MRTD: man versus machine

Arthur D. van Rheenen, Petter Taule, Jan Brede Thomassen, Eirik Blix Madsen

Arthur D. van Rheenen, Petter Taule, Jan Brede Thomassen, Eirik Blix Madsen, "MRTD: man versus machine," Proc. SPIE 10625, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIX, 106250N (26 April 2018); doi: 10.1117/12.2304581

SPIE.

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

MRTD – Man vs. Machine

Arthur D. van Rheenen, Petter Taule*, Jan Brede Thomassen, and Eirik Blix Madsen

Norwegian Defence Research Establishment, P.O. Box 25, N-2027 Kjeller, Norway

*Norwegian University for Science and Technology, N-7491, Trondheim, Norway

ABSTRACT

We present Minimum-Resolvable Temperature Difference (MRTD) curves obtained by letting an ensemble of observers judge how many of the six four-bar patterns they can "see" in a set of images taken with different bar-to-background contrasts. The same images are analyzed using elemental signal analysis algorithms and machine-analysis based MRTD curves are obtained. We show that by adjusting the minimum required signal-to-noise ratio the machine-based MRTDs are very similar to the ones obtained with the help of the human observers.

Keywords: MRTD, observer trial, four-bar pattern

1. INTRODUCTION

The Minimum Resolvable Temperature Difference (MRTD) is an important infrared sensor characteristic that is obtained by letting an ensemble of observers view images of patterns with different spatial frequency and determining what the minimum temperature contrast is for which the pattern is just visible. The obtained MRTD curve divides the plane spanned by the spatial frequency axis and the temperature contrast axis into two: the area above the curve where the contrast is sufficient to observe details with a certain spatial frequency and the area under the curve where the contrast is insufficient and such detail cannot be discerned.

Apart from being a characteristic that can be used to determine which sensor is better suited for a certain task, the MRTD curve is a link in the process of estimating, for example, recognition ranges once it is established what level of detail (spatial frequency) is required for the recognition task. For a given spatial frequency, the MRTD curve specifies the minimum required apparent temperature contrast. For a target with given contrast to its background, and a specified atmosphere between target and sensor, the maximum range may be calculated.

For autonomous systems with an image-forming sensor MRTD curves obtained with a man-in-the-loop are not very useful. When object recognition is performed automatically, it is necessary for estimating recognition ranges that they are based on a characteristic that is measured without a man in the loop. In this work we address this issue by comparing MRTD curves that are obtained with the help of an ensemble of observers with MRTD curves that are obtained by applying simple signal analysis algorithms to analyze image sequences of standard four-bar pattern with different spatial frequencies and at different temperature contrasts.

Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIX, edited by Gerald C. Holst, Keith A. Krapels,
Proc. of SPIE Vol. 10625, 106250N · © 2018 SPIE · CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2304581

2. EXPERIMENTS

During the Minotauros trial in September 2016 on Crete, organized under the auspices of NATO SET-211, the U. S. Naval Research Laboratory offered the opportunity to measure the MRTD curves of IR sensors. Our group gratefully accepted the opportunity and our two IrCam cameras, LWIR and MWIR, were used to make recordings of a plate with six four-slit patterns with varying spatial frequency, mounted in front of a black body. A controller maintained a constant temperature difference between plate and black body. The temperature difference (contrast) was stepped from 0 to 0.3 K in small steps. During this process one of us was used as a human observer declaring how many of the six patterns he was able to see (all four bars visible) for each of the temperature contrasts. This procedure gave us a first indication of what the MRTD curves might look like. Simultaneously, recordings were made (100 frames at 100 fps) of the plate for each contrast for later analysis.

Figure 1. Set-up for MRTD measurement: left, black body with plate with slit patterns; middle, source and sensors; right, picture of PC screen during recording

During these experiments 200-mm lenses were mounted on the cameras resulting in instantaneous fields of view (IFOV) of 80 µrad for the LWIR camera and 75 µrad for the MWIR camera. Recordings were made at two ranges, 8.35 m and 13.00 m, to get two sets of six spatial frequencies for the same plate, thereby populating the spatial frequency axis with more data points.

At the shortest range recordings were made, both with the LWIR and the MWIR sensor, for 30 different temperature contrasts ranging from -0.2 to +0.27 K. At the longest range recordings were made for 18 different temperature contrasts ranging from -0.1 to +0.2 K.

The slit widths, equal to the spacing between slits, were 10.16, 7.925, 5.08, 2.54, 1.27, and 0.889 mm. The slit lengths were seven times their widths.

It may be a disadvantage to have such large ranges (due to the long focal length of the lenses) if the main purpose is to obtain as accurate an MRTD curve as possible, but for our analysis where human observer performance is compared with machine-based analysis this is not an important issue.

3. HUMAN OBSERVER TRIALS

The recordings are naturally divided into four sets: LWIR and MWIR recordings made at two ranges. For the human observer trials these sets are treated separately. Since the human-observer trials are time consuming we limit them here to the evaluation of the two LWIR sets only and plan to perform further analysis on the MWIR sets in the near future. Each recording was cropped to include all six patterns in the frame, leaving a margin of about the width of the largest slit

around the pattern, resulting in a square frame. Next the median intensity in camera units (LSB) over the 100 frames was calculated for each pixel in the frame. No further processing was applied to the frame and the data was stored as a matrix. In the next step 16 representations of the same frame were generated by rotating and mirroring the original image resulting in 8 versions, and by inverting the contrast of these 8 versions (new pixel intensity = 2 x average intensity of frame – original pixel intensity). The purpose of this step is to produce both horizontal and vertical stripe patterns, switching the location of the patterns in the frame, and produce both positive and negative contrast. From image to image the observer will see a new "order" of the patterns, forcing the observer to search in different locations in the image. The images are stored as matrices (floating point), recalled and presented to the observers through the Matlab® function `imagesc()` without any limits on dynamic range.

For the set recorded at the largest range this procedure resulted in an extended set of 18 x 16 = 288 images with only 18 different temperature contrasts. From this set of 288 images a subset of 100 images was randomly selected to represent recordings made at this range.

A group of observers, colleagues who are not especially trained as observers, was asked to individually view the subset (same subset for all observers) of images and record how many of the patterns they could resolve in each image. Twenty observers were used for the LWIR set recorded at the largest range and ten were used to evaluate the set recorded at the shortest range. The only instruction provided was that in order to "claim" that they see a pattern they should be able to resolve all four bars of the pattern, not just three or two. Their response was limited to seven choices: a number in the range from 0 and 6. There was no time constraint on the test. Typically, an observer used 10 – 15 minutes to perform the task, although there is a large variation in the amount of time used.

The trial took place in an office with the outside blinds closed and the curtains closed, however, some daylight seeped in. The single source of light was desk lamp in order for the observers to see the keys on the keyboard for their responses. The size of the image on the ThinkVision screen was 20 x 20 cm.

The analysis proceeds as follows: for each observer, their responses were binned in the seven categories (0 – 6) and for each bin the lowest temperature contrast is recorded. Each image has an associated temperature contrast, this information is retained when the random sequence is generated, and the responses are coupled to the image. Hence, for each observer a "curve" is produced with a minimum temperature contrast (difference) for each bin. Each bin corresponds to a spatial frequency. If an observer records a "4" for a certain image, implying that she/he can resolve the largest 4 patterns, this means that she/he can resolve the spatial frequency corresponding to pattern 4 but not the one corresponding to pattern 5.

Finally the minimum observed temperature difference for each bin or spatial frequency is averaged over all observers to provide an ensemble averaged MRTD curve. An estimate of the uncertainty in each temperature difference value is obtained by calculating the standard deviation.

4. MACHINE ANALYSIS

After calculating the median intensity over the 100 frames for each pixel the resulting image was parsed into six sub-images, one for each of the six four-bar patterns. The size of each of these sub-images was chosen such that the width was one bar width larger than the pattern width, half a bar width on either side, and the height was chosen to be four pixels less than the pattern height, cutting two pixels from the top and the bottom each. Figure 2 shows an example of what was analyzed for each four-bar pattern.

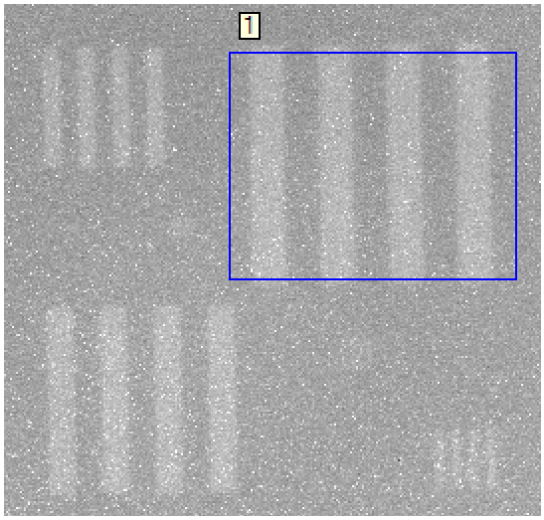

Figure 2. One frame from a recording (cropped) indicating what part of each pattern (largest only, here) is used for machine analysis.

Next, the average for each column in the sub-image was calculated as well as the standard deviation. The averaged column intensity forms a modulation signal and the standard deviation an estimate for the noise level: typically 5 LSB for the recordings with the sensor at 8.35 m from the object and 2 LSB for the recordings at 13.0 m. The modulation signal is shifted so that the average is zero. Next the part of the signal that is larger than zero is averaged, as is the part that is less than zero, and the second average is subtracted from the first to obtain an estimate for the modulation signal. Hence, for each recording (each temperature contrast ΔT) the modulation in LSB is calculated for each of the patterns. Obviously, when the contrast decreases the modulation signal will drown in the noise.

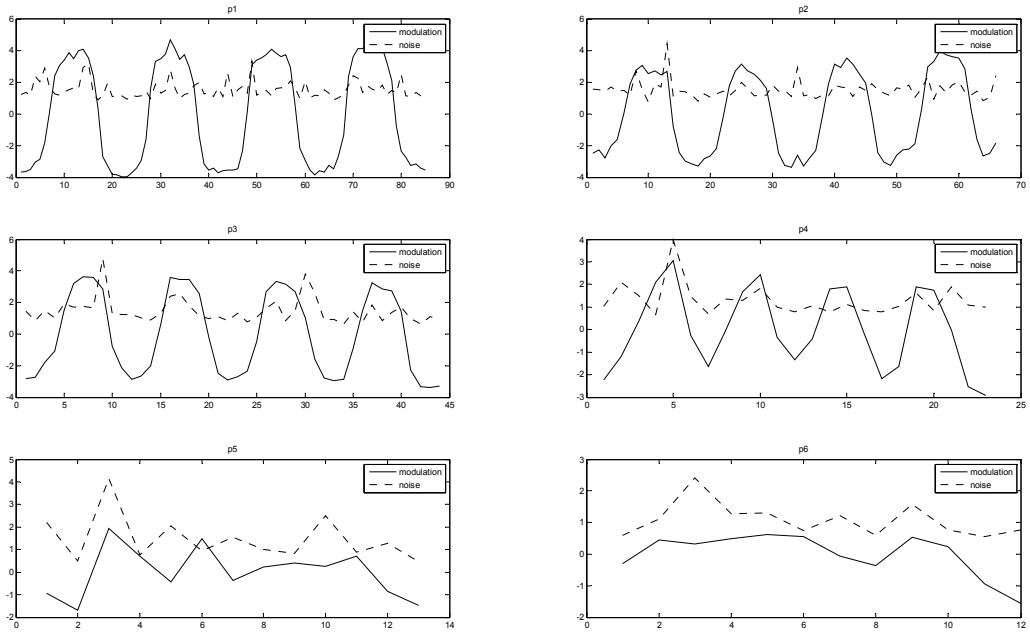

Figure 3. Modulation signal (solid line) and noise (dotted line) both in LSB as a function column number for each pattern. For this recording the smallest two patterns do not yield a clear modulation signal.

The modulation signal is plotted as a function of temperature contrast, such as in Figure 4.

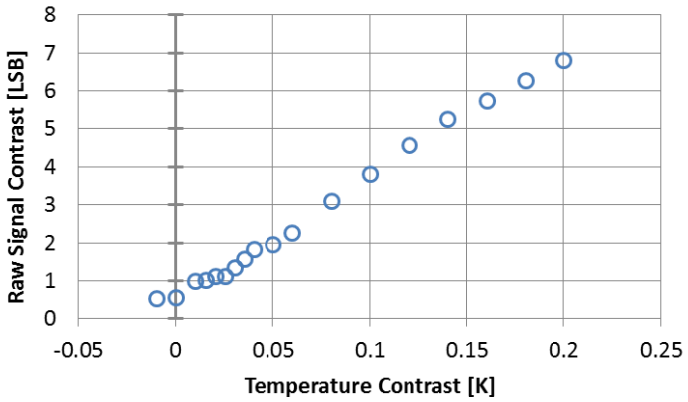

Figure 4. Example of a plot of strength of the modulation signal as a function of temperature contrast, here for the largest pattern in recordings made at range 13.0 m.

Figure 4 shows a steady decrease of the modulation signal with decreasing temperature contrast until the noise floor (0.5 LSB) is reached. The transition point between these two behaviors may be used to estimate the lowest observable temperature contrast for this pattern (spatial frequency). Note that the observed noise floor in Fig. 4 is notably smaller than the noisiness (2 LSB) suggested in Fig. 3. Extracting the modulation signal is apparently surprisingly robust.

In addition to calculating the modulation of the signal, we have Fourier transformed the modulation signal to calculate the amplitude spectrum. Rather than looking at the amplitude of the signal we look at the frequency content here. By choosing the width W of the sub-image one bar width wider than the four-bar pattern, the image is precisely 4 wavelengths wide, one wavelength consisting of one light and one dark stripe width. The Fourier transform will then have a peak at the fifth component: $4\Delta f = 4/W = 1/\lambda$, where λ is the wavelength of the pattern measured in number of pixels. This allows for easy selection of the interesting Fourier coefficient. This coefficient is then calculated for each of the six patterns in the recording, i.e. for each temperature contrast. The calculated spectra from one recording are shown in Fig. 5.

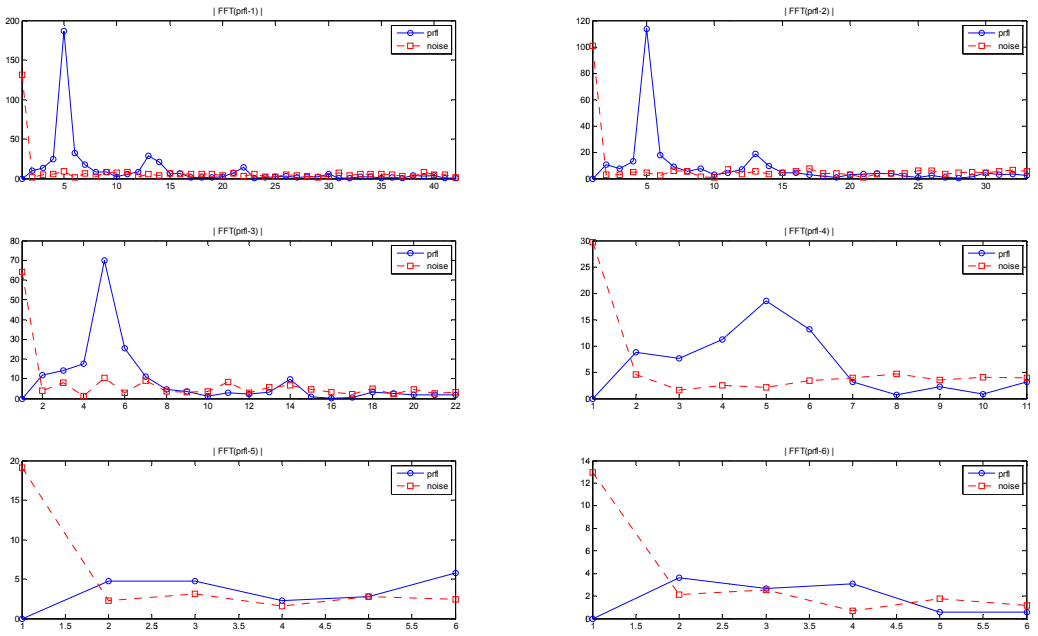

Figure 5. Fourier components of both the modulation signal (solid line) and the noise (dotted line) for the six patterns. The fifth component is strongest for the largest four patterns.

We observe clear peaks at the fifth Fourier component for the largest four patterns, as predicted. Clear to see for the largest two patterns are also peaks at the 13th Fourier component, this corresponds to half-wavelength periodicity. Also visible for the largest pattern is a peak at the 22nd component, which is half the maximum frequency – this corresponds to a two-pixel wavelength. Apparently there is some cross talk between neighboring pixels or columns. The value of the fifth component is plotted versus temperature contrast for each pattern. An example is shown in Figure 6.

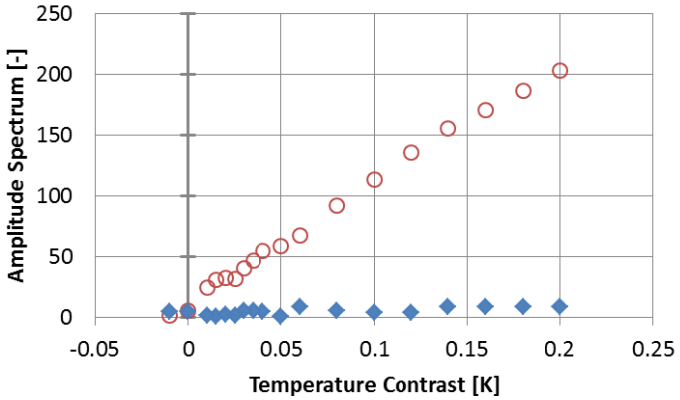

Figure 6. Example of a plot of the fifth Fourier component versus temperature contrast, here for the largest pattern in recordings made at range 13.0 m. The circles indicate the signal and the solid symbols indicate the noise.

The minimum required contrast may be found where the steadily declining spectral component approaches the noise floor.

5. DISCUSSION OF RESULTS

5.1 Observer test

An example of the response from a single observer is shown in Fig. 7. For each of the 100 images presented to the observer he/she has to respond with a number ranging from 0 to 6, the number of patterns that he/she can resolve (horizontal axis). Each image has a known temperature contrast (vertical axis). For each category the lowest temperature contrast is selected since we are interested in the minimum resolvable temperature difference, resulting in a sequence of seven minimum temperature contrasts, one for each category.

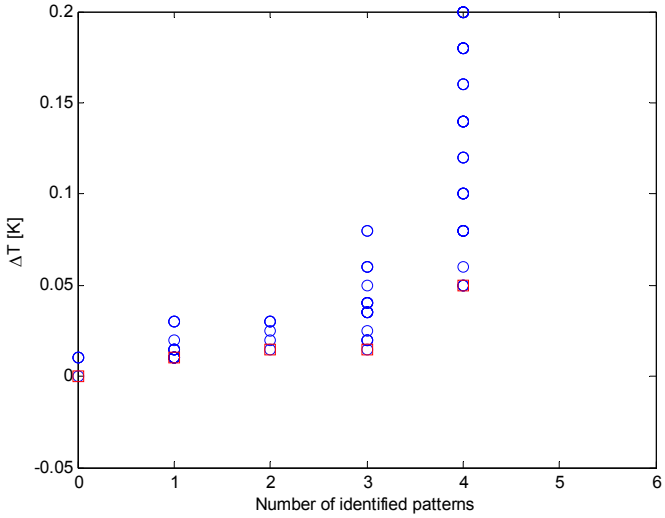

Figure 7 Responses from one observer, circles are responses for all 100 images and the squares represent the lowest contrast value for each category.

Results such as presented in Fig. 7 are averaged over the ensemble of observers and this then results in the MRTD curve displayed in Figure 8 for recordings made at range equal to 13.00 m. In addition to the average also the standard deviation is calculated, indicated by the error bars in Fig. 8. Instead of taking the lowest temperature contrast for each category we considered averaging the lowest two temperature contrast values, assuming this would yield a more robust result. However, neither the average nor the standard deviation changed significantly. The results are listed in Table I.

The general trend of the curve is as expected: to identify smaller and smaller details (higher spatial frequency), better contrast is required. Some observers reported to be able to identify the smallest pattern (highest spatial frequency). However, this frequency is higher than the cut-off frequency, which is 6.2 cycles/mrad for this camera with a detector pitch of 16 μm and a 200-mm lens. The data point for the highest frequency is therefore deemed unreliable, possibly caused by “wishful thinking” rather than true observation.

When the average of the lowest two temperature contrasts is taken as the minimum observable contrast, by definition this has to be at least as large as the minimum value itself. In Table I we observe that the MRTD values calculated in the former way have generally a slightly higher value, whereas the uncertainty not necessarily increases. This approach could be classified as slightly more robust, albeit yielding a more conservative MRTD curve.

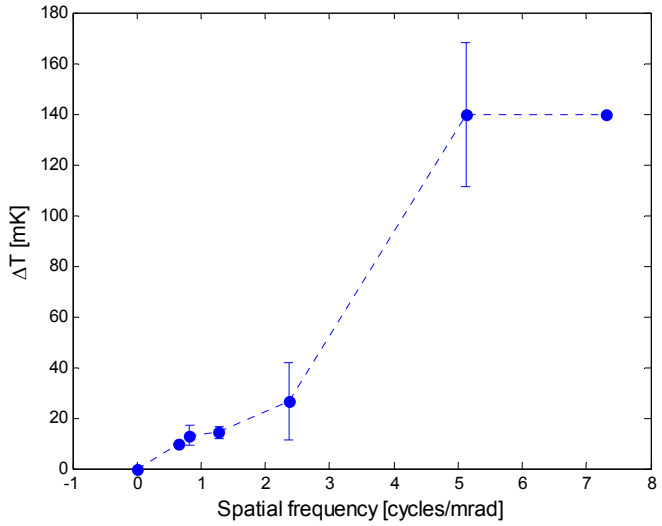

Figure 8. MRTD based on recordings made at 13.00 m range

Spatial frequency [cycles/mrad]	MRTD [mK], based on lowest contrast value	MRTD [mK], based on mean of lowest two contrast values
0	0	3 ± 2
0.64	10 ± 0	10 ± 0
0.82	13 ± 4	14 ± 5
1.28	14 ± 2	15 ± 2
2.36	27 ± 15	32 ± 12
5.12	140 ± 28	146 ± 20
7.31	140 ± 0	140 ± 0

Table I. MRTD values calculated by using two estimates for the minimum contrast: the lowest value and the mean of the lowest two values. Shown are average values, averaged over the observer ensemble, and their standard deviations. Range is 13.00 m.

Analysis of observer responses to a set of images recorded at range 8.35 m yields the results plotted in Fig. 9. Regarding the results for this set of images the same comments can be made as for the previous set. For this case the spatial frequencies for all patterns are smaller than the cut-off frequency and therefore all patterns should be identifiable provided that the contrast is high enough. We observe that the curve in Fig. 9 flattens out for the highest spatial frequency, where one would expect a stronger increase. Currently we do not have an explanation for this possibly unexpected behavior.

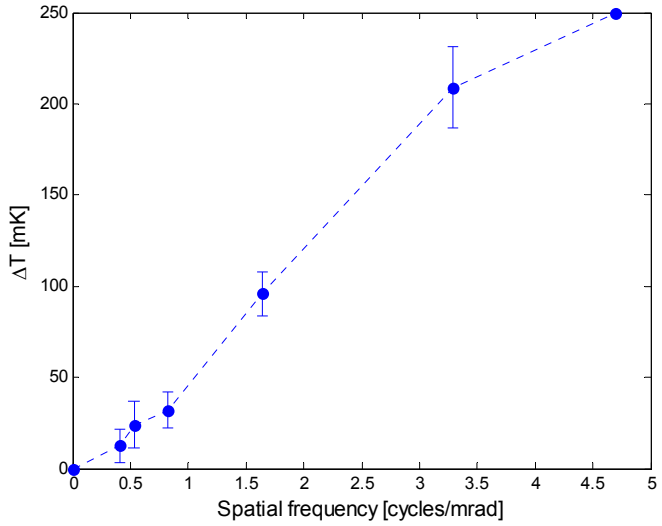

Figure 9. MRTD based on recordings made at 8.35 m range

Spatial frequency [cycles/mrad]	MRTD [mK], based on lowest contrast value	MRTD [mK], based on mean of lowest two contrast values
0	0	0.2 ± 0.7
0.41	12 ± 9	14 ± 9
0.53	24 ± 13	26 ±12
0.82	32 ±10	33 ± 11
1.64	96 ±12	98 ±12
3.29	209 ±22	216 ± 27
4.70	250 ± 0	250 ± 0

Table II. MRTD values calculated by using two estimates for the minimum contrast: the lowest value and the mean of the lowest two values. Shown are average values, averaged over the observer ensemble, and their standard deviations. Range is 8.35 m.

5.2 Machine analysis

The noise floor for each of the curves similar to the ones presented in Figs. 4 and 6 is determined by taking the average of the lowest three values. For lack of a model, the parts of the curves for temperature contrast larger than 0.04 K were modeled as straight lines. The intercept of the noise floor with the straight-line models yields an estimate for the minimum required temperature contrast to resolve the particular pattern. This procedure is repeated for each pattern. Unfortunately, it turned out that only the largest four patterns could be analyzed this way. Analysis of the smallest two patterns resulted only in noise. The analysis results based on the modulation signal and the fifth Fourier component are summed up in Figures 10 and 11. We investigated the effect of increasing the robustness of the analysis by requiring a minimum signal strength that was not just 1× but 1.25× or 1.5× the noise level for the case of the modulation signal analysis, and 1.5×, 2.0× or 2.5× for the Fourier component analysis. Both analyses suggest that signal-to-noise ratios of 1 yield poor results, but that a marginal increase in SNR yields results that are comparable with the ones obtained from the observer test.

The second largest pattern was "located" in a noisy part of detector focal plane array. As a result, the machine analysis of the images yield high noise floor and the minimum required temperature to identify this pattern is also high, resulting in a local peak in the MRTD curves. Apparently, the observers are less sensitive to this noise (solid circles). A possible explanation could be that when the observers could resolve the third largest pattern they assumed or did not check whether the second largest pattern could be resolved, whereas the machine analysis reveals a difficulty with this pattern. A better non-uniformity correction of the detector array might have alleviated this discrepancy.

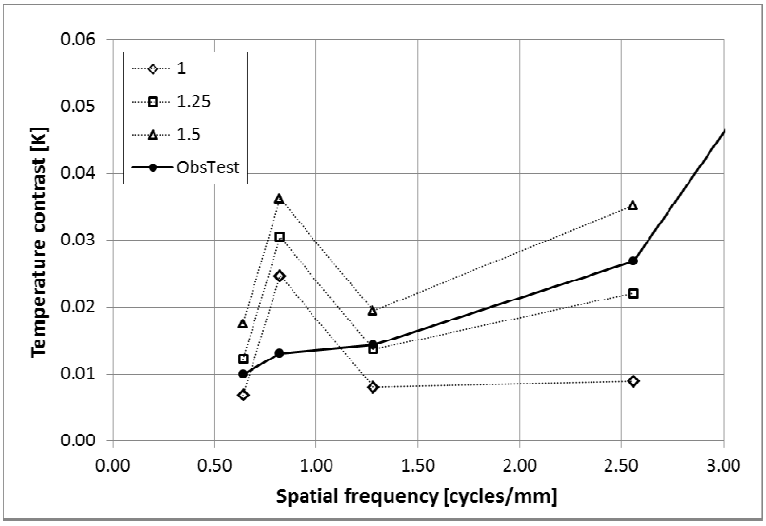

Figure 10. MRTD curves obtained by machine analysis of recordings of the six four-bar patterns at 13.00 m range for three different signal thresholds: 1 (diamonds), 1.25 (squares), and 1.5 (triangles) times the noise level. The curves are compared with the one (solid circles) obtained by the observer test. Analysis is based on the modulation signals from the patterns.

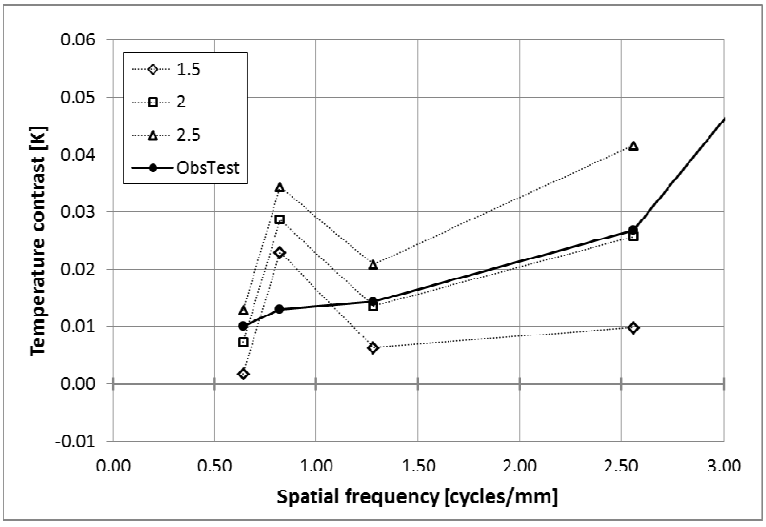

Figure 11. MRTD curves obtained by machine analysis of recordings of the six four-bar patterns at 13.00 m range for three different signal thresholds: 1.5 (diamonds), 2 (squares), and 2.5 (triangles) times the noise level. The curves are compared with the one (solid circles) obtained by the observer test. Analysis is based on the 5th Fourier component of the frequency analysis of the signals from the patterns.

The MRTD curves obtained from the recordings made at 8.35 m range are presented in Figure 12 and 13. Also for these recordings machine analysis of the smallest patterns proved to be difficult. This is unfortunate since it limits the spatial

frequency range of the analysis. A comparison between the observer-test MRTD and the machine-analysis MRTD in Fig. 12, shows similar behavior for larger frequencies when the threshold is set higher, but with quite a difference in the absolute values for the required temperature contrast. For the smaller spatial frequencies the human observer seems more sensitive than the machine, at least for the simple analysis we have implemented here.

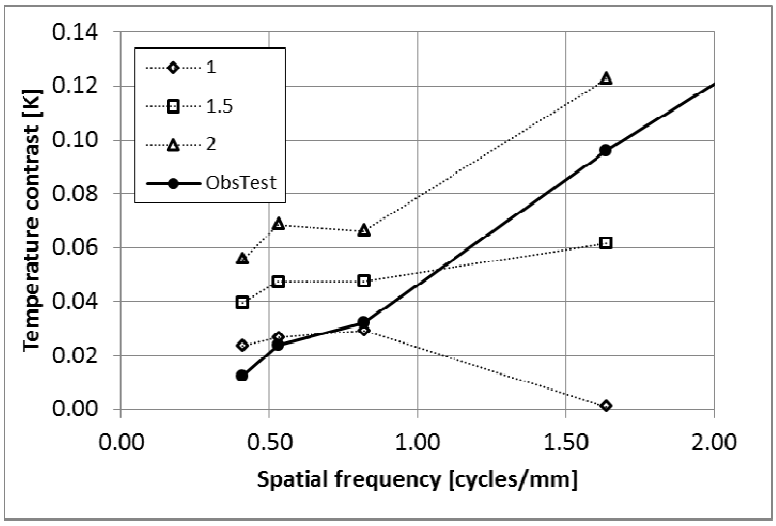

Figure 12. MRTD curves obtained by machine analysis of recordings of the six four-bar patterns at 8.35 m range for three different signal thresholds: 1 (diamonds), 1.5 (squares), and 2 (triangles) times the noise level. The curves are compared with the one (solid circles) obtained by the observer test. Analysis is based on the modulation signals from the patterns.

Considering the analysis based on the pertinent Fourier component we observe quite a good match between the observer-test MRTD and the machine analysis, by choosing the "right" threshold, 2x the noise floor. In this case it seems that the analysis based on the frequency content of the signal gives better results than the analysis based on the modulation amplitude.

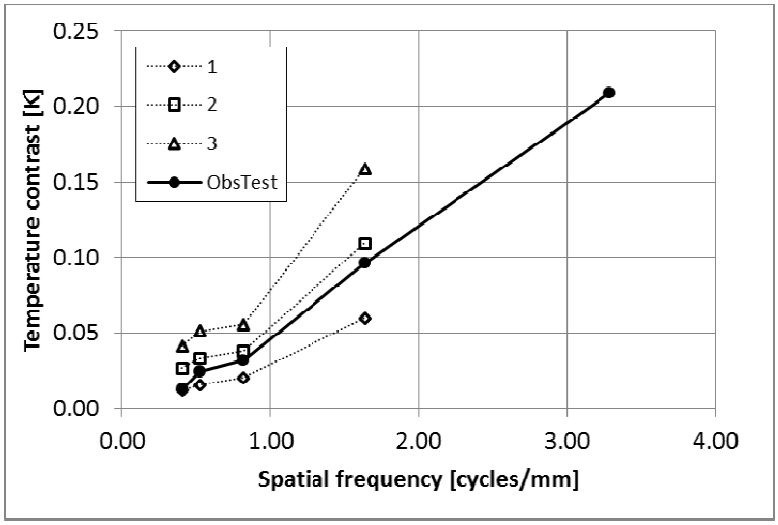

Figure 13. MRTD curves obtained by machine analysis of recordings of the six four-bar patterns at 8.35 m range for three different signal thresholds: 1 (diamonds), 2 (squares), and 3 (triangles) times the noise level. The curves are compared with the one (solid circles) obtained by the observer test. Analysis is based on the 5th Fourier component of the frequency analysis of the signals from the patterns.

6. CONCLUSIONS

Based on the analysis of two sets of LWIR recordings of six four-bar patterns with different spatial frequencies, each set consisting of recordings with different temperature contrast settings between black body and the plate with the bar patterns, we conclude that machine-based analysis of these recordings using very elemental signal analysis techniques yields MRTD curves that are very similar to the ones obtained by an observer test of the same image sets. The observer test made use of an ensemble of observers, either 20 or 10 observers. Whereas observers use an internal strictness or conscience to judge whether they can discern, for example, the fourth largest pattern or just the third largest, for machine analysis an objective threshold (signal-to-noise ratio) must be set. Interestingly, the human observer MRTD can be replicated by setting the required SNR to 1 – 2, depending on which pattern characteristic is used for analysis: the modulation or the frequency content.

As to the question of which yields the better MRTD curve, the human-observer base experiment or the machine analysis, this may be hard to answer. We observe that by tweaking the threshold the machine MRTD may be more or less optimistic than the human-observer based one. But whether this gives a more reliable result cannot be inferred from this work.

Obviously, this is a limited experiment, too limited to make broad claims, but it is interesting to repeat these kinds of experiments to see whether observer trials may be replaced by simple signal analysis bar-pattern recordings.

7. ACKNOWLEDGMENT

We gratefully acknowledge the help we got from Rebekah Wilson and Greg Keating from the Navy Research Laboratory who made their source available and controlled the source while we recorded the image sequences.