Detection metrics and ship [D]RI

Arthur D. van Rheenen, Lars T. Heen, Eirik B. Madsen, Erik Brendhagen, Kristin H. Løkken, et al.

Arthur D. van Rheenen, Lars T. Heen, Eirik B. Madsen, Erik Brendhagen, Kristin H. Løkken, Bernt Almklov, Eirik Glimsdal, "Detection metrics and ship [D]RI," Proc. SPIE 10625, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIX, 106250V (26 April 2018); doi: 10.1117/12.2304583

SPIE.

Event: SPIE Defense + Security, 2018, Orlando, Florida, United States

Detection metrics and ship [D]RI

Arthur D. van Rheenen, Lars T. Heen, Eirik B. Madsen, Erik Brendhagen, Kristin H. Løkken, Bernt Almklov, and Eirik Glimsdal

Norwegian Defence Research Establishment, P. O. Box 25, NO-2027 Kjeller, Norway

ABSTRACT

Well-known detection metrics based on Johnson criteria or Target Task Performance (TTP) models were developed for land-based targets [1,2]. In this paper we investigate how (whether) we can apply these metrics to especially recognition and identification of ships at sea. Large sea targets distinguish themselves from land-based targets by their large aspect ratio, when seen broad side, and their relatively large and hot plume. We shall only address the second of these two issues here. First, however, we shall investigate how the simple Johnson approach to recognition and identification stacks up against a TTP approach. The Johnson approach has clear and simple criteria to measure the target task performance. To apply the TTP model N50 (V50) values need to be found through observer trials. We avoid these trials here but estimate the criteria based on a comparison of the models. From analysis of LWIR and MWIR recordings of a multipurpose ship running outbound and inbound tracks, we find little difference between the two metrics. As mentioned, we study the effect of the plume on task performance ranges, by considering two different estimates for the target contrast: the average contrast and the root of the squares of this contrast and the standard deviation of the contrast. We argue that the plume skews the recognition and identification ranges to much too optimistic values when the standard deviation is included. In other words, although the plume helps to detect the target, it does not help the recognition or identification task. It seems a more careful definition of the temperature contrast needs to be applied when these models are used.

Keyword list: target detection, ship DRI, recognition range, identification range

1. INTRODUCTION

For detection of a target it is sufficient that one pixel or a small collection of pixels shows sufficient contrast against its background and the spatial resolution requirement is lax. The sensitivity of the sensor is crucial. For recognition and identification tasks spatial resolution of the sensor defines at what target range these tasks can be performed, provided some minimum contrast requirement is satisfied. The minimum resolvable temperature difference (MRTD) curve, an important sensor characteristic, divides the plane spanned by the apparent temperature contrast and the spatial resolution into a “see” part above the curve and a “can’t see’ part below the curve. Above the curve the contrast is sufficient to be able to resolve a detail with a given spatial resolution or put differently, for a given apparent contrast the curve indicates the highest possible frequency that can be resolved. Hence, the MRTD curve plays an important role when task performance ranges are estimated.

As an example, consider the case that in order to perform a certain target task with a given level of confidence it is required to have N cycles on target. If the target has a critical size S and is at range R, then the spatial frequency is $N/(S/R)$. Through the MRTD curve this specifies a minimum apparent temperature contrast between target and background. For a given atmosphere, which defines the IR transmission between target and sensor, and the considered range R the minimum inherent temperature contrast between target and background is defined. By rearranging the problem, for a given inherent target contrast, the maximum range can be calculated at which the target task can be performed satisfactorily.

Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIX, edited by Gerald C. Holst, Keith A. Krapels,
Proc. of SPIE Vol. 10625, 106250V · © 2018 SPIE · CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2304583

As to how many cycles on target are required to perform a target task, the work by Johnson was possibly the first in a series of developments. Without reciting the history, which is well documented in [1], we like to mention the original Johnson criteria. They are listed in Table I.

Task	Cycles on target
Detection	1
Orientation	1.4
Recognition	4.5
Identification	6.5

Table I. Johnson criteria, defining how many cycles on target are required to perform a certain target task.

When an image of a target is shown to an ensemble of observers, then in order for 50% of the observers to recognize the target, the target size in the picture should cover 4.5 cycles.

It was observed that using the Johnson criteria resulted in target task ranges that were too pessimistic [2]. For a target at range R, with an inherent target-background contrast ΔT_I and a sensor with a specified spectral response, the apparent temperature contrast ΔT_A can be calculated. The MRTD curve for the sensor then specifies a maximum spatial frequency $f_{X,max}$ that can be resolved by the sensor, see Fig. 1 [3].

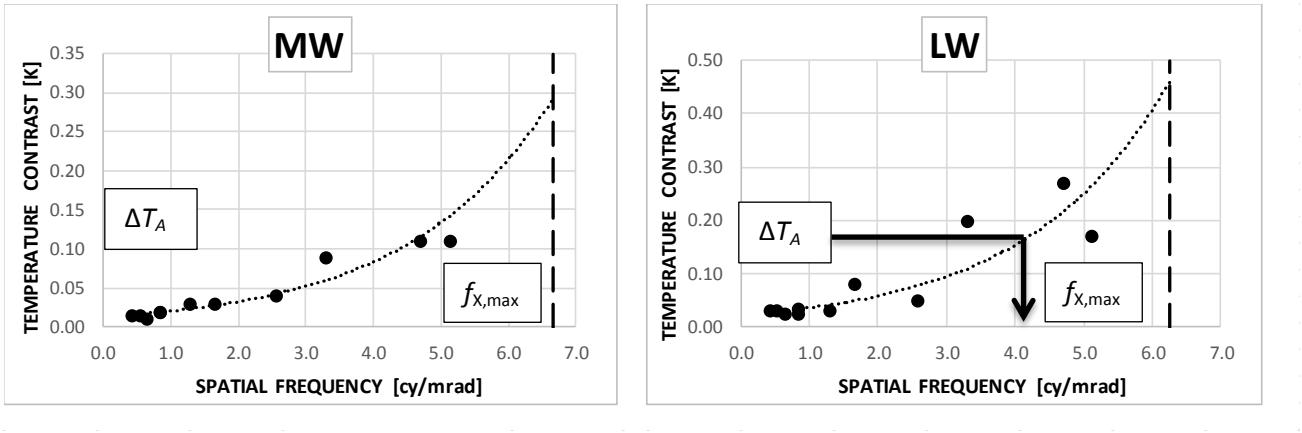

Figure 1. MRTD curve for our MWIR IrCam camera (left) and LWIR camera (right), measured with a 200-mm lens. The dotted line is an exponential fit to the data and the dashed line indicates the spatial cut-off frequency of the sensor.

When this method of estimating the maximum spatial resolution of the sensor yields a pessimistic (low) value, the estimate for the target task range will be pessimistic (low) as well. It was suggested that the excess of contrast available at lower spatial frequencies actually helps the human brain resolve higher spatial frequency details on the target. This effect was modeled as in Eq. 1 [2].

$$f_{X,crit} = \int_0^{f_{X,max}} \sqrt{\frac{\Delta T_A}{MRTD}}\, df_x \geq f_{X,max} \qquad (1)$$

In this work we apply these two metrics, estimates for $f_{X.max}$ and $f_{X,crit}$, to estimate recognition ranges for a research vessel, equipped with two black bodies, traveling on inbound and outbound tracks during the SQUIRREL trial in 2011, just east of Kiel in Germany. It is important to realize that since $f_{X,crit} \geq f_{X,max}$ its value could be larger than the cut-off frequency. This may not be a problem in itself, because it is not a physical quantity any more, but it suggests that physical insight may not be applied any more.

After a short description of the trial we present the procedure that was used to analyze the sequences of long-wave infrared (LWIR) and mid-wave infrared (MWIR) recordings we made of the runs and extract estimates for the apparent ship-background temperature contrast. Using our measured MRTD curves we find the number of cycles on target, based on the two metrics for highest resolvable spatial frequency. Finally, we judge how realistic the estimated recognition and identification ranges are by viewing the recordings at these ranges.

2. THE SQUIRREL TRIAL

A multinational measurement campaign (SQUIRREL) was performed at Eckernförder Bucht, near Kiel (Germany), in a corner of the Baltic Sea where Germany borders on Denmark, September 11 - September 23, 2011. Measurement teams from 13 nations, some with several teams, participated in the campaign. One goal of the campaign was to obtain infrared ship modeling validation measurements in cool conditions, while another was to perform infrared propagation measurements in the marine boundary layer under such conditions. Several measurement teams participated in each part of the trial. In addition to these IR experiments, measurements involving radar where performed at the same location, during the same period of time. The following sections give a brief summary of the measurement setup.

Measurements were performed from a German Naval Laboratory field site located at Surendorf, on the south shore of the Baltic Sea. For propagation measurements, two black-body sources were mounted on the German research ship Mittelgrund. Mittelgrund started its runs ca 3 km from the shore where the measurement teams were located, and sailed at a speed of 10 - 11 kts in a straight line away from the shore, solid line in Fig. 2 (left). After an hour the ship turned around, made sure the black bodies were radiating towards the shore, and sailed back towards shore. Recordings made during these runs are used in the analysis here.

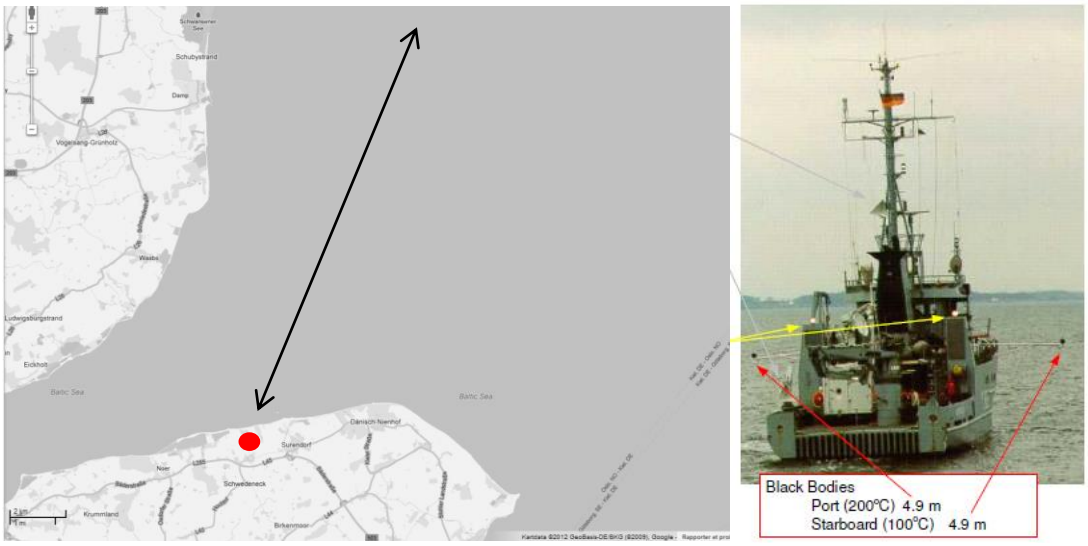

Figure 2. Map of the south shore of the Baltic Sea (left) with ship tracks. The ship track here is shown by the double arrow: a straight inbound/outbound track. The measurement location is indicated by the red dot. On the right the multipurpose vessel Mittelgrund with the two black bodies.

The Mittelgrund (length 38.5 m and beam 9.2 m) was equipped with two black bodies (100 ^{o}C and 200 ^{o}C) mounted on outboard riggers, one mounted on either side of the ship (Fig. 2), extending 4 m out and at a height of 4.9 m above the water.

3. ANALYSIS APPROACH

While on its outbound or inbound course recordings were made with two IrCam cameras in the two bands, LWIR and MWIR, using 100-mm lenses. The recordings consisted of 100 frames taken at a rate of 100 fps at roughly 1-km range intervals. Estimation of the apparent target-background contrast proceeded as follows. First, the images are averaged in time (over the 100 frames). Second a simple segmentation algorithm is applied: subtract the median vertical profiles from the image resulting in a contrast image, sort the pixels according to contrast, selecting the hottest P pixels (P is range dependent), and finally remove pixels that do not have at least two hot nearest neighbors. Figure 3 illustrates this procedure, showing intermediate results. The large dynamic range, mostly due to the hot exhaust, makes is hard to see the hull pixels, therefore the target mask is included in the bottom row. The mean contrast of the left over pixels is converted to an apparent temperature contrast by applying a calibration technique based on separate recordings of two black bodies at close range, about 15 cm.

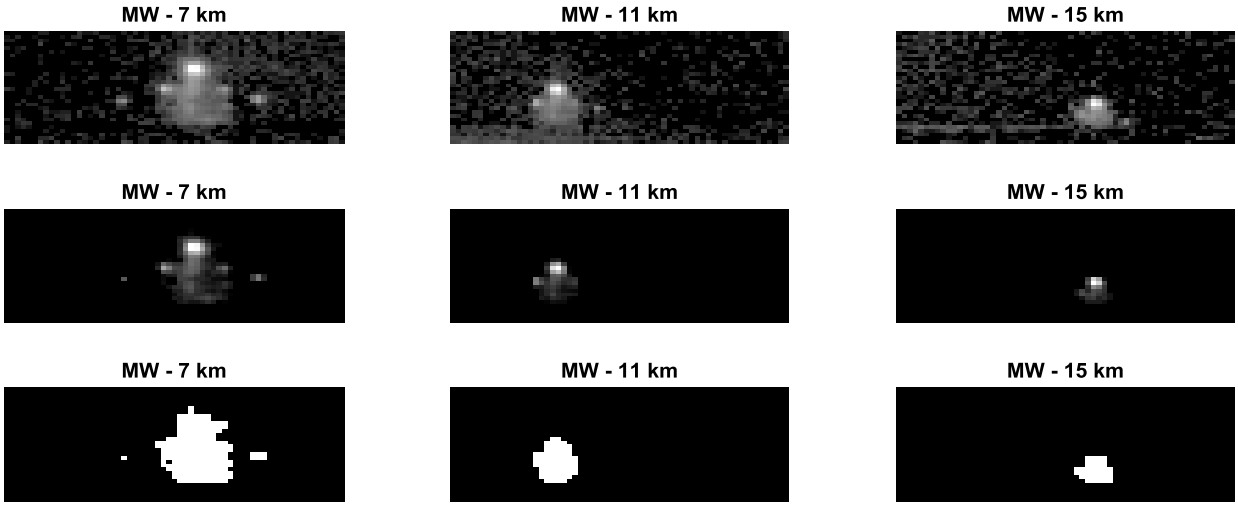

Figure 3. Illustration of the simple segmentation procedure: top row, average image of Mittelgrund at three different ranges after subtraction of median background profile; middle row: after the hottest N pixels are retained and lonely pixels are removed; bottom rom: target pixels. In this sequence the Mittelgrund is on an outbound course.

Given the apparent temperature contrast, the measured MRTD curves (Fig. 1) are used to calculate both $f_{X,max}$ and $f_{X,crit}$ (see section 1).

The dotted lines in Figs. 1 suggest exponential relationships,

$$\Delta T_A = 0.0125e^{0.47 f_X} \text{ for } f_X \leq 6.67 \text{ cycles/mrad (MWIR)}$$

$$\Delta T_A = 0.022e^{0.49 f_X} \text{ for } f_X \leq 6.25 \text{ cycles/mrad (LWIR),}$$

where f_X is the spatial frequency in cycles/mrad and ΔT_A is measured in K. The MWIR camera was used with a 100-mm lens, hence for this camera the spatial frequency axis of its MRTD curve needs to be scaled by a factor two. The factor in the exponent becomes 0.94 rather than 0.47 for this camera. While we don't have a model that describes the exponential relationship, the dashed line seems to capture the data, although also other expressions could have been used. We did not investigate this.

The square root of P, the number of target pixels, multiplied by the instantaneous field of view (IFOV) of the sensor-lens combination gives a measure of the geometric mean size of the target in mrad. The IFOVs are 0.15 and 0.08 mrad for the MWIR and LWIR camera, respectively.

The number of cycles on target is found by multiplying this size measure by the extracted maximum or critical spatial frequency:

$$N_1 = IFOV\ \sqrt{P}\ f_{X,max}$$

$$N_2 = IFOV\ \sqrt{P}\ f_{X,crit}$$

4. ANALYSIS RESULTS

In all, 14 runs were performed with the Mittelgrund, both during the night and during the day. In Figure 4 we present the calculated number of cycles on target for six of those runs recording during the early hours of 12 September 2011. Each of the data points represents a recording that was made during the runs. The outbound runs (odd run numbers) yield different results compared to the inbound runs (even run numbers). Most significantly the average temperature contrast for the ship seen from the front (bow aspect) is lower than that for the ship seen from the rear (stern aspect). This is not unexpected; seen from the rear the plume is better visible and possibly other heat sources are prominent. Similar plots were produced for the other eight runs. Especially in the LWIR recordings we observe consistency in the data between inbound runs and between outbound runs. For MWIR the spread for the outbound runs is larger.

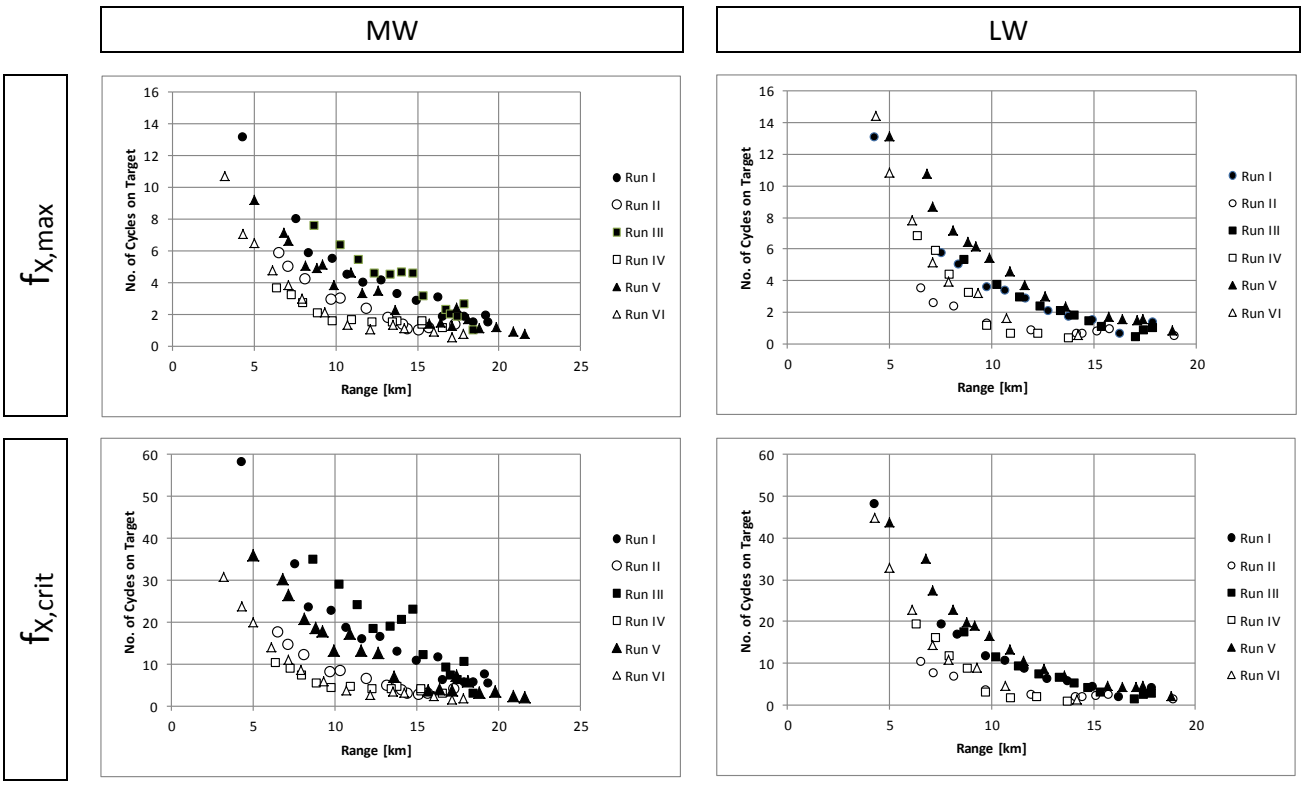

Figure 4. Number of cycles on target (Mittelgrund) for six runs made during the early hours of 12 September 2011: MWIR in the left column and LWIR in the right column. Top row, calculated using $f_{X,max}$ and bottom row calculated using $f_{X,crit}$. The odd run numbers are for outbound runs and the even run numbers for inbound runs.

The final step in the analysis is to estimate ranges at which target tasks can be performed. Where $f_{X.max}$ is used to estimate number cycles on target we apply the Johnson criteria: 4.5 cycles on target to recognize a target and 6.5 cycles on target to identify it. For each of the fourteen runs and for each of the IR bands the curves for the number of cycles on target

versus range (Fig. 4) were used to estimate for what range they crossed the intercept at 4.5 or 6.5 cycles. These ranges are plotted in Fig. 5 (top row) versus run number.

Where $f_{X.crit}$ is used to estimate the number of cycles on target it does not make sense to use the Johnson criteria. At this point in the TTP approach, observers are used to develop a probability curve: a curve of the probability to successfully perform a target task versus the number of cycles on target. The 50%-probability point defines N50, equivalent to the Johnson criteria. Here, we avoid using observer trials. Rather we try to estimate what those criteria are. For the apparent contrasts measured in our recordings, 0.2 – 1.0 K, $f_{X,crit}$ is about three times larger than $f_{X.max}$ for the MW camera and about 2.5 times larger for the LW camera. This can be shown analytically by considering the MRTD curves for this contrast range and calculate both $f_{X.max}$ and $f_{X,crit}$ for these values. It seems therefore reasonable to multiply the Johnson criteria by respectively, 3 and 2.5 for the MWIR and LWIR recordings. This may seem like a rough approach to use a global multiplication factor, but the results are not too bad. The estimated target task ranges are shown in the bottom row of Fig. 5.

Considering Fig. 5 we notice that there is not a big difference between using $f_{X.max}$ and $f_{X,crit}$ when the recognition and identification ranges are calculated, both in MWIR and LWIR, although for LWIR there may be somewhat larger difference, but there is also a larger spread in the data. Whereas the Jonson criteria are clear and predefined, similar criteria for the TTP-like analysis are more diffuse, depending on both the actual temperature contrast and actual number of target pixels, resulting in a factor 2.5 – 3 times the Johnson criteria. In this sense there is not much to gain in using $f_{X,crit}$ rather than $f_{X.max}$.

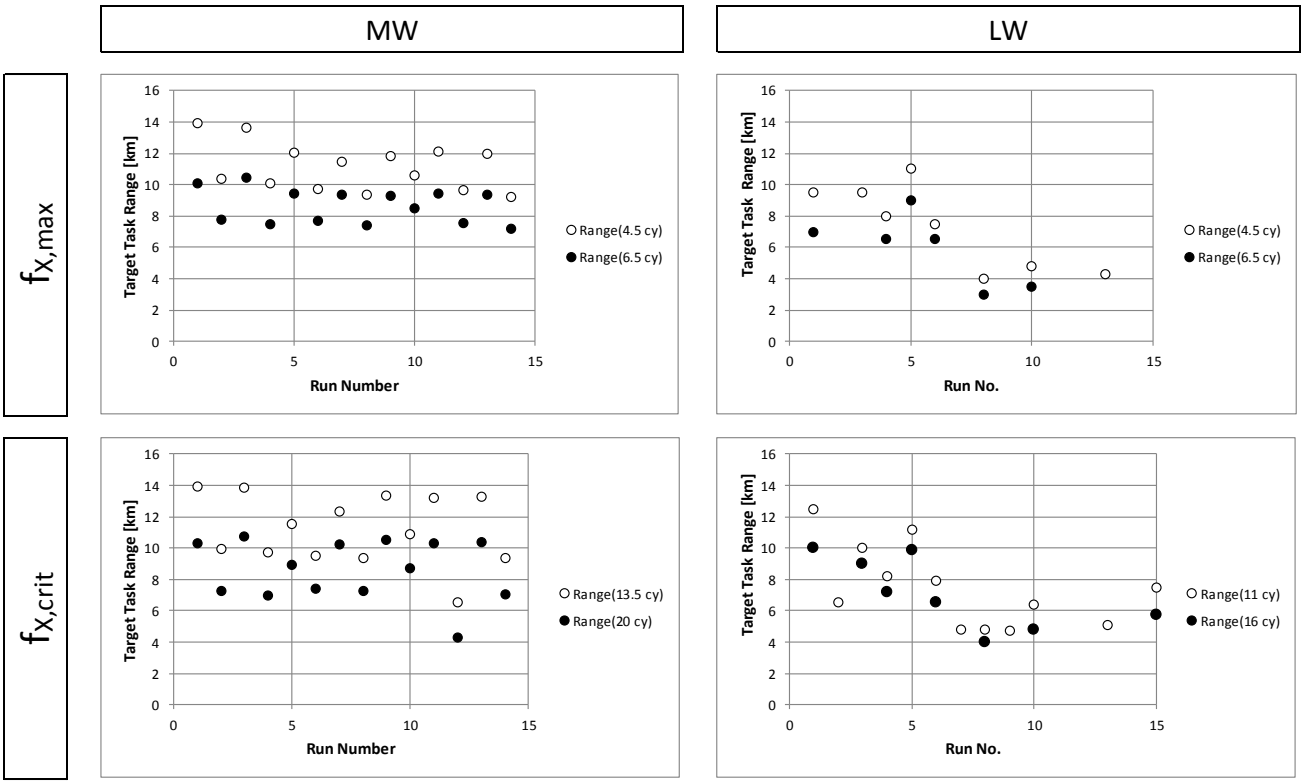

Figure 5. Recognition (open circles) and identification (filled circle) ranges (MWIR left column, LWIR right column) using the two estimates for maximum spatial resolution: top row - $f_{X,max}$, bottom row- $f_{X,crit}$.

Especially for MWIR relatively robust estimates for the recognition and identification ranges are obtained, about 11 km and 9 km, respectively, when averaged over both inbound and outbound runs. Note that for outbound runs (odd run

numbers) the ranges are consistently 2 km longer. For LWIR, the data spread is too large to make such a claim. We observe there are fewer LWIR than MWIR data points; this is due to the relatively poor LWIR transmission conditions.

Are these recognition and identification ranges realistic? To get an impression we show MWIR images (Fig. 6) of the outbound Mittelgrund at ranges 7, 10, and 13 km, the ranges that we found for the recognition and identification tasks. We realize that this is speculative, but the image taken at 7 km is probably good enough to identify the ship and the image at 13 km is probably good enough to distinguish it from other ships. We observe that the estimated recognition and identification ranges are realistic.

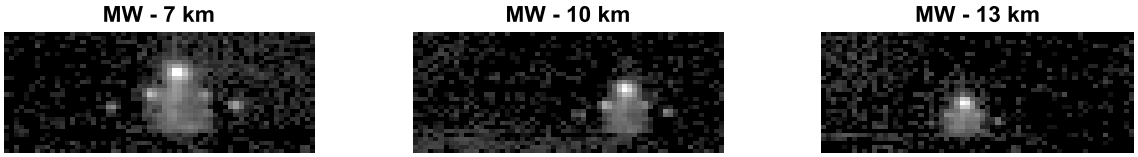

Figure 6. Mittelgrund MWIR recordings made at three different ranges: 7, 10, and 13 km.

Similarly we show LWIR images (Fig. 7) of the outbound Mittelgrund at ranges 2.4, 4.2, 7.5, and 8.3 km. The images taken at the two shortest ranges are probably good enough to identify the ship but the images taken at 7.5 km and longer ranges cannot be used for anything else but detection of the ship. Images presented here are from recordings made during the first run, where recognition and identification ranges are 7 km or better. Here, we observe that the recognition and identification ranges are overestimated.

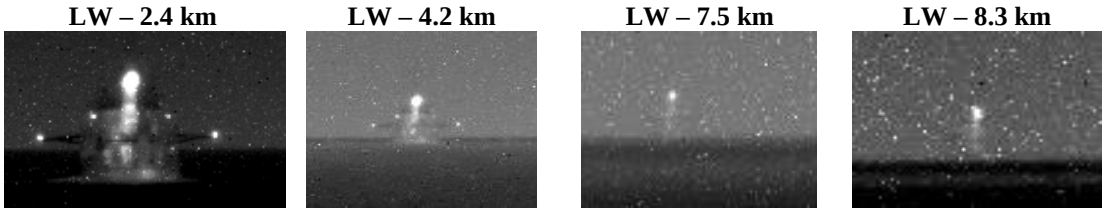

Figure 7. Mittelgrund LWIR recordings made at five different ranges: 2.4, 4.2, 7.5, and 8.3 km

5. DEFINITION OF CONTRAST

In this work we have defined the temperature contrast as the difference between the average target temperature and the average background temperature. In [2] it is suggested that variation of the contrast over the target may help to perform the target tasks and the following definition of temperature contrast is proposed:

$$\Delta T = \sqrt{\left(T_{tgt} - T_{bkg}\right)^2 + \sigma_{tgt}^2}$$

Here, T_{tgt} is the average target temperature, T_{bkg} the average background temperature and σ_{tgt} the standard variation of the target temperature. Clearly, including the standard deviation over the target in the calculation of the target contrast yields a larger contrast value, which in turn yields a larger maximum spatial frequency and a larger number of cycles on target (N_3 to distinguish it from the earlier defined N_1 and N_2), suggesting an easier target task. It would also require an adjustment of the cycle criteria for the task performance, similar to the adjustment (a factor 2.5 – 3) multiplying the Johnson criteria when $f_{X,crit}$ were used rather than $f_{X,max}$.

In Fig. 8 we plotted the fractional increase ($N_3/N_2 - 1$) of the number of cycles on target when the standard deviation of the contrast is included, for the first six runs. There is a clear difference between the outbound runs (odd run numbers, solid symbols) and the inbound runs (even run numbers, open symbols). On the outbound runs the plume is clearly visible, whereas it is hidden behind a mast on the inbound runs.

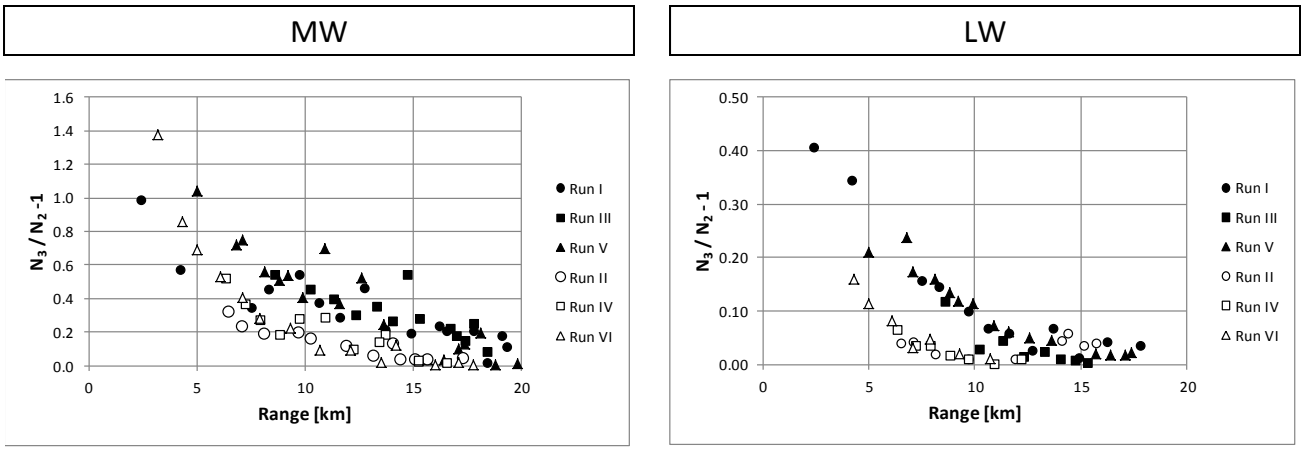

Figure 8. Fractional increase of the number of cycles on target when the standard deviation of the contrast is included in the contrast definition. The increase is plotted versus the range for both outbound (solid symbols) and inbound runs (open symbols).

The high temperature of the exhaust plume results in a larger standard deviation of the contrast for the outbound runs, both in MWIR and LWIR and as a result in a larger increase in the number of cycles on target. The larger number of cycles on target should simplify the target task. But does it? The increased contrast is due to the plume while the contrast of the rest of the ship, superstructure and hull, has not changed. Hence, the increased contrast does not help recognize or identify the ship at longer ranges. Although the hot plume helps to detect the ship at longer ranges, it does not help in the recognition or identification tasks.

6. CONCLUSIONS

A straight forward application of Johnson's criteria seems to yield reasonable results for estimates of the recognition and identification ranges for the MWIR imagery considered here. For the LWIR imagery this resulted in overestimation of these ranges.

Replacing the maximum resolvable spatial frequency ($f_{X,max}$) by an estimate ($f_{X,crit}$) that better accounts for how the brain (human observer) exploits both spatial information and contrast does not necessarily give better results. The new spatial frequency estimate requires an adjustment of the criteria for declaring recognition and identification. A simple analysis of the measured MRTD curves gives a means to estimate these new criteria: roughly 2.5 – 3 times the original Johnson criteria. Using these criteria yields the same estimates for recognition and identification ranges.

A refinement of the definition of target contrast by not only considering the average difference between target and background but including the variation of this contrast over the target will probably result in an overestimation of the target task ranges when the plume of the ship is part of its signature. The hot plume contributes considerably to the temperature contrast (large variation over target, in addition to the large value itself), but this increased contrast does not help resolve finer details.

Based on the example presented here, a series of recordings of an outbound and inbound multipurpose ship, it seems that the "old" Johnson criteria may be applied to estimate recognition and identification ranges of ships. A big advantage is that the criteria are fixed quantities and the maximum spatial frequency is a physical quantity. The TTP approach needs

N50 criteria that need to be established through extensive testing, requiring a new test for a new target task. We were not able to demonstrate better results using the latter approach.

7. REFERENCES

[1] Sjaardema, T. A., C. S. Smith, and G. C. Birch, “History and evolution of the Johnson criteria”, Sandia Report SAND2015-6368

[2] Vollmerhausen, R. H. and E. Jacobs, “The targeting task performance (TTP) metric – A model for predicting target acquisition performance”, Technical report AMSEL-NV-TR-230

[3] van Rheenen, A. D., P. Taule, J. B. Thomassen, and E. Blix Madsen, “MRTD- man versus machine, SPIE –DCS, this conference, paper number 10625-23