Applicability of TOD, MTDP, MRT and DMRT for dynamic image enhancement techniques

Piet Bijl[a], Klamer Schutte[b] and Maarten A. Hogervorst[a]
[a]TNO Defense, Security & Safety: Human Factors, P.O. Box 23, Soesterberg, The Netherlands
[b]TNO Defense, Security & Safety: Observation Systems, The Hague, The Netherlands
Phone: +31 346 356 277, fax: +31 346 353 977
E-mail: piet.bijl@tno.nl, klamer.schutte@tno.nl, maarten.hogervorst@tno.nl

ABSTRACT

Current end-to-end sensor performance measures such as the TOD, MRT, DMRT and MTDP were developed to describe Target Acquisition performance for static imaging. Recent developments in sensor technology (e.g. micro-scan) and image enhancement techniques (e.g. Super Resolution and Scene-Based Non-Uniformity Correction) require that a sensor performance measure can be applied to dynamic imaging as well. We evaluated the above-mentioned measures using static, dynamic (moving) and different types of enhanced imagery of thermal 4-bar and triangle tests patterns. Both theoretical and empirical evidence is provided that the bar-pattern based methods are not suited for dynamic imaging. On the other hand, the TOD method can be applied easily without adaptation to any of the above-mentioned conditions, and the resulting TOD data are in correspondence with the expectations. We conclude that the TOD is the only current end-to-end measure that is able to quantify sensor performance for dynamic imaging and dynamic image enhancement techniques.

Keywords: Target Acquisition, dynamic imaging, sensor performance, image enhancement, super resolution, scene-based non-uniformity correction

1. INTRODUCTION

There is an increasing interest in the use of thermal imagers under dynamic conditions. There are several reasons. First, the military application has undergone a shift from a long range, static theatre towards more dynamic situations, such as the use in urban environments, and the increased speed of operation asks for imaging a moving target or from a moving platform. Second, when using an undersampled pixel-array camera, motion can be used to improve sensor performance because effectively the number of samples over the scene increases. Some sensors make use of this by having a built-in micro-scan. Other applications are signal processing techniques such as Super Resolution[1,2] or Scene-Based Non-Uniformity Correction[2]. These techniques now become of practical importance since they nowadays can be applied in real-time[3]. With hand-held systems, range performance can be higher when operated hand-held than from a tripod[9].

These developments inevitably ask for an end-to-end measure that is able to quantify dynamic sensor performance or system performance including signal processing, and to compare this performance to that under conventional static imaging. Currently available test methods, such as the MRT (Minimum Resolvable Temperature Difference)[4], TOD (Triangle Orientation Discrimination)[5], MTDP (Minimum Temperature Difference Perceived)[6], and DMRT (Dynamic MRT)[7] have all been developed to determine static image performance. The question is how robust these measures are when applied to dynamic imaging with and without signal processing techniques. Applicability means that the method is theoretically sound, that the test method is practically feasible, and most importantly, that the result has a direct relationship with field performance.

The goal of the present study is twofold. First, we consider the applicability of the measures that are currently most applied: MRT (the current NATO standard), TOD, MTDP and DMRT. Second, we wish to quantify the effect of dynamic imaging and the enhancement techniques applied. We perform a theoretical analysis to the test methods, and those methods that pass this analysis are applied experimentally to static and dynamic conditions with and without

processing at high, medium and low thermal contrast. Finally, the experimental results are compared to predictions made with the Target Acquisition models NVThermIP[8] and TRM3[6].

In Chapter 2, some examples of dynamic signal processing techniques are given. A theoretical analysis of the test methods is given in Chapter 3. The experimental method is described in Chapter 4. The results of the experiments are given in Chapter 5, and compared to model predictions in Chapter 6. The results are discussed in Chapter 7.

2. DYNAMIC SIGNAL PROCESSING TECHNIQUES

Dynamic signal processing techniques such as super-resolution reconstruction (SR) or scene-based non-uniformity correction (SBNUC) and many others have been widely reported in the literature[1,2,3,8]. These techniques make use of motion in the image in order to improve pixel resolution and/or to reduce temporal noise (SR) in an undersampled system or to improve contrast sensitivity (reduce the fixed pattern noise in real time) by comparing the responses from one pixel to the other using the scene (SBNUC). An example result of SR is shown in Figure 1. In this study we will use DSR and SBNUC processing from a comprehensive signal processing software package developed by TNO.
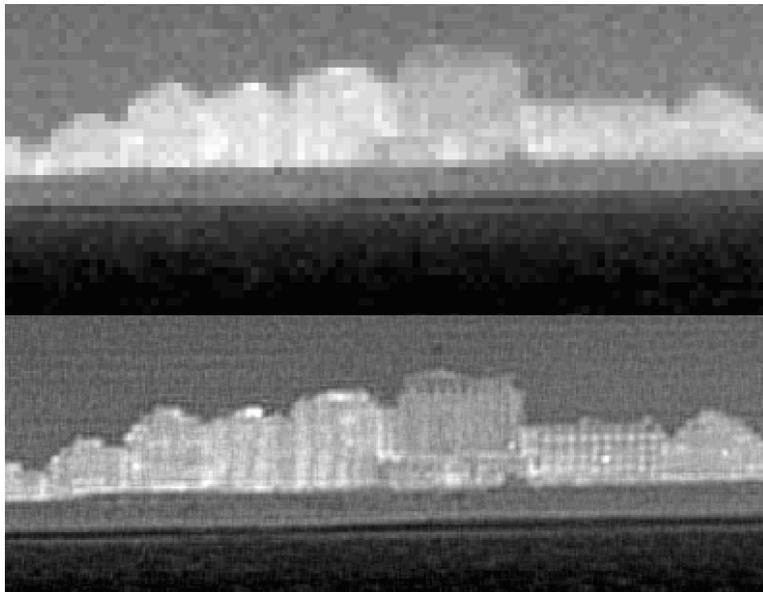

Figure 1. Example result of super-resolution. Upper picture: original image. Lower picture: image after application of the TNO Dynamic Super Resolution algorithm (DSR).

3. THEORETICAL ANALYSIS

In this chapter, we consider theoretical and practical problems and/or weak points of the different sensor performance measures when used for evaluating dynamic signal processing techniques.

3.1 MRT

The MRT is defined as the thermal threshold contrast at which the four bars of a standard bar test pattern can just be seen, and is measured as a function of spatial frequency[4]. The observer indicates this threshold while the contrast of the test pattern is temporally modulated, although many varieties of this procedure exist[10]. The MRT was designed for scanning thermal imagers that can be considered as nearly linear systems. With undersampled systems, the appearance and number of visible bars of high spatial frequency bar patterns depends largely on the position with respect to the pixel array (phase), and the MRT should not be measured beyond the half sample (Nyquist) frequency. No serious problems occur at low spatial frequencies.

Although the MRT is still the standard, it has been reported many times that the MRT, when applied to undersampled imagers, does not result in correct range predictions for real targets[11]. This makes the MRT essentially not useful for dynamic imaging, since we are interested in the effects of dynamic imaging and enhancement techniques which are mainly applied to undersampled imagers. The US NVESD range performance model NVThermIP[8] calculates the MRT but includes a warning that this MRT is only for comparison with a lab test and should not be used for range predictions. There are several other problems with the MRT. For example, MRT for an undersampled system should stop at the Nyquist frequency but with dynamic image enhancement techniques it is a priori unknown whether a system is well-sampled or undersampled. Also, the Nyquist frequency is unknown if the samples are irregularly distributed. Some other problems apply to the MTDP and DMRT as well and will be discussed in sections 3.2 or 3.3.

In conclusion, the MRT is not useful to quantify dynamic imaging. Hence, this measure is excluded from a further study.

3.2   MTDP

The MTDP[6] is one of the approaches to bypass the problems of the MRT with undersampled imaging systems. The definition and measurement procedure is almost equal to that of the MRT, with the exception that the number of visible bars should not necessarily be 4, but may also be 3 or 2. The observer is allowed to move the bar pattern to an optimum phase. This enables a threshold measurement beyond the Nyquist frequency. In practice, MTDP for an undersampled system typically cuts off around 1.7 times the Nyquist frequency and the range predictions with the MTDP for undersampled vs well-sampled systems are in closer agreement than with the MRT[12].

A practical problem with the MTDP for dynamic imaging is that during the measurement two things are varied at the same time: the test pattern position is varied while the thermal contrast is modulated (see definition in section 3.1). Since the appearance of a high spatial frequency bar target depends heavily on the phase and the optimum is only available at certain positions, it is an impossible task for the observer to find the threshold. With a still image (a picture), temporal modulation is impossible and therefore the MTDP is not defined. We propose a modified MTDP measurement procedure in which the patterns are presented at a range of fixed contrasts. With dynamic imaging, the observer may wait until the optimum phase is presented and then he judges whether the pattern is above or below threshold.

Given the definition of the MTDP (threshold at optimum phase), a priori it is expected that motion does not result in a much better MTDP curve.

3.3   DMRT

The DMRT[7] is another MRT variation to avoid the problems associated with undersampled imaging systems. In this method, the bar pattern is moved relative to the sampling lattice with optimum speed (about ¼ pixel per frame). In this way, the human observer is able to integrate over the various positions and a four-bar pattern is perceived at spatial frequencies considerably higher than the Nyquist frequency.

Of course, this is exactly the subsampling mechanism that enables the observer to obtain a better natural image with motion. Therefore, the DMRT should be a measure for dynamic rather than static performance. In addition, by using a moving pattern, the method bypasses many characteristics of sampling that play a role in static imaging. Another fundamental drawback is that the test procedure itself depends on the sensor characteristics (the optimum speed of the bar pattern is a function of frame rate and pixel pitch) and this is unacceptable. For example, consider two identical sensors: one operates at 50 Hz, the other at 500 Hz. Assume that, for the first one, the display decay is fast enough and the bar pattern travels at a convenient speed to the observer. The second sensor requires a 10 times higher test pattern speed and this may result in a much poorer performance curve.

There are additional problems when applying the DMRT to still and dynamic imagery. As with the MTDP and the MRT, temporal modulation and motion at the same time makes the task impossible to the observer. However, with the DMRT, motion is essential and this makes the method inapplicable to still imagery. Further, the method uses optimum speed to predict static sensor performance, which means that the dynamic performance measured with the DMRT can

never exceed that for the static case. Finally, with arbitrary signal processing on samples that are irregularly distributed (which is the case with super resolution applied to a hand-held sensor, for example), it is impossible to a priori know the optimum speed. This means that for each and every case thresholds at a range of speeds need to be measured and the optimum speed has to be determined afterwards. This we consider unworkable.

In conclusion, the DMRT contains a large number of theoretical and practical problems, and no useful outcome with respect to real sensor performance is expected. Hence, this measure is excluded from the remainder of this study.

3.4   TOD

The final test method in this study that was developed to cope with undersampled imagers, is the TOD[4,9]. In the TOD method, a triangular test pattern in four possible orientations (Up, Down, Left or Right) is presented to the observer who has to judge the orientation. The test patterns are presented at a range of fixed contrasts and sizes. Threshold contrast (at a certain size) is defined as the contrast at which the observer correctly judges 75% of the presentations. This results in a thermal threshold contrast curve similar to the MRT and its derivatives.

There are two important differences with the MRT family: 1) the method uses a non-periodic test pattern, and 2) the observer has to choose between several alternatives. The non-periodic test pattern is essential whenever under-sampling plays a role. The basic idea behind the test pattern is that it represents features of the objects (e.g. military targets) that need to be recognized. In this way the method is very close to the field application it represents and there is a high probability that new sensor developments will act on both in a similar way. Using more than one alternative yields a bias-free threshold, because judgement errors can be objectively determined. In addition such a procedure is close to real target acquisition where the observer has to choose between more than one target.

With dynamic imaging, no practical problems are expected. Driggers et al.[14] report on a super resolution reconstruction performance study in which they applied the TOD test method at high contrast in the field.

4. EXPERIMENT

4.1   EXPERIMENTAL SET-UP AND IMAGE COLLECTION

4.1.1.   Sensor

We used a FLIR SC2000 undersampled, uncooled microbolometer sensor with a focal plane array of 320 by 240 pixels (see Figure 2a). FOV is 24 by 18 degrees. The camera gives a calibrated output of the temperatures in the scene. Data was 14 bit digital recorded at a frame rate of 50 Hz.

4.1.2.   Test pattern generation

Thermal test patterns were generated with the TCAT[15] (Thermal Camera Acuity Tester, see Figure 2b). The standard test plate consists of 5 lines with 4 thermal triangle test patterns of arbitrary orientation on each line. The test patterns at the top line are the largest (triangle base = 2 cm, or triangle square-root area S = 1.32 cm) and every next line the test pattern size decreases by a factor of $\sqrt[4]{2}$,so that the test pattern size decreases by a factor of two over the plate. The TCAT is used like an eye chart but then with a thermal imager in the loop. The test plate can be placed in the apparatus in four different orientations, which enhances the number of possible test pattern presentations and makes learning from the heart more difficult. Standard contrast settings are approximately $\Delta T = 0, \pm 2, \pm 5, \pm 10$ and $\pm 20K$ but by applying an offset to the zero condition, stable small contrasts in the order of 0.2 K can be produced. The actual TCAT contrasts are not calibrated and are different from the above values. Therefore the actual thermal contrasts were measured using the FLIR camera.

A second test plate (not shown) consists of a standard four-bar pattern (cycle width 1 cm).

A surface mirror was placed in front of the camera objective under an angle of approximately 45 degrees (see Figure 2a.) so that a TCAT image was obtained (this image can be seen on the small display on top of the camera). The mirror was mounted on the axis of an electric motor and slightly tilted with respect to the rotation axis in order to produce a circular motion of the image. Diameter and speed are set by the changing the tilt angle and rotation frequency.

Figure 2. Left: Camera and the rotating tilted mirror used to generate a dynamic image. Right: the TCAT used to generate the thermal test patterns. See text for details.

4.1.3   Triangle test pattern recordings

The TCAT was placed at four ranges from the sensor differing by a factor of 2. Hence, test patterns varied in angular size by a factor of 16. At each range, recordings were made with 4 test plate orientations.

Three contrasts were used: $\Delta T = 0.23K$, 1.93 K and 16.9 K. These values were determined with the thermal imager. For the lowest contrast, values differed slightly with the orientation of the test plate: $\Delta T = 0.19$ K for the two test plate conditions with the largest triangle sizes on top, and $\Delta T = 0.26$ K for the two test plate conditions with the smallest triangle sizes on top. All observer scores were analyzed separately for these two different contrasts and no significant differences were found. Hence, the two values were taken together. Example images are shown in Figure 3.

Both static (50 frames) and dynamic recordings (250 frames) were made. For the dynamic recordings, the diameter of the circular motion was 0.23º and rotation frequency was 1.0 Hz. This corresponds to a test pattern speed of approximately ¼ pixel per frame.
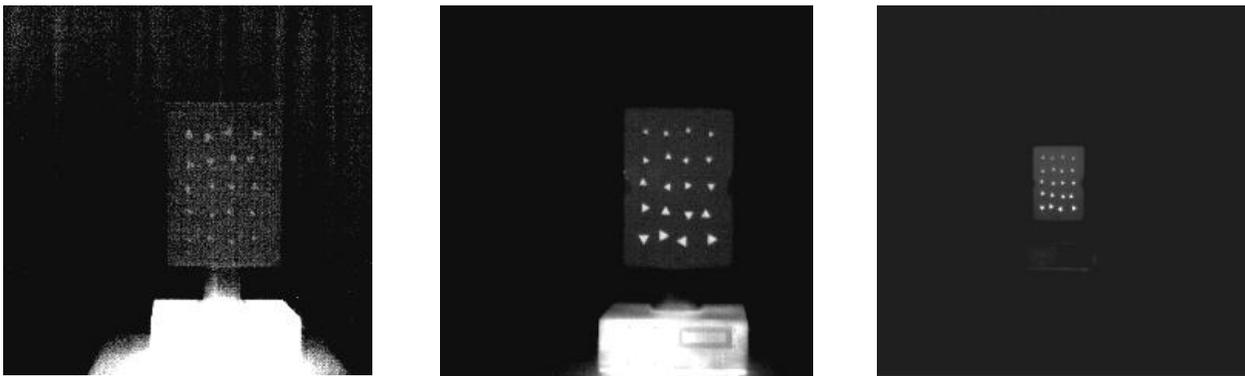

Figure 3. Examples of the triangle test pattern images. Left: $\Delta T = 0.23K$. Center: $\Delta T = 1.93$ K. Right: $\Delta T = 16.9$ K

4.1.4   Bar test pattern recordings

Bar test patterns were recorded at 25 different target ranges. Bar spatial frequencies are 0.25, 0.50, 0.75, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9 and 3.0 times the Nyquist spatial frequency of the FLIR camera. Only one (high) thermal contrast was used: $\Delta T = 4.3$ K.

Both static (50 frames) and dynamic recordings (250 frames) were made. Static recordings were made with the bar pattern in optimum phase and out-of-phase. For the dynamic recordings, the diameter of the circular motion was 0.94º and rotation frequency was 0.32 Hz. This corresponds to a test pattern speed of approximately ¼ pixel per frame.
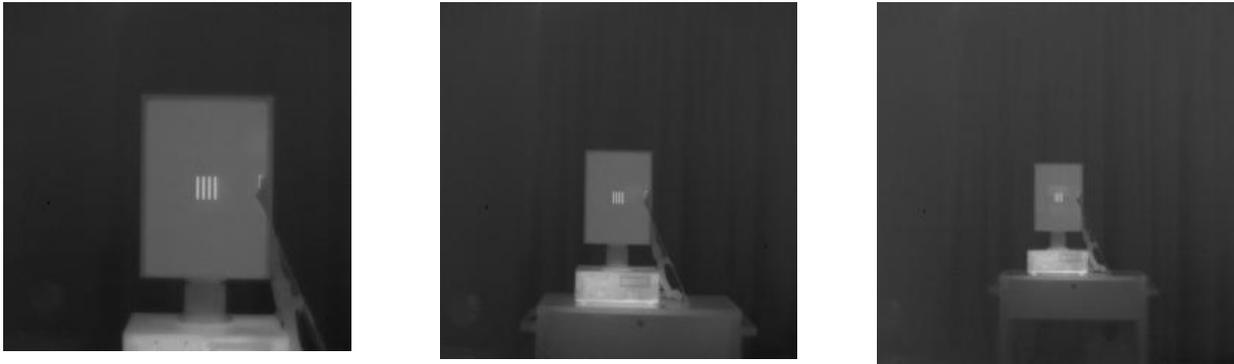

Figure 4. Examples of the collected bar test pattern images at spatial frequencies f_N (Nyquist), $1.3 \cdot f_N$ and $1.7 \cdot f_N$

4.2   SIGNAL PROCESSING

Basically, seven different conditions are used in this study. These conditions are listed in the left-hand column of Table 1. Three of these conditions are unprocessed: 'still' (which is a single frame taken from the static recordings), 'static' and 'moving' (these are described in the previous section). In the next four conditions signal processing has been applied to the dynamic recordings using the package developed at TNO (see Chapter 2): SBNUC, DSR 1, DSR 2 and DSR 4. As mentioned in Chapter 2, these techniques are only useful when applied to dynamic scenes. DSR 1 is a condition in which resolution has not been increased but temporal noise is integrated over a number of frames. DSR 2 effectively results in a reduction of the detector pitch by a factor of 2 and a doubling of the pixel resolution. DSR 4 reduces the pitch by ¼ and results in a fourfold pixel resolution. In the TNO algorithm DSR is always preceded by SBNUC in order to avoid artefacts due to the fixed pattern noise.

Table 1: Normal and signal processing conditions used in this study. Some conditions were shown at different sizes on the display. See text for details.

Basic conditions	Condition number	Additional magnification factor	Total magnification
Still	1	1	1
Static	2	1, 2	1, 2
Moving	3	1, 2, 4	1, 2, 4
Moving + SBNUC	4	1	1
Moving + DSR1	5	1, 2	1, 2
Moving + DSR2	6	1,2	2, 4
Moving + DSR4	7	1	4

In addition, magnified imagery of the basic imagery were generated. Since DSR 2 and DSR 4 result in a higher resolution, they are shown 2 and 4 times larger on the display than the unprocessed imagery. Improvement ascribed to the DSR algorithm may be partly due to the magnification of the test patterns on the display. In order to be able to disentangle signal processing and size effects, 5 magnified versions of the basic conditions were added, resulting in a total of 12 conditions listed in Table 1.

4.3 IMAGE PRESENTATION

4.3.1 General

The experiments were carried out in a dimly lit room. Test patterns were shown on a 17 inch computer CRT set at low resolution (800 by 600 pixels) in order to be sure that pattern size on the display is not a primary limiting factor. Subjects were free to choose the optimum distance from the display (most of the time distance was approximately 50 cm).

4.3.2 TOD image selection and presentation

Prior to the experiment, sequences from two of the four ranges containing a total of 10 triangle sizes (see section 4.1.3) were selected for each condition and contrast. This results in a total of 5760 responses observer: 2 (ranges) * 5 (rows) * 4 (test patterns per row) * 4 (test plate orientations) * 3 (thermal contrasts) * 12 (conditions). The total number of sequences was 288.

The test procedure was as follows: after a key stroke, an image sequence (either with a stationary or rotating test plate containing the 5 rows with 4 test patterns each) was continuously present. The computer randomly selected the line the observer had to read (this was indicated to the observer by a number of beeps but also shown as a number on the display). In this way, learning effects are much lower than when reading the test plate from top to bottom. The observer responded using the arrow keys on the computer keyboard (triangle orientation Up, Down , Left or Right). After four responses, a next line was selected and after the test plate was completely read, a low beep sounded and a new image sequence started. There was no time limit to the experiment.

In order to equally divide learning effects over the different conditions, the sequences were divided into 4 blocks. Each block contained the same set of sequences in random order except that the test plate orientations are different (these were also randomly divided over the four blocks). The block presentation order was different for each observer according to a 4 by 4 Latin square design[17,18].

Four observers participated in the experiment. The total experiment took about 4 hours per observer.

4.3.3 MTDP image selection and presentation

As argued in section 3.2, the original MTDP measurement procedure has to be modified slightly in order to be able to determine a threshold for dynamic imaging and still pictures: the pattern needs to be presented at fixed contrasts. With a dynamic image, the observer waits until the pattern is at optimum phase and then judges whether it is above or below threshold.

In the present study we limit ourselves to determining the cut-off at a single, high contrast. This is the most critical situation. MTDP cut-off frequency may be reached in two different appearances: for well-sampled systems the modulation of the four bars of the test pattern becomes no longer visible. For undersampled systems, near cut-off the four bars reduce to two bars, and cut-off is reached whenever the modulation between the two bars is no longer visible in optimum phase (and thus in any phase).

Prior to the experiment, for all conditions six sequences ranging from a spatial frequency clearly below and clearly above the cut-off frequency were selected for presentation (a total of 120 sequences). For the still and stationary condition, optimum bar phase was selected. All remaining sequences up to the maximum spatial frequencies were studied in order to determine any unexpected phenomena that may be induced by the combination of the bar pattern and signal processing. No such phenomena were observed.

The test procedure was as follows: after a key stroke, a randomly selected image sequence (containing either a stationary test plate with the bar pattern in optimum phase or rotating test plate) was continuously presented to the observer. In a Yes/No task, the observer had to indicate whether a modulation was visible at any time (i.e. at optimum phase). There was no time limit to the experiment. After another key stroke, a new sequence was presented.

One observer participated in the bar pattern experiment. However, as will be shown in the next chapter, the results are very clear.

5. RESULTS

5.1 TOD RESULTS

5.1.1 Raw data analysis

In total, 144 thresholds were obtained from the measurements: 12 (conditions) * 3 (contrasts) * 4 (observers), resulting in 12 TOD curves with three points each. Each threshold estimate is a result of a best fit of a Weibull-shaped psychometric curve[13] through the fraction correct data from 16 triangle presentations at 10 different sizes. The fit yields a 75% correct triangle angular size and slope (including a standard error). In Figure 5, examples of the data and the curve fits are shown. Thresholds were very similar between observers. Averages and standard error in the mean over the observers were taken and used in the plots in this paper.
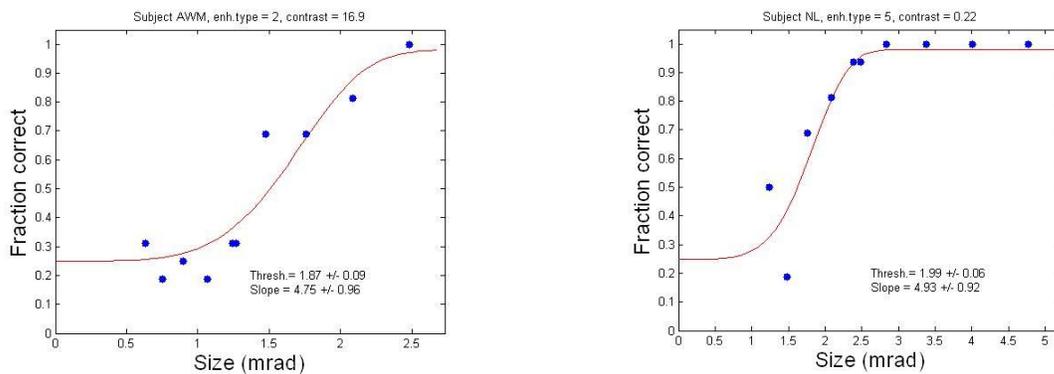

Figure 5: Examples of the observer response data and the best fits resulting in a 75% correct triangle angular size.

5.1.2 Effect of image magnification

In Figure 6, TOD acuity is plotted as a function of the magnification on the display (see Table 1) at three different contrasts. The curves are flat or almost flat (sometimes even decrease), showing that under the present conditions (e.g. a low resolution display) magnification has little or no effect on performance. This means that any performance increase with DSR can be ascribed solely to the effects of the signal processing algorithm.

In the rest of this study, we will only consider the 7 conditions with the standard magnification.
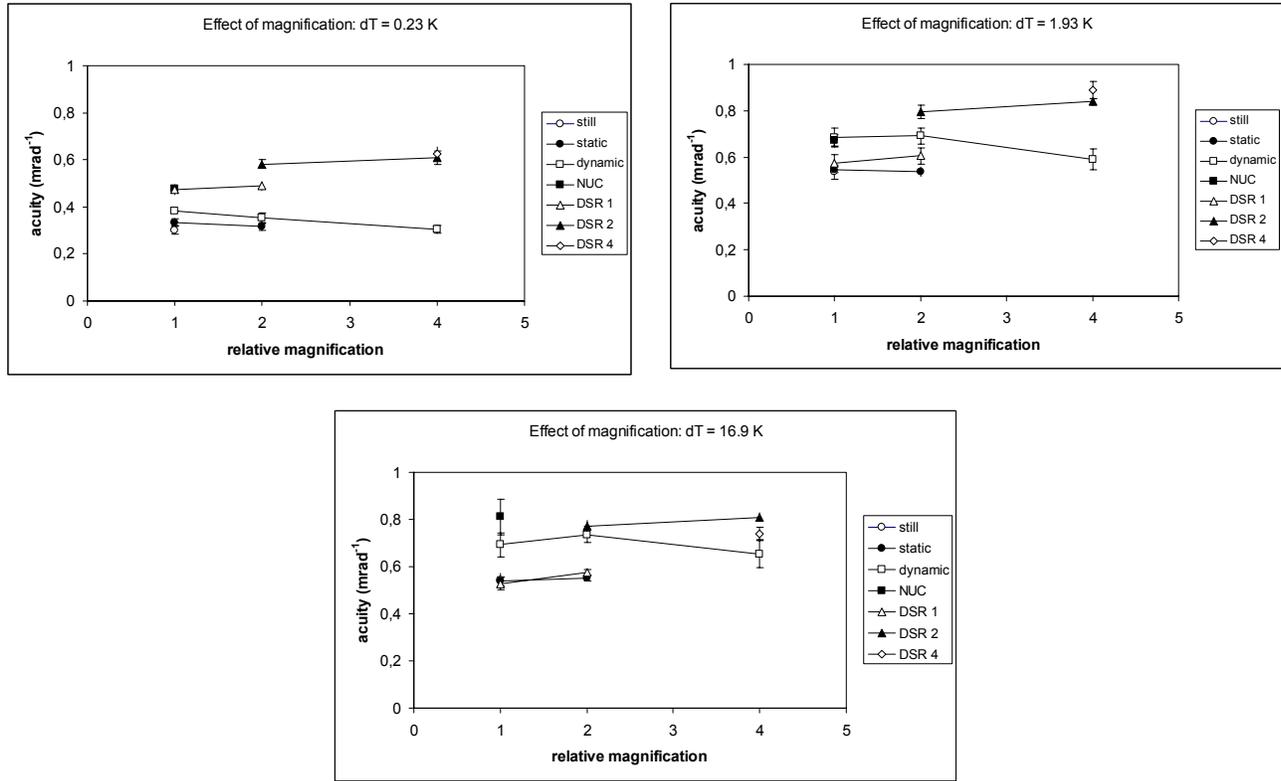
Figure 6: Effect of magnification on acuity. Top left: ΔT = 0.23 K, Top right: ΔT = 1.93 K, Bottom: ΔT =16.9 K. The curves are flat or almost flat, showing that magnification has little or no effect on performance in this study. This means that any performance increase with DSR can be ascribed solely to the effects of the signal processing algorithm.

5.1.3   Effect of image processing

In Figure 7, TOD acuity is plotted for all seven conditions mentioned in Table 1 at three different contrasts. The plot clearly shows effects of image processing on performance. From Figure 7 we see that:

- The results for ΔT = 1.93 and 16.9 K are very similar, except for the data with DSR 4. This indicates that maximum sensor acuity is already obtained at 2K: a further contrast increase does not lead to a higher performance.
- As expected at high contrasts, the differences between still and static are small: these conditions only differ in temporal noise which is negligible.
- A pronounced increase in acuity (+ 25%) is found when the target is moving.
- Application of SBNUC has no effect at high contrast, since fixed pattern noise is relatively small.
- The results with DSR 1 at high contrasts are comparable to those for the static conditions
- DSR 2 and DSR 4 give a significant improvement over static acuity and a reasonable improvement over dynamic acuity.
- The results for ΔT = 0.23 K show an even higher relative improvement than at higher contrasts. This is the result of the static noise reduction of the SBNUC and the dynamic noise reduction of DSR. There is also a small improvement from still to static due to the possibility to integrate over several frames.

In conclusion: at high contrasts performance is resolution limited. Acuity improvement (and thus range improvement) can be obtained by image motion (so that the observer is able to integrate over the different focal plane array positions) or by super resolution reconstruction. At low contrast, performance is also limited by spatial and temporal noise. Temporal integration by the human eye and spatial noise reduction by SBNUC improve performance.
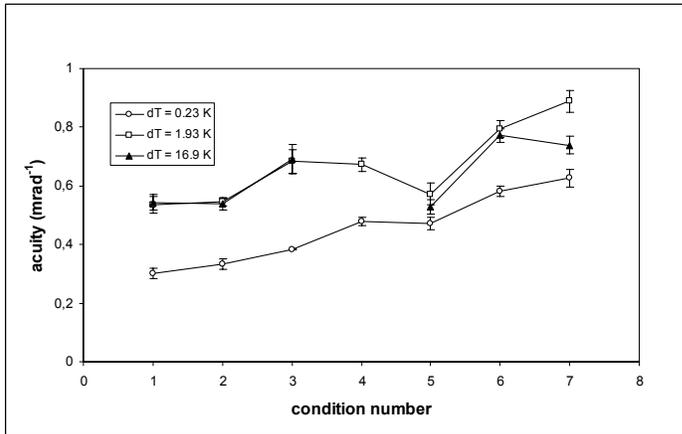
1: still

2: static

3: moving

4: moving + SBNUC

5: moving + DSR 1

6: moving + DSR 2

7: moving + DSR 4

Figure 7: Image enhancement techniques clearly show an effect on TOD acuity. See text for details.
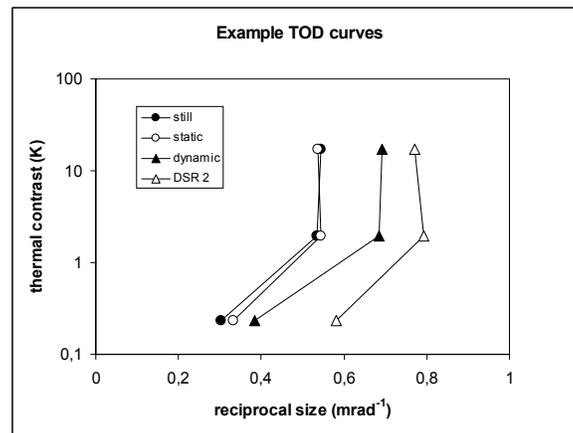

Figure 8. Example TOD curves for four conditions. Dynamic imaging gives a 25% acuity increase, and DSR 2 gives a 45% improvement. From these curve range performance can directly be calculated using ACQUIRE and the TOD criteria. See text for details.

In Figure 8, the same results are plotted as regular TOD curves (similar to MRT curves), with reciprocal triangle size on the abscissa and thermal contrast on the ordinate. Only four curves are plotted: still, static, dynamic and DSR 2.

5.2 MTDP RESULTS

The results for the MTDP were very constant: cut-off frequency was 1.7 times the original Nyquist frequency f_N for all seven basic conditions except for the DSR 4 condition which cuts off at $1.9 \cdot f_N$. Analysis of the imagery shows that the system is well-sampled for DSR 4, is in a transition area for DSR 2 (shows 3 or 4 bars at $1.7 \cdot f_N$ but phasing effects as well) and is undersampled for all other conditions.

In Figure 9a, the performance increase relative to the static condition is plotted for both MTDP and TOD (at 2 K). There is a pronounced difference in the effects of motion and signal processing on the two sensor performance measures. This means that they will make significantly different range predictions when applied to these techniques. See also Chapter 6.
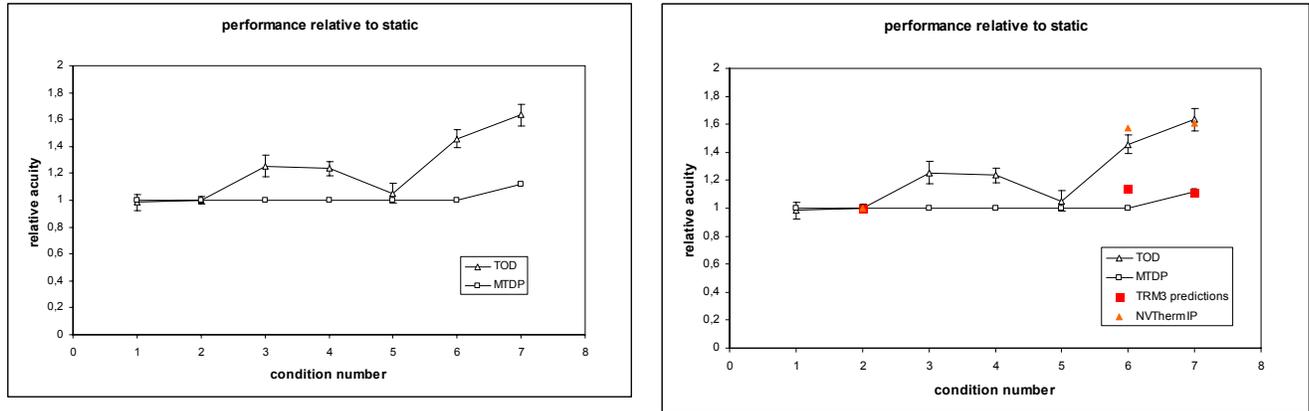
Figure 9. a: Performance increase relative to the static condition for the MTDP and TOD. MTDP hardly predicts any range improvement with motion or DSR. b: same plot, but with NVThermIP and TRM3 model predictions for DSR 2 (condition 6) and DSR 4 (condition 7). See text for details.

6. MODEL PREDICTIONS

Model predictions for the static case and for 2 by 2 and 4 by 4 super resolution algorithms were performed with NVThermIP and TRM3. The FLIR camera (section 4.1.1) for which the predictions were made, was assumed to have a fill factor of 1.0. With NVThermIP, the procedure as described by Jacobs and Driggers (2005) was used, resulting in range predictions for an optimal algorithm. TRM3 provides MTDP predictions. TRM3 is unable to directly predict the effects of super resolution reconstruction, but it contains a module that calculates the effects of 2 by 2 or 4 by 4 microscan which is a controlled type of super resolution reconstruction.

The results of the calculations were normalized to the static condition and plotted in Figure 9b. Despite of the limited number of observers in the experiment, the plot shows a very good agreement between the predicted and measured MTDP. In addition, the TOD data are excellently supported by the NVThermIP calculations, which differ significantly from the TRM3 predictions with super resolution reconstruction.

7. DISCUSSION AND CONCLUSIONS

We investigated the applicability of four static end-to-end sensor performance measures (MRT, TOD, MTDP and DMRT) to dynamic imaging and signal processing algorithms. The MRT and the DMRT are rejected on theoretical grounds. The MTDP needs a small adaptation before it can be used. The TOD can be applied without any problem or adaptation. The latter result was expected since the TOD has already been applied succesfully in an earlier field study[14]. The present study was more extensive and performed in the laboratory under controlled conditions.

The laboratory experiment shows important differences in results for the TOD and MTDP. Whereas the TOD shows a significant performance improvement due to motion and signal processing, the MTDP is rather insensitive to this. Given the experience with super resolution techniques, the results with the TOD are more plausible although an experimental validation with real targets needs to be performed.

The results obtained with the TOD agree very well with the NVThermIP model predictions for DSR. These model predictions are based on observer performance experiments with real targets and on an ideal super resolution algorithm (given the limitations due to sensor optics, detector size and noise). The agreement indicates that the improvement obtained with the TNO Dynamic Super Resolution algorithm is close to the maximum that may be expected.

The measured MTDP is predicted very well by TRM3. With this method, little or no improvement due to motion can be expected because the bar pattern is already in optimum phase for the static situation. Only when super-resolution results in a well-sampled system (DSR 4), improvement can be obtained.

The study shows that a reliable sensor test method is needed more than ever. First, analytical models can never cover all conditions or techniques that are being developed. For now, the dynamic unprocessed condition has not yet been modeled. Second, current models only predict the ideal situation. This means for example that the difference between two competing algorithms cannot be calculated, and that errors in algorithms may remain unnoticed. Another example is the DSR 1 condition for which we expected a performance improvement, but the data show that it completely nullified the acuity improvement induced by motion. This can be understood afterwards but might have been unnoticed for a long time without doing a test. Third, a test may be used to optimize or tune the parameters of the algorithm for specific sensor, preferably in combination with a vision model[18,19]. In our case, we may use the recordings from this study over and over again to test new algorithms.

The TOD is a good candidate for such a test. This is because the task is very close to real target acquisition, and yet very simple. The method does not suffer from observer bias as the MRT family does, and results show that the experimental errors are small (Figure 6). It is often assumed that TOD measurement is more time consuming than MRT measurement. In this study a total of 12 TOD curves (consisting of 3 points each) were collected in 16 hours. Normally, a TOD consists of 7 points, which takes about 3 hours. This is considerably less than an MRT takes.

REFERENCES

1. Young, S.S. & Driggers, R. G. (2005). Super-resolution image reconstruction from a sequence of aliased imagery. *SPIE Proceedings* 5784-18, 51-62.
2. Park, S.C., Park, M.K. & Kang, M.G. (2003). Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 21-36.
3. Schutte, K, de Lange, D.J., van den Broek, S.P., (2003). Signal conditioning algorithms for enhanced tactical sensor imagery, *SPIE Proceedings*, Vol. 5076, 92-100.
4. Ratches, J.A. (1976). Static Performance Model for Thermal Imaging Systems. *Optical Engineering*, 15, 6, 525-530, 1976.
5. Bijl, P.& Valeton, J.M. (1998). TOD, the alternative to MRTD and MRC. Optical Engineering 37, 7, 1976 - 1983.
6. Wittenstein, W. (1999). Minimum temperature difference perceived – a new approach to assess undersampled thermal imagers. Optical Engineering 38, 5, 773 – 781.
7. Webb, C.M. (1993). Dynamic MRTD for Staring Focal Plane Arrays. *Proceedings IRIS, Vol. II*, ERIM, Ann Arbor.
8. Jacobs, E. & Driggers, R. G. (2005). NVThermIP modeling of super-resolution algorithms. *SPIE Proceedings* 5784-19, 125-135.
9. Valeton, J.M., Bijl, P. & Varkevisser J. (1997). Waarnemingsprestaties en ergonomie van de Lion hand held *warmtebeeldkijker* [Observer performance and ergonomic evaluation of the Lion hand-held thermal viewer] (Report TNO-TM 1997 A071). Soesterberg, The Netherlands: TNO Human Factors Research Institute.
10. Bijl, P., Toet, A. & Valeton, J.M. (2003). Electro-optical imaging system performance measurement. In R.G. Driggers (Ed.), Encyclopedia of optical engineering (pp. 443-450). New York, USA: Marcel Dekker Inc.
11. Driggers, R.G., Vollmerhausen, R., Wittenstein, W., Bijl, P., Valeton, J.M. (2000). Infrared Imager Models for Undersampled Imaging Systems. *Proc. Fourth Joint International Military Sensing Symposium*, 45, 1, 335-346.
12. Bijl, P. (2001) *Validation of the TOD and MTDP Sensor Performance Measures for staring and scanning thermal imagers* (Report TNO-TM-01-A020). Soesterberg, The Netherlands: TNO Human Factors.
13. Bijl, P.& Valeton, J.M. (1999). Guidelines for accurate TOD measurement. *SPIE Proceedings, Vol. 3701* 14 - 25.
14. Driggers, R.G., Krapels, K., Murrill, S., Young, S.S., Thielke, M. & Schuler, J. (2004). Superresolution performance for undersampled imagers. *Optical Engineering 44, 1*.
15. Valeton, J.M., Bijl, P., Agterhuis, E. & Kriekaard, S. (2000). T-CAT, a new Thermal Camera Acuity Tester. *SPIE Proceedings Vol. 4030*, 232 – 238.
16. Bijl, P., Toet, A., & Valeton, J.M. (2003). Psychophysics and psychophysical measurement procedures. Encyclopedia of Optical Engineering, Volume, 2176-2187.
17. Wagenaar, W.A. Note on the construction of diagram-balanced Latin Squares. Psychol. Bull. 1969, 72, 384–386.
18. 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. *SPIE Proceedings, Vol. 3701*, 104-111.
19. Hogervorst, M.A., Bijl, P. & Valeton, J.M. (2001). Capturing the sampling effects: a TOD sensor performance model. *SPIE Proceedings Vol. 4372*, 62-73.