8

'Real' baggage inspection

It is expected that an inspector's performance will improve when he is presented with a spatial impression of baggage instead of a single view. The experiments in the previous chapters, which tested various tasks related to baggage inspection, indicated that the inspector's performance can be improved by enhancing his spatial impression with the DVWS. It is possible that baggage inspectors can benefit from such a spatial impression without further training. If so, this might be a convincing argument for airports to use, and manufacturers to build, an x-ray scanner based on the DVWS. The experiment described in this chapter tests the feasibility of incorporating the DVWS into current baggage inspection procedures.

The Delft Virtual Window System

The Delft Virtual Window System (DVWS) will be used to give the observer a spatial impression of the baggage (Smets, Overbeeke and Stratmann, 1987; Overbeeke, Smets and Stratmann, 1987). Based on economical, technological, ergonomical and perceptual considerations (see Chapter 3 and 'Previous work' below), the views are restricted to the horizontal arc, and only a small number of images in this range are made available (Figure 1.3a). The required view is selected via a turning knob. Figure 8.1 shows a scheme of the setup. It consists of a monitor, a turning knob and the stored (available) views.

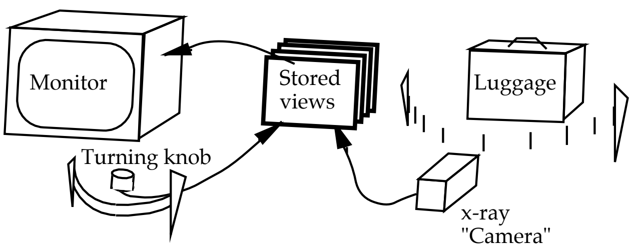

Figure 8.1. Scheme of the setup. A turning knob is used to select the closest available view, which will be displayed. The available views were taken earlier.

The experiment described in Chapter 5 indicated that observer performance can benefit from a camera range (angle between left- and rightmost available view) of 180°, but due to the limited size of the available x-ray scanner the range in the present experiment was ±45°. Providing a camera range of more than 180° seems not useful, since this will give only mirrored images because x-ray images are see-through images.

Previous work

Many airports are working towards a *semi-automatic* scanning system (Attree, 1996; den Ouden, 1995; ACI, 1995; Heimann, 1996, 1997). Such a system consists of an automatic scanner and a human inspector. The automatic scanner picks out suspicious suitcases and marks suspicious items on the display. In the near future, many major airports can be expected to use such a system, partly because purely human inspection is becoming extremely expensive, partly because air traffic and thus the amount of baggage to be inspected is growing fast, and partly because a 100% check of baggage of all kinds will be required in the near future (see Heimann, 1996). In such a semi-automatic system, the baggage inspector will be confronted only with suitcases containing suspicious items, so that his task will be more difficult than with current inspection where all suitcases are checked by a human. This development may make the need for an enhanced spatial impression more urgent.

Several attempts have been made before to improve baggage inspection by giving the inspector a spatial impression of the baggage. For example, Evans, Godber and Robinson (1994) and Scanray (Wooley, 1986) tried to give a spatial impression by scanning two images from slightly different viewpoints and presenting them stereoscopically. This gives the inspector an additional depth cue and may improve his performance. However, a stereoscopic view does not allow the inspector to look around x-ray blocking objects. At the other extreme there are scanners which take two images with a large angle between the views. Two views wide apart will allow looking behind an x-ray blocking object, and may resolve ambiguities in one of the images and camouflaging effects (Nodine and Kundel, 1987). For many tasks, two such views improve observer performance as compared to a single view. This was shown for detecting wires between objects (Chapter 5), for mammography (Wald, Murphy, Major, Parkes, Townsend and Frost, 1995) and for a module replacement task in space (Martin Marietta Aerospace, 1988).

On the other hand, x-ray images of real baggage look more complex than a box with two objects and a wire, and for complex scenes a large angle between the views will disturb the spatial impression of the baggage. Several other complex spatial tasks benefit from more than two available views, for example object recognition (Edelman and Bülthoff, 1992), spatial shape matching (Braunstein, Hoffman and Shapiro, 1987; Andersen and Bennett, 1987) and tracing a path in a tree (Arthur, Booth and Ware, 1993). But much previous work on the effect of number of available views on observer performance was done with non-transparent scenes, and therefore may not hold for transparent x-ray scenes. For example, Kersten, Bülthoff, Schwartz and Kurtz (1992) showed that depth from transparency and opacity can override the bias to see rigid motion. For tracing a wire through a transparent knot (Chapter 6) I showed that observer performance increases with the number of available views, up to at least 33 views. Furthermore I showed that adding views vertically to horizontally continuous views does not increase the percentage of correct responses, and I concluded that the effect of the number of available views on observer performance depends on the spatial complexity of the scene. In the future, semi-automatic inspection will cause an increase in the complexity of suitcases that are inspected by baggage inspectors. I expect that the complexity of this baggage falls somewhere between the complexity of our connected-objects experiment (Chapter 5) and the knot tracing experiment (Chapter 6). Concluding, I expect that more than two views of each suitcase will be useful in order to improve human baggage inspection.

Sometimes x-ray CT scanners are used to make a complete spatial reconstruction of the baggage (Imatron, 1991; Attree, 1996; Henderson, 1990; InVision, 1997), allowing the inspector to inspect arbitrary views and cross-sections. But CT scanners are rarely used, as they are slow, bulky and expensive. An x-ray baggage scanner giving a spatial impression via a moderate number of available views, say 16, may improve baggage inspection as compared to the existing one- and two-view scanners, while being less expensive and faster than the CT scanners (see also Chapter 3).

Taking N images does not necessarily mean an N times increase of the x-ray dose the baggage is exposed to. In the experiment described in Chapter 5, I found that active parallax can compensate for low resolution and a small number of grey levels. Therefore, each of the N images can be of lower quality than if only a single view is presented to the inspector. For taking such lower-quality images, a lower x-ray dose may be sufficient. And with current scan technologies, up to 25 high quality images (Europscan, 1993) can be taken without damage to the baggage.

An inspector's performance is affected by the way the views are presented to him. A spatial impression can be given by presenting the views to the inspector in sequence, as in film. But coupling the images to the eye position of the observer can improve his performance as compared to such a 'film' presentation (Smets and Overbeeke, 1995; Overbeeke and Stratmann, 1988; Arthur, Booth and Ware, 1993; Durgin, Proffitt, Olson and Reinke, 1995). Thus, observer control of the view is essential. The way the observer has control over the view is less important, as I showed that for detecting wires between objects (Chapter 5) selection of the view via a knob works as well as selection by the eye position of the observer. Many inspectors indicated that they preferred manual viewpoint selection over an eye-position-coupled mechanism, principally because they are reluctant to wear markers for the head trackers on their head. Finally, manual viewpoint selection seems preferable over an eye-position coupled mechanism for ergonomic reasons: it seems ergonomically unacceptable to have inspectors move their heads around the display all day. For example McVey (1970) indicated that for watching normal television, viewpoints more than 15 degrees oblique require a head rotation of the observer in order to look at the screen, which is visually fatiguing and therefore may decrease observer performance.

An important problem with the baggage inspection task is that there is no formal system of decisions that leads the inspector from the cues in the image towards a judgement. For example, most inspectors claim to look for objects that may be part of a bomb, such as batteries, electronics and detonators. However, most of these objects have no definite appearance and are not always present in a bomb. Similar problems exist in medical x-ray reading (e.g., Bass and Chiles, 1990). Although the results of Chapters 4 to 7 indicated that performance may increase with increasing numbers of properly chosen viewpoints, I cannot determine the relevance of the tasks used there for x-ray baggage inspection.

Experiment

Method

Stimuli

The stimuli consisted of x-ray images of real baggage. Figure 8.2 shows a view of one of the stimuli from the training series. Two bomb experts working for the responsible police authorities packed 68 suitcases that would give an alarm on an automatic x-ray scanner, and they hid 15 complete bombs in the baggage. The images were digitized in 24-bit colour by a Heimann 7555 Hi-view machine with OTS extension. The colours indicate the materials: orange for organic material, green for aluminium-like materials and blue for heavy metals (see also Heimann, 1997).

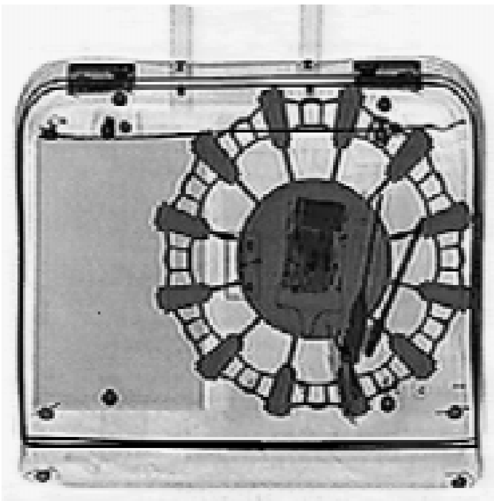

Figure 8.2 (**see colour figure 1.1 on left cover flap**). Example of x-ray scan of suitcase containing a clock, a book and a tin opener. Colours (see text) indicate the materials.

17 images were taken of each suitcase, each with a 90˚/16 rotated suitcase orientation (see Figure 8.4). The suitcase was kept in a rotated orientation with a foam construction that was nearly invisible on the x-ray image. The way of rotation was chosen to minimize distortions due to the perspective-parallel perspective of the scans (see page 55 of Chapter 3, 'Acquiring multiple views with a conventional scanner'). Because of the height of the scanning tunnel (55 cm) and the rotation of the suitcases, the maximum suitcase size was 53x36x20 cm. As the suitcases did not use the full length and width of the scanning machine, only the relevant part containing the image of the suitcase (400x383 pixels) was selected for storage on a hard disc.

Apparatus

Figure 8.3 shows an overview of the experimental setup. For inspection of a suitcase, its 17 images were read from the hard disc into the main memory of an Acorn Risc-PC 702. During the inspection, some or all of these images could be selected with the turning knob. The display was updated at 29 Hz to the latest knob position. The box with two buttons, a green button labelled 'safe' and a red button labelled 'unsafe', allowed the participant to make his judgement. The 15 inch screen (MicroScan 4V/ADI model LM-1564; dot pitch 0.28mm) had a refresh rate of 100Hz. It was warmed up at least 30 minutes in advance of each trial to prevent colour changes during the experiment.

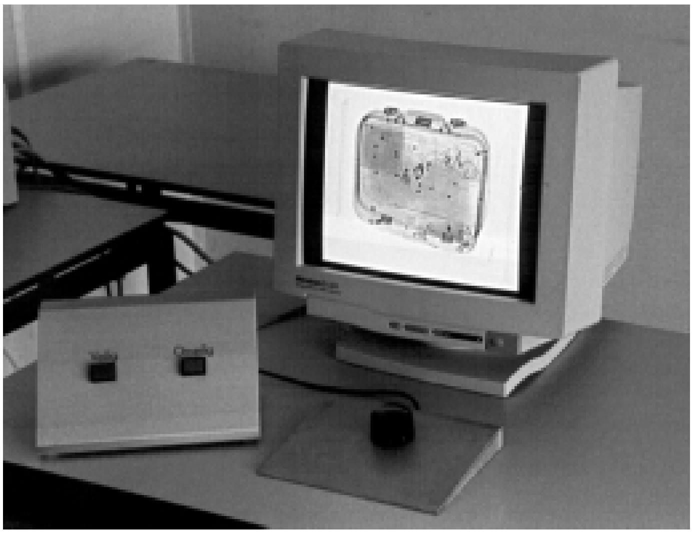

Figure 8.3. Overview of the experimental setup. One of the available views of the suitcase could be selected with the turning knob. The decision for 'safe' or 'unsafe' was made with the button box.

Variables, participants, design

The independent variables were the number of available views N with possible values 1, 2, 5, 9 and 17 (see Figure 8.4) and baggage type L (with or without bomb). The number of views is roughly doubled with each step; the number of available views has to be odd as both one front view and a symmetric camera range are required. The two views condition replaces a three-views condition. This replacement was done because in daily practice inspectors sometimes make a sideview of the baggage with a normal x-ray apparatus by placing a piece of foam under one side of the the suitcase. This corresponds with the 2-views condition that replaces the 3 views that would be required for a symmetric camera range. The dependent variables were the responses R ('safe' and 'unsafe') and the response time T.

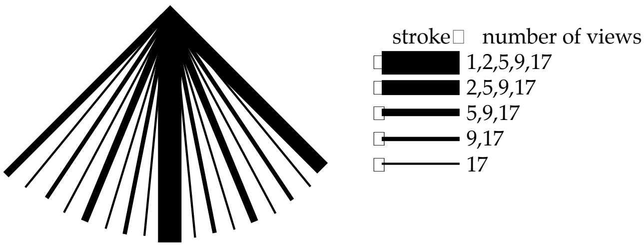

Figure 8.4. Available viewing directions for each number of available views.

The participants were 62 inspectors from the security staff of an airport (20 women, 42 men). All participants were volunteers with at least 2 months of inspection experience and normal or corrected-to-normal vision. They knew the Heimann machines and were able to work with the images, but the airport does not (and will not) use these machines for its regular inspection.

All participants judged the same 55 suitcases, but the suitcases were presented in random order. In total, each suitcase was inspected 12 times in each of the 5 conditions, and these 60 conditions per suitcase were randomized over the participants.

Procedure

The participants were told that their task was to search for complete bombs only, and that they had to press the 'unsafe' button if they judged that a bomb might be present in the baggage, and 'safe' otherwise. They were told that the baggage to be inspected had given an alarm on an automatic x-ray scanner. The colours of the materials (see 'Apparatus') were explained, and they were warned that explosives are not necessarily organic, but can be orange-green and even blue. It was explained that they could select a view with the turning knob, and that some suitcases would turn smoothly, some jerkily and that some could not be turned at all. They were asked to turn the knob at the start of each trial, to find out whether the control was of use to them, because previous experiments had indicated that a trial with only one available view tends to demotivate participants from moving in subsequent trials. They were also asked to rely on their own judgement and not to use hints from other participants. Most participants had not spoken with other participants about the experiment.

They were informed that they could inspect each suitcase for up to 25 seconds, after which the screen would go blank, and that a beep would warn for the time limit after 23 seconds' viewing. Participants did not need to choose before this 25 s limit. They were asked to try to make a correct judgement in the first place, and a fast judgement in the second place.

The participants were trained with 13 suitcases before the experiment. One suitcase of the training session contained a bomb. They were not told about the number of bombs in the training series, but immediately after they pressed a button they were informed whether they had made the right choice (that is, 'safe' if it did not contain a bomb and 'unsafe' if it contained a bomb), and how long they took to make the choice. The response times were shown to encourage them to work fast. The next suitcase appeared on the screen two seconds after a response.

Before the experiment they were informed that they had to judge 55 suitcases, and that they would not receive direct feedback now, but that the number of correct responses was to be shown after the experiment. Again, they were not told about the number of bombs in the series, and in contrast with the training series they could not deduce this number from their final result. Between two suitcases the screen went blank for two seconds. At the very beginning of the experiment they were told that nobody would be given their individual scores, as the aim of the experiment was to test a system and not the performance of individual inspectors, but that they would be informed about their results by a personal letter. In total, each trial lasted about 30 minutes.

Results

Many participants made enthusiastic comments about the 3D impression, improved recognition of objects and the operational comfort offered by the system. Most participants were not used to inspecting colour images, and many expected wires to be more visible in black and white images. Some remarked that they would like to rotate the suitcase up to ±90°.

One female inspector stopped after the training because she was reluctant to participate. She had had bad experience with another experiment, could not work with the colours and found the image quality so poor that she would respond 'unsafe' in all cases. Furthermore the results of one male inspector were excluded because he kept the 'safe' button depressed for about 5 suitcases, i.e. judged the suitcases without having seen them. This left the results of 60 participants for analysis.

Figure 8.5 shows the percentage of 'unsafe' judgements made by the participants and the 95% confidence interval (Loosen, 1994). An analysis of variance (Hays, 1981) was done to test for the effects of the number of available views N and the baggage type L on the 'safe' responses. The number of available views N was not found to be significant: $F(4,3290)=1.09$, $p=0.359$. The baggage type L was found to be significant: $F(1,3290)=27.07$, $p<0.001$. Suitcases containing a bomb were considered unsafe significantly more often than suitcases without a bomb. The interaction of N and L was found not significant: $F(4,3290)=0.20$, $p=0.936$. Although the interaction is not significant, the confidence intervals in the Figure 8.5 suggest that the judgements for baggage with and without bomb differ only when more than one view is available. The effect might prove significant when the number of measurements are increased, but the practical use of such a small difference is dubious.

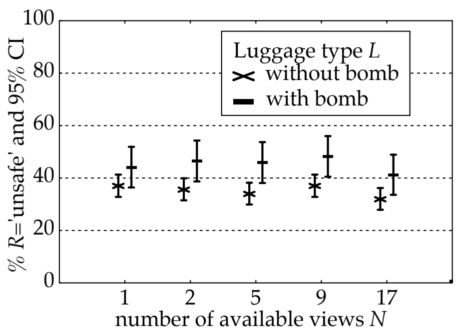

Figure 8.5. Percentage responses R='unsafe' of the participants for different baggage types and number of available views.

Figure 8.6 shows the mean response time and standard deviation. An analysis of variance (Hays, 1981) was done to find the effects of the variables. The baggage type L was found to be not significant: $F(1,3290)=.15$, $p=0.702$. The number of available views N is significant: $F(4,3290)=10.14$, $p<0.001$. A post-hoc Tukey HSD test (Kirk, 1968) indicated that for one available view the response time is significantly shorter than when more views are available. No significant difference in the response time was shown between the 2, 5, 9 and 17 views condition. The interaction between L and N was found to be not significant: $F(4,3290)=0.46$, $p=0.769$.

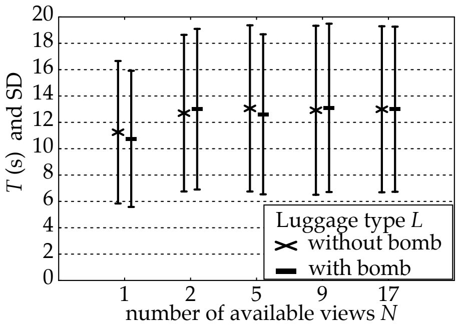

Figure 8.6. Average response times of the participants in the 5 conditions for different baggage types and number of available views.

Concluding, inspectors make a significantly different judgement of baggage with and without a bomb, but the difference is small. The number of available views only has an effect on the response time: as compared to a single available view, operators work more slowly when more than one view is available.

Discussion and conclusions

About 45% of the suitcases containing a bomb were judged unsafe. The responses show a bias of the inspectors towards a 'safe' judgement. This result is not extreme, as den Ouden (1995) showed that, with an automatic scanner in a realistic setup, detection of bombs with a general alarm is about 30%. A false alarm rate of 35% (see Figure 8.5) is high when compared with a false alarm rate during normal baggage inspection of about 5% (den Ouden, 1995), but may be plausible as normal baggage contains lots of 'easy' baggage that is not marked as suspicious on an automatic scanner, while I only examined more complex suitcases which, according to our expert, would be marked as suspicious. Automatic x-ray scanners, such as the Z-scan, even have a false alarm rate of about 35% (InVision and EG&G, 1997).

The results do not support our hypothesis that baggage inspection can be improved by providing the inspectors with multiple views of the baggage with the DVWS without further training of the inspectors. The only effect of the number of available views is an increase of the response time from one to two available views, but the increase is quite small. These results seems to contradict earlier results that showed that, for well-defined tasks, performance increases with the number of available views. Why does the ability to look around baggage not improve the inspection?

A first explanation is that the task is ill-defined. This may cause the inspectors to rely heavily on hints from the machine, such as explosive or detonator indications, when looking for bombs. Results of den Ouden (1995) support this possibility.

The second explanation is that the inspectors are unable to use the extra depth cues provided by the DVWS, probably because of their extensive training with single x-ray images. For medical x-ray reading, the cognitive abilities, which are related to training play a major role (Bass and Chiles, 1990; Kundel and Follette, 1972). I could not give very extensive training because of limited time and limited financial possibilities, and because I have insufficient knowledge to know what capabilities should be trained and how. For example, Drascic (1991) found long learning effects for parallax displays, and training medical x-ray readers also takes a large number of trials (Nodine, Kundel, Lauver and Toto, 1996).

A less plausible explanation is that the resolution is too low for the task. Some participants complained that it would be impossible to see wires given the resolution of my images (400 x 383 pixels). This seems to contradict the findings of Chapter 5, which showed that wires are perfectly visible even at a resolution of 256x128 with 16 grey levels. However, in that experiment the 'suitcase' contained only two objects and a wire, and real baggage is more complex. Another possibility is that the resolution is not too low for wire detection, but that it is too low to recognize critical parts, for example a detonator. Still, the explanation that the resolution is too low seems not very plausible as the scanner I used was the latest Heimann scanner, a commercial scanner optimized to recognize suspicious parts with a single x-ray image. Instead, the complaints of the participants suggest that they based their judgement on a single view.

A last explanation is that the scene is of limited complexity, comparable with a connection-judgement task (Chapter 5). For such tasks, it was shown that it suffices to have only a front view and an extreme side-view; more views do not improve the performance of the inspector. But in the present experiment, only the response times suggest a difference between the one- and two-view conditions. And it seems implausible that real baggage is even less complex than this connection-judgement task, as inspectors indicate that recognition of connections and relations between objects are important.

Concluding, I was unable to show an advantage in providing inspectors with a spatial impression of the baggage without training them thoroughly to use the additional depth cues. It seems that the DVWS did provide extra depth cues that can improve the performance of the inspectors, but that insufficient training of the inspectors to use these cues caused the ineffectiveness of the DVWS. But the primary problem with the baggage inspection task is that the task is hardly operationalized. Future work on baggage inspection should start by operationalizing the task.