A test method for multi-band imaging sensors

Piet Bijl and Maarten A. Hogervorst
TNO Human Factors
P.O. Box 23
3769 ZG Soesterberg
The Netherlands
Phone: +31 346 356 277, fax: +31 346 353 977
e-mail: bijl@tm.tno.nl, hogervorst@tm.tno.nl

ABSTRACT

The emergence of multi-band sensor technology, e.g. in the thermal infrared, promises significant improvements in TA (Target Acquisition) performance. With these new sensor systems, targets may be distinguished from their background not only on the basis of differences in radiation magnitude in the sensor's spectral range (as is the case with single-band systems), but also on differences in spectral properties. However, existing end-to-end sensor performance measures, such as the MRTD, MTDP and TOD laboratory tests or the NVTherm model, produce threshold curves of resolution vs. thermal or luminance contrast and do not take spectral difference into account. Until now no test methodology exists to characterize or quantify the additional benefits of a multi-band sensor above a single-band system. We propose an extension to the current end-to-end test methods that may overcome this shortcoming. The method yields a 2-D threshold surface of resolution, contrast *and* spectral difference between a test pattern and its background. This surface may be used in TA models to predict the ability of a human observer, using the sensor system, to recognize or identify a target given its size, radiance difference and spectral difference with the background. The extension can be incorporated in the TOD, but in other sensor performance measures as well.

Keywords: multi-band, sensor performance, TOD, test method, target acquisition

1. INTRODUCTION

1.1 SINGLE-BAND SYSTEMS

Until now, thermal IR image forming systems are usually *single-* or *broad-band*, i.e. they have one channel with a spectral sensitivity in the Mid Wave (MW, 3-5 μm) or Long Wave (LW, 8-12 μm) infrared band. These systems usually produce a grayscale image that is related to the amount of radiation emitted (or reflected) by the scene in the spectral region of the sensor. If two objects produce the same amount of radiance on the Focal Plane Array in the spectral region of the sensor, regardless their actual temperature and emissivity (and thus their spectral properties), then they have the same 'equivalent blackbody temperature' and this will result in the same detector output.

For single-band IR systems, several measures and models exist to characterize performance with the human-in-the-loop. Examples of these so-called end-to-end measures are: the TOD[1,2,3,4] (Triangle Orientation Discrimination threshold), the MRTD[5] (Minimum Resolvable Temperature Difference) and MTDP[6] (Minimum Temperature Difference Perceived). The TOD[7] model, NVTherm[8,9], and TRM3[6] are sensor models that predict these measures. The TOD is also defined for cameras operating in the visual wavelength domain and Image Intensifier systems. The equivalent for the MRTD in this spectral region is the MRC (Minimum Resolvable Contrast), and equivalents for NVTherm have recently been developed. All these methods have in common that they yield a threshold curve of test pattern size (or resolution) vs. thermal or luminance contrast. This curve is used in Target Acquisition (TA) models[2,6,9] to predict identification and recognition performance for a human observer using the sensor system. As an example, a short description of the TOD method is given in section 3.2 and a TOD curve is given in Fig. 3.

1.2 MULTI-BAND AND MULTI-COLOR SYSTEMS

Currently, *multi-band* or *multi-color* IR imaging systems are under development. Multi-band and multi-color systems have two or more channels that are sensitive in different spectral regions. This can be realized for example by producing Focal Plane Arrays with several layers of detectors, or by using a filter wheel with band-filters that rotates in front of the sensor. The term *multi-band* is often used for a sensor that has channels in different bands (e.g. a MW and a LW channel) and the term *multi-color* is used if the sensor has several sub-bands *within* the MW or LW region. In the present paper, we will use the term multi-band for both types of system. If a sensor has many channels (e.g. more than 12), we will call the system *hyperspectral*. These systems, which usually have a low spatial or temporal resolution, will not be discussed here.

Multi-band systems have the potential advantage over single-band systems that targets may be distinguished on the basis of spectral differences aside from radiation differences. An example of a multi-color system in the visual region is an RGB (color) CCD-camera; in the X-ray region, dual-energy systems are equivalent to multi-band systems. These systems are able to discriminate between materials such as metals or plastics.

Until now no adequate test methodology exists to characterize multi-band sensor performance, to compare two competing multi-band sensors, or to quantify the anticipated benefits of a multi-band sensor above a single-band system.

1.3 APPROACH

In this paper, we describe a method to test multi-band sensor systems. First, we will consider the analogy between multi-band systems and the Human Visual System (HVS): in fact the HVS is a multi-band system, a single-band system can be considered as a multi-band system with a color deficiency. Then, we will look at existing HVS performance tests, for example to determine color deficiencies, and see how we can apply these methods to a general multi-band system.

This paper is organized as follows. In Chapter 2, we will describe some advantages and disadvantages of multi-band systems, and we will discuss the challenges related with the development of multi-band systems. In Chapter 3, HVS testing and color vision are described. In Chapter 4, we propose the new test. Discussion and conclusions are presented in Chapters 5 and 6.

2. CHALLENGES WITH MULTI-BAND SYSTEMS

2.1 ADVANTAGES AND DISADVANTAGES OF MULTI-BAND SYSTEMS

Multi-band systems promise advantages over single-band systems. Possible advantages of multi-band systems over single-band systems are:

- operability in a wider range of climates and weather conditions
- improved search and detection performance (clutter reduction, camouflage reduction, mine detection)
- improved recognition and identification performance (including the distinction from decoys)
- improved visual comfort
- improved automatic detection of hot targets
- estimation of actual temperature and material emissivity; with single-band systems only the equivalent blackbody temperature can be assessed.

Disadvantages may be:

- significant higher costs
- lower SNR (Signal-to-Noise Ratio) per channel
- more complex use
- more complex interpretation of the image

2.2 CHALLENGES

Characteristic to the multi-band systems is the high degree of freedom at several stages. First, the number of channels and their spectral ranges can in principle be chosen freely. The spectral ranges of the bands can even be variable and optimized depending on the circumstances, e.g. by taking a range of hyperspectral bands together. Second, a wide variety of image processing and fusion techniques can be applied to the channel outputs. Third, the resulting imagery can be presented to the observer in many different ways.

In order to optimize multi-band system performance for a certain application, many choices will have to be made during the design. It is further expected that performance will depend largely on the specific task and the circumstances under which the task has to be performed. We can distinguish several challenges:

1. The *Front-end* problem

The Front-end problem deals with the question how to optimize the number of channels and their spectral ranges, depending on the application and task, in order to get maximum information out of the scene. This is mainly a physical problem. For example, if the task is detection and recognition of man-made targets in a natural background, than spectral characteristics of man-made objects and natural scenes are required[10,11]. From these characteristics it can be determined which spectral ranges yield complementary (and therefore additional) information, and which yield similar information (and therefore can be combined).

2. The *Back-end* problem

The Back-end problem deals with the question how to optimize the processing of information from the various channels and how to present that information to the observer. This is a Human Factors and image processing problem, and further depends on the technical display possibilities. There are many possibilities, each with their advantages and disadvantages. A few examples: (i) present each channel image separately on a gray-scale display or let the operator use a switch to select the desired channel, (ii) present only the optimum channel(s), depending on task, weather and climate, (iii) use an image fusion algorithm to obtain maximum contrast in the displayed image, (iv) use false-color to combine the images, and (v) use an algorithm that transforms thermal images into a natural daylight image[12,13].

3. Sensor performance characterization

Sensor performance characterization is crucial to quantify the advantages and disadvantages of different systems or different settings and options as listed above in terms of task performance. Without a systematic measure, one can only subjectively compare systems or optimize system settings. In principle, such a measure needs to be applicable to the wide variety of possible systems mentioned above. On the other hand, the measure only has to be representative for the relevant task that needs to be performed. We will limit ourselves to the target identification and recognition task. Search, for example, requires a different approach[14,15].

3. HUMAN VISUAL SYSTEM PERFORMANCE

3.1 HUMAN VISUAL PERFORMANCE TESTS

Human visual performance tests usually include:

- Visual Acuity (or resolution). An acuity test determines the smallest (angular) detail of high-contrast that a human observer is able to resolve. An example is the Snellen acuity test.
- Contrast Detection. A contrast detection test determines the lowest contrast at which a large test pattern can be detected from the background. An example is the Pelli-Robson test chart.
- Color vision or color deficiency. An example of a color deficiency test is the Ishihara test[21], decribed briefly in section 3.3.1.

Remarkably, Visual Acuity and Contrast Detection represent the two ends of the contrast threshold vs. resolution curves of the TOD and MRC (see Introduction). In this respect, TOD and MRC are an extension of these two HVS tests. In the next section, we will shortly describe the TOD. In section 3.3, we will describe some aspects of human color vision and two methods to measure human color vision performance.

3.2 THE TOD TEST METHOD

This section gives a summarized description of the TOD test method. For a detailed description we refer to Bijl and Valeton[1,4].

In the TOD method, test patterns on a uniform background are shown to an observer using an Electro-Optical sensor system such as a CCD camera or a thermal imager. The tests patterns are equilateral triangles and they have one of four possible orientations: apex Up (▲), Down (▼), Right (►), or Left (◄). See Fig. 1.

In the procedure, test patterns of different contrast with the background are presented for a range of test pattern sizes. After each presentation, the observer has to indicate which triangle orientation he sees, even if he is not sure. This is called a Four-Alternative Forced-Choice (4AFC) measurement procedure[16]. For each strength, the test pattern is presented many times. This yields a relationship between stimulus contrast (or strength) and the probability of correct responses, ranging from 25% (a complete guess) to 100% (the task is very easy). This relationship, shown in Fig. 2, is not a step function but a gradual increase and is called a *'psychometric function'*. The continuous S-shaped curve represents the best fit to the measured data. From this curve the 75% correct point ('*threshold'*) can be obtained. A strong advantage of the 4AFC procedure is that this threshold is independent from a subjective decision criterion of the observer. This as opposed to a Yes/No procedure[16,17] used in other measures such as the MRTD, MRC or MTDP.

A complete TOD curve is obtained by measuring 75% correct contrast thresholds for a range of different test pattern sizes and plotting these thresholds as a function of the reciprocal of the triangle angular size (see Fig. 3). This curve can directly be used in Target Acquisition models such as ACQUIRE[2]. The method has been succesfully applied to thermal imagers[18], visual CCD cameras[2] and recently to X-ray baggage screening systems[19,20].

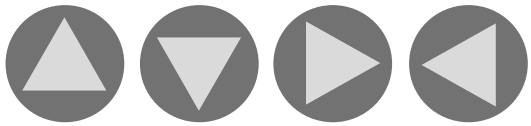

Fig. 1. The test patterns in the TOD method are equilateral triangles with four possible orientations: apex Up, Down, Right or Left. Test patterns of different sizes and contrasts with the background are presented to an observer using an Electro-Optical (EO) viewing system. After each presentation, the observer has to judge the orientation of the triangle, even if he is not sure.

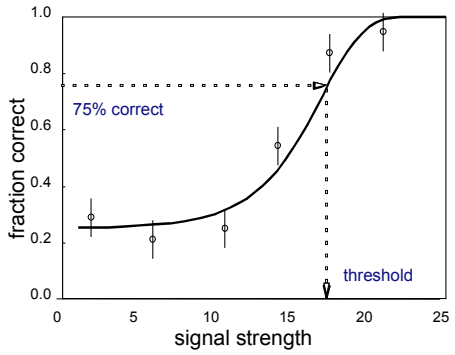

Fig. 2. The probability of correct responses increases gradually with stimulus strength (size or contrast) from 25% (a complete guess) to 100% (the task is very easy). This relationship is called a *'psychometric function'*. The continuous S-shaped curve represents the best fit to the measured data. From this curve the 75% correct point ('*threshold'*) can be obtained which is independent from a subjective decision criterion of the observer.

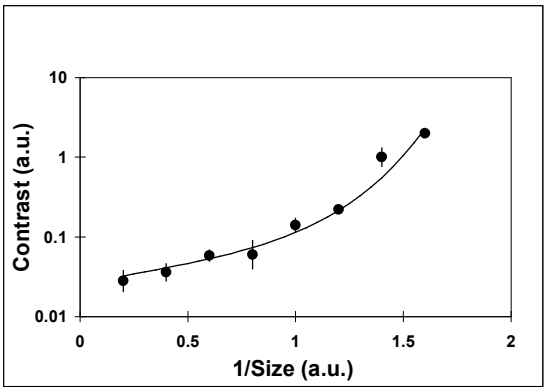

Fig. 3: Example of a TOD curve. In this plot, 75% correct contrast thresholds are plotted as a function of the reciprocal of the triangle size (the reciprocal is used to be compatible to conventional measures in which resolution is plotted on the ordinate). The TOD curve is strongly related to real TA and can directly be used in Target Acquisition models[2].

3.3 HUMAN COLOR VISION AND COLOR VISION TESTS

Under daylight conditions the Human Visual System is able to perceive color due to three types of cones with different spectral sensitivities: L (Long wave sensitive), M (Middle wave sensitive) and S (Short wave sensitive). Especially the sensitivities of the L and M (often incorrectly named 'Red' and 'Green') cones largely overlap. Nevertheless, the HVS is able to distinguish between Red and Green very well by subtraction of their responses.

Since the color (and luminance) coding is limited to three numbers (the responses of the L, M and S cones), all colors can be represented in a 3-D color space or color chromaticity diagram. Several representations of color space exist, e.g. CIE XYZ, LHS, or L*u*v*[26]. In the LHS (Luminance, Hue, Saturation) representation, for example, one axis represents luminance; The colors (hue) are ordered in a circle around this axis, and saturation increases from the achromatic point (along the luminance axis) radially outward to full saturation.

Defects in color vision occur when one of the three types of cones fails to function properly. This leads to problems in *differentiating* between colors. As a result, a range of colors in the chromaticity diagram will coincide, depending on the type of color deficiency. In the western world, about 8% of the males and 0.5% of the females suffer from some kind of color deficiency. In most cases, problems with the L (protanomaly or mild red weakness; protanopia or strong red weakness) or M (deuteranomaly or mild green weakness; deuteranopia or strong green weakness) occur, both giving problems in discriminating between reddish and greenish colors.

Color vision can be tested or measured in several ways. In the next subsections, we will discuss two methods.

3.3.1 Test A: the Ishihara Color Perception Test

Most people are familiar with the Ishihara Color Perception Test[21]. The test plates in this test consist of dots slightly varying in size, luminance and color. The test pattern is a one or two-digit number (e.g. 97) composed of dots of a different color (for example, red dots in a background of green dots). The observer task is to read the number. This is easy for a human with normal vision, but for a color-blind person the number may be invisible or an incorrect number may be seen. The observer has to go through a sequence of test plates. Depending on this responses, the type of color deficiency can be assessed quickly. Normally, there is a possibility that a color-blind person can discriminate between two colors because they appear as a different luminance. In the Ishihara test, this is avoided by the variation in dot size, luminance and color. Alternative tests based on the same principle exist, e.g. in the TNO Vision Screener[22].

3.3.2 Test B: Measuring JND in color space

In color space, one can measure the minimum difference in luminance, hue or saturation (or a combination), required for a human observer to detect a test pattern from its background (see for example Watson[23] or Driggers et al.[24]). This minimum amount or threshold is sometimes called JND: Just Noticeable Difference. At each point in color space, this

JND has the shape of an ellipsoid. The JND can be determined for different types of test patterns and tasks (e.g. detection of discs of a certain size), and with different psychophysical measurement procedures, for example a Yes/No method[16] or a Forced-choice method[16] (see 3.2). Making a map of JND ellipsoids for the entire color space is an extensive experimental study and takes much more time than doing an Ishihara screening.

4. PROPOSAL FOR THE TEST METHOD

4.1 USEFULNESS OF COLOR TESTS FOR MULTI-BAND SENSORS

The Ishihara test (3.3.1) is very appropriate for quick screening of human vision. Usually, it is sufficient to diagnose a color vision deficiency. Although an Ishihara type of test might be applied more generally to multi-band systems (e.g. to an RGB CCD camera), there are a number of disadvantages for our purposes. First, the test cannot directly be included in an existing end-to-end sensor performance measure. Second, because of the many degrees of freedom in design, the number of 'color deficiencies' for a general multi-band system is virtually infinite which means a much more extensive test. Third, due to the variation of the dots in size, luminance and color, the definition of test pattern and background is not very precise. Fourth, sensor systems are often under-sampled and this may cause problems with the dots used in the test.

The JND approach (3.3.2) seems more appropriate for multi-band testing. It can easily be combined with the TOD methodology: for a certain test pattern size, luminance, hue and saturation difference with the background, the fraction correct responses can be measured and the exact 75% correct thresholds or JNDs for triangle orientation discrimination can be determined in the same way as shown in section 3.2 for size and luminance difference (or contrast) only. Combination with other test methods such as the MRTD, MTDP or MRC is possible as well. Further, the test pattern is well-defined (in contrast with the Ishihara test pattern) and especially suited for under-sampled imagers when the TOD methodology is applied[1,25].

Using the JND approach has a practical limitation. Doing such a test for the human vision color space is very extensive. In general, a multi-band system will be sensitive to differences in a sub-space of a virtually infinite multi-dimensional space. Therefore, making a map of JND ellipsoids over this entire space is practically impossible. In order to reduce this space, a possibility is to use a scenario-based approach: the method can be limited to a set of relevant target and background spectra.

In the next subsection, we will work out the latter method.

4.2 THE METHOD IN STEPS

The method consists of five steps, described below for thermal imagers. The same method, however, can be applied to visual cameras. In the next section, an example will clarify the method. The steps are:

Step 1: Start with a standard end-to-end sensor performance measure

The TOD method is taken as starting point. In this measure, both test pattern and background are (approximately) blackbodies. Size and thermal contrast are varied and fraction correct is measured. This yields a 75% correct threshold for positive contrast (T^+) and negative contrast (T^-) at each test pattern size. The average threshold is defined as $\Theta = (T^+ - T^-)/2$.

Step 2: Choose relevant target and background properties

Instead of taking blackbodies, choose a background with spectral emissivity $\varepsilon_B(\lambda)$ and a target with emissivity $\varepsilon_T(\lambda)$.

Step 3: Vary the test pattern spectral properties gradually from background to target

Define the emissivity of the test pattern as follows:

$$\varepsilon_{Test\ Pattern}(\lambda) = \alpha \cdot \varepsilon_T(\lambda) + (1-\alpha) \cdot \varepsilon_B(\lambda),$$

α being an additional variable. If $\alpha = 1$, the test pattern spectral properties equal those of the selected target. If $\alpha = 0$, the spectral emissivities of test pattern and background are equal, yielding a test very similar to the standard test. Now, instead of varying test pattern size and temperature only, also the spectral emissivity of the test pattern can be varied.

Step 4: The test

Keep the background temperature and emissivity constant. Measure the fraction correct as a function of test pattern temperature for a range of test pattern sizes and for a range of values of α. If possible, determine the 75% correct T^+, T^- and Θ. If, for a certain size and α, the fraction correct does not fall below 75%, then the test pattern can be detected on the basis of emissivity differences and $\Theta = 0$.

Step 5: Results

Construct a 2-D threshold surface in the 3-D space with size, thermal contrast and α as ordinates, as in Fig. 6.

4.3 ILLUSTRATION OF THE TEST METHOD

Test pattern variables

In this section, an illustration of the test method is shown for the unaided human eye (HVS) in the visual range. The reason is that the method is easier to understand in the visual than in the thermal range.

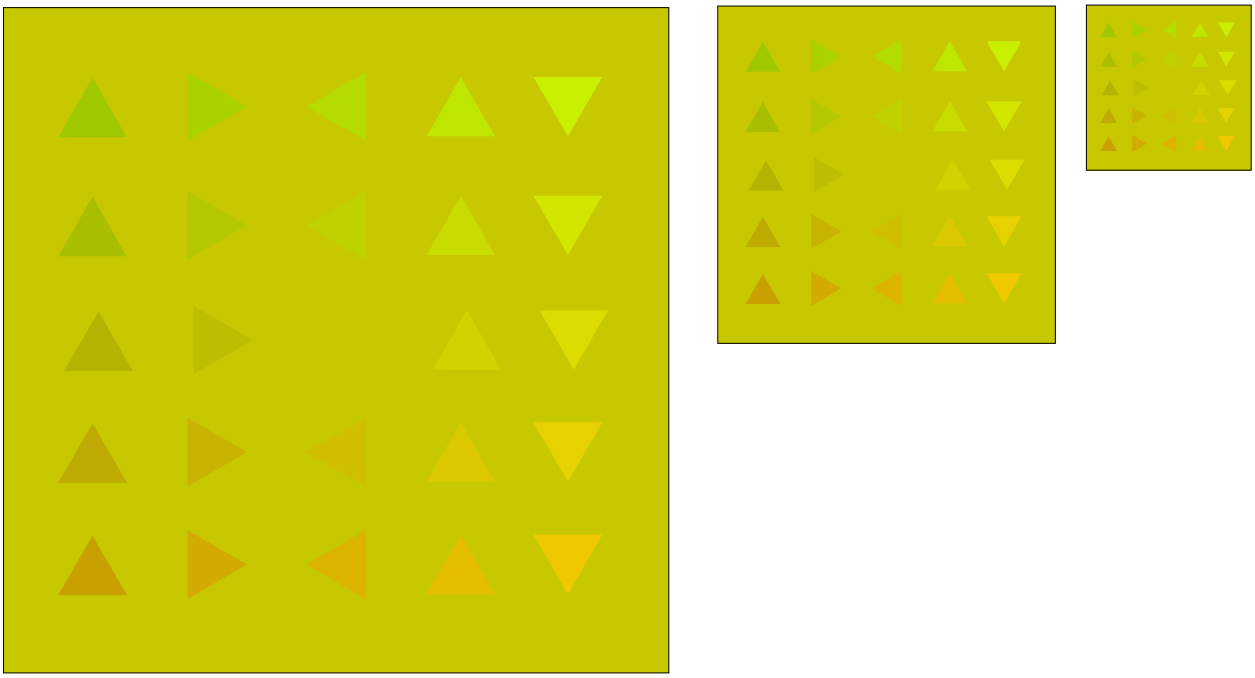

Fig. 4. Illustration of the variables in the multi-band test pattern space (for the HVS): size, luminance (temperature in the thermal range), and spectral difference. The three squares show a set of 5 by 5 test patterns on a uniform yellow background. The different squares represent test patterns of different size. Inside each square, the test pattern luminance is varied in the horizontal direction. In the vertical direction spectral difference is varied, yielding a greenish (top row) to reddish (bottom row) hue for the HVS. In the middle row, the spectral properties of test pattern and background are identical ($\alpha = 0$). Contrast is zero for the center test pattern.

Fig. 4 shows the test pattern variables in the method: size, luminance and spectral difference. This figure shows three squares with 5 by 5 test patterns on a uniform yellow background. The different squares represent test patterns of different size. Inside each square, the test pattern luminance is varied in the horizontal direction. In the vertical direction spectral difference is varied, yielding a greenish (top row) to reddish (bottom row) hue for the HVS. In the middle row,

the spectral properties of test pattern and background are identical ($\alpha = 0$), and contrast is zero for the center test pattern. Note that the appearance of the test patterns and the contrast with the background depend critically on the system that is used, as will be shown later.

Test results for the HVS

In the test method, the fraction of correctly judged triangle orientations is measured as a function of test pattern size, luminance and spectral difference. As an example, Fig. 5 shows the results of these measurements for the largest triangles in Fig. 4. Note that these results are an illustration. Depending on the printer or monitor type, the presented results may disagree with the visual impression of the reader. This is certainly the case if the test patterns are printed in black and white.

The right-hand side plots in Fig. Yy show the fraction correct as a function of luminance for each spectral difference. For the middle row, test pattern and background are of the same color, and the test is similar to a standard TOD measurement. Contrast is exactly 0 in the middle, and fraction correct increases with positive and negative contrast according to the psychometric function introduced in Fig. 2. The 75% correct thresholds T^+ and T^- are approximately symmetrical around zero contrast.

For the top row, the test pattern is always judged correctly in more than 75% of the presentations, irrespective of the test pattern luminance. This is because the test pattern can be discriminated from the background on the basis of its color difference. Thus, the luminance contrast threshold $\Theta = 0$ in this case. The same result is found for the bottom row. For the two intermediate rows, color difference partly contributes to the judgement of the test patterns and Θ is lower than for the case without color difference.

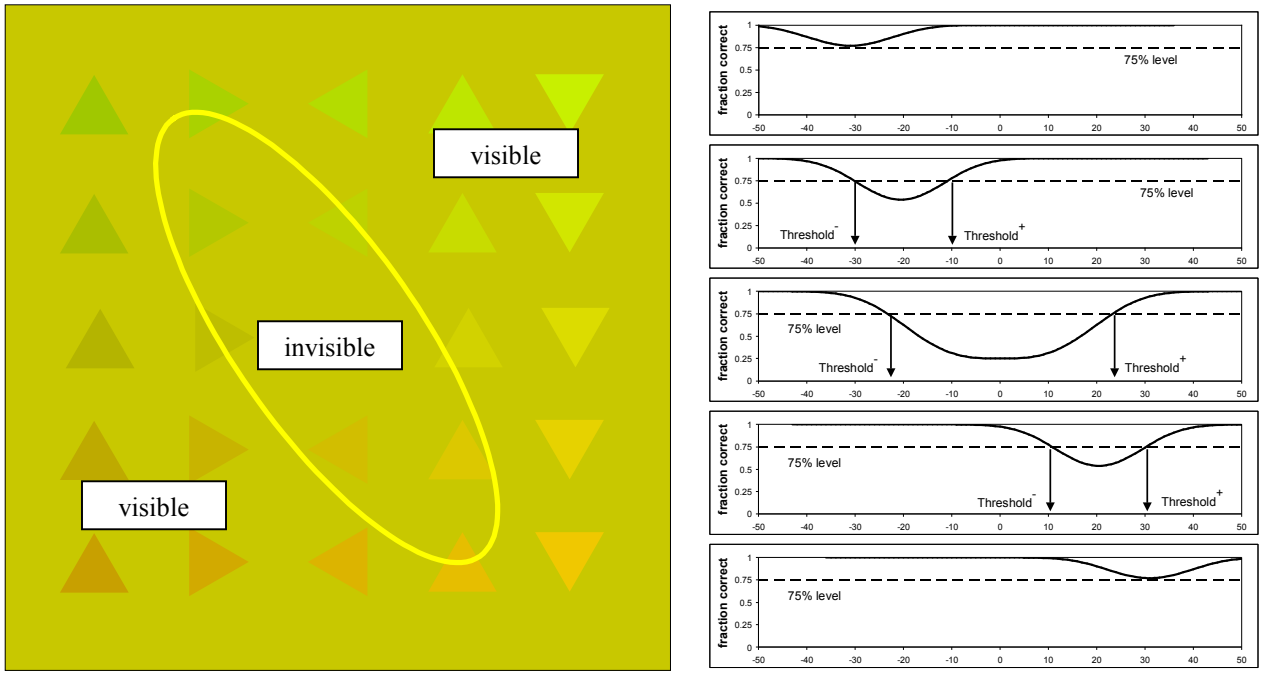

Fig. 5. Results of the threshold measurements with the new test method for the HVS. The right-hand side plots show the fraction correct as a function of luminance for each spectral difference. In the left-hand side picture, an ellipse indicates the 75% correct threshold. See text for details. Note: depending on the printer or monitor type used to show this picture, the presented results may disagree with the visual impression of the reader.

Note that the position of the lowest fraction correct in the curves shifts with spectral difference. This point corresponds to equiluminance of test pattern and background: there is only a color difference. The position of the equiluminance point depends largely on the system, both on the spectral sensitivities of the channels (as will be shown below) and on the processing. This is however an important feature of the proposed test method: it does not only yield the threshold, but also the equiluminance point without knowledge of the system!

In the left-hand side picture of Fig. 5, an ellipse indicates the 75% correct thresholds T^+ and T^-, derived from the curves in the right-hand side plots. Inside the ellipse, fraction correct is below 75% (indicated by 'invisible'), outside the ellipse (indicated by 'visible') the test patterns are judged correctly in more than 75% of the cases.

Similar results can be found for other test pattern sizes (as long as the acuity limit has not been reached, see sections 3.1 and 3.2), except that the threshold curves and the size of the ellipse will be different. In general, the size of the ellipse will be larger for smaller test pattern sizes (similar to the fact that contrast threshold increases for smaller test patterns with a conventional TOD or MRC/MRTD measurement).

The thresholds can now be plotted as a *2-D threshold surface in a 3-D graph* with size, spectral difference and contrast as ordinates. An example graph is shown for a multi-band system in the left-hand side plot of Fig. 6. The right-hand side plot in Fig. 6 represents a threshold surface for a single-band system and will be explained below.

The threshold surface indicates how well a human observer, using a sensor system, is able to discriminate a test pattern. Inside the surface, indicated by 'invisible', the test pattern is below threshold; outside the surface, indicated by 'visible', the test pattern is above threshold. If a system has a higher resolution, better contrast detection performance or better spectral discrimination, the threshold surface will move accordingly towards the origin of the graph.

The cross-section with the zero spectral difference plane ($\alpha = 0$) yields a conventional contrast threshold vs. size (or resolution) curve like the TOD or MRC/MRTD (by convention, these curves are plotted as a function of spatial frequency or 1/size instead of size). Note that the 2-D threshold surface can possibly replace the conventional TOD or MRC/MRTD curve in a model such as ACQUIRE to predict TA performance for multi- and single-band sensor systems.

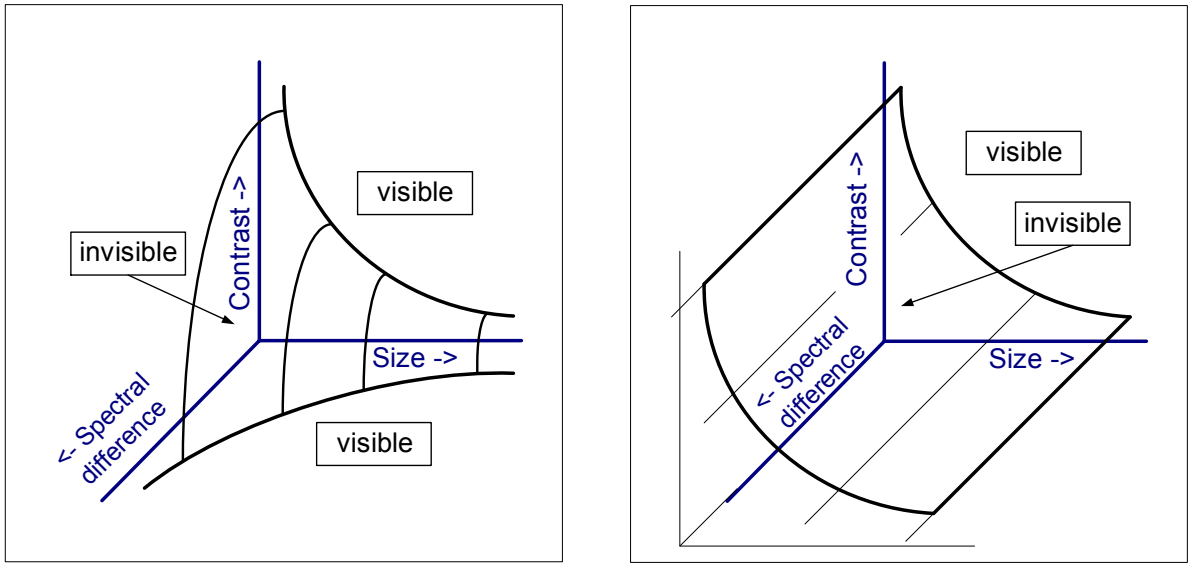

Fig. 6. 3-D multi-band test pattern space with a 2-D threshold threshold surface. Left graph: example for a multi-band system that is able to discriminate between colors or spectral differences. Right graph : threshold surface for a single-band system, unable to discriminate between spectral differences. For both systems, the cross-section with the zero spectral difference plane yields a conventional contrast threshold vs. size (or resolution) curve like the TOD or MRC/MRTD. See text for details.

Test results for a single-band system

In Fig. 7, the same test patterns are shown as in Fig. 4, but now seen through a red filter (left picture) and a green filter (right picture), respectively. The red and green filters reduce the HVS into a *single-band system* with different spectral sensitivities. The parallel lines indicate the 75% correct thresholds T^+ and T^- for the HVS including the filter.

Fig. 7 illustrates two important findings. First, the invisible region (within the 75% correct boundaries) is not limited by an ellipse but by two parallel lines. Thus, for each spectral difference there is a region (around the equiluminance point) where the test pattern orientation cannot be discriminated. This is because, with a single-band system, the test pattern cannot be distinguished on the basis of spectral differences.

Second, the position of equiluminance and the threshold lines as a function of spectral difference is entirely different for the red and the green filter. This again shows that the equiluminance point depends largely on the spectral sensitivity of the channel on the system.

Note that for the middle row (no spectral difference) the width of the ellipse in Fig. 5 approximately equals the distance between the lines for the filtered test patterns in Fig. 7. This indicates that these contrast thresholds are equal and that the TOD and MRC for the multi-band HVS and the two single-band filtered HVS are (approximately) the same. The parallel lines indicate that the threshold Θ is independent of the spectral properties of the test pattern under this condition, although the position of these lines is entirely different for red and green.

In the right-hand side plot in Fig. 6, the 2-D threshold surface for a single-band system is plotted in a 3-D graph. The difference with the results for the normal HVS shown in the left-hand side figure is that the threshold surface is in fact a threshold curve (e.g. given by the cross-section with the zero spectral difference plane) that is independent of spectral difference or hue. In practice, this will only be the case in a first order approximation: factors such as chromatic abberation may affect performance if test pattern and background have largely different spectral properties.

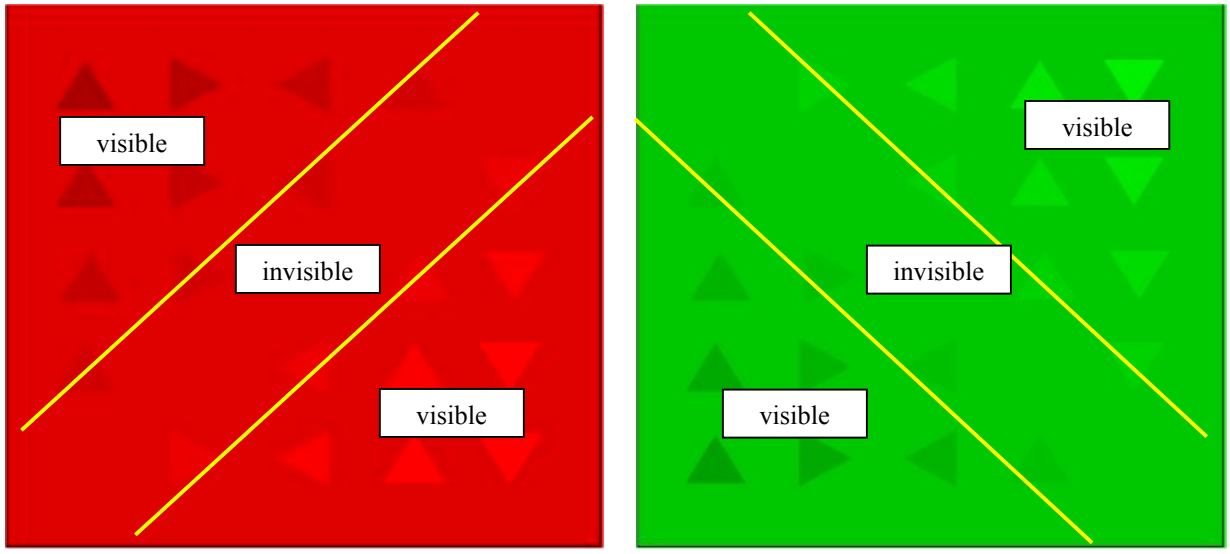

Fig. 7. Test patterns from Fig. 4, now shown to a human observer through a color filter, simulating a single-band system. Left picture: Red filter. Right picture: Green filter. The parallel lines indicate the 75% correct thresholds. See text for details. Note: depending on the printer or monitor type used to show this picture, the presented results may disagree with the visual impression of the reader.

5. DISCUSSION

We propose a new end-to-end test method to characterize performance with multi-band and multi-color imaging systems. The method is an extension of an existing end-to-end measure that is used for single-band systems. It yields a 2-D threshold surface in a 3-D graph with size, spectral difference and contrast as ordinates, that could be used in a TA model such as ACQUIRE. In addition, it renders the point of equiluminance for a test pattern that is spectrally different from the background, without knowledge of the system. This point depends on the spectral sensitivities of the channels, on the processing and presentation to the observer.

A number of considerations have to be made. For a single-band test a blackbody source or a single color for test pattern and background is sufficient. Results for a multi-band system may depend strongly on target and background spectral properties, and this has several consequences. First, performance will be strongly scenario dependent which means that there is a need for representative spectra for targets and backgrounds in certain scenarios (for TA purposes e.g. trees, grass, sand, stone, concrete, tank paint, low-emissive paint, car paint, camouflage nets). Second, for practical reasons it is impossible to test all these relevant combinations of targets and backgrounds. Some kind of test set reduction has to be applied. This might be achieved by, e.g.: (i) Defining a limited set of well-selected characteristic spectra (e.g. by Principle Component Analysis), (ii) Measuring the cut-offs only, (iii) Combination of the test with physical measurements, or (iv) (combination with) Automated testing using a HVS model. Third, each different combination of target and background in principle leads to a different 2-D threshold surface. This means that some kind of averaging is needed to obtain a single threshold surface. Fourth, the practical realization of test patterns is much more complex than for the single-band test. Possible solutions for a thermal test setup are the use of test plates with paints that are mixed in different proportions to obtain a range of test pattern emissivities, the use of filter sets in front of a test plate or background, or the use of mirror array systems that reflect radiation from different thermal sources.

Finally, we propose to start studying the new test method and validation by simulation of both the test and the multi-band systems, preferably including a HVS model[7] in order to avoid lengthy observer testing.

6. CONCLUSIONS

1. A first attempt is made to define an end-to-end test method to characterize multi-band sensor performance. It yields a 2-D threshold surface in a 3-D space which might replace the TOD or MRC/MRTD in a TA model.

2. The method can be combined easily with the TOD method. Combination with MRC/MRTD is possible as well but might be more difficult due to aliasing effects and its subjective threshold measurement procedure

3. In order to be used properly, representative spectra for targets and backgrounds in certain scenarios are required

4. It is impossible to test all relevant combinations of targets and backgrounds, which means that reduction of the number of test combinations is necessary. In addition, averaging of the results for different spectra is needed to obtain a single threshold surface.

5. Thorough validation with field performance or performance on simulated target imagery is required to ensure that the test really characterizes sensor performance in operational situations.

6. The test is an extension to the TOD or MRC/MRTD, which are mainly representative for target identification and recognition. Characterization of visual search performance or other tasks needs to be treated in a different way.

REFERENCES

1. Bijl, P.& Valeton, J.M. (1998). TOD, the alternative to MRTD and MRC. Optical Engineering 37, 7, 1976 - 1983.
2. 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.
3. Bijl, P., Valeton, J.M., (1998) TOD, a new method to characterize electro-optical system performance. SPIE Proceedings, Vol. 3377, 182-193.
4. Bijl, P.& Valeton, J.M. (1999). Guidelines for accurate TOD measurement. SPIE Proceedings, Vol. 3701 14 - 25.
5. STANAG 4349.
6. Wittenstein, W. (1999). Minimum temperature difference perceived – a new approach to assess undersampled thermal imagers. Optical Engineering 38, 5, 773 – 781.
7. Hogervorst, M.A., & Bijl, P. (2001). Capturing the sampling effects: a TOD sensor performance model. SPIE Proceedings Vol. 4372, 62-73.
8. Night Vision Thermal Imaging Systems Performance Model NVTherm. Version 1.0. March 27, 2001.
9. Vollmerhausen, R., & Driggers, R.G (1999). NVTherm: next generation night vision model. Proc. IRIS Passive Sensors, 1, 121-134.
10. Scribner, D.A., Schuler, J. & Kruer, M.R. (1998). Multi-spectral Imaging: Band Selection and Performance Predictions. *IRIS Target and Backgrounds*.
11. Moyer, S., Driggers, R.G., Vollmerhausen, R., Soel, M., Warren, P., Welch, G. & Rhodes, W.T. (2002). Mid-wave Infrared Target Source Characteristics for Focal Plane Applications. *SPIE Proceedings Vol. 4719,* 63-74.
12. Toet, A. (2002). Paint the night: applying daylight colours to nighttime imagery (Report TM-02-B006). Soesterberg, The Netherlands: TNO Human Factors.
13. Toet, A. (2003). Natural colour mapping for multiband nightvision imagery. Information Fusion, In press.
14. Toet, A., Kooi, F.L., Bijl, P.& Valeton, J.M. (1998). Visual conspicuity determines human target acquisition performance. *Optical Engineering 37, 7,* 1969 - 1975.
15. Toet, A. & Bijl, P. (2003). Visual search. Topic: Psychophysics. *In: The Encyclopedia of Optical Engineering. Marcel Dekker, Inc.* (in press).
16. Bijl, P., Toet, A. & Valeton, J.M. (2003). Psychophysics and psychophysical measurement procedures - Introduction. Topic: Psychophysics. *In: The Encyclopedia of Optical Engineering. Marcel Dekker, Inc.* (in press).
17. Bijl, P., Valeton, J.M. & Toet, A. (2003) Electro-optical imaging system performance measurements. Topic: Psychophysics. *In: The Encyclopedia of Optical Engineering. Marcel Dekker, Inc.* (in press).
18. 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.
19. Bijl, P., Hogervorst, M.A., Varkevisser, J. & van de Kooij, J.G.S. (2003). BAXSTER User Guide. Report TNO-TM, Soesterberg, The Netherlands: TNO Human Factors (in preparation).
20. Bijl, P., Hogervorst, M.A., Valeton, J.M. & Ruiter, C.J. de (2003). BAXSTER: An Image Quality Tester for X-ray Baggage Screening Systems. *SPIE Proceedings Vol. 5071* (in press).
21. Ishihara, S. (1962). *Tests for color blindness*. Tokyo, JP: Kanehara Shupper Corp. Ltd.
22. Walraven, J., Lucassen, M.P., Alferdinck, J.W.A.M., Kooi, F.L. (2002). Geautomatiseerde visuele keuring; ontwikkeling van gecomputeriseerde tests voor de TNO VisionScreener [Automated visual screening; development of computerised vision tests for the TNO VisionScreener] (Report TM-02-A044). Soesterberg, The Netherlands: TNO Human Factors. In Dutch. Company confidential.
23. Watson, A.B. (1994). Perceptual optimization of DCT color quantization matrices. Proceedings of the IEEE Intern. Conf. On Image Processing, Austin (TX).
24. Driggers, R.G., Krapels, K., Vollmerhausen, R., Warren, P., Scribner, D., Howard, G., Tsou, B.H. & Krebs, W.K. (2001). Target detection threshold in noisy color imagery. *SPIE Proceedings Vol. 4372,* 162-169.
25. Bijl, P., Valeton, J.M. & Hogervorst, M.A. (2001). A critical evaluation of test patterns for EO system characterization. *SPIE Proceedings Vol. 4372,*27-38.
26. Wyszecki, G. & Stiles, W.S. (1982). *Color Science, Concepts and Methods, Quantitative Data and Formulae*, 2nd. Edition, New York: Wiley.