TOD to TTP calibration

Piet Bijl[1], Joseph. P. Reynolds[2], Wouter K. Vos[1], Maarten A. Hogervorst[1] & Jonathan D. Fanning[1,2]
[1]TNO Behavioural & Societal Sciences, P.O. Box 23, Soesterberg,The Netherlands
[2]US Army Night Vision and Electronic Sensors Directorate, Ft. Belvoir, VA, USA 22060
Phone: +31 888665863, E-mail: piet.bijl@tno.nl

ABSTRACT

The TTP (Targeting Task Performance) metric, developed at NVESD, is the current standard US Army model to predict EO/IR Target Acquisition performance. This model however does not have a corresponding lab or field test to empirically assess the performance of a camera system. The TOD (Triangle Orientation Discrimination) method, developed at TNO in The Netherlands, provides such a measurement. In this study, we make a direct comparison between TOD performance for a range of sensors and the extensive historical US observer performance database built to develop and calibrate the TTP metric. The US perception data were collected doing an identification task by military personnel on a standard 12 target, 12 aspect tactical vehicle image set that was processed through simulated sensors for which the most fundamental sensor parameters such as blur, sampling, spatial and temporal noise were varied. In the present study, we measured TOD sensor performance using exactly the same sensors processing a set of TOD triangle test patterns. The study shows that good overall agreement is obtained when the ratio between target characteristic size and TOD test pattern size at threshold equals 6.3. Note that this number is purely based on empirical data without any intermediate modeling. The calibration of the TOD to the TTP is highly beneficial to the sensor modeling and testing community for a variety of reasons. These include: i) a connection between requirement specification and acceptance testing, and ii) a very efficient method to quickly validate or extend the TTP range prediction model to new systems and tasks.

Keywords:TTP, TOD, Target Acquisition, perception, identification, range prediction, test method

1. INTRODUCTION

1.1 General Introduction

When new Electro-Optical or Infrared (EO/IR) cameras are required by the DoD, their performance required to carry out the intended task needs to be specified. An important military observer task is Target Acquisition (TA): the Detection, Recognition and Identification of military relevant targets. Hence, the range at which military targets can be distinguished when using the camera is a key requirement parameter in the procurement process of any new military sensor system.

Theoretical models and sensor tests that are able to supply TA range performance with camera systems are both inevitable elements in procurement processes and acceptance testing TA models are used to theoretically predict if a certain camera system will meet the requirements. They calculate the expected Target Acquisition performance on the basis of the physical parameters of the sensor.

Sensor performance tests actually measure the performance with a real camera system, often with a human-in-the-loop. They are required for (lab and field) acceptance of the delivered systems and their maintenance. Acceptance testing is essential since the sensor parameters required for the model may not all be available. Another reason is that sensor systems may not meet their expected performance under practical circumstances.

Ideally, a sensor performance model and a sensor performance test are two complementary parts of a single methodology that are strongly connected and predict the same thing: field performance. This is illustrated in Figure 1 (left graph). Such a methodology was developed in the eighties for thermal imagers and Image Intensifiers and is described in a set of standard NATO Agreements or STANAGs[1,2,3].

In the second half of the nineties it became apparent that the standard models and end-to-end performance tests (such as the MRTD or Minimum Resolvable Temperature Difference for thermal imagers) described in the STANAGs were

Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXII, edited by Gerald C. Holst, Keith A. Krapels,
Proc. of SPIE Vol. 8014, 80140L · © 2011 SPIE · CCC code: 0277-786X/11/$18 · doi: 10.1117/12.887219

unable to provide accurate characterization for most modern pixel-array imaging systems. This stimulated the sensor performance testing and modeling community to develop alternative methods[4].

Figure 1. Left: Required methodology for the procurement and acceptance of new sensor systems. The model theoretically predicts whether a camera will meet the performance requirements. The sensor test measures its actual performance. The two elements are complementary and predict the same thing: field performance. Right: Current situation: The US model and NL test were developed and validated along separate lines. In order to be useful for the complete sensor procurement and acceptance process, the TOD method needs to be calibrated against the same data set as the TTP model. This will repair the desired situation sketched in the left graph.

1.2 The US approach

The US Army Night Vision and Electronic Sensors Directorate (NVESD) developed the theoretical TTP target acquisition model[5,6] that predicts Detection, Recognition and Identification range performance for military vehicles. A widely-used software package based on the TTP is NVThermIP (for IR sensors). Over a decade the TTP model has been continuously improved and extended using hundreds of hours of real observer performance data collected on real target imagery processed with simulated sensors. In addition, the TTP uses physically measured sensor blur and noise characterization. The TTP does not include a direct field measurement that correlates to range.

1.3 The Dutch approach

TNO in the Netherlands proposed a sensor test method that was able to cope with the particular testing problems associated with sampled imaging systems, the Triangle Orientation Discrimination (TOD) method[7,8]. Basically, a human observer using the sensor has to judge the orientation of triangular test patterns (apex up, down, left or right) of different sizes and contrasts. The result is a TOD curve of (thermal or visual) contrast threshold versus size. Specific features of the method include ease-of-use, accuracy, similarity to the TA observer task, and easy transfer to existing TA models.

Over the years it was shown that the method can be applied to a very wide range of sensors, systems and platforms, and for specific applications the method is being used by major laboratories such as NVESD, ARL, NRL, and MESA. In addition to the human observer test, the method has been extended with a Human Vision System (HVS) model that allows automated and objective testing without a human observer[9,10].

Although the TOD has been successfully tested against many types of image degradations[11-16], the method has not been systematically validated against the US military target set.

1.4 Current stage and problem

The US theoretical model and the Dutch sensor test are both complementary components of a full and feasible methodology required for procurement and acceptance. However, the two approaches followed a different path and are tested against a different set of field perception data, and currently cannot be used together to cover the full acquisition process (see Figure 1, right graph). By limiting the procurement process to theoretical predictions, there exists a potential risk that fielded sensor systems do not meet the specified performance under operational circumstances. Field tests with real targets may reduce that risk but are very expensive and often not feasible.

By validating the TOD method against the extensive US data set the two components may be integrated, thus repairing the desired situation as sketched in Figure 1 (left graph).

1.5 Present study

In the present study, we will assess the relationship between tactical vehicle perception data and TOD data for a wide range of sensors, and investigate how robust this relationship is under variation of the most fundamental parameters: noise, contrast, blur, and sampling. The key parameter describing this relationship will be the ratio between the TOD triangle pattern size and the target characteristic size at the same target contrast and a specified performance level. The better TOD represents real target ID, the more constant this ratio will be across the sensor parameter space.

We will use a selection of the extensive NVESD human vehicle perception database. In order to ensure a direct comparison, the corresponding simulated sensors will process the TOD test patterns for the TOD measurements in exactly the same way as the tactical vehicle imagery were processed for the US perception experiments.

The paper is organized as follows. In Chapter 2, we will describe the NVESD data used in this study and reanalyze them according to the standard TOD analysis procedure. In Chapter 3, we will describe the corresponding TOD experiment. The results of this experiment and the comparison are described in Chapter 4. Finally, the results will be discussed in Chapter 5 and conclusions will be provided in Chapter 6.

2. ANALYSIS OF THE NVESD PERCEPTION DATA

2.1 Overview of selected experiments

Sensor simulation scripts (MATLAB code) and corresponding raw observer data were provided by NVESD for a selection of 9 experiments (public release only). An overview of the experiments and their most relevant independent variables are provided in Perception test

The images were presented to a number of active military personnel.The observer task was to name the target, i.e. a 12 alternative forced choice (12AFC) vehicle identification task. Prior to the test, the observers had gone through an extensive training phase (using the ROC-V training package) after which they scored at least 95% correct on the undegraded targets.

Two different displays were used: an 8-bit color CRT with a 0.381 mm dot pitch, and a 10-bit, high resolution black and white display with a 0.14 mm dot pitch. Background luminance was 17.1 cd/m2, and no gamma correction was applied. The color CRT was used in experiments 6a, 6b, 9, 13 and 19 at a viewing distance of approximately 45cm, and the black and white display was used in experiments with a viewing distance of approximately 38 cm.

Table 1.The experiments are described in several papers and reports[5,6,17]. For a detailed description we refer to those papers and reports.

In each experiment, a number of sensors (2, 3, 4) were simulated. We will name these after the experiment number adding the letters A,B,C, D. For example, the two sensors in Experiment 6a will be named sensor 6aA and 6aB.

In each experiment, six difficulty levels per sensor (named a-f) were introduced by either changing the MTF-cut-off spatial frequency (in Experiments 6a, 6b, 9, 13, 19, 34, 35) or simulated target range (Experiments 36 and 36a).

2.2 Image generation

A pristine thermal image set of 12 military targets and 12 aspects, recorded in the field with an Agema 1000 in NFOV (5 x 3.3 degrees), served as the basis of all experiments. Target distance was 125 m, characteristic size was 3.11 m (angular size = 24.88 mrad) and contrast (grey level contrast of the recorded imagery) was 0.205 according to the definition given by Vollmerhausen[5] except in experiment 34 where contrast served as a variable. All imagery were processed using the MATLAB sensor simulation scripts, resulting in a set of degraded images.

2.3 Perception test

The images were presented to a number of active military personnel.The observer task was to name the target, i.e. a 12 alternative forced choice (12AFC) vehicle identification task. Prior to the test, the observers had gone through an extensive training phase (using the ROC-V training package) after which they scored at least 95% correct on the undegraded targets.

Two different displays were used: an 8-bit color CRT with a 0.381 mm dot pitch, and a 10-bit, high resolution black and white display with a 0.14 mm dot pitch. Background luminance was 17.1 cd/m2, and no gamma correction was applied. The color CRT was used in experiments 6a, 6b, 9, 13 and 19 at a viewing distance of approximately 45cm, and the black and white display was used in experiments with a viewing distance of approximately 38 cm.

Table 1 Overview of perception experiment scripts and sensors provided by NVESD.

Experiment	# sensors	Degradation type	variable
6a	2	blur	Linear filter type (Gaussian, Exponential)
6b	3	Blur	Linear filter type (Gaussian, Exponential, Exponential)
9	4	Blur	Linear filter type (Gaussian, Exponential, DOG, Rectangle)
13	4	Noise	White noise level
19b	4	Noise	White noise level
34	4	Contrast, boost	Contrast (2 levels) and boost (2 levels)
35	4	Noise	Noise type: No noise, White noise, Low pass noise, High pass noise
36	4	Sampling	Fill factor, display interpolation
36a	4	Sampling	Fill factor, display interpolation

2.4 Analysis used in this study

In order to make the best possible comparison between vehicle and TOD data, they are both processed using the regular TOD analysis procedure. This procedure differs from the procedure that is used by NVESD. The advantage of the present procedure is that it yields an accuracy estimate of the thresholds based on i) the internal error in the individual observer thresholds and ii) the external error determined by the differences between the observer scores. Such an error estimate is essential in order to judge the significance of differences between the TOD and NVESD vehicle identification performance. Note that the resulting error is determined by statistics only; systematic errors due to changes in setup, viewing distance or eye correction may significantly affect the result but are not included.

In the first step of the analysis, for each observer and for the entire observer group in an experiment, we calculated the maximum likelihood fit of a Weibull function (an s-shaped function) and confidence interval of the fraction correct versus MTF cut-offs (6a, 6b, 9, 13, 19, 34, 35) or ranges (36, 36a) for each of the sensors A,B, C and D. the same fit was applied to the overall observer scores. The Weibull function we use is of the form:

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

where x = stimulus strength (either contrast or size), α is a stimulus strength threshold value (see below), β determines the steepness of the function, γ is the guess rate, i.e. in a 4AFC task γ = 0.25 and in the NVESD task γ = 1/12 = 0.083, and 1- δ is the final correct level that can be reached at infinite stimulus strength. The parameter A is introduced to achieve that a chosen threshold level θ is reached when the signal strength x equals α, independent from the values of β, γ and δ. It is given by

(2) $$A = \frac{(1-\delta-\gamma)}{(1-\delta-\theta)}$$

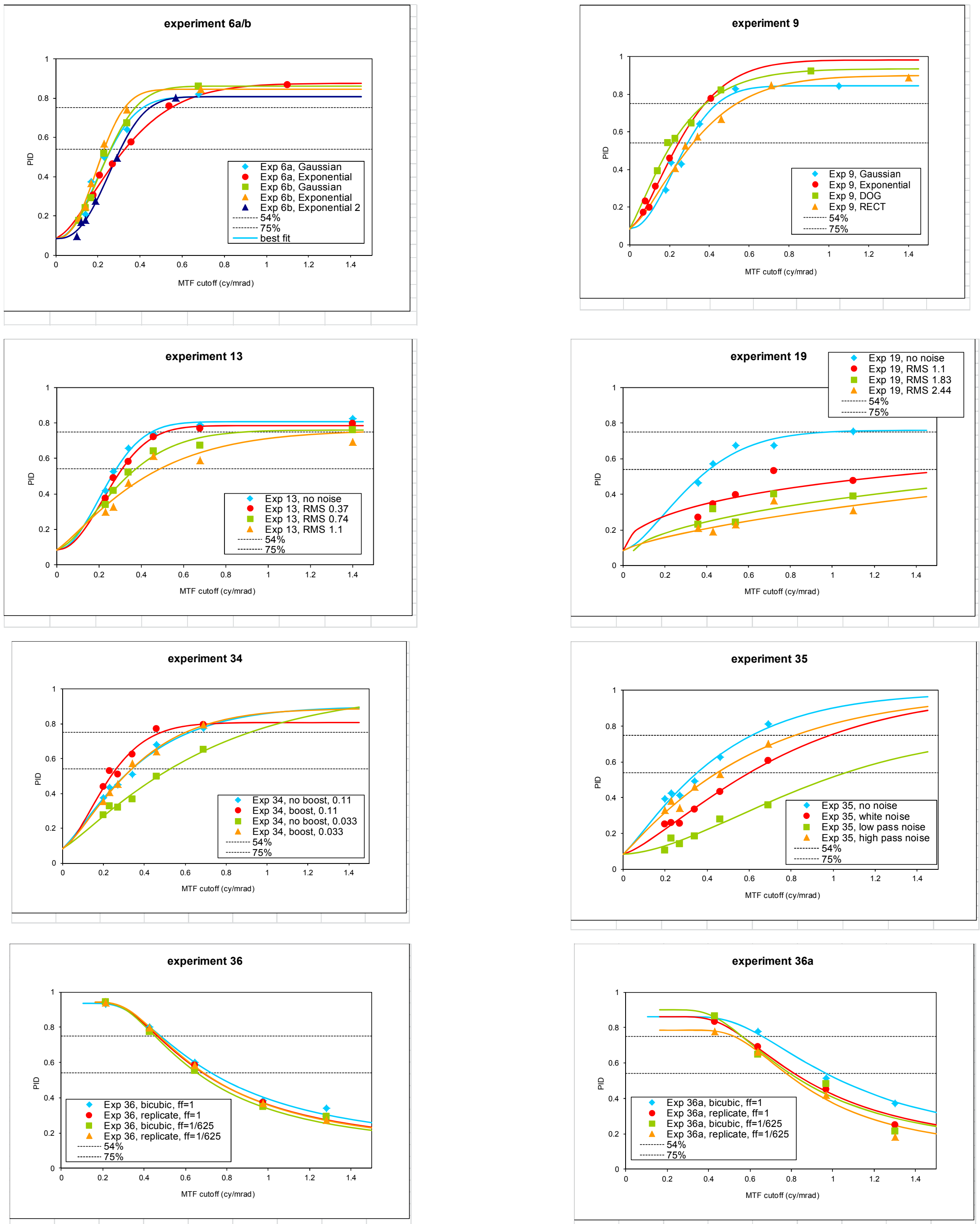

Figure 2: Probability versus MTF cut-off frequency or range relationships for tactical vehicle ID with the simulated sensors used in the NVESD historical experiments, together with the maximum likelihood fits.

For the TOD test, δ is usually set at 0.02 to take into account the probability that the observer erroneously pushes a wrong button or misses a presentation because he blinked with his eyes at that moment.

In the NVESD experiments, the correct fraction often deviates significantly from 1 even at high MTF cut-off frequencies or at short range. For this reason, δ was taken a free parameter in these fits.

The second step was to exclude the data from a few observers who performed extremely poorly (around chance level) compared to the others. Since they all were able to pass the training phase, the most probable reason is lack of motivation. Two additional observers were excluded because his/her data were incomplete. The total number of excluded observers on the basis of this analysis was 5 out of 150.

In the third step, the final threshold values were calculated by taking the weighted geometric (i.e. logarithmic) average of the threshold values over all observers. The logarithmic average is taken because observer performance values are often log normally distributed. By taking the weighted average, the observer thresholds with higher accuracy contribute more to the final threshold than values with lower accuracy. The maximum of internal and external error was taken as the standard error in the final thresholds. The internal error is determined by the accuracy of all individual observer thresholds, while the external threshold is governed by observer differences.

Thresholds were calculated at two correct levels: 75% (i.e. the regular TOD threshold level used in the TOD procedure) and 54.15% (i.e. the 50% correct level used by NVESD, but not corrected for chance).

2.5 Results of the analysis

The vehicle identification data for all 33 conditions (all sensors in all experiments) including the overall maximum likelihood Weibull fits are provided in Figure 2a-h. In Figure 2a-f, fraction correct increases with MTF cut-off frequency. In Figure 2g-h, fraction correct de creases with range which is the independent variable in these experiments. The graphs show that the Weibull nicely follows the shape of the data.

Not all data fits could be completed. For a number of conditions, some or all observers did not reach the required threshold level even at the highest MTF cut-off or smallest range. Exclusion of data from one or more observers to the final threshold estimate in some condition while including them for other conditions where they reach the threshold introduces bias. So we accepted the following rules: i) in order to always obtain a fit to the data, we limited the value of 1- δ to just above the threshold (0.76 for the 75% threshold and 0.55 for the 54% threshold, keeping in mind that these threshold values may not be realistic, ii) if three or more observers in an experiment do not reach the required threshold level at a certain condition, this threshold value (at 75% or 54%) is considered to be outside the measurable range of the NVESD set and is excluded from further analysis, and iii) If only one or two observers do not reach the threshold level, the effect of their threshold values on the weighted geometric average over the observers is observed. It turned out that the final results were minimally affected, so we decided to take this as a practical approach to treat conditions with a maximum around threshold level.

In this way, the analysis resulted in 27 thresholds at 54.15% and only 10 out of 33 at 75%. The statistical error (the maximum of internal and external error in the weighted geometric average) was 5-14% (at 54%) and 10-17% (at 75%).

3. TOD EXPERIMENT

3.1 Comparison procedure

The real target set is characterized by two physical parameters: characteristic angular target size and characteristic target contrast. These are provided in section 2.2. The fastest way to make the comparison would be to take the 75% and 54.15% correct thresholds (MTF cut-offs for experiments 6-35 and ranges for experiments 36 and 36a) obtained in section 2.5, process for those threshold levels images of a range of TOD test patterns with the same contrast and determine the corresponding triangle threshold sizes in a human observer experiment. However, we will calculate TOD images with the original scripts and measure TOD thresholds for each of the six difficulty levels. The triangle threshold size corresponding with the 75% and 54.15% MTF cut-offs and ranges will be obtained by interpolation. This is more elaborate but in this case the original sensor scripts are unchanged as far as possible, which means a more direct comparison between vehicle and TOD experiments. In addition, the thresholds received this way will be based on more estimates and be more accurate. Finally, we receive a full set of TOD thresholds that can be used in other applications, for instance to validate simulation models.

3.2 Test pattern generation and selection

TOD test patterns were generated ranging from 0.5 – 8 mrad in 20 sizes (equidistant on a log scale). For each size, 16 triangle were generated, varying in orientation, horizontal, and vertical position. Contrast was 0.41 using the definition as used in the TOD method[9] (i.e. 0.205 in Vollmerhausen's definition[5]).

All images were processed through the simulated sensors using the MATLAB codes provided by NVESD. This resulted in a total of 20 (sizes) * 16 (per size) * 33 (sensors) * 6 (MTF cut-offs resp. ranges) = 63360 images with a 16-bit depth. Example pictures are shown in Figure 3.

In a pilot experiment, the approximate threshold was estimated for each of the 198 conditions and 7 triangle sizes around this threshold were selected for the final experiment (a total of approximately 20,000 images).

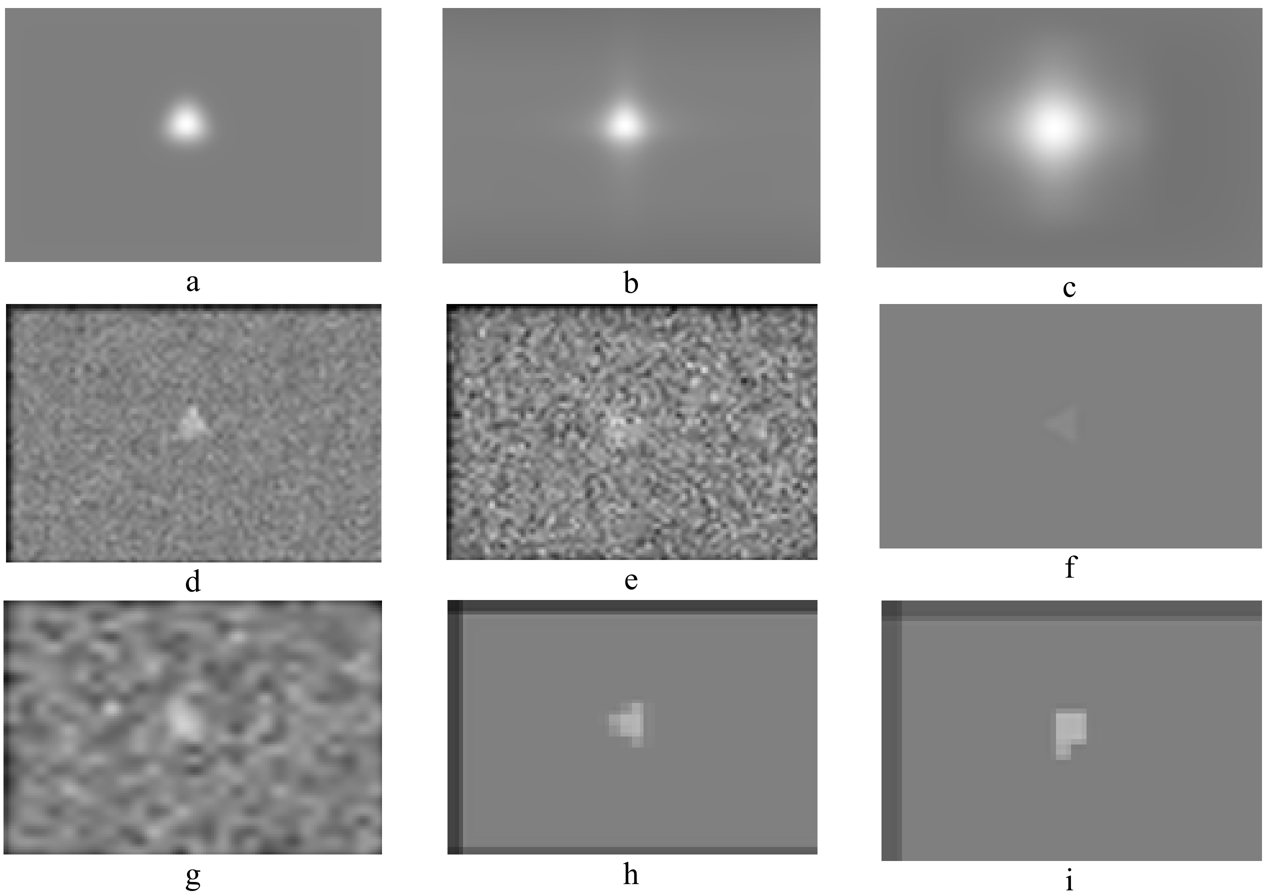

Figure 3 Example images of the TOD test patterns processed by the simulated sensors. a: Gaussian blur, b. Exponential blur. c: DOG filter. D,e: white noise, f: low contrast, g: band-limited noise. h, i: sampling.

3.3 Experimental setup and test procedure

An experimental setup was built that mimics the technical specifications of the original NVESD setup as closely as possible. It consists of a Dell OptiPlex GX270,a Bits++ Digital Video Processor from Cambridge Research Systems and

22” Superbright Mitsubishi Diamond Pro 2070SB CRT display (20”viewable), enabling the display of monochrome images in 14 bits. The screen is calibrated using the automated ColorCAL colorimeter.

For the experiment, background luminance is set to 17.8 cd/m^2 . Minimum and maximum luminance are 2.8e-3 cd/m^2 and 35.6 cd/m^2, respectively. The room is dimly lit. The video card is set to 1024 x 768 pixels at a refresh rate of 75 Hz. For experiment 13 and 19 the dynamic noise is presented with a frame rate of 60 Hz; one of the 16 frames is randomly chosen.

During the experiment the observer is seated at 45.7 cm from the screen for the experiments that need to mimic the original 8-bits monitor and 106 cm for the experiments that need to mimic the original 10-bits gray scale CRT. See Figure 4. The observer can check the correct distance by using one of the two wires with the correct length attached to the side of the display. Image presentation is fully computer-controlled. After each presentation, the observer responds with on of the four arrow keys on a keypad. After each response the next stimulus is presented. This enables response collection at fast pace.

Five observers between 30 and 57 participated in the experiment: WV, JA, MH, PB and JF. Each selected image was presented to each observer once, resulting in a total of approximately 100,000 responses. The images were presented in 8 separate sessions corresponding to the experiments 6a/b, 9, 13, 19, 34, 35, 36 and 36a. The stimulus order within a session was randomized, an the order or the experiments were randomized for each observer. Each session took approximately one hour or a little more. The total amount of measurement time over all observers was about 50 hrs.

Figure 4 Experimental setup with observer at 106 cm from the screen.

3.4 Data analysis

Threshold calculation is described in section 2.4. Thresholds were calculated at the 75% correct level only. For the maximum likelihood fit, a Weibull with fixed variables $\gamma = 0.25$ and $\delta = 0.02$ was taken. There were no observers that needed to be excluded from the analysis. For each sensor and each MTF cut-off/range, the weighted geometric average of the triangle angular threshold size (in mrad) is calculated.

3.5 Results of the TOD experiments

Triangle thresholds for all 33 conditions are plotted as a function of 1/MTF cut-off frequency respectively range are plotted in Figure 5. Experimental errors are small in most cases. By plotting the data this way, most plots show a linear relationship over a wide region. In those regions, a least-square linear fit was applied to the data. There were two conditions for which no thresholds could be estimated (above 8 mrad for all conditions (a-f) for at least two observers), and those were excluded from further analysis. One of these conditions corresponded with one of the excluded conditions in the NVESD experiment (see 2.5). This leaves a total of 26 (out of 33) threshold values at the 54% level and 10 at the 75% level for the final comparison (see Chapter 4).

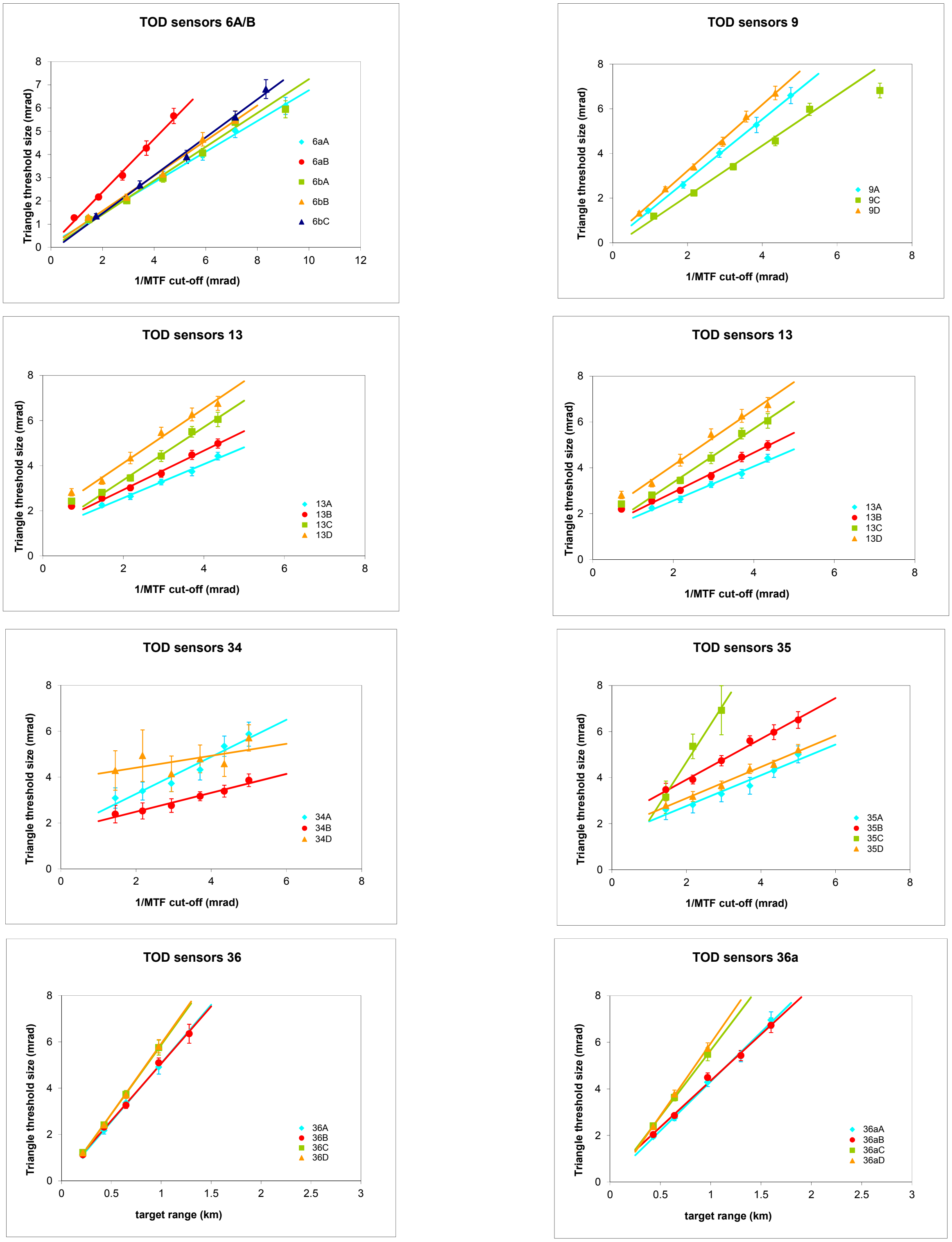

Figure 5: Threshold triangle size as a function of 1/MTF cut-off frequency (in mrad/cy) or range (in km). A least-square linear fit is applied where the relationship does not significantly deviate from linear.

4. COMPARISON

4.1 Procedure

For each of the 33 sensors in the experiments, the triangle threshold sizes corresponding to the 75% and 54.15% correct MTF cut-off or range for tactical vehicle ID were calculated as follows. First, the threshold MTF cut-off (in cy/mrad) of the range (in km) and the experimental error for the NVESD data is calculated using the procedure described in section 2.4 (Figure 6, left graph). Next, the corresponding triangle threshold (and error) are found by calculating the y-value of the linear least-square fit to the TOD data shown in section 3.5 see right graph of Figure 6.

Finally, the ratio between tactical vehicle characteristic angular size (i.e. 24.88 mrad) and corresponding TOD 75% correct threshold size (in mrad, including error) is calculated

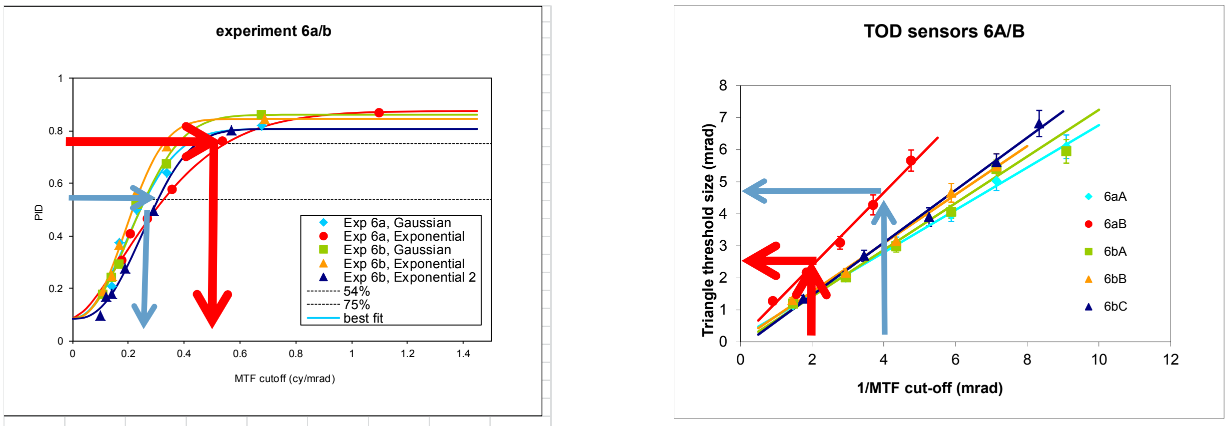

Figure 6: Threshold triangle size as a function of MTF cut-off frequency or range. A least-square linear fit is applied where the relationship does not significantly deviate from linear.

4.2 Results of the comparison

The overall results are plotted in Figure 7. This graph shows the Tank/TOD size for tactical vehicle identification at the 54.15% and the 75% correct level for all the sensors for which a threshold estimate could be calculated. Taking the average over all values results in the following estimates:

- $M_{54} = 6.3 \pm 0.2$
- $M_{75} = 8.8 \pm 0.3$

where M denotes the multiplication or scaling factor required to convert the 75% correct triangle threshold size into the tactical vehicle angular size at 54 (i.e. 50% after correction for chance) and 75% correct level.

For the 75% correct level, no data points deviate significantly from the average. For the 54% level there are significant deviations (see also section 5.2). However, subdividing the data into groups for blur (experiments 6a, 6b, 9), noise (experiments 13, 19, 35), contrast and boost (experiment 34) and sampling (experiments 36 and 36a) results in very small differences between the ratios.

In words:

- "the 50% correct ID range on the standard NVESD 12 tactical vehicle/12 aspect target set (3.11 m) is equal to the 75% correct orientation discrimination range on a TOD triangle set (triangle size is 3.11/6.3 = 0.49 m), regardless of the sensor system (blur, sampling, spatial and temporal noise) used."
- "the 75% correct ID range on the standard NVESD 12 tactical vehicle/12 aspect target set (3.11 m) is equal to the 75% correct orientation discrimination range on a TOD triangle set (triangle size is 3.11/8.8 = 0.35 m), regardless of the sensor system (blur, sampling, spatial and temporal noise) used."

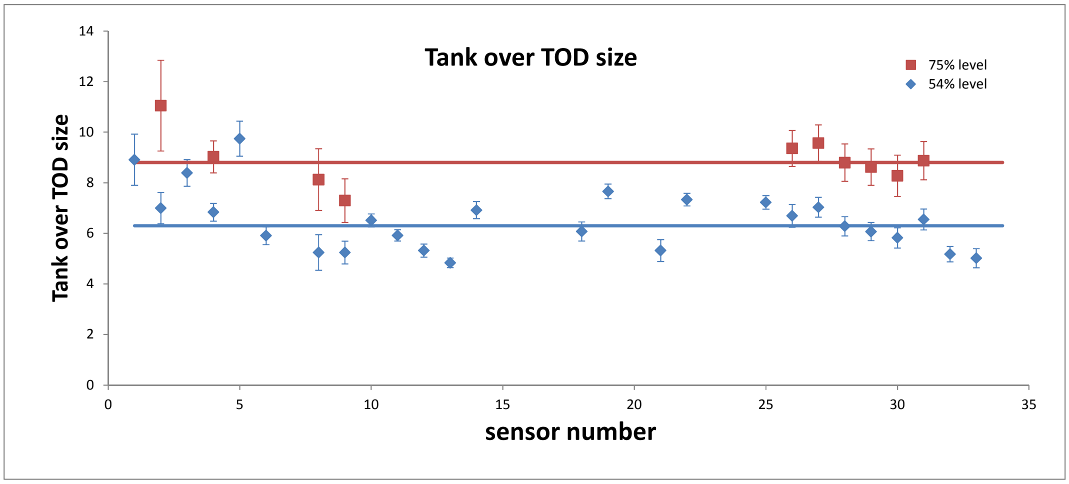

Figure 7: Tank/TOD size at threshold for all sensors. The average value at 54% (50% after correction for chance) vehicle ID is M_{54} = 6.2± 0.2, at 75% it is M_{75} = 8.8 ± 0.3.

5. DISCUSSION

We have made a direct comparison between the tactical vehicle ID size and a triangle orientation discrimination size at threshold, when the image is degraded by a variety of different sensor types. The result is that the ratio depends very little on the type of sensor degradation. Note that the results are purely based on the statistical analysis of empirical data and this require no underlying model or assumption except that the vehicles and the triangle test patterns have the same (RSS) contrast value.

In this Chapter, we make some additional comments to the comparison study.

5.1 Additional analyses

In some cases, a 75% or even 54% threshold could not be obtained with the current analysis. With the noise experiments, an alternative analysis will be carried out using the noise floor as independent variable, which may result in additional thresholds for the comparison.

5.2 Error sources

In the analysis and error calculation, we have limited ourselves to the statistical uncertainties in the data. However, there are many factors that may introduce (systematic) errors and have not been taken into account. To name a few: i) the NVESD experiments have been collected over years. This means e.g. that the setup and test procedures may have changed, or that the observer population may have shifted in qualification and motivation. Some setup parameters are difficult to trace back, e.g. the gamma of the CRT. In addition, Vollmerhausen[5] reports that some experiments depend critically upon the viewing distance, but these were not always controlled. Given these factors, the variation of the ratio M_{54} over the conditions is quite satisfying.

5.3 Efficiency of the TOD method

The tactical vehicle experiments for 33 sensors approximately took 480 hours (i.e. 15 observers * [2 hours training + 2 hours test] * 8 experiments). The TOD experiment for 6 * 33 = 198 sensors took 50 hours, while the accuracy is two times better. Assuming that an accuracy increase by two requires a 4 times larger experiment, collecting TOD thresholds is approximately 240 times more efficient.

5.4 Human Visual System model

With the set collected, we will assess the accuracy of the HVS (Human Visual System) model developed to perform the TOD test[9,10]. A validated model further increases the possibilities and efficiency of the method, e.g. enabling automated sensor performance testing and automatic characterization and optimization of signal processing

6. CONCLUSIONS

1. The standard 12/12 NVESD tacticle vehicle test, used to setup and calibrate the TTP metric, has an operational equivalent for lab and field applications We have empirically shown that the vehicle set can be replaced by a TOD triangle test set of 1/6.3 times the characteristic size of the target set and the same contrast
2. Using the TOD instead of the 12 target set, TTP validation and extension can be sped up by a factor of 240.
3. More benefits can be expected with a validated Human Visual System model.

ACKNOWLEDGMENTS

This study was supported by USAITC-A.

REFERENCES

[1] STANAG 4347.
[2] STANAG 4349.
[3] STANAG 4350.
[4] Driggers, R.G., Wittenstein, W., Williams, W., Bijl, P., Lopez, J.M. (2003). Modeling of Undersampled Imaging Systems. NATO Technical Report RTO-TR-075(I), AC/323(SET-015)TP/25.
[5] Vollmerhausen, R.H., Jacobs, E., Hixson, J. & Friedman, M. (2006). The Targeting Task Performance (TTP) Metric: A New Model for Predicting Target Acquisition Performance. Technical Report AMSEL-NV-TR-230, NVESD, Fort Belvoir, VA 22060 (January 2006).
[6] Vollmerhausen, Richard H., Eddie Jacobs, and Ronald Driggers (2004), New metric for predicting target acquisition performance. Optical Engineering, Vol.43, No.11, pp 2806-2818.
[7] Bijl, P.& Valeton, J.M. (1998). TOD, the alternative to MRTD and MRC. Optical Engineering 37, 7, 1976 - 1983.
[8] Bijl, P.& Valeton, J.M. (1999). Guidelines for accurate TOD measurement. SPIE Proceedings, Vol. 3701 14 - 25.
[9] De Lange, D.J., Valeton, J.M. & Bijl, P. (2000). Automatic characterization of electro-optical sensors with image-processing, using the Triangle Orientation Discrimination (TOD) method.
[10] Hogervorst, M.A., Bijl, P. & Valeton, J.M. (2001). Capturing the sampling effects: a TOD sensor performance model. SPIE Proceedings Vol. 4372, 62-73.
[11] Bijl, P., Beintema J. A., Dijk, J. & Leden, N. van der. (2010). Effectiveness assessment of signal processing in the presence of smear. In: Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXI, 7662, 76620E-76620E-10.
[12] Bijl, P. (2010). Visual Image Quality Assessment with sensor motion: Effect of Recording and Presentation Velocity. Applied Optics, Vol. 49, Issue 3, pp. 343-349.
[13] Bijl P. & Hogervorst M.A. (2009). Sensor performance as a function of sampling (d) and optical blur (Fλ). In: Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XX, 7300, 7300, 73000C.
[14] Bijl, P., Hogervorst, M.A. & Toet, A. (2004). Identification of military targets and simple laboratory test patterns in band-limited noise. SPIE Proceedings Vol. 5407, 15, 104-115.
[15] Bijl, P., Valeton, J.M. & de Jong, A.N. (2000). TOD predicts target acquisition performance for staring and scanning thermal imagers, SPIE Proceeding Vol. 4030, 96-103.
[16] Bijl, P.& Valeton, J.M. (1998). Validation of the new TOD method and ACQUIRE model predictions using observer performance data for ship targets. Optical Engineering 37, 7, 1984 - 1994.
[17] O'Connor, J.D., Driggers, R.G., Vollmerhausen, R.H., Devitt, N., & Olson, J. (2003). Fifty-percent probability of identification (N50) comparison for targets in the visible and infrared spectral bands. Op. Eng. Vol. 42, 3047.