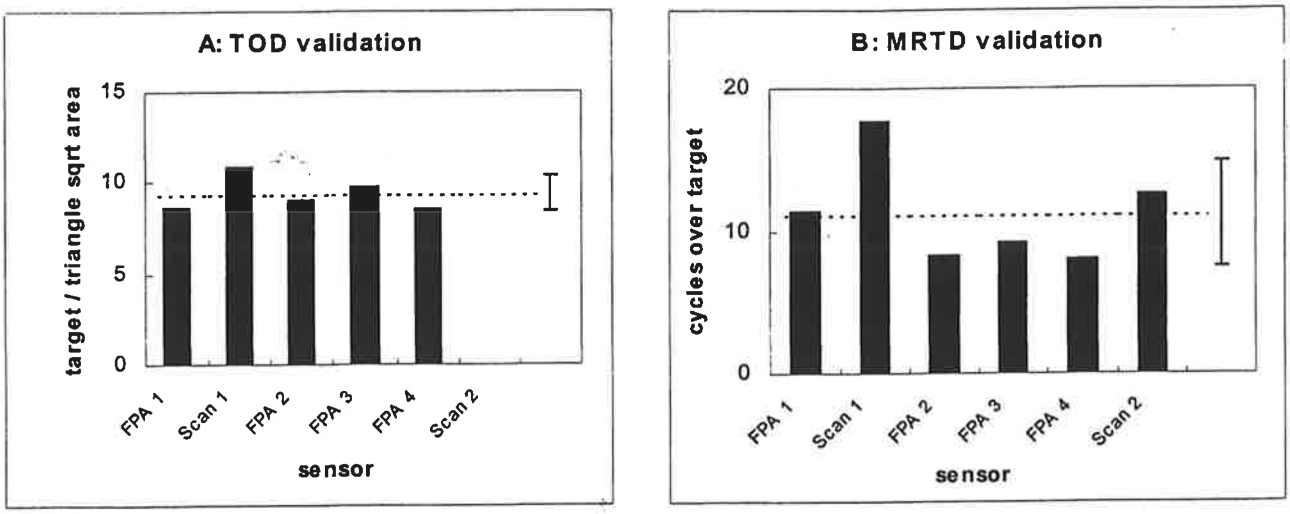

Fig. 3. A: TOD validation. The dark bars show the ratio between the 75% correct identification target angular square-root area $S_{ang,75}$ and the 75% correct angular triangle square-root area S_{75} for each sensor (these are listed in Table 1) separately. B: MRTD validation. The dark bars represent the required number of cycles over the target for a 75% correct identification. The results for recognition (not shown) are similar. Dotted lines: average value over all sensors. Error bars: average +/- 1 standard deviation.

With the TOD (Fig. 3A), the ratio is *independent* of the sensor (small standard deviation), which means that the TOD predicts range performance for different types of sensor very well. The average ratio is an estimate of the TOD criterion M_{75} (equivalent to the cycle criterion in ACQUIRE). Values are 9 for identification, and 4 for recognition. The required number of MRTD bar cycles over the target (Fig. 3B) is *largely dependent* on the sensor type used (large standard deviation), confirming that the MRTD *does not* predict range performance for staring array cameras.

7. CONCLUSIONS

1. In this study, a field performance experiment was used for the validation of the TOD and the MRTD. The data may also be used for validation of other thermal imager performance measures such as the MTDP and the MTF squeeze model.
2. The results support the TOD: a single translation factor suffices to predict target acquisition ranges for different types of sensors: scanners, cooled and uncooled FPA's in the MW and LW. The data confirm that the MRTD is *not* suited to predict range performance for staring or under-sampled cameras. In an earlier study[4], it was already shown that the shape of the TOD for a CCD camera matches the relationship between contrast and TA range for ship targets very well, while the MRC was much too steep.
3. The TOD criterion M_{75} (or translation factor between target and triangle angular square-root area) is approximately 9 for 75% correct identification and 4 for recognition of military targets. Quantitatively, these values come close to the cycle criteria that are used in ACQUIRE.
4. The ACQUIRE model can be greatly improved by simply replacing the MRTD and the cycle criteria with the TOD and the TOD criteria.

ACKNOWLEDGMENTS

The author wish to thank the NATO RTO AC/323SET TG.12 group for providing their imagery, Hans Winkel (TNO-FEL) for digitizing and converting the imagery, and the two experienced observers military from OCMan for their participation in the experiment.

REFERENCES

1. P. Bijl and J.M. Valeton, "TOD, a new method to characterize electro-optical system performance". *Proc. SPIE* 3377, 182-193 (1998).
2. P. Bijl and J.M. Valeton, "TOD, the alternative to MRTD and MRC". *Optical Engineering* 37(7), 1984-1994 (1998).
3. Driggers, R.G., Vollmerhausen, R. & Halford, C.E. (1999). Sampled Imaging Systems. *Optical Engineering 38, 5*, 740 – 741.
4. Bijl, P.& Valeton, J.M. (1998c). Validation of the new TOD method and ACQUIRE model predictions using observer performance data for ship targets. Optical Engineering 37, 7, 1984 - 1994.
5. Bijl, P.& Valeton, J.M. (1999). Guidelines for accurate TOD measurement, SPIE Proceedings, Vol. 3701, 14-25.
6. Wittenstein, W. (1999). Minimum temperature difference perceived – a new approach to assess undersampled thermal imagers. *Optical Engineering 38, 5*, 773 – 781.
7. Vollmerhausen, R., Driggers, R.G. & O'Kane, B.L. (1999). Influence of sampling on target recognition and identification. *Optical Engineering 38, 5*, 763 – 771.
8. Ratches, J.A. (1997). Static Performance Model for Thermal Imaging Systems. *Optical Engineering 15, 6,* 525-530.
9. De Jong, A.N. and Winkel, J. (1999*). The Nettuno trials of TG.12; results of Experiment B; July 1998.* (Report FEL-99-A189). The Hague, The Netherlands: TNO Physics and Electronics Laboratory.
10. Williams, D., Olson, R., Bunney, L., Johnson, P. (1999). NATO-RTO-SET/TG.12; Final report Nettuno field trials of 1998.
11. Valeton, J.M. & Bijl, P. (1994). *Target aquisition: human observer performance studies and TARGAC model validation* (Report TNO-TM 1994 B-15). Soesterberg, The Netherlands: TNO Human Factors Research Institute.
12. Wagenaar, W.A. (1969). Note on the construction of digram-balanced Latin-squares. *Psychological Bulletin, 72*, 384-386.
13. Wittenstein, W. (1999) Private communication.