BAXSTER: An Image Quality Tester for X-ray Baggage Screening Systems

Piet Bijl[a], Maarten A. Hogervorst[a], J. Mathieu Valeton[a] and C. Jaap de Ruiter[b]

[a]TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
[b]TNO Prins Maurits Laboratory, P.O. Box 45, 2280 AA Rijswijk, The Netherlands
Phone: +31 346 3567 277, fax: +31 346 353 977
e-mail: bijl@tm.tno.nl, hogervorst@tm.tno.nl

ABSTRACT

TNO Human Factors in the Netherlands developed a prototype apparatus for testing the image quality of X-ray Baggage Screening Systems: BAXSTER. BAXSTER stands for BAggage X-ray Screening TesteR. The test has a variety of applications, e.g. support by the choice of optimal screening systems for airport security, comparison of competing X-ray systems, delivery tests and routine tests. Advantages over existing tests are i) a strong relationship with real object recognition, ii) accurate and objective test results, and iii) ease of use: performing a test is almost as easy as doing an eye test at the optometrist. BAXSTER is based on the patented TOD (Triangle Orientation Discrimination) test method, which is well-suited for standardization, and development of the test apparatus was funded by the US FAA/TSA (Transportation Security Administration). The apparatus consists of two parts. The first part is a set of test charts containing triangular test patterns of various metals of different sizes and thicknesses. These charts are placed in a frame that is scanned by the X-ray system like a regular baggage item. The operator has to judge the orientation of these patterns on the X-ray image. The second part is a laptop with peripherals and software that controls and analyses the test. The result of a test is a set of performance indicators (relating to detection, resolution, penetration, wire detection and wire penetration) for the entire system, including the display, the operator and the effect of environmental conditions. No (electronic) connection with the X-ray system is required. The effectiveness of automatic object detection and material discrimination through dual-energy X-ray analysis cannot be tested with BAXSTER. In conclusion: with BAXSTER the image quality of X-ray Baggage Screening Systems can be tested easily and objectively.

Keywords: BAXSTER, X-ray imaging, baggage screening, airport, aviation safety, TOD, Electro-Optical system performance

1. INTRODUCTION

X-ray Baggage Screening Systems at Airports are a valuable tool for maintaining aviation safety and to prevent or avoid terroristic acts. One family of systems produces a still image of a baggage item that has to be judged by a human operator. It is of great importance that these X-ray systems perform optimally, and are regularly tested. Therefore, there is a need for an objective and easy test method to measure the image quality of X-ray Baggage Screening Systems.

We developed a prototype apparatus to test the image quality of X-ray Baggage Screening Systems with the human-in-the-loop. The apparatus is named BAXSTER: BAggage X-ray Screening TesteR. A picture of the BAXSTER test apparatus is shown in Fig. 1.

The test is based on the TOD (Triangle Orientation Discrimination) methodology developed by Bijl and Valeton[1,2,3,4] and patented in the US, Europe and Israel[5,6,7]. This method was developed especially to characterize object recognition performance of human observers using an Electro-Optical system. The TOD method has a number of advantages over other performance measures that are applied to sensor systems operating in different regions of the spectrum. Examples of such measures are the MRTD[8] (Minimum Resolvable Temperature Difference) and MTDP[9] (Minimum Temperature Difference Perceived) used for thermal imagers, the MRC (Minimum Resolvable Contrast) used for visual cameras and Image Intensifier systems, and the ASTM (American Society for Testing and Materials) and ECAC (European Civil Aviation Conference) test pieces used to test X-ray systems. Several advantages are given in Chapter 3 and are described in detail in

several papers[1,10]. In particular for the specific tasks required for the application of threat detection with X-ray systems, the TOD methodology has been extended.

This paper is organized as follows. In Chapter 2, some basic principles of X-ray imaging and the security operator task are given. In Chapter 3, the TOD method is described. In Chapter 4, the BAXSTER design is described. Chapter 5 gives a discussion and conclusions.

Fig. 1. Photograph of the BAXSTER test apparatus. The apparatus consists of the following components: a set of test charts with test patterns stored in two small flight cases, two test frames that are integrated in the bottom flight case, and a laptop with control software, joystick, laser barcode reader and headset in a fourth flight case. The flight cases can be stacked and the bottom case contains (removable) wheels to move the test system around easily. Details of the BAXSTER design are described in chapter 4.

2. X-RAY IMAGING AND BAGGAGE SCREENING

X-ray imaging is based on the principle that high-energy radiation can travel through materials that are non-transparent for visible light. The amount of X-ray absorption depends on the type of material, the amount of material and on the energy of the radiation. In general, radiation travels easily through low density materials such as plastics or human tissue, but less well through high density materials such as metals or bones. This makes X-ray imaging especially useful for baggage screening (detection of threats such as guns, knives or explosives inside a suitcase without opening it) and medical imaging (e.g. detection of bone fractures inside a human body).

A typical X-ray Baggage Screening System consists of the following components[11]: a high-energy X-ray source producing a fan-shaped beam that shines on a one-dimensional array of X-ray detectors, and a moving belt. If an object is put on the belt, it moves through the beam and is scanned line-by-line. This yields a 2-D gray-scale still image of the amount of absorption by the object. Fig. 2 gives an example of a grey-scale image of a baggage item. In addition, there exist dual-energy X-ray systems that create a false-color image and discriminate between different types of material.

The operator task at airports is to screen baggage for threats such as knives, guns or explosives. This means that recognition of metal objects (or details of metal objects) and wires is of special importance (wires may indicate the presence of a detonator, for instance). These objects may be obscured by other materials, which reduces the quality of the image and hinders the operator task.

For this application, it is obvious to define image quality in terms of (human) performance on a controlled set of measurements related to detection and recognition of metal details and wires, under non-obscured (ideal) and obscured conditions. In section 4.2, the types of measurement designed for this application will be described.
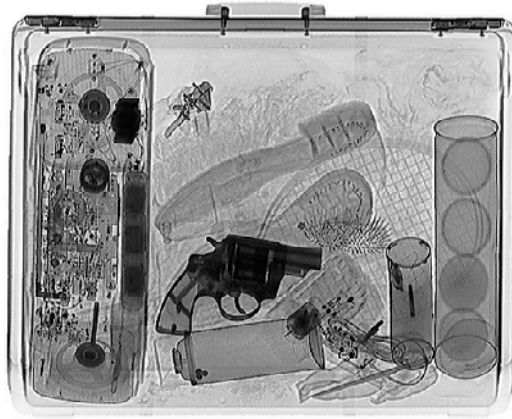

Fig. 2. Example of a gray scale X-ray image of a baggage item (from[12]).

3. BACKGROUND: THE TOD TEST METHOD

This Chapter 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 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 on a uniform background and they have one of four possible orientations: apex Up (▲), Down (▼), Right (►), or Left (◄). See Fig. 3.
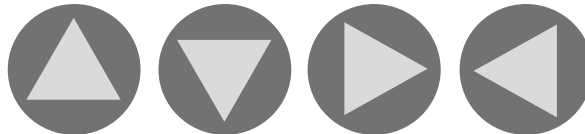

Fig. 3. The test patterns in the TOD method are equilateral triangles with four possible orientations: apex Up, Down, Right or Left. Test patterns of different sizes and contrasts with the background are presented to an observer using an Electro-Optical (EO) viewing system. After each presentation, the observer has to judge the orientation of the triangle, even if he is not sure. This is called a *Four-Alternative Forced-Choice* (4AFC) or *Regular* measurement procedure.

In the procedure, test patterns of different stimulus strengths (sizes or contrasts with the background) are presented. After each presentation, the observer has to indicate which triangle orientation he sees, even if he is not sure. This is called a Regular or Four-Alternative Forced-Choice (4AFC) measurement procedure. For each strength, the test pattern is presented many times. This yields a relationship between stimulus strength and the probability of correct responses, ranging from 25% (a complete guess) to 100% (when the task is very easy). See Fig. 4. The relationship 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 the subjective decision criterion of the observer.

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. 5). This curve can directly be used in Target Acquisition models such as ACQUIRE[2,13,14].

Often, the measurements are limited to two cut-offs:

· Acuity (or resolution): the 75% correct threshold triangle size for a high contrast test pattern

· Contrast detection: the 75% correct threshold contrast for a large test pattern
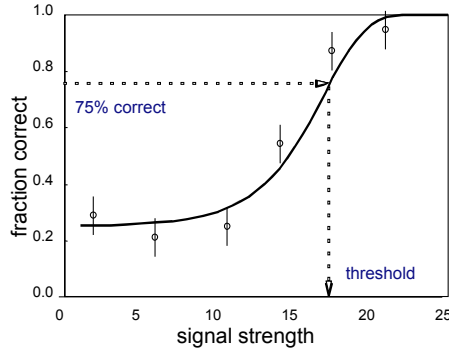

Fig. 4. The probability of correct responses increases with stimulus strength (size or contrast) from 25% (a complete guess) to 100% (when the task is very easy). This relationship is not a step function but a gradual increase and is called a ' psychometric function'. The continuous curve S-shaped curve 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.
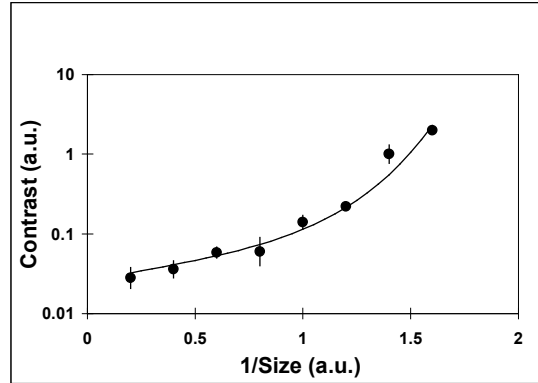

Fig. 5: Example of a TOD curve. 75% correct contrast thresholds are plotted as a function of the reciprocal of the triangle size. This curve can directly be used in Target Acquisition models[2,13,14]. Two important cut-offs are: (i) Acuity or resolution (i.e. the 75% correct threshold triangle size for a high contrast test pattern) and (ii) Contrast detection (i.e. the 75% correct threshold contrast for a large test pattern).

The rationale of the TOD method is that the test patterns represent features or the relation between features of real targets. If an observer is able to discriminate between the different triangle orientations, he also has information about the target features necessary to identify a target. It is also obvious that this ability is strongly related to physical properties of the sensor system, such as blur, pixel size, and the amount of noise in the system.

In conclusion, the TOD method has a large number of theoretical and practical advantages: it is suitable for under-sampled and well-sampled (electro)-optical imaging systems operating in different regions of the spectrum, it has a close relationship to real Target Acquisition, and the observer task is easy. The results are free from observer bias and allow statistical significance tests. Validation studies[2,15] with real targets show that the TOD curve predicts Target Acquisition (TA) performance for under-sampled and well-sampled imaging systems very well. In addition, a theoretical sensor model with a separate vision model to predict the TOD has been developed[16].

4. BAXSTER DESIGN

4.1 GENERAL DESIGN

The BAXSTER test apparatus consists of two parts, the Test Pattern Unit (TPU) and the Control & Analysis Unit (CAU).

The TPU consists of a large set of test plates with test patterns and two test frames. For a test run, a subset of the test plates are placed in a frame that is scanned by the X-ray system like a regular baggage item. The operator views the X-ray image of the frame and has to judge the test patterns. The TPU is described in section 4.4.

The CAU is a laptop with peripherals and software that instructs the operator how to fill a test frame with test plates, controls the test and analyses the results. Further, the software can be used to compose tests. The CAU is described in section 4.5.

An extensive description of the use and possibilities of BAXSTER are given in the BAXSTER User Guide[17].

4.2 MEASUREMENT TYPES

In order to apply the TOD methodology (Chapter 3) optimally to the baggage screener task (see Chapter 2), three additional *measurement types* have been added to the existing *Acuity* and *Contrast detection* measurement types: *Penetration, Wire detection* and *Wire penetration*. The five measurement types are listed in the left column of Table 1. The middle column gives the specific measure that is determined.

The right column in Table 1 describes the test patterns that are used for the different measurement types. For the Acuity measurements, we use *lead* triangle test patterns of different sizes and a fixed thickness. By using lead, the test patterns appear as high contrast dark objects against the background (just as we need for acuity measurements, see Chapter 3) because lead has a very low transmission for X-ray radiation. For the Contrast detection measurements, we use *steel* test patterns of different thicknesses and fixed sizes. Steel is much more transparent to X-ray radiation than lead. For the Wire detection measurements, we use *copper* wires of different diameters. The reason for using copper is that most electric wires are primarily made of copper. For Penetration and Wire penetration, we use *steel* uniform obscurant plates of different thicknesses that we put on top of the plates with test patterns.

Table 1: Overview of the measurement types in BAXSTER

Measurement type	Measure	Test pattern description
Acuity	The smallest detail that can be discriminated (using the X-ray system)	Lead triangles of a range of sizes
Contrast detection	The thinnest piece of material that can be discriminated	Steel triangles of a range of thicknesses
Penetration	The maximum amount of material through which a dark object can still be discriminated	Steel uniform obscurant plates of different thicknesses on top of acuity test plates
Wire detection	The thinnest wire that can be discriminated	Copper wire frames of a range of diameters
Wire penetration	The maximum amount of material through which a wire of certain diameter can still be discriminated	Steel uniform obscurant plates of different thicknesses on top of wire detection plates

Fig. 6 A-C show a range of Acuity, Contrast detection, and Wire detection test patterns, respectively. Fig. 6D-F show their X-ray images. Fig. 6 shows that below a certain size (A, D), a certain steel thickness (B,E) or below a certain wire diameter (C,F) it becomes difficult to judge the orientation of the triangles. Similarly, by covering the test patterns with obscurant

plates of increasing thickness, it becomes difficult to judge even large, high contrast (lead) triangle test patterns (Penetration) or thick copper wires (Wire penetration). With BAXSTER, we determine the size, thickness, wire diameter or obscurant thickness at which a human observer with an X-ray system can just judge the triangle orientation correctly for 75% of the test patterns.

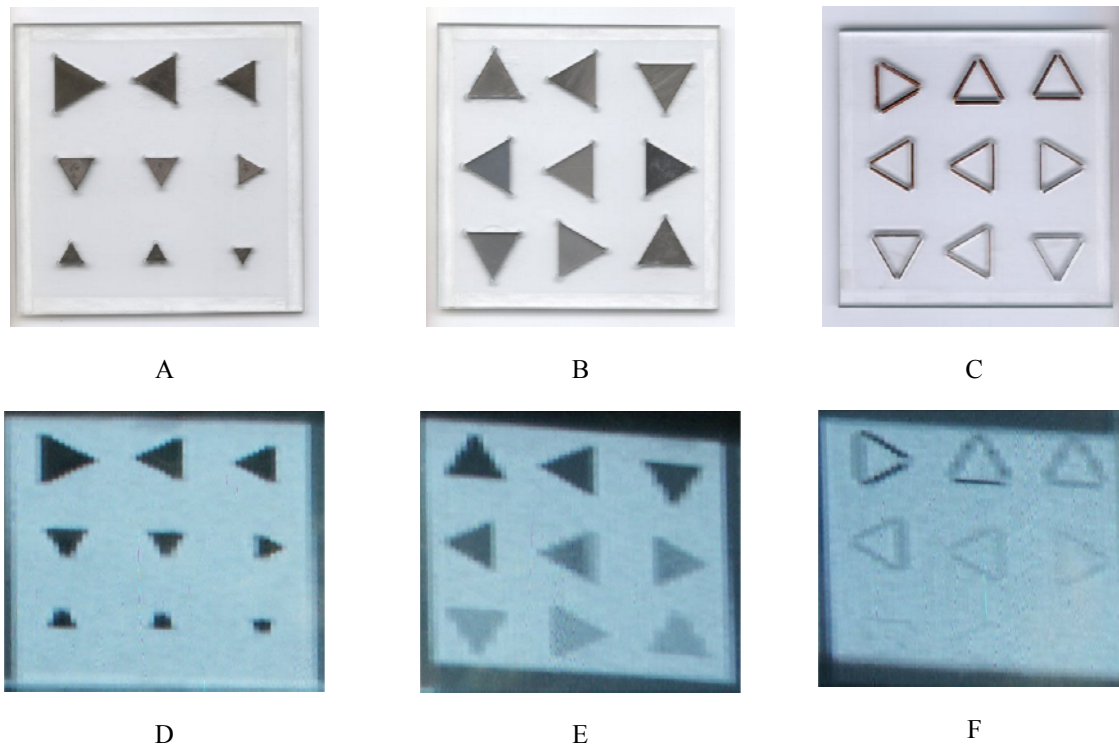| A	B	C

D	E	F

Fig. 6. Examples of different types of test patterns and their X-ray images. A: Lead triangle test patterns. B: Steel triangle test patterns (thickness is varied). C: Copper wire frame test patterns. D-F: X-ray images of A-C. Below a certain pattern size or thickness, it becomes difficult to judge the orientation of the test patterns. With BAXSTER, we measure how well the X-ray system (including the human observer) is able to detect and recognize small details, details with low contrast, thin wires or obscured details.

4.3   MEASUREMENT PROCEDURES

In BAXSTER, two different measurement *procedures* are implemented:

·   The Regular or 4AFC (Four-Alternative Forced-Choice) procedure, already described in Chapter 3

·   The Q&D (Quick and Dirty) procedure

The Regular procedure is the primary measurement procedure, because it yields the most reliable, stable and bias-free 75% correct thresholds. The Q&D procedure however is fast and can be applied to obtain a first rough estimate of the threshold. In BAXSTER a Regular measurement procedure is preceded by a Q&D measurement if the user has no estimate of the threshold at all. From the result of the Q&D measurements, a range of test patterns around the initial threshold estimate is selected to perform the Regular measurements. In that case, lengthy testing using a range of too easy of too difficult test patterns is avoided.

In the Q&D procedure, a range of test patterns of different strengths (e.g. different sizes, thicknesses or obscurant thicknesses) is presented to the observer, and he or she decides for which of the test patterns he is just able to see the orientation. The procedure is illustrated in Fig. 7 for acuity. The size of the test pattern decreases (in two 3 by 3 blocks) from top-left to bottom-right. The top-left test pattern is easily visible while the bottom-right pattern is too small to determine its orientation. The threshold is near the three test patterns indicated by circles, and most observers will select one of these patterns as the initial threshold.

Note that the estimate from the Q&D procedure is not very accurate and may be subject to discussion. This is not a problem if the result is only used as an initial estimate. Many other sensor performance tests however, such as the ASTM, ECAC, MRTD, MTDP and MRC (see the Introduction), use a measurement procedure similar to the Q&D method to obtain a final threshold estimate. Although the accuracy in these tests is increased by repeating the measurements several times, a large disadvantage of these methods is that the results are subjective and observer training is required to avoid observer differences as much as possible.
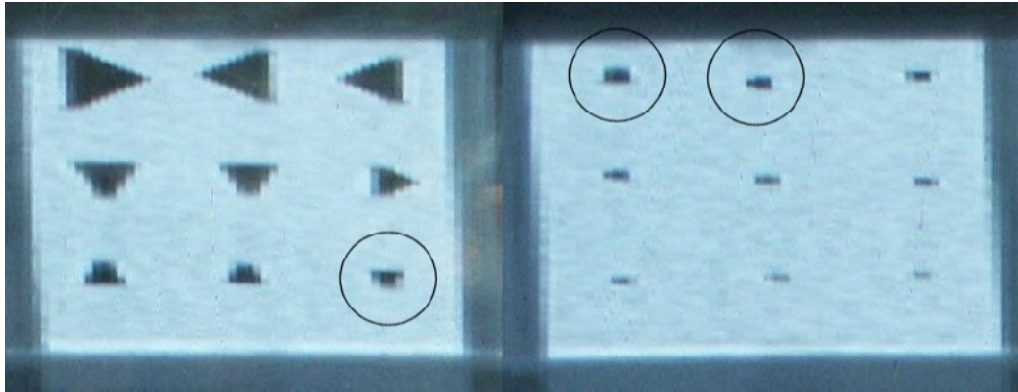

Fig. 7. Illustration of the Quick & Dirty (Q&D) procedure, a fast method used in BAXSTER to obtain an *initial* estimate of the threshold. The size of the test pattern decreases (in two 3 by 3 blocks) from top-left to bottom-right. The top-left test pattern is easily visible while the bottom-right pattern is too small to determine its orientation. Note that the estimate from the Q&D procedure is not very accurate and may be subject to discussion: the orientation is just visible for the bottom-right pattern of the left block, or the top-left or top-center pattern of the right block (indicated by circles).

4.4   TEST PATTERN UNIT (TPU)

4.4.1   *Test plates*

In the present version of BAXSTER, there are a total of 144 square *test plates* (size 5 by 5 in). They can be placed in four possible orientations at any location in a test frame (see 4.4.2). Orientation and location are selected randomly by the control software (see 4.5). So, learning the test pattern orientations by heart is virtually impossible.

Fig. 8. A. Test Plate containing test patterns, with a number, an arrow and four barcodes. B. Obscurant Plate, with a number and one barcode only.

There are three different types of test plates: *Regular test plates*, *Q&D test plates* and *Obscurant plates*. The *Regular and Q&D Test Plates* contain 9 (3 by 3) test patterns. The test patterns are mounted in a perspex template with triangle cut-outs and perspex covers. An arrow indicates the orientation of the test plate, and 4 barcodes identify the plate and orientation. Examples of Q&D test plates without arrow and barcodes are given in Fig. 6A-C, and Fig. 8A gives an example of a Q&D test plate with arrow and barcodes. The *Obscurant Plates* are uniform steel plates without test patterns. The thinnest plates are mounted in perspex covers. The plate orientation is irrelevant, and Obscurant plates have only 1 barcode. Fig. 8B gives an example of an Obscurant plate.

4.4.2   *Test frames*

Test frames are used to hold the test plates during an X-ray scan. The present version of BAXSTER contains two test frames. A picture of an empty test frame is given in Fig. 9.

Each test frame consists of 16 (4 by 4) boxes. Each box has a barcode and can contain 1 test plate and up to 4 obscurant plates. The test frame can be put on the X-ray belt in four possible orientations. A yellow strip indicates the orientation of the Frame.

Fig. 10A shows a photograph of a test frame filled with a number of test plates. Fig. 10B shows the X-ray image of a frame with test plates.

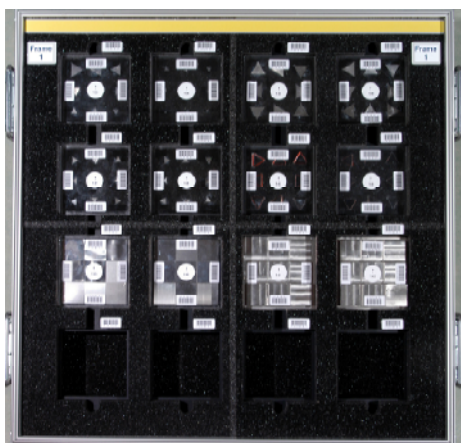Fig. 9. Photograph of an empty test frame

Fig. 10. A: Photograph of a test frame filled with a number of test plates. B: X-ray image of a frame with test plates

4.5   CONTROL & ANALYSIS UNIT

The CAU is a laptop computer with control software, joystick, laser barcode reader and headphones. The control software has several functions (see 4.5.2), and BAXSTER has different user levels: *Test Designer, Supervisor* and *Operator* (4.5.1). Table 2 gives an overview of the functions and their availability to the three types of users.

4.5.1   *User levels*

The *Test Designer* has the highest user level and has access to all functions. He has a good knowledge of X-Ray systems, of the BAXSTER test and the test methods that are used in BAXSTER. He is authorized to design tests, enter X-ray system specifications, add or remove Users of the Supervisor and Operator level, perform tests and analyze the test results. There is only one Test designer.

The *Supervisor* has a good knowledge of X-ray systems and of the BAXSTER test. He is authorized to view the tests and system specifications entered by the Test Designer, add or remove Users of the Operator level, perform tests and analyze the test results.

The *Operator* knows how to operate X-ray systems. He is authorized to perform tests and gets a summarized overview of the test results. At this level, the use of BAXSTER is self-explanatory.

The advantages of the division into three levels are: (i) operation for normal users is as simple as possible, (ii) one person keeps the overview and (iii) improper use or changes of test settings by unqualified personnel is avoided.

Table 2: Overview of the software functions and user levels

Function \ User level	Test Designer	Supervisor	Operator
Test setup	X	View only	-
Add X-ray systems	X	View only	-
Add Users	X	X	-
Run tests	X	X	X
View data (limited)	X	X	X
View data (detailed)	X	X	-

4.5.2   *Software functions 1: Test setup, Add X-ray systems, and Add users*

The Test designer has a GUI available to combine different measurements into a test (test setup), to enter the specifications of an X-ray system (Add X-ray systems) or to add new users (Add users). As an example, the Test setup tab is shown in Fig. 11. The GUI is very user-friendly: many standard tests with default values can be selected, but it is also possible to compose User-defined tests. The Supervisor has limited access to these functions, and the operator has no access.

4.5.3   *Software functions 2: Run test and limited data view*

Run test is the basic function of BAXSTER and is available to all users and is self-explanatory. As soon as Run test is selected, the user selects a X-ray system and a test from a list, and the CAU guides the user through the test, following these steps:

·   It indicates how to fill the test frames with test plates
·   It checks whether the test plates are in the correct position (the user scans the barcode of the box and of the test plate)
·   It indicates how the test frame has to be entered into the X-ray system

- During the measurements, it instructs auditively (using the headphones) and visually (on the laptop screen) which test patterns should be read. This is a very convenient feature: most of the time the observer can keep his eyes on the X-ray image during the measurements, but he can use the laptop information as backup. Fig. 12 shows an example of an X-ray image and the corresponding instructions on the CAU display.
- It collects the user responses entered with the joystick
- It analyses the response data
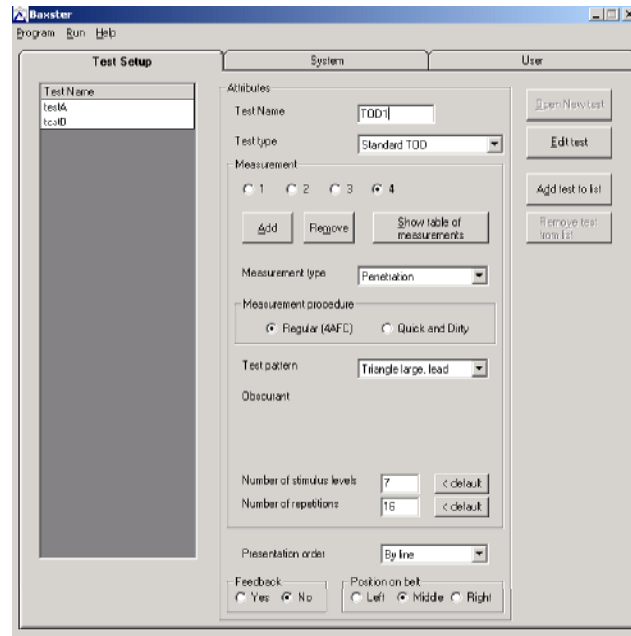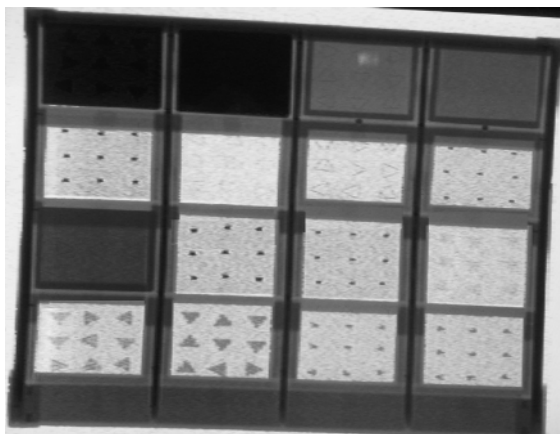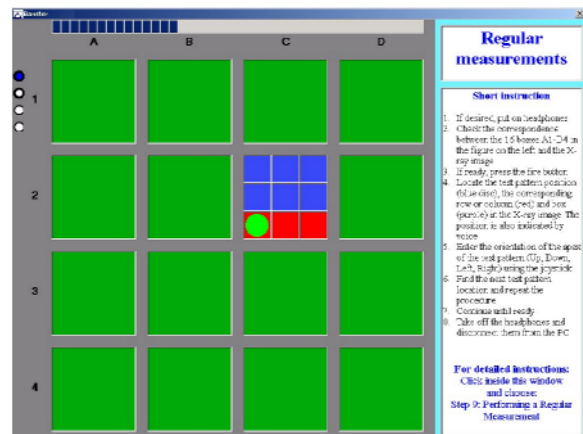- It present a summary of the results

Fig. 11: The Test setup test can be used to compose tests

A

B

Fig. 12. A: example of an X-ray image of a test frame during a test. B: corresponding instructions on the CAU display. The filled circle in the square with boxes indicates the location of the test pattern in the frame. Short instructions are given in the right-hand side window. Online Help with detailed instructions will be activated by a mouse click inside this window. The upper bar shows the progress. The circles in the top-left corner indicate the number of scans in a Test.

4.5.4   *Software functions 3: detailed data view*

A detailed overview of the data is stored in an Excel sheet for further analysis.

5. DISCUSSION AND CONCLUSIONS

We developed a prototype test apparatus for X-ray baggage screening systems. The test has a variety applications, e.g. support by the choice of optimal screening systems for airport security, comparison of competing X-ray systems, delivery tests and routine tests. The test is based on the TOD method which yields results that are closely related to object recognition and is suited for standardization. Especially for the baggage screening application the test has been extended with several new measurement types. Further, the test is easy to use and the results are free from observer bias. The apparatus is flexible: it is very easy to add different test patterns or to extend the ranges of existing test patterns if desired.

The next step with the BAXSTER system is operational use and data collection with X-ray systems. This may lead to improvements of the design, and to the development of dedicated (smaller and faster) tests for specific X-ray machines. Other developments may be the extension of tests for image enhancement techniques, for dual-energy systems, or the extension of functionality (e.g. Go-NoGo testing). Finally, the development of an X-ray sensor model with a separate vision model (similar to the model developed for thermal imagers[16]) can be useful as X-ray design tool, whereas automatic X-ray performance measurement can be achieved by using BAXSTER in combination with the vision model instead of a human observer.

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. Bijl, P.& Valeton, J.M. (1999). Test system for optical and electro-optical viewing systems. United States Patent no. 5926279.
6. Bijl, P.& Valeton, J.M. (2001). Test system for optical and electro-optical viewing systems. Patent No. 120504, State of Israel, Patent Office.
7. Bijl, P.& Valeton, J.M. (1997). Test system for optical and electro-optical viewing systems. European patent application no. 97200852.8, European Patent Office, Bulletin 1997/39.
8. STANAG 4349.
9. Wittenstein, W. (1999). Minimum temperature difference perceived – a new approach to assess undersampled thermal imagers. Optical Engineering 38, 5, 773 – 781.
10. 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.
11. Pasman, W. (1997). Enhancing X-ray baggage inspection by interactive viewpoint selection. PhD Thesis, Delft University of Technology, The Netherlands.
12. Website: www.heimann.com
13. Night Vision Thermal Imaging Systems Performance Model NVTherm. Version 1.0. March 27, 2001.
14. Vollmerhausen, R., & Driggers, R.G (1999). NVTherm: next generation night vision model. Proc. IRIS Passive Sensors, 1, 121-134.
15. Valeton, J.M. & de Jong, A.N. (2000). TOD predicts target acquisition performance for staring and scanning thermal imagers, SPIE Proceeding Vol. 4030, 96-103.
16. Hogervorst, M.A., & Bijl, P. (2001). Capturing the sampling effects: a TOD sensor performance model. SPIE Proceedings Vol. 4372, 62-73.
17. 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).