Enhancing x-ray baggage inspection by interactive viewpoint selection

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr.ir. J. Blaauwendraad,
in het openbaar te verdedigen ten overstaan van een commissie,
door het College van Dekanen aangewezen,
op maandag 17 november 1997 te 10:30 uur
door

Wouter PASMAN
doctorandus in de informatica

geboren te Soest

Dit proefschrift is goedgekeurd door de promotor:
Prof.dr. G. J. F. Smets

Samenstelling promotiecommissie:

Rector Magnificus, voorzitter	
Prof.dr. G. J. F. Smets,	Technische Universiteit Delft, promotor
Dr. P. J. Stappers,	Technische Universiteit Delft, toegevoegd promotor
Prof.ir. G. Honderd,	Technische Universiteit Delft
Prof.dr.ir. I. Horváth,	Technische Universiteit Delft
Prof.dr.ir. J. van Katwijk,	Technische Universiteit Delft
Prof.dr. J. J. Koenderink,	Universiteit Utrecht
Prof.dr. C. Nodine,	University of Pennsylvania

This work is supported by the Dutch Technology Foundation (STW)
under grant DIO22-2732.

ISBN-nummer 90-9010959-5

Contents

1

Introduction

When we travel by air or boat our baggage has to be checked because the safety of other passengers may be at stake. To preserve the passengers' privacy and because of the high cost of checking baggage manually, the preferred option is usually to use x-ray scanners. Figure 1.1 shows a typical x-ray image. Most objects are transparent to x-rays, but very dense metals, such as lead, completely absorb x-rays.

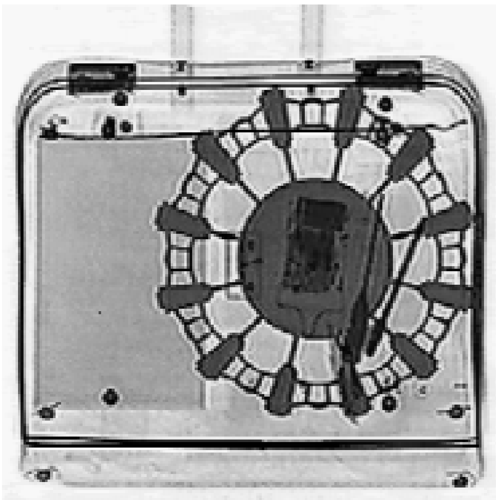


Figure 1.1 (**see colour figure on left cover flap**). Example of x-ray scan of suitcase containing a clock, a book and a tin opener. Colours (see Figure 8.2 on page 134) indicate the materials.

The baggage inspector has a hard job, as he has a large responsibility but little time: usually only about 6 seconds to decide whether a suitcase is safe or not. Guns and knives are relatively easy to detect, as they partly absorb and partly scatter x-rays and have a characteristic shape. But bombs are much harder to detect, as the typical parts of a bomb can be hidden easily: plastic explosives can have any shape, a delay mechanism can be extremely small, the detonator may be placed in such an orientation that it is hardly visible and very thin wires may not be visible at all. Furthermore, X-ray images get cluttered if many items are stacked on top of each other. Finally, some objects may completely absorb x-rays, thereby making things behind them invisible.

The Delft Virtual Window System (DVWS)

In order to unclutter stacked objects, to allow looking behind objects which absorb x-rays and to enhance image resolution, a spatial impression of the baggage contents may be given to the inspector instead of a single x-ray image. This thesis is about the optimization of baggage inspection by means of the Delft Virtual Window System (*DVWS*), a display technique that can give the observer a spatial impression of a scene. The DVWS was developed and patented by Smets, Overbeeke and Stratmann, and is described in detail in the patents (Smets, Stratmann and Overbeeke, 1988, 1990) and in the doctoral dissertation of Overbeeke and Stratmann (1988). Its potential for applications is described by Smets (1995).

The DVWS consists of a monitor, a viewpoint position tracker and a camera that looks at a scene (Figure 1.2). The system gives the observer a spatial impression of the scene on the monitor display by coupling the camera position to the viewing position of the observer. I will refer to an image belonging to a camera position as a *view*. For single pictures out of their 3D context I will use the word *image*, while I will use the word *view* for a single picture within the 3D context. To prevent the point of interest, the *fixation point*, from shifting off the screen as the observer moves, the camera rotates in such a way that it remains aimed at the fixation point. This coupling can be made for the left-right, up-down and forward-backward movements of the observer. In practice, the precise coupling between eye position and camera position differs slightly over implementations, and will be described for each experiment.

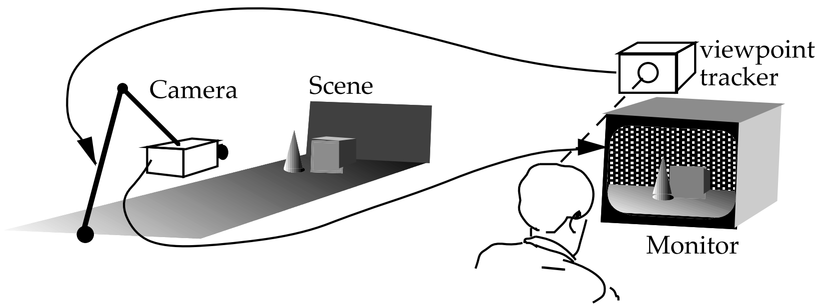


Figure 1.2. Principle of virtual window displays. The camera position and orientation are coupled to the observer's viewpoints. For the observer, the scene on the display becomes spatial.

Applying the DVWS to baggage inspection

For x-ray baggage inspection, the setup of Figure 1.2 is not practical: for each new view a new x-ray photo has to be taken. The observer can take an infinite number of positions (Figure 1.3a), and therefore the number of views is potentially infinite, but the baggage contents may be damaged if more than about 25 x-ray photos are taken. In this thesis, an attempt is made to solve this problem by restricting the selectable views (A) to the horizontal arc (Figure 1.3b) and (B) to certain discrete positions along that arc (Figure 1.3c). As the number of selectable views is reduced, a mismatch is introduced between the actual observer's position and the displayed view. Either the distance between extreme viewpoints (*camera range*) will decrease or the angular distance between the available views (*angle between views*) will increase. This raises questions about the consequences of this reduction on the quality of the inspection.

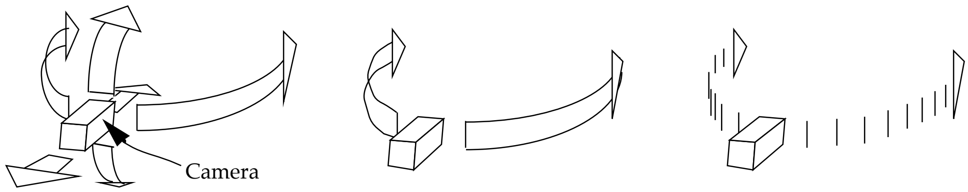


Figure 1.3a. In the unrestricted DVWS the number of available views is infinite in 3 dimensions.

Figure 1.3b. To limit the number of views, camera motion is restricted to the horizontal arc.

Figure 1.3c. To make the number of views finite, camera motion is made discrete.

In the building of a prototype system, other factors are also important, particularly the cost of the system and the best way of achieving interaction between the machine and the operator. In order to find the most cost-effective solution, Chapter 3 outlines a number of technical solutions that are able to record the required views. These solutions differ in the time required to record the required views, the number of x-ray cameras and x-ray sources, the size of the apparatus, the number of moving mechanical parts and, most important for the inspection quality, how the views are taken. For the man-machine interface, ergonomic aspects have to be considered. The ergonomic aspects of the eye position tracking mechanism need special attention, as it is not ergonomically acceptable to have operators moving around the display all day. Furthermore, baggage inspectors prefer not to wear special reflectors or mechanical tracking constructions on their head. An attempt was therefore made to replace viewpoint selection via eye position with viewpoint selection by a knob. As with restricting the number of available views, this may affect the quality of the inspection.

Previous work

The task of the inspector is different for inspecting hand baggage and inspecting hold baggage. These differences became clear after the first experiment (Chapter 4) and in the course of contacts with an airport. When inspecting hand baggage, the inspector tries to recognise all objects in order to detect the presence of dangerous items such as guns or knives. When inspecting hold baggage he tries to find parts of a bomb. For checking hand baggage, cheaper inspection machines are used that give little information to the inspector and thus necessitate manual inspection, because manually inspecting baggage of the passengers is part of security. A suitcase may be opened only in the presence of its owner, and therefore inspectors are not allowed to open hold baggage. For these reasons it was decided to concentrate on hold baggage in this thesis.

According to the inspectors, a ‘standard’ bomb consists of a battery, a delay mechanism, a detonator, explosives and wires connecting these parts. The operator tries to match these components to the actual baggage contents, and if they do not match the suitcase may be safe. This is quite a ‘fuzzy’ process, as the components of a bomb have no standard shape, and because some components are not essential. Therefore there is a large variation among inspectors. There is similar uncertainty about what makes a good inspector for medical x-ray screening (Bass and Chiles, 1990). Nodine, Kundel, Lauver and Toto (1996) have shown that in the case of medical x-ray images it takes a huge number of trials before radiologists know what a normal situation looks like. For baggage inspection, much more variation exists in normal situations, thus suggesting an even longer learning curve. Nevertheless, I assume that wires between objects provide an important clue. In 1983, the FAA required scanners to be able to display wires of 0.5mm diameter (Tsacoumis, 1983), and at that time most x-ray equipment was able to display wires with a diameter of 0.16 mm diameter (Dorey, 1983). Currently, most machines are able to display wires down to 0.1 mm diameter.

Much research on x-ray baggage inspection is classified, as depriving potential terrorists from information about the inspection process is part of the security. This gives problems for the precise description of the baggage inspection task. Furthermore, the inspection task itself seems to be not strictly defined (*operationalized*) either. It is not clear precisely what baggage inspectors are looking for: the baggage inspection task is not well *operationalized*. Most baggage inspectors will explain that they are looking for parts of a bomb, but often items resembling such parts are present, while the inspectors do not mark them as parts of bombs. This problem with the operationalisation of the task seemed quite irrelevant at the beginning of the project. However, this problem makes the implications of research on tasks that seem related to baggage inspection uncertain, and as the project progressed it became clear that it was of crucial importance. Given this knowledge, an analysis of the baggage inspection task would be an appropriate first step, but as observed above the importance of this question was underestimated.

Below, previous work that set the direction for the research of this thesis is discussed. The work in question is split into two parts: a technological part that describes how x-ray images can be made, and a perceptual part that describes the aspects of the display system that were expected to affect the performance of the inspector. The way the images are made does not dictate the way they are presented. In principle, a conventional scanner can be adapted in a straighforward way to provide views for presentation with the DVWS, by rotating the baggage in the required orientation and storing the acquired views. The important questions here are about the trade-off between scanner price, the amount of exposure of the baggage to x-rays, and the inspectors' performance.

Previous work - technological

The technological aspects of x-ray baggage inspection can be split into three categories: the technology to record the required number of selectable views, the technology for recognising aspects of the material in the baggage and the technology that analyses the visual cues in the images.

Currently, conventional x-ray baggage scanners use a fan-shaped x-ray beam shining on a row of x-ray sensors: the *sensor line* (Figure 1.4a). The baggage is pulled through the fan-shaped beam, and the baggage is scanned line by line. These lines are put together on the display to form the x-ray image. To present the inspector a spatial image of the baggage, a conventional scanner can be adapted in a straightforward way (Evans, Godber and Robinson, 1994). Figure 1.4b shows a scanner that can record two images for *stereoscopic* viewing (where images taken from slightly different viewpoints are presented to the left and the right eye).

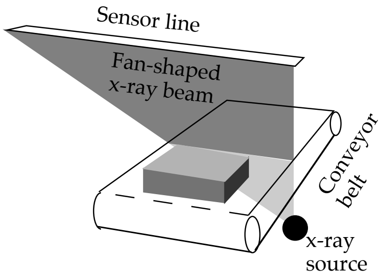


Figure 1.4a. Conventional x-ray baggage scanner. The baggage is pulled through a fan-shaped x-ray beam, and the scan lines are collected to form the x-ray image.

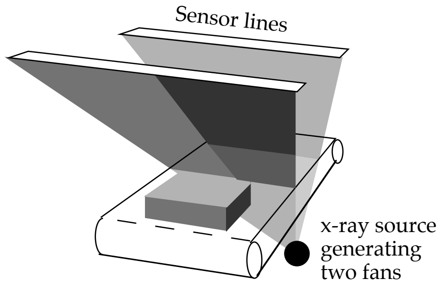


Figure 1.4b. As Figure 1.4a, but two images are scanned simultaneously.

With the machine made by Isorad (Isorad, 1988) the baggage can be rotated during scanning. This enables the inspector to look around x-ray absorbing objects and to select a view that provides sufficient cues to decide whether a suitcase is safe. This system is rarely used, probably because it is designed for careful inspection of a few suitcases and because it exposes the baggage to a continuous beam of x-rays. In current practice, sometimes the baggage is rescanned in a rotated orientation with a conventional scanner by placing a piece of foam under the baggage, but the amount of rotation is limited by the height of the scanning tunnel. Rescanning usually takes too much time, and therefore a scanning machine exists (EG&G Astrophysics, 1996) that always takes two images: one front view and one at a 60° angle (see Figure 3.7 on page 43). Finally, a complete spatial reconstruction of the baggage can be made with a CT scanner (Henderson, 1990; Invision, 1997). With such a reconstruction any desired view or cross-section of the baggage can be inspected. However these machines are rare, as they are expensive, bulky and slow, and the scanning tunnel is too small to fit all baggage (Henderson, 1990). Currently, Invision and EG&G Astrophysics are cooperating (Dotzler, 1996; InVision & EG&G, 1997) to improve the scanning speed of the CT scanners, and hope to attain the same speed as the fastest line scanners (1500 bags/hour).

The second category is the way the properties of the materials in the baggage are derived. The density of the material can be derived from the amount of x-ray absorbance and scattering, and suspicious densities (near the density of nitrogen, the principal constituent of explosives) can be detected. By scanning with two x-rays with different energy levels, organic, inorganic and metallic materials can be distinguished. This material information is usually merged into the image by using pseudo-colours, where each colour represents a particular class of material (Figure 1.1, see also Figure 8.2 on page 134).

The third category is automatic object recognition. Especially for bomb detection it is useful to detect automatically the presence of objects resembling detonators and batteries. Such information is usually merged into the image by adding arrows or other signs pointing to the suspicious items, and flashing a warning sign or sounding a beep.

Usually, baggage inspection is not done solely based on x-ray images. For detecting explosives, nitrogen-detection can be carried out more precisely with neutron scanners or gamma ray scanners than with x-ray scanners. Neutron scanners determine the way neutrons are absorbed and what radiation is generated when the baggage is exposed to neutrons. Gamma ray machines expose the baggage to gamma rays with an energy that is absorbed highly by nitrogen. These machines are expensive because of the necessary shielding and short-lived neutron- or gamma-ray sources. Another important technique is vapour detection. Even a few evaporated atoms from explosives can be detected, but explosives evaporate very little. Finally, in high-risk flights the owners of the baggage are screened and judged on the basis of an interview.

Previous work - perceptual

I am looking for the combination of viewpoints and image quality that gives optimal inspector performance. This section discusses relevant literature to sketch our major expectations and gaps in our knowledge that have to be filled.

Image resolution and number of grey levels are known to have an effect on the response time when manipulating remote objects (Ranadivé, 1979). Observers controlling the camera motion can cope with much lower resolution than observers not having this control (Smets and Overbeeke, 1995). These results suggest an interaction between image resolution, number of grey levels and amount of control over the camera. For several inspection tasks, performance has been shown to improve with increasing numbers of available views. However, Kersten and Bülthoff (1991) reported that transparency can interfere with the human bias to see objects as being rigid, and therefore the results mentioned above may not hold for transparent scenes such as x-ray images.

I expect that providing multiple views improves the spatial impression of the baggage. Geometrically, it is possible to reconstruct a scene from two distinct views (Longuet-Higgins, 1981) although some assumptions may be required (Ullman, 1979). With perspective assumptions even a single view may contain sufficient spatial cues to reconstruct the scene up to a scaling factor (Halloran, 1989). However, experiments testing human performance with sparse spatial scenes (e.g., Braunstein, Hoffman, Shapiro, Andersen and Bennett, 1987) suggest that human accuracy in reconstructing the 3D scene from distinct views will increase with more distinct views, more points and with increased constraints, even when fewer or more viewpoints are provided than are geometrically necessary. Furthermore, providing multiple views may enhance scene rigidity (Todd, Akerstrom, Reichel and Hayes, 1988; Todd and Bressan, 1990).

For tasks with real scenes, an increasing number of available views can also increase observer performance. For instance, multiple images may resolve ambiguities in cluttered parts of the image such as the so-called 'camouflage effects' (Nodine and Kundel, 1987; Vyborny, 1997). For breast cancer screening with x-ray images, Wald, Murphy, Major, Parkes, Townsend and Frost (1995) showed that, compared with a single view, two mammographs increase the chances of detecting a cancer. Evidence exists that humans remember a number of views of a spatial object, and use them for recognition (Perrett, Harries and Loker, 1992). Bülthoff and Edelman (1992) showed an increase in recognition errors as the available view gets further away from the learned view. Making multiple views available will increase the chance that a view close to such a learned view is available, and may therefore improve observer performance. Multiple views are expected to compensate for low image resolution (Smets and Overbeeke, 1995). When presenting multiple views, an intuitive way to select a desired view seems essential to improve performance (e.g., Diner and Venema, 1989), and a low number of fixed viewpoints always seems to be suboptimal (Gaver, Sellen, Heath and Luff, 1993).

The accuracy of the coupling between the camera position and the actual viewing position of the observer may also have an effect on performance. If these positions do not match, the apparent sizes in the scene may not match the sizes as might be measured with a measuring rod (*distortion*). For static images, Lumsden (1980) showed that the apparent layout is affected by the viewing distance. Halloran (1989) showed that viewpoint displacements parallel to the display may also cause distortion. In the case of film, Meister (1966) made an analysis of places with 'acceptable' distortion, but he bases his results on geometry and not on experimental results. To prevent such distortions caused by an inappropriate viewpoint, Gibson (1971) indicated that perspective pictures should be looked at with one eye, and that a reduction screen should be placed in front of the display to hide the rest of the world. All these results implicitly assume that human perception relies strongly on the geometry of the scene. However, perspective is only one of the many available depth cues, and distortion will not depend on geometry alone. In searching the literature on distortion I found no studies where the picture's perspective is coupled to movements of the observer (as is the case with the DVWS). Chapter 7 decribes an experiment which tested whether and to what extent perspective distortion occurs in virtual window systems.

The aim of this thesis is to find the requirements for enhancing x-ray baggage inspection with the DVWS, and to propose technical solutions that fulfill the requirements. There are a lot of cross-relations between the technical possibilities, technical solutions and investigated research questions. In order to arrive at an overview, it may be useful to read the abstract at the back of this thesis.

The effects of static image quality and the number of available views on the observer performance will be investigated for transparent scenes in experiments described in Chapters 4 to 6. The effects of the accuracy of the coupling between the camera position and the actual viewing position, and other parameters that might introduce distortion, were investigated in the experiment described in Chapter 7. Because looking with one eye, as suggested by Gibson (1971), may meet resistance from baggage inspectors, the implications of looking with both eyes was also investigated in the experiment described in Chapter 6. For similar reasons, replacement of viewpoint selection via the eye position with manual viewpoint selection was investigated in the experiment described in Chapter 5.

With the results of the experiments of Chapters 4 to 7 and the technical discussion of Chapter 3, technical choices were made for a prototype system for baggage x-ray inspection with the DVWS. X-ray photos of real baggage were made in accordance with these choices. Chapter 8 describes an experimental test of the prototype system. Chapter 9 concludes with a discussion of the relevance of the results for x-ray baggage inspection, virtual window displays, Industrial Design Engineering and perceptual theories.

2

Information and task performance

The x-ray baggage inspection task is difficult, compared with situations where the inspector can open the baggage and do all the checks he wants (*natural inspection*). First, with x-ray baggage inspection the inspector only gets visual cues[1] to the baggage contents. Second, x-ray images are the usual kind of see-through image but with an unusual perspective. Nevertheless, the inspector is restricted to such x-ray images. We need to understand the implications of such a restriction on his ability to find suspicious items. We therefore need a theory about the relationship between the available depth cues in the display and human task performance (i.e. a *perceptual* theory).

Two perceptual theories are distinguished here: the direct theory (Gibson, 1986 is the basic work on direct theory) and the indirect theory (the basic literature on reconstructional indirect theory is Marr, 1982). The first section describes and compares these theories. In the second section I propose some tasks that I believe to be relevant for baggage inspection, although the baggage inspection task is not well operationalised (see Chapter 1). For these tasks, the theories suggest ways the observer can attain the required information (his *exploratory behaviour*). The third section uses the theories to predict task performance and alternative exploratory behaviour when the inspection is done via a monitor instead of natural inspection.

The direct and indirect perception theory

Perceiving is extracting information from the light from the environment. Here, I use the term information in the Gibsonian sense: *information* is what the light from the environment of the observer means to that observer. The direct and indirect theory differ in the way this information is extracted from the environment.

The direct theory (see Gibson, 1986) states that the observer interacts with the environment, e.g. by moving through it and by manipulating objects, in order to get the information needed for his task. This information is extracted directly from the light in the environment (Figure 2.1), i.e. without the need for intermediate representations (Gibson, 1986, p.147). The information is acquired in such a way that it is related to the actions the observer wants and can undertake.

The indirect theory of perception (see Marr, 1982) states that a complete spatial map (a *3D reconstruction*) is always made from the image properties indicating depth (the *cues*) in the views. Next, the required information is extracted from this reconstruction. Figure 2.2 shows the information recovery process according to the indirect perception theory.

---

[1] Without taking any theoretical stand, I use 'cue' to denote structural hints to the layout of the environment given in the structure in the light reflected by the environment. I use 'information' for properties which are relevant to the observer's task.

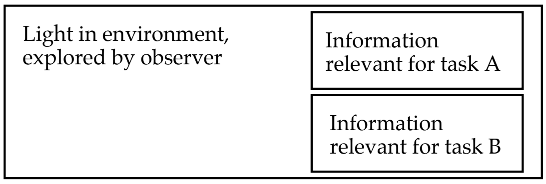


Figure 2.1. The direct theory proposes that the light in the environment contains all information necessary for the task.

Building a spatial reconstruction from the light reflected by the environment is difficult and can be done in many ways. The indirect approach has paid little attention to the question about what the important information is, and to the task dependency of the required information, but instead concentrates on the study of elementary sensations. The direct approach skips the reconstruction part entirely, thus forcing attention to the goal of perception: acquiring the required information.

In the following subsections the direct and indirect theory will be explained in more detail, and an example of a stair climbing task will illustrate the approach of both theories.

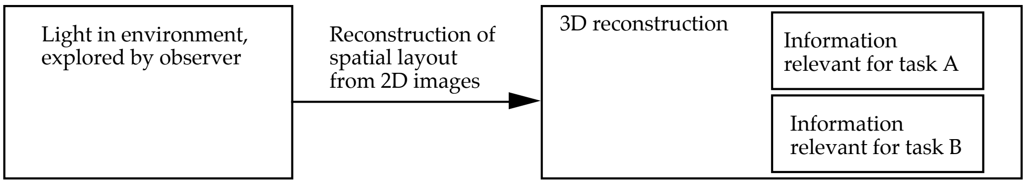


Figure 2.2. The indirect theory proposes that part of the information is lost and has to be recovered into a 3D reconstruction.

The direct theory

In this subsection the ecological approach to visual perception (Gibson, 1986) is outlined. The central idea of this theory of direct perception is that observers directly perceive the relevant information, for example their potential actions. To do this they pick up the constant and changing structures in the light reflected from all directions in that environment, as related to their own body size where relevant. Gibson (1986) postulates that observers can extract the information immediately from their interaction with their environment, and that no difficult processing of the input stimuli, like that described under the indirect theory, is required. Perception is seen as a two-way interaction between the observer and his environment: the observer manipulates his environment to acquire information and the environment elicits exploration. In this interaction process, not only the eyes but the whole body of the observer is involved in a nested way: the eye and head movements are for local exploration, and locomotion of the body for a more global exploration.

Example - stair climbing

The following example illustrates a direct specification of the ability of the observer to comfortably use a stair (Figure 2.3). This is the case when the height of the stair H is smaller than the observer's leg length L.

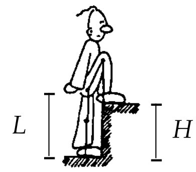


Figure 2.3. Observer stepping on a stair. H must be smaller than his leg length L.

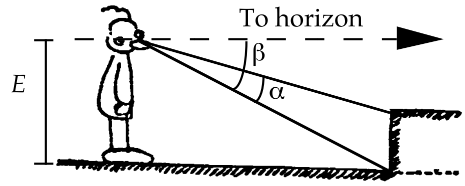


Figure 2.4. The ability to climb a stair may be detected by checking whether $\alpha/\beta < L/E$. E is the eye height of the observer (from Stappers, 1992).

The direct approach might suggest that the observer can check his ability to step on the stair by checking whether the ratio of optical angles $\alpha/\beta < C$ (Figure 2.4). C is a constant depending on the observer's eye height E and his leg length L, both of which are relatively constant during normal walking. C can be derived as in Equation (1), but this does not imply that the observer has to make this calculation. Note that the observer has to know neither the distance to the stair nor the absolute height of stair in order to judge his ability to step on it. Figure 2.5 illustrates this test for a realistic situation.

$$H < L \Leftrightarrow \frac{H}{L} < 1 \Leftrightarrow \frac{E}{L} \cdot \frac{H}{E} \approx \frac{E}{L} \cdot \frac{\alpha}{\beta} < 1 \Leftrightarrow \frac{\alpha}{\beta} < \frac{L}{E} = C \quad (1)$$

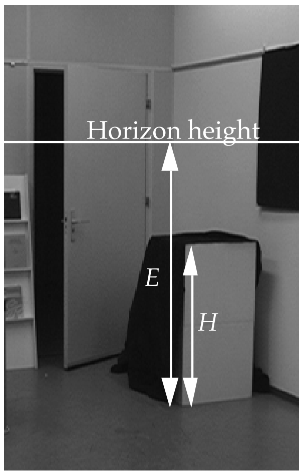


Figure 2.5. According to the direct theory, the observer may determine his ability to step on a stair from the ratio H/E. The horizon is indicated by gravitational, perspective and texture cues.

The direct approach may also use other cues, for example those discussed in the example for the indirect theory below. But the difference between the theories is that the direct theory supposes that human perception uses some of these cues directly, while the indirect theory combines all depth cues in a 3D reconstruction, from which the task-relevant information is extracted.

The indirect theory

This subsection sketches the indirect theory of visual perception as described by Marr (1982). The central idea is that the observer must construct three-dimensional representationsof objects and of the space they occupy in order to recognise and handle them.

The indirect theory distinguishes a number of steps, each with a higher level of abstraction of the forms in the image. Each higher level of abstraction is calculated from the lower levels. Each of these calculations is specified as a computational problem with input and output constraints. Next, specific algorithms solving the computational problem are proposed, and the ones matching human behaviour are selected. Finally, an attempt is made to match the algorithm against the human neural cells that are expected to implement the algorithm.

Figure 2.6 shows a more detailed plan of the steps leading to a 3D reconstructionof the scene. The first step splits the retinal image into several levels of detail, using bandpass filters. All subsequent steps are performed in parallel for these levels of detail, and finally they are joined in the 3D reconstruction. The second step is to recover the brightness edges (the zero-crossings) and blobs that are expected to contain spatial information. Thirdly, the orientation and endpoints of these edges and blobs are extracted. These oriented blobs form the raw primal sketch. A full primal sketch is formed by distinguishing these boundaries and by grouping the blobs by form. All the data from the first steps can be used by the subsequent steps, such as 'depth from shading', 'depth from perspective', 'depth from stereopsis', 'depth from contour' and so on. Some of these 'depth from X' steps are discussed below.

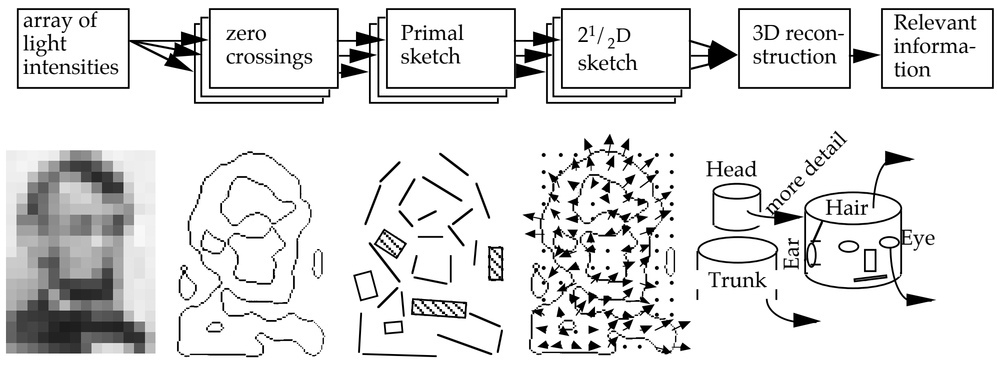


Figure 2.6. Outline of the computational model of vision as described by Marr. The image of the eyes is converted in a number of steps to a hierarchical 3D reconstruction. The first row shows these steps, the second typical structures in these steps

Most depth cues are discussed by Hochberg (1986) and Sedgwick (1986), but shorter overviews can be found in Wickens (1990) and in de Beurs (1994). The results from these 'depth from X' algorithms are compared, conflicting results may be eliminated and the results are combined to form the $2^1/_2$D sketch. The results for the several levels of detail are combined into a hierarchical *3D reconstruction*. The final step acquires the relevant information from the 3D reconstruction.

Example - stair climbing and 'depth from X'

There are numerous depth cues in the array of light intensities reaching the eyes of the observer (e.g., Sedgwick, 1986). All these cues are expected to influence the final 3D reconstruction. To get an idea of them, I will discuss image matching, depth from the difference between the images in the two eyes (*stereoscopic* cues) and structure from parallax shifts. Some of these depth from X cues will be discussed for the stair climbing example that was discussed above for the direct theory.

Matching two images

Depth cues from image pairs require that scene points are located in both images: the images have to be *matched*. Both for depth from stereoscopic cues and for depth from parallax, a matching process can be used. Assume that two images are acquired from the light reflected from the stair to be climbed (Figure 2.7). In the Figure, they are images from the left and right eye, but essentially they are images taken from different viewpoints. *L* and *R* in Figure 2.7 indicate one such matched position.

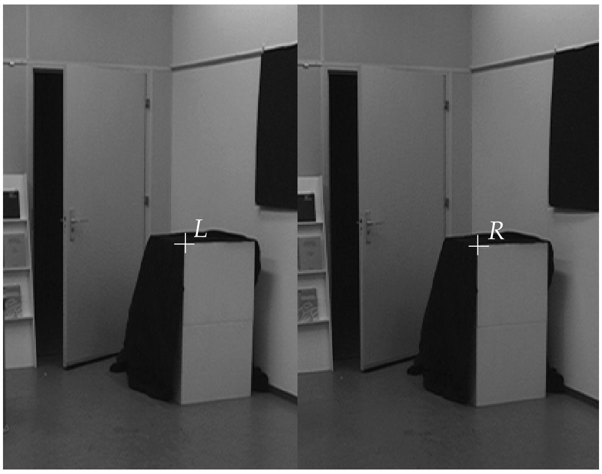


Figure 2.7. Image taken at left and right eye position. *L* and *R* are a matched position in these images, and they are expected to indicate a single position *S* in the spatial scene. This is a stereoscopic image, and can be viewed by looking at the left image with the left eye and at the right image with the right eye.

Depth from stereoscopic cues

If the views from the left and right eye have been matched, the directions λ and ρ to a point *S* in the real scene are known from both eyes (Figure 2.8). Given the constant distance between the two eyes, the intersection of direction λ and ρ uniquely specifies the position of point *S*.

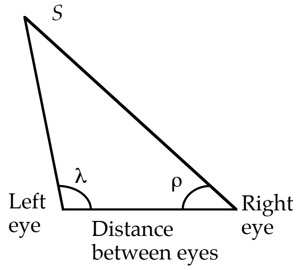


Figure 2.8. Depth reconstruction of point *S* from the direction from both eyes to that point.

Depth from parallax shifts

The term *parallax shift* indicates that two points in space move synchronously but with different angular speed relative to the observer. Such a relative movement can be caused either by a motion of the points in space (*motion* parallax) or by a movement of the observer through space (*movement* parallax).

Geometrically, there are a large number of ways to reconstruct the depth of the points from their parallax shifts, for example by using their speed, their acceleration and/or multiple views (Braunstein, Hoffman, Shapiro, Andersen and Bennett, 1987). However, Todd and Bressan (1990) show that human shape judgments typically do not involve the use of higher order temporal relations such as acceleration or comparisons of more than two views of the scene. Most theories assume that the shape of the objects in the scene does not change (the spatial scene is rigid), but it seems that human perception does not use such assumptions (e.g., Todd, Akerstrom, Reichel and Hayes, 1988), as humans are able to recognise when objects are deformed.

If the motions of two points in space can be related to each other or to the movement of the observer, the extraction of depth from parallax is relatively straightforward. For movement parallax, the same reconstruction process as with depth from stereoscopic cues can be used. Knowing the position and orientation of the two views may simplify the reconstruction. To illustrate motion parallax, Figure 2.9 shows a truck passing by an observer. Here, points having a higher angular speed are nearer to the observer, thus the relative distance between the near and far side of the truck can be determined from the angular speeds alone.

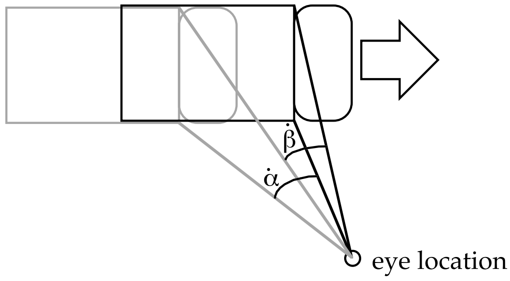


Figure 2.9. Top view of a truck passing an observer, as an example of motion parallax as a depth cue. Relative to the observer's eye location, the side of the truck near the observer has a higher angular speed than the side further away.

Next to stereoscopic cues and parallax shift cues, other cues such as contours, texture gradients and shading will also allow a spatial reconstruction of the scene. These reconstructions are not necessarily equal. Conflicts may be solved by giving each cue a weighting. Surprisingly, little research has been done to determine the relative strength of all depth cues (Hochberg, 1986). From the resulting 3D reconstruction, the observer is expected to extract the required information. In the stair climbing task, this will mean extracting the height of the stair *H* and comparing it with his leg length *L* (which is known by experience).

Comparison of the theories

Our criterion for comparing these theories is their usefulness in suggesting what cues help the inspector with his task.

The direct theory emphasises that explorative behaviour is guided by the task of the observer. It does not require a complete spatial reconstruction, which seems unnecessary and not done by human observers for most tasks such as stair climbing, picking up a cup or opening a window. Therefore its suggestions about the required explorative actions seems more to the point than the indirect theory. Thus, the direct theory is especially useful for estimating the effectiveness of multiple or particular views.

The indirect theory has a much longer tradition, and therefore it is much further developed. Mathematical models exist to describe how depth may be reconstructed from a single view or multiple views. Detection theory, which usually assumes human perception to be as suggested by the indirect theory, can predict effects of technical limitations such as a limited frame rate of displays and an image consisting of pixels. Furthermore, the indirect theory allows us to do predictions of the kind and size of distortions that may appear in virtual window displays. For making hypotheses, such precise predictions are more useful than the vague suggestions made by the direct theory.

The direct and indirect theory complement each other. The direct theory can be used to predict the explorative behaviour of the observer, thus giving us clues about how to design an intuitive and effective user interface. The next section will use both theories to determine reasonable requirements for a number of tasks that seem relevant for baggage inspection.

Theoretical usefulness of multiple and particular views

This section discusses how multiple views may be useful for inspection tasks, according to the theories. Between the extreme approaches of the direct and the indirect theory, less extreme theories exist. These theories are of interest to us because they indicate other reasons why multiple views may improve observer performance.

According to the direct theory, in a natural situation the observer interacts with his environment, and he is able to acquire those views that are relevant for his task. The required information is extracted directly from the views, without building a spatial reconstruction. Because the required information is task dependent, we have to investigate tasks related to the baggage inspection task in order to find out what views may be relevant. The indirect theory would suggest that multiple views will enhance the apparent depth as compared to a single view because of the extra depth cues given by parallax shifts, but it does not suggest particular useful views. Geometrically, depth can be reconstructed from 2 to 5 views, depending on the number of matched points in the views and additional constraints on the viewpoints (Braunstein et al., 1987). Humans are apparently unable to do such a perfect geometric reconstruction, as Braunstein et al. (1987) showed that, for same or different judgements of a few dots moving through space, human performance may improve even with a larger number of available views (see also Chapter 6).

Two other, less general, theories are related to the number of available views: the multiple viewpoint theory and the geon theory. These theories are neither completely direct, as they involve some processing of the visual stimulus, nor completely indirect, as they do not suggest that human perception builds a complete spatial reconstruction of the viewed scene. The multiple viewpoint theory assumes that humans learn several views of an object (Bülthoff and Edelman, 1992). Humans might compare an available view with learned or with prototypical views, possibly doing some mental image processing to compare the learned and the available view. There exist some clues about the properties of these learned views. For example Perrett, Harries and Looker (1992) had observers inspect a widget (a 3D object resembling a photo camera). They found that orthogonal views are inspected more frequently and brought more easily to mind than other views. The multiple viewpoint theory implies that presenting a view close to the learned view of objects might improve observer performance. Therefore, the observer may recognise the object faster and more precisely if a view close to the learned view is available, as the observer does not need to use mental image processing for recognising.

Geon theory assumes that object recognition is done by acquiring some characteristic viewpoint-independent features (geons) from the object, and matching the reconstructed geons with the geons of known objects stored in a database (Biederman, 1987; Biederman and Gerhardstein, 1995). A geon is a volume that might be made by extruding a cross-section over a straight or curved path (usually called ‘generalised cone’). Objects usually consist of several geons, and the boundaries of these geons can be found from cues in a view, such as parallel line segments, symmetry and terminations of line segments in a common point. It is claimed that the geons can be extracted from 2D cues and that they are largely viewpoint independent. In other words, this theory suggests that multiple views are useful only insofar as they make new geons visible. The geons will depend on the available stimuli, and for x-ray baggage inspection it is not clear what combinations of geons might be relevant for distinguishing a bomb from the other contents of a bag. As discussed before, there is no finite list of ‘possible bombs’, but in practice inspectors might use the list of bombs seen at their training sessions.

Requirements for relevant tasks

This section discusses five tasks that seem to be relevant for baggage inspection. It is assumed that the inspector can explore the object as with natural inspection. The relevant cues for these tasks and the required visual cues suggested by the discussed theories are described. These tasks will return in the experiments later in this thesis.

A problem with selecting tasks is the absence of knowledge about what baggage inspectors are actually looking for (see Chapter 1). In my research I investigated what I thought to be plausible subtasks: detection of specific shapes (task 1 and 2), specific relations (task 3 and 4) and specific sizes (task 5). The relevance of these tasks for x-ray baggage inspection was agreed on during discussion with experts on x-ray baggage inspection and visual perception researchers.

Task 1 - detecting sharp edges

The first task is detecting objects with a sharp edge. This task is relevant for baggage inspection as sharp objects, such as knives, are suspicious items [2]. In baggage inspection, such suspicious objects will usually be recognised by similarity with familiar suspicious items, but I investigated unfamiliar sharp objects like the one in Figure 2.10. The objects were made of transparent potting resin, to match the x-ray inspection task.

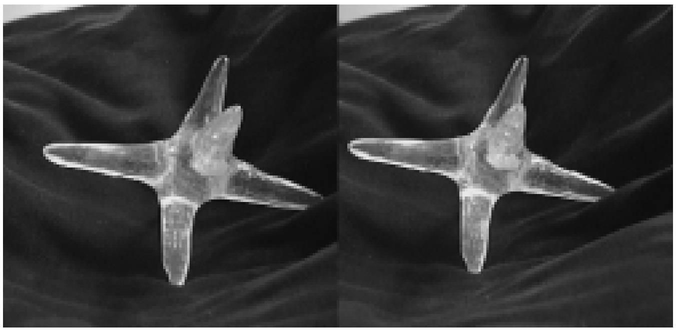


Figure 2.10. Stereoscopic image of an object with possibly sharp edges.

Sharpness cues are most prominent in a view orthogonal to a sharp edge. A view perpendicular to the sharp edge gives direct visibility of its sharpness, and thus makes a judgement of the sharpness easy. For example in Figure 2.10, the sharpness of points with their angle parallel to the paper is clearly visible, but for the point protruding from the paper its sharpness is less clear. So it can be expected that the exploratory behaviour of the observer is such that suspicious edges are viewed perpendicularly. A single, optimal view of a suspicious edge suffices, independent of the size, distance or 3D shape of the edge.

This expected exploratory behaviour suggests that the task will be more difficult if no view perpendicular to the edge exists. For example the object in Figure 2.11 has no such view, as the edge is occluded at the required viewpoint. In such a case, the observer may

[2] This task may be less relevant to the inspection of hold baggage, where sharp objects are not allowed. However, I was unaware of the differences in the inspection of hand and hold baggage, so that the decision to concentrate on hold baggage was made after an investigation of this task (see also Chapter 1, 'Previous work').

have to revert to another strategy to determine the sharpness of the edge. There are many other cues about the sharpness of the edge, for example stereoscopic, parallax and contour cues can be used. These cues have no specific difficulties with embedded edges like Figure 2.11. This task is investigated in the experiment described in Chapter 4.

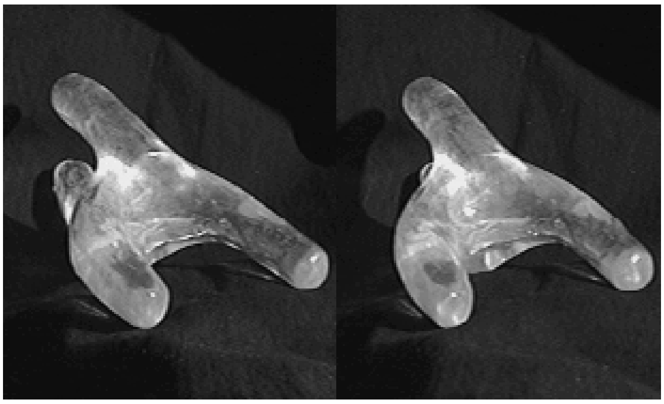


Figure 2.11. Stereoscopic image of an object with a sharp edge in its inner curve.

Task 2 - detecting wires

The second task is wire detection. This task is important for baggage inspection: wires may connect the typical parts of a bomb (the detonator, the battery and the delay mechanism). Chapter 5 investigates the effect of the number of available views on the performance of this task. The wires used there had a diameter of 0.3 mm . For an average viewing distance of 50 cm and assuming an acuity of 1 arcmin (e.g., see Olzak and Thomas, 1975), wires down to a diameter of 0.1 mm should be visible. Only if the wire is straight and the viewpoint is near the line through the wire, the wire will occupy a small visual angle, and this may reduce its visibility. If other objects are present, they may occlude the wire. This thesis deals with x-ray inspection, in which objects are transparent and therefore cannot occlude other objects. But if a contour of an object coincides with the wire, the wire may be camouflaged by that contour (Nodine and Kundel, 1987; Vyborny, 1997). In any case, two perpendicular views will be sufficient to detect the wire and to resolve most camouflage effects.

For wire detection, contour cues seem sufficient. The indirect theory may suggest that a complete reconstruction is still required, but to make such a reconstruction seems to involve much more work than to judge whether a line in the picture connects the objects.

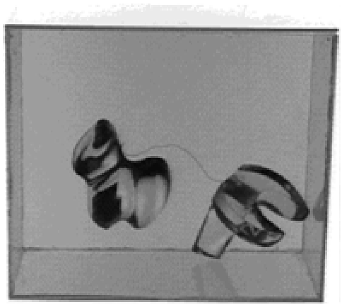


Figure 2.12. Wire connecting two
objects in a box.

Task 3 - detecting connecting wires

The third task relevant for baggage inspection is checking whether a wire connects objects. A wire in a suitcase makes the suitcase suspicious only if the wire connects to suspicious objects. There is an area from which it can be seen directly whether there is a space between the wire and the object (Figure 2.13). This area is in the plane touching the object on the place where the wire seems to hit the object, and preferably perpendicular to the local direction of the wire. As with the sharp edges, these views may be occluded, but my research deals with transparent objects. This task is also investigated in Chapter 5.

Stereoscopic cues and parallax shift cues can also be used to detect a connection between the object and the wire. The direct theory might suggest that parallax shifts directly group the parts belonging to one wire. For example, two stacked transparent sheets filled with a random dot pattern are perceived as a single random dot pattern, but the dot patterns are perceived as two separate groups as soon as the layers are moved independently (e.g., Metzger, 1975). For the indirect theory, an interesting question might be whether the 3D reconstruction is fine enough to contain such thin wires.

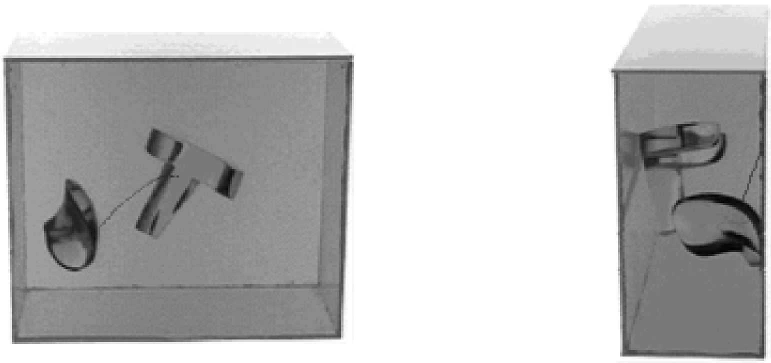


Figure 2.13. Front view of wire (left), possibly connecting to the objects. From an appropriate viewpoint (right) the wire is found not to touch the object on the right.

Task 4 - tracing a wire through a knot

As discussed above for task 2 and task 3, tracing a wire is relevant for baggage inspection. However, x-ray images of real baggage are more complex than two objects and a wire in the box. Compare Figure 1.1 with Figure 2.12 to get an impression of the difference in complexity, and the suitcase of Figure 1.1 still contains less items than average baggage. Therefore a more complex scene was created by making a knot of three wires (Figure 2.14). Observers were asked to trace one of the wires through the knot from one of the top spheres and indicate its lower end. An experiment with this task is described in Chapter 6.

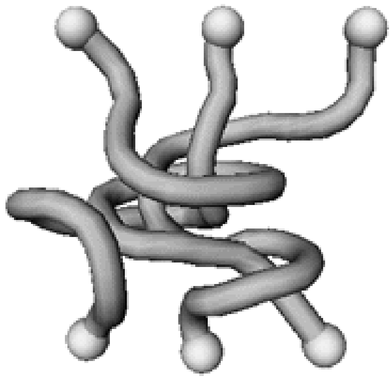


Figure 2.14. For tracing a knot through a wire, one view is sufficient.

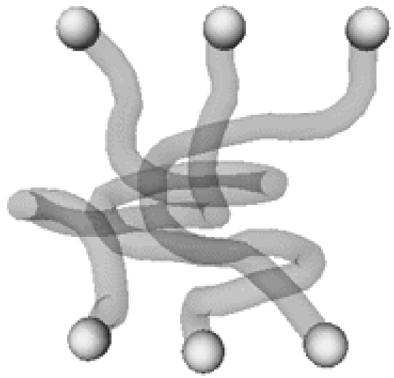


Figure 2.15. If the knot is transparent, additional depth cues are required.

If the wires were non-transparent, wires closer to the viewer would hide parts of other wires further away (*occlusion*). Occlusion would be an important cue for this task, but in x-ray images objects are transparent, thus occlusion cues are absent. Without occlusion cues other depth cues are required in order to be able to follow a wire through a knot (Figure 2.15).

Stereoscopic, parallax and contour cues become more important in the absence of occlusion cues. Perspective cues may be of little help, but the perspective effects are very small and the scene lacks vanishing points. Again, the direct approach might suggest that movement parallax directly groups the parts belonging to one wire, but the grouping in the knot tracing task is not as straightforward as the task with the superimposed layers with random dots. The wire can be followed in a single view by means of contour cues. Where overlapping with another wire occurs, the contour cues are not sufficient to follow the wire. Here, stereoscopic viewing can specify directly whether there is a difference between the depth of the wires. Parallax may also provide a direct cue about depth differences. Matching is easy, as it can use the sharp contours of the wires.

Task 5 - matching bump heights

For this task, observers adjust the height of bumps in a landscape (Figure 2.16), to match the height of the bumps of a second landscape that has bumps on the same position as the adjusted landscape. This task is not directly relevant for baggage inspection, as suspicious items exist in any size. However, this task is relevant more generally for virtual window displays. An important difference with the previous tasks is that the judgement of the observers is continuous, where previous tasks only allowed discrete choices. Chapter 7 describes an experiment with this task.

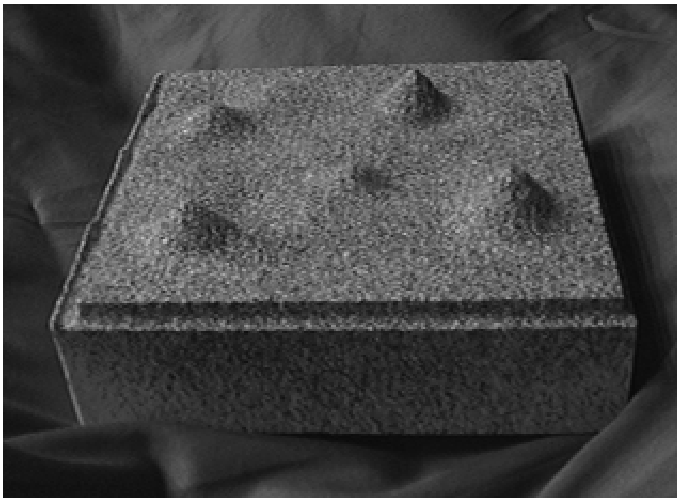


Figure 2.16. **See also figures on right cover flap.** Bumps in a landscape. Subjects had to adjust the height of the bumps to match a second landscape that had bumps in the same positions.

From a viewpoint perpendicular to a bump, the heights can be matched directly from visual angle H or with the angle ratio H/W (Figure 2.17). Furthermore, at such a viewing position the top of the bumps will no longer be camouflaged by the background. If the observer uses H directly, he must take care to compare the bumps from the same distance, as H will decrease with observer distance. In the H/W ratio this distance effect is divided out. Bump heights could also be matched directly by comparing the shadows. Thus, I expect exploratory behaviour to try to attain viewpoints perpendicular to the bumps.

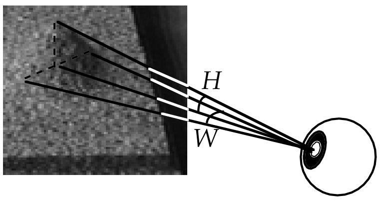


Figure 2.17. From a viewpoint perpendicular to the bumps, the bump heights can be compared by comparing the visual angle H or the ratio H/W.

Alternatively, the observer may view slightly over the top of the bumps, to compare the movement parallax of the texture of the plane behind the tops of the bumps relative to that top. To do this, he should compare the bump heights from the same viewing height. He may control his viewing height by checking the ratio of the visual angles of the width and the height of the entire landscape.

Stereoscopic, parallax and contour cues can also be used. A view perpendicular to the bumps is not specifically advantageous for these cues. Perspective cues may be of limited use as there are few parallel lines in the landscapes, but a perpendicular view is not useful as it lacks vanishing points.

I have discussed five tasks that seem to be relevant for x-ray baggage inspection. The required information, the depth cues that can be used most efectively to obtain this information and the resulting explorative behaviour were discussed. What happens with these cues, and how they may be compensated with other cues when the natural inspection is replaced with inspection via a monitor will be discussed in the next section.

Cues via a monitor

In current baggage inspection, usually a single x-ray view of the baggage is displayed on a monitor. This causes the observer to receive fewer visual cues, while the cues he does receive are of lower quality than with natural inspection, and the cues in the image may be inconsistent in themselves or with the environment around the monitor. The explorative actions are restricted: only a single view is available. To sum up the most important restrictions and inconsistencies of inspection via a monitor as compared with natural inspection (see also Edgar and Bex, 1995):

1. The spatial and colour resolution of the image will be lower (see Chapter 4 and 5).
2. The contrast is lower.
3. Noise will be added to the image.
4. There is a visible break between the environment in the monitor and the environment around it. For example, the horizon in the x-ray scan and the horizon in the world the monitor is in will usually not match.
5. The observer can not manipulate the objects.
6. The observer can not look around the baggage: parallax shifts and multiple views are absent (see Chapter 4, 5 and 8).
7. Stereoscopic cues may be in conflict with other cues, as the observer is viewing with both eyes while no stereoscopic image is provided (see Chapter 6 and 7).
8. Perspective cues may conflict with other cues, as the viewing position may be geometrically inequivalent to the recording position (see Chapter 7).
9. Perspective cues may be unusual, as usual x-ray scanners give perspective effects that are never encountered in natural perception (this is discussed in Chapter 3).
10. Accomodation cues due to focal distance indicate the flatness of the display: for all objects, the focus distance of the eye's lens is at screen distance.

The absence of stereoscopic and parallax cues is expected to have a large impact on the observer's task performance for the tasks discussed above. According to the indirect theory, absence of perspective and occlusion cues will hinder an accurate 3D reconstruction or even make it impossible. A more serious drawback, according to the direct approach, is that a static view of the scene does not allow the observer to explore the scene by taking other viewpoints. The reduction of the spatial and colour resolution, the extra noise and the lowered contrast are less important, although they should have a minimal value, depending on the smallest part in the baggage that should be detected or recognised.

Improving the shortcomings

Most of these shortcomings can be partially or wholly resolved. An x-ray scanner providing a higher resolution and number of grey levels can be developed. The visible break between the environment in the monitor and the environment around the monitor can be lowered by placing a reduction screen around the monitor, thus hiding part of the environment. Objects may be manipulated via a turning knob or other manipulation tools. Stereoscopic cues can be provided to the observer, and the Delft Virtual Window System can be used to recover the movement parallax and lookaround possibilities. Technical solutions may be developed to make x-ray images with perspective information appear less distorted.

The indirect theory would suggest enhancing the cues that contain relevant information. This can be done by building a scanner that provides stereoscopic images or parallax, or gives a higher contrast or a higher spatial or colour resolution in the image. If the available views are not stereoscopic but the observer still looks at the views with both eyes, stereoscopic cues only tell the observer about the flatness of the display. If these conflicting cues are not eliminated, the reconstructed depth will be flattened because of this indication of flatness. Alternatively, the direct theory might suggest that observers can eliminate the stereoscopic cues (for example by closing one eye or by using the images in both eyes to construct a single image with reduced noise), depending on their task, as it is known that they can judge both distances suggested by photographs and their distances to the photographs (Gibson, 1986).

The direct theory would suggest providing just those cues that are necessary to get the required information. For most tasks discussed above an appropriate view would suffice. Such an optimal view usually depends on the orientation of the objects, and this orientation is unknown before the x-ray photographs have been taken. Providing movement parallax would enable the observer to choose a useful viewpoint himself, as in natural inspection. Besides the ability to choose a useful view, multiple views may compensate for a low image resolution or a low number of grey levels, for example a frozen television image looks far worse than moving television images.

This thesis concentrates on providing movement parallax to improve inspector performance with the Delft Virtual Window System. Only a limited number of views can be made available (see Figure 1.3c), as each x-ray photo will expose the baggage to an x-ray dose. The rest of this section discusses the effects of a restricted number of available views on task performance in the light of the theories.

Task 1 - detecting sharp edges

As discussed above, for detecting sharp edges the exploratory behaviour of the observer will be such that suspicious edges are viewed perpendicularly. If no such viewpoint is available his performance will drop. In the absence of such a viewpoint, the observer is forced to adopt other ways of estimating the sharpness of the edge, for example parallax shift cues and contour cues. With low resolutions, sharp edges may appear rounded, but this may be compensated by viewpoint multiplicity (see Chapter 5). Therefore, with the DVWS this task may elicit observer movements both in order to find a useful view and in order to compensate for low resolution.

Task 2 - detecting wires

If the resolution and number of grey levels of the views is reduced, wires may become invisible due to noise and numerical rounding (Figure 2.18). For example, with 4 grey levels and against a background of a grey level of 74%, wires have to screen 34% of the remaining light in order to be displayed, while they have to screen only 8% with 16 grey levels. For wires in front of darker backgrounds these percentages will be even higher.

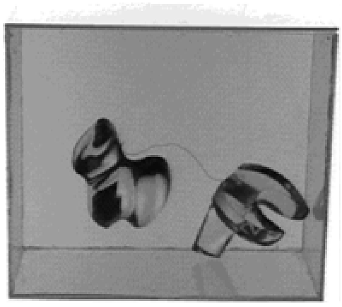


Figure 2.18a. A wire is visible when enough grey levels are displayed.

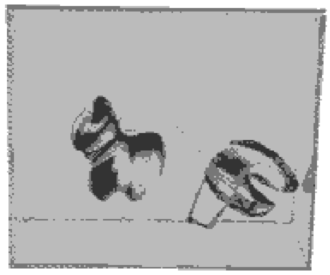


Figure 2.18b. With reduced number of grey levels, the wire may become invisible.

More specifically for baggage inspection, if a wire is scanned with 1 sensor per mm^2, and the wire has a diameter of 0.1 mm, the wire occludes at most 10% of the sensor. If the wire screened the light perfectly, this would reduce the light falling on the sensor with 10%, giving a contrast ratio of 0.1 (Olzak and Thomas, 1975). If such a wire were displayed unscaled at a distance of 50 cm, the threshold for visibility lies at a contrast of about 0.03. However, with real baggage inspection the wire does not screen the x-rays perfectly, the image is scaled down, the background is cluttered and viewing conditions are not optimal, thus visibility will be marginal.

The extra noise, the low number of grey levels and the resolution may be compensated by providing multiple views. If we assume that some kind of method averaging over several views forms the basis of this compensation, a compensation can be expected to work only if the angle between these views is small; otherwise the views may not be averaged by the observer in a simple way. Alternatively, the background colour may alternate over different views, for example because an object shifts behind the wire. This may make the wire visible. In fact, some noise in the image, in this case a shifting object, can increase the visibility of the wire.

A 3D reconstruction required according to the indirect theory may be difficult. With a low number of grey levels, matching of a small number of pixels originating from a wire (Figure 2.18b) in adjacent views may fail, and consequently depth from parallax shifts may fail. As pixels get larger, the 3D reconstruction will become less precise. Contour cues may be hard to use since the contours are incomplete in separate views, but this problem may be overcome if the curvature of the wire is small. Because the views are not presented stereoscopically, observers using both eyes will receive stereoscopic cues specifying the flatness of the display. All these problems distort the 3D reconstruction, and thus may degrade observer performance.

Task 3 - detecting connections

For detecting connections, the wire has to be visible in the first place, giving similar basic requirements as for Task 2. For detecting connections, the analysis for the natural inspection case (see under 'Requirements for relevant tasks') suggested that viewpoints within a specific area are especially useful. But with inspection via a restricted DVWS, such a view may be not available, causing a drop in performance. Alternatively, motion parallax may be used to detect connections, but as with task 2, matching may pose problems. Because it is not enough to see only parts of the wire for this task, the requirements for the resolution, number of grey levels and for a 3D reconstruction (if humans make such a reconstruction) will be higher than for Task 2.

Task 4 - tracing a wire through a knot

For the knot-tracing task, parallax shifts are expected to be an essential depth cue in the absence of stereoscopic and occlusion cues. If only a finite number of views are available, the angle between the views (see Figure 1.3c) may cause matching problems. For example, the wires seen in Figure 2.19a are difficult to relate to the wires in Figure 2.19b. The direct theory might argue in such a case that it is difficult to pick up the constant structures from such views with a large angular distance between them.

Decreased image resolution will make it more difficult to use contour and perspective cues than with natural inspection. Furthermore, the indirect theory may indicate that the 3D reconstruction will be flattened if the observers use both eyes while the views are not stereoscopic. This may lower the distance between the wires in the reconstructed knot, thus making the task more difficult.

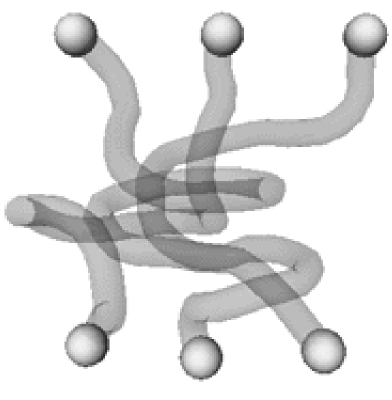


Figure 2.19a. Front view of a transparent knot.

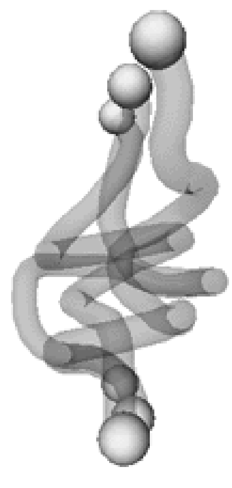


Figure 2.19b. Side view of the same knot. It is difficult to match the two views.

Task 5 - matching bump heights

Instead of presenting two real landscapes with bumps, one of the landscapes was presented on a monitor. Observers were asked to adjust the height of the displayed bumps to match the real bump. This task was designed to investigate distortions in virtual window displays caused by cues to the flatness of the display and by displaying objects subtending visual angles not matching those of a real scene (*geometric inequivalence*).

Geometric inequivalence may be caused by inaccurate measurement of the viewing position of the observer, the way the camera is coupled to the movements of the observer, and by delays in the system. This issue is complex, and is discussed in length in Chapter 7.

The distortions are expected to increase as the observer moves away from the middle of the image. Therefore, observers had to be induced to take such viewpoints. As the direct theory suggests that observers will prefer a viewpoint perpendicular to a bump, observers could be elicited to take extreme viewpoints relative to the display by placing the bumps so that they protruded out of the display (Figure 2.20). The bumps and the plane between the bumps were sprayed with the same fine texture, and no shadows were added in the virtual scene. Therefore, the bumps were visible against their background only when the observer moved. In this way observer movements were elicited by preventing observers from matching the shadow and contour cues.

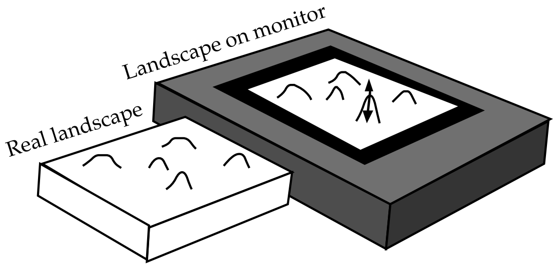


Figure 2.20: The plane between the bumps is placed parallel to the monitor display to induce the observer to take viewpoints that are oblique relative to the display.

The analysis of exploratory behaviour in natural inspection suggested that observers will prefer views perpendicular to a bump, but with this setup observers are unable to see the monitor display in these positions. In this case, a position slightly above the plane between the bumps can also be used. Parallax and texture cues alone may be sufficient for a spatial reconstruction, and therefore the direct and indirect theory may disagree about the need for views perpendicular to a bump.

Conclusions

In natural inspection the inspector can do all the checks he wants, from any viewpoint he likes and at any level of detail. When inspection is done via x-ray images displayed on a monitor with the DVWS, the number of available views is restricted, and image quality is degraded. Both the direct and the indirect theory were used to predict the consequences of such restrictions on task performance for five tasks relevant to baggage inspection (Chapter 4-7).

Both the direct and the indirect theory were used, in order to determine the usefulness of multiple and particular views. A rough estimation was made of the image resolution and number of grey levels required to make wires visible. It is expected that multiple views can compensate for both a low spatial resolution and a low number of grey levels. However, as far as I know no precise data exists describing the compensatory effect of the number of available views on a low static image quality. I described how specific viewpoints may be especially useful, depending on the task and the layout of the scene. If such a specific viewpoint is not available, observer performance may drop as observers lack important cues. Distortions may be caused by cues to the flatness of the display and by a geometrically inequivalent viewpoint, and such distortions may also diminish observer performance. The experiments of Chapters 4 to 7 test the expected effects for the tasks that were described in this section.

3

Shooting multiple x-ray images efficiently

This chapter works towards a technical concept for shooting multiple x-ray images efficiently. The plan is to improve the ability of the baggage inspector to detect suspicious items in the baggage, by giving him a spatial impression of the baggage and by enabling him to look around occluding objects. To do this, the Delft Virtual Window System (see Chapter 1) will be used. The DVWS works by updating the x-ray image on the display to match the viewing position of the observer. With current technologies it is hardly possible to acquire the appropriate x-ray image in real time at the moment the observer moves. Therefore a number of available views have to be recorded prior to the presentation via a virtual window display. How can we take useful multiple images of a suitcase efficiently, and what do potential manufacturers and users expect from such a scanner?

To answer the question about how to shoot x-ray images, we need to understand the recording techniques for shooting x-ray images. Next to x-ray sources and x-ray sensors, x-ray mirrors may be useful for shooting multiple x-ray images. The first section discusses these components.

It can be expected that the manufacturer wants to integrate his own know-how in the concept that I will suggest. In order to make this feasible, the state of the art in baggage scanners and their working principle has to be understood. To find out what potential users and manufacturers will expect from an x-ray scanner, and to find out how efficient the implementation of the concept really has to be, the state of the art in baggage x-ray inspection will be discussed in section two.

The most difficult question is to determine how many and which images the inspector needs. All experiments described in this thesis deal with this question. Reasonable answers can be given about what views are useful, and about what the image quality should be. However, no conclusive answer can be given about the required perspective properties of the images. This question becomes urgent here, as x-ray images have a very unusual perspective. Therefore, the third section mainly sketches the possible perspective combinations, in order to make a reasonable choice for the perspective.

Given these answers, a number of concepts are proposed in the last section. Two concepts are worked out in more detail, optimising scanning speed, price, and operational safety.

X-Ray components

The components used for x-ray image generation have to be understood to find an optimal solution for shooting a series of x-ray images. Generation and detection of x-rays and x-ray mirrors are discussed.

Generating and sensing x-rays

An x-ray source consists of a cathode and an anode with a high voltage (140 kV is typical for baggage inspection) between them (Figure 3.1). The cathode is heated so that electrons leave the cathode. They are accelerated by the high voltage between the cathode and the anode, and crash into the anode. In the anode, 99% percent of the energy the electrons have at the time of impact is converted into heat. Therefore the anode has to be cooled, usually with oil. About 1% of the energy is turned into photons. Given the energy of the electrons of 140kV, the shortest wavelength of these photons is about 9 pm (see Schweers and Vianen, 1982).

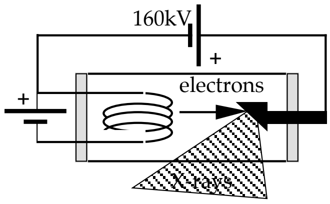


Figure 3.1. Diagram of an X-ray source. Electrons are emitted from the cathode and crash into the anode. This causes photons with high energy, i.e. x-rays, to be emitted from the anode.

The x-rays spread in all directions. In order to get a narrow x-ray beam, the source is shielded with lead, and a slit in the lead gives the x-ray beam the desired fan-like shape. The lead and the cooling make an x-ray source expensive and heavy. Typically, an x-ray source costs about NLG 20,000 (USD 10,000).

The x-rays are sent through the baggage. The higher the density of the baggage contents, and the more material it contains, the more x-rays it will absorb and scatter. Furthermore, x-rays with a higher energy are able to penetrate denser material than x-rays with lower energy. The scattering behaviour is characteristic to the material, and can thus be used to identify the materials in the baggage (Hughes, 1989).

The x-rays that remain after their journey through the material have to be made visible. X-rays can be converted to visible light if they hit zinc sulphide: the zinc sulphide will emit a green light when hit. Alternatively, scintillation crystals are used to convert the x-rays to light. These scintillation crystals have a higher efficiency than zinc sulphide. The visible light can be converted to a voltage difference with a photo diode. To optimise the signal-to-noise ratio of such a two-step detector that converts the amount of x-rays to a voltage difference, the scintillation crystal and the photo diode are usually integrated in one electronic part: an *x-ray sensor*. Typically one x-ray sensor costs about NLG 10 (USD 5). Usually a row of such x-ray sensors is used (a *sensor line*), to scan one x-ray line at once. Thus a typical sensor line with 576 sensors will cost about NLG 6,000 (USD 3,000).

X-ray mirrors

Reflecting x-rays may be a way to multiply the number of virtual x-ray sources and/or sensors without requiring additional real sources and sensors. Such x-ray mirrors consist of a large number (typically 150) of layers of two alternating materials, one with a high and one with a low density. The distance between two of these layers is typically 0.1 - 20 nm (Figure 3.2), and should match the wavelength of the x-rays to be reflected.

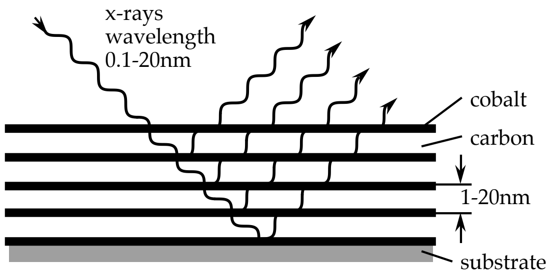


Figure 3.2. An x-ray mirror consists of a large number of layers, each reflecting a small fraction of incoming x-rays. The distance between the layers is related to the wavelength and the angle of incidence of the x-rays. Because of the multiple layers the reflected x-rays will be blurred slightly.

Each layer reflects only a small percentage of x-rays, but combined together these mirrors have an efficiency of about 10%. This may be sufficient for use in x-ray scanners, as the loss arising in the mirrors can be compensated by generating a higher x-ray dose. This higher dose should not be sent through the baggage, thus the x-rays should be reflected before they are sent through the baggage.

X-ray mirrors have a number of properties that may be critical when they are used in a baggage scanner. First, the amount of reflection is not constant for different angles of incidence. Second, such mirrors are still expensive, as the layers have to be extremely highly polished. Accurate prices for x-ray mirrors are not known, because they have to be custom-made. Third, I have no knowledge of whether x-ray mirrors exist which are capable of reflecting the high energy x-rays used with baggage inspection (typically, 0.01 nm).

In conclusion, x-ray mirrors may be an interesting way to multiply the number of x-ray sources and sensors, which is a requirement for multiple-view x-ray baggage inspection. However, it is not clear whether x-ray mirrors with properties suited for our purpose exist. The next section discusses the various ways of building a scanner given the components, and sketches the state of the art in x-ray scanning.

X-ray scanners- state of the art

In order to know what potential users and manufacturers will expect from an x-ray scanner, and to understand how efficient the concept has to be, this section discusses the state of the art in x-ray baggage inspection. Important points are scanning speed, image resolution , image processing, the amount of x-rays to which the baggage is exposed and the availability of multiple views.

Instead of discussing these points separately, I will group the machines into three types of sensor mechanism: (1) those using a fluorescent screen, (2) those using a sensor line and (3) Computer Tomography (CT) scanners (Figure 3.3).

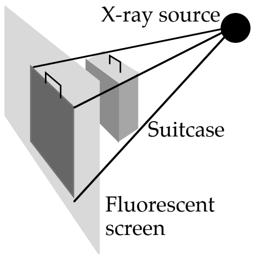


Figure 3.3a. A Fluorescent screen can convert the x-rays directly into a visible image. Because the x-rays are partly absorbed by the baggage, the image is the shadow of the suitcase.

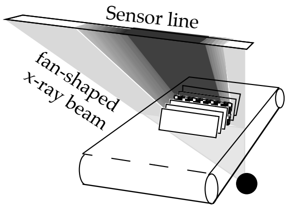


Figure 3.3b. Scanners with a sensor line scan the baggage slice by slice. Each slice is projected with point perspective, while all slices are scanned in a parallel direction.

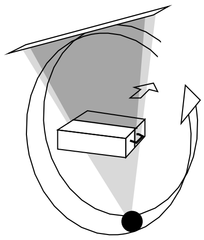


Figure 3.3c. CT scanners scan a large number of slices, using a source and sensor line rotating around the baggage. From these images a 3D reconstruction is made.

In the discussion of these three types of machines, a large number of scanners will be mentioned. Table 3.1 gives an overview of them (see also Macrae and Taverna, 1990). In the text, I will refer to the machine name.

Table 3.1. The x-ray scanners discussed in this section.

Machine name	Manufacturer	Sensors	Article	Image processing
Heimann GPA8014	Heimann Systems GmbH	fluorescent screen	Linkenbach & Stein (1981)	
SDS400/P	Isorad	fluorescent screen	Isorad (1987)	
Heimann 6040, 9080	Heimann Systems GmbH	sensor line	Heimann (1987)	Pseudocolouring for material identification
Heimann 10050EDS	Heimann Systems GmbH	sensor line	Heimann (1996)	As Heimann 6040; markers indicate suspicious items
Controllix-Vision	Europscan	sensor line	Europscan (1993)	Pseudocolouring for material identification
Vivid	Vivid Technologies Inc.	sensor line	Vivid (1990)	Pseudocolouring for explosives identification
Aisys 370B	Magal Security Systems	sensor line	Magal (1994)	Markers indicate suspicious items
Z-Scan	EG&G Astrophysics	sensor line	EG&G Astrophysics (1996)	Pseudocolouring for material identification; Gives side and bottom view of suitcase
Scanray	Scanray	double sensor line	Evans, Godber & Robinson (1994)	
CTX5000	InVision Technologies	CT scan	Invision (1997)	Pseudocolouring for detonator and explosives identification

Scanners with a fluorescent screen

One way of making an x-ray image of a suitcase is by exposing the entire suitcase to x-rays and by converting the x-rays that went through the baggage to visible light with a fluorescent screen. Figure 3.4 shows an SDS 400/P, a commercial system which works this way. A glass plate containing lead has to be present between the fluorescent screen and the inspector, to screen him from radiation. The images on the fluorescent screen are similar to a normal perspective photograph, as if the x-ray source replaces the camera lens, and the fluorescent screen replaces the sensitive plate that is normally behind the lens. The SDS400/P allows the inspector multidirectional real-time viewing by rotation of the inspected object with a knob. Although scanners with a fluorescent screen are very fast in generating an x-ray image, none of the commercial machines with such a screen are designed for fast throughput.

An image on a fluorescent screen has a low light intensity due to the low amount of radiation that can be sent through the baggage. Furthermore, because the image will decay if the x-ray beam is turned off, such machines will require a constant exposure of the baggage to x-ray. This continuous exposure may give the baggage an unacceptable dose of x-rays. To avoid continuous exposure and to amplify the brightness of the image, some systems record the image on the fluorescent screen with a light-sensitive video camera, whose image is stored in computer memory. Figure 3.5 shows such a system, the Heimann GPA8014. Usually the video camera views the fluorescent screen via a mirror, to prevent x-rays that went through the fluorescent screen from hitting the video camera. Finally, the image is displayed on a monitor display (Linkenbach and Stein, 1981).

Such a system with fluorescent screen, mirror and camera has two disadvantages. First, it requires a large empty space between the fluorescent screen, the mirror and the camera, and space is expensive. Second, it is quite an indirect way of displaying an x-ray image, and will introduce additional noise and blur.

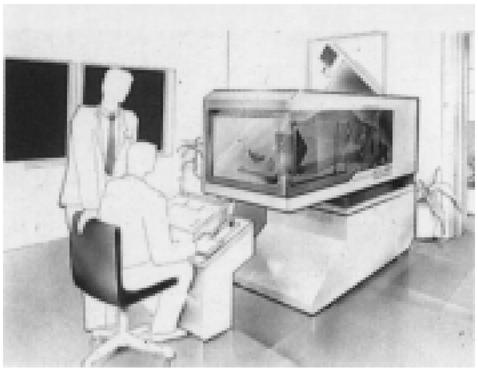


Figure 3.4. The SDS400/P, a machine using a fluorescent screen. Such machines give images with very low light intensity, and continuously expose the suitcase to x-rays.

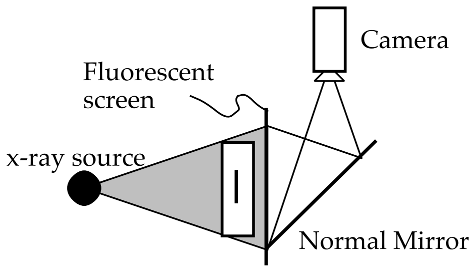


Figure 3.5. To improve light intensity and to shorten exposure time, the image can be recorded with a light-sensitive TV camera, stored in computer memory and displayed via a monitor.

Scanners with a sensor line

In systems with a fluorescent screen, noise in the image is caused by the scattering of the x-rays by the baggage and the low contrast of the fluorescent screen. Image contrast can be improved and scanning noise reduced by scanning the baggage slice by slice in stead of in one shot (Kotowski, 1986), as shown in Figure 3.3b. To accomplish this, a thin fan-shaped x-ray beam and a sensor line are required. The baggage is scanned slice by slice, and each scanned slice gives one line for the x-ray image. A conveyor belt moves the baggage through the scanner. The thinner the x-ray beam the sharper the image. Maharay (1989) reported that these fans have a thickness of about 5 mm. Current beams are even thinner than this. The need for a mirror and empty space can be avoided by using highly optimised x-ray sensors.

A further advantage of scanning the baggage line by line is that the scattered x-rays can be registered as such. The way the x-rays are scattered provides information about the materials in the baggage. In particular, the scattering characteristics of explosives may be of interest. If the whole suitcase were exposed at the same time, it would be much more difficult to determine what part of the baggage contains the material associated with to the scattering detected.

Because of these advantages over scanners with a fluorescent screen, most current x-ray scanners scan the baggage with a sensor line. This gives the typical x-ray scanner configuration as showed in Figure 3.6. The speed of the conveyor belt is usually 0.24 m/s (about 600 suitcases per hour). The belt of the Z-scan and of the HI-Scan 10065EDS (Heimann, 1996) have a speed of 0.5 m/s (up to 1500 suitcases per hour). Many systems allow the inspector to review the last few images, when in doubt after an earlier approval of a suitcase. Some systems can store all scanned images on a disk or tape. For example, the Aisys allows storage of a few hundred scans on a 525 Mb data tape. A complete scan gives the suitcase an x-ray dose between 0.9 μSv and 2 μSv. This allows for 10 (Heimann, 1987) or 25 (Europscan, 1993) scans to be made without exposing photographic films (1600 ASA) that may be in the baggage.

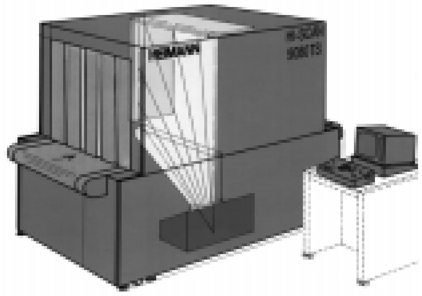


Figure 3.6. Exploded view of a typical x-ray scanner with sensor line. The x-ray source in the bottom of the machine sends a fan-shaped x-ray beam through the scanning tunnel. The sensor line is folded against the side of the tunnel.

The x-rays are sent through the baggage, and the baggage partly absorbs the x-rays. The amount of remaining x-rays is measured with the sensor line. The sensor line is often folded, but the distortions caused by this folding can be compensated easily. Usually the sensor line contains between 576 and 2048 x-ray sensors. In most modern machines two x-ray sensors are used for one pixel in the final image. The analogue voltages coming from the detector array are digitised to 8 to 12 bit digital numbers, and combined in a computer to form the x-ray image. This gives images of up to 1280x1024 pixels, with up to 4096 grey levels per pixel (Europscan). Next to absorbing, the baggage also scatters the x-rays, so the tunnel and its entrance and exit hole are shielded with lead.

Some objects completely absorb x-rays, thus occluding objects lying behind them (a 'black hole'). The Z-scan (EG&G Astrophysics, 1996) solves this by making two images in stead of one, one side view and one bottom view. This bottom view may be used to see the objects that were hidden by the x-ray absorbing object. Figure 3.7 and 3.8 illustrate the principle. In stead of one x-ray beam, two fan-shaped x-ray beams are sent through the baggage. This will, of course, double the x-ray dose on the baggage. For another system, Evans, Godber and Robinson (1994) proposed displaying two such views stereoscopically. Scanray has been developing such a system (Wooley, 1986). For such a stereoscopic system the angle between the views has to be small. However, such a stereoscopic image does not enable the observer to look around an x-ray absorbing object.

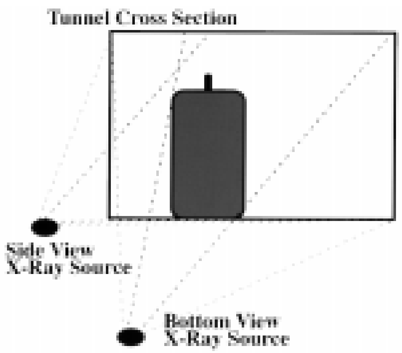


Figure 3.7. Cross section of a Z-scan by EG&G Astrophysics. This machine makes two scans of the baggage in one pass.

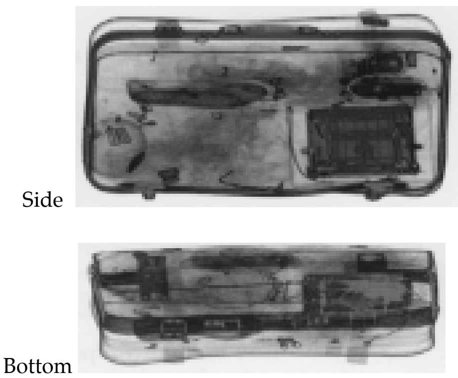


Figure 3.8 (**see colour figure on left cover flap**). Typical side and bottom view acquired with the Z-scan.

The sensor-line x-ray scanners mentioned above (Table 3.1) alter the raw x-ray images so that the information that is expected to be relevant for the inspector is emphasised. First, they increase the contrast in the image. Humans can distinguish up to about 256 grey levels in a monitor display, but to enhance the visibility of objects this range is usually limited to about 22 levels. A consequence of this contrast enhancement is that the inspector may have to adjust the contrast and brightness, depending on the part of the baggage he is inspecting. Second, most machines are able to magnify part of the image, to aid detailed inspection. For example the Heimann 9075 enables up to 16 times magnification of the image (Heimann, 1994).

Some machines provide information about the materials the suitcase contains. The Heimann machines distinguish organic, aluminium-like and metallic materials. This information is displayed by pseudo-colouring the grey-level image: organic parts are rendered orange, aluminium-like materials green and metallic parts blue (Figure 1.1 and

3.13; see also Heimann, 1997b). The Z-scan uses similar pseudo-colouring (Figure 3.8). The Vivid detects areas with a density close to the density of explosives, and pseudo-colours these areas red. The inspectors in the real-baggage experiment did not appreciate the pseudo-colouring of the Heimann 9075. The inspectors in that experiment suggested that the pseudo-colouring of the Heimann machines gives material information not relevant to their task.

Automatic recognition of suspicious shapes is of increasing importance. The Aisys automatically recognises parts of a detonator and places markers indicating those parts in the display (Figure 3.9). The Z-Scan places an ellipse around suspect areas (Figure 3.8).

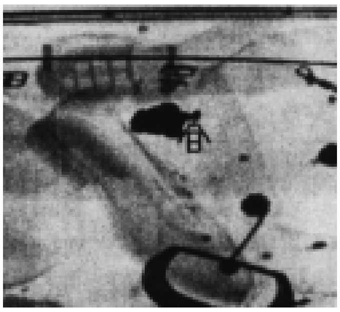


Figure 3.9. Detail of an image from the Aisys 370B. An arrow (just above the B) indicates a suspect item.

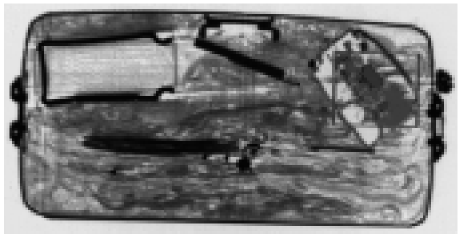


Figure 3.10 (**see colour figure on left cover flap**). Image from the CTX5000. Red indicates possible explosives, green parts of a detonator.

Computer Tomography scanners

A Computer Tomography (CT) scanner can make a 3D representation of the suitcase and its contents in computer memory. From this 3D representation, arbitrary views such as slices, perspective renderings and cut-through images, can be made. CT scanners are rarely used for baggage inspection because their use has practical problems. The CTX5000 (InVision, 1997) can scan 350 suitcases per hour, while scanners with a sensor line can scan up to 1500 bags per hour. Furthermore its 26-inch wide opening is too small for some baggage. The image resolution is 512x512 pixels, quite low compared to scanners with a sensor line. CT scanners have a complete x-ray scanner rotate around the baggage, and the reconstruction process requires a vast amount of calculation and computer storage, causing these machines to be bulky and expensive. Optimising strategies use 2D x-ray images to decide what part of the baggage needs more precise CT scans (Imatron, 1991, InVision & EG&G, 1997). A complete 3D reconstruction of the suitcase allows sophisticated object recognition. The CTX5000 tries to mark explosives and detonators with pseudocolouring (Figure 3.10).

Conclusion

Currently, most baggage inspection is done with scanners using a sensor line. Such scanners provide a single high-resolution x-ray view of the baggage. These machines usually enhance the x-ray images with material information. The more recent machines also add markers indicating suspicious items.

Perceptual requirements for x-ray imaging are another important factor for the design of an x-ray baggage scanner. These perceptual requirements are discussed in the next section.

What images are useful?

The question about the usefulness of particular views is difficult. All the experiments in this thesis deal with this question. However, as long as the baggage inspection task is not precisely operationalised no definitive answer can be given about the requiredvisual and other information (see Chapter 1 and 2). In this section I will try to make a reasonable choice with the present knowledge and the results of the experiments described in Chapter 4-8. I start with choices about static image quality, number of views and the required viewpoints. Next, the perspective properties of the views are discussed. There are a large number of possibilities for the perspective, each having its own problems.

What views and what quality?

I start with the easier choices, concerning image quality, number of images and required viewpoints. For static image quality it is necessary to consider the size of the critical objects. Consider the typical parts of a bomb: a battery, timing mechanism, detonator, explosives and wires. Visibility of wires up to 24 gauge (24 AWG = 0.5 mm Ø, Dorey, 1983) was required by the Federal Aviation Administration (FAA) (Tsacoumis, 1983). Currently, a resolution of 34 AWG (0.16 mm Ø) is usual. The timing mechanism can be extremely small, as a few transistors on a chip will suffice. Therefore, a timing mechanism can be invisible in x-ray images. The battery and detonator can be detected easily: typical detonators are cylindrical: about 5 cm long and 0.5 cm diameter. Batteries are usually also cylindrical, usually ranging from 0.5 to 5 cm long and 1 cm diameter. A number of articles suggest that 16 by 16 image pixels are sufficient to recognise an object. For example, face recognition is possible with the image of the face reduced to only 16 x 16 image pixels (Harmon, 1973) and aircraft silhouettes are reasonably identified with about the same number of pixels (Uttal, Baruch and Allen, 1995). Given the typical sizes of a battery and a detonator, about 15 x-ray sensors per cm seems reasonable for x-ray baggage inspection. For wire detection, fewer sensors per cm are sufficient: the Heimann 9075 has fewer than 5 x-ray sensors per cm and can detect up to 38 AWG (0.1 mm Ø). When multiple views are available, the results of the experiment described in Chapter 5 suggest that even fewer sensors per cm can be used.

For the required number of views, the complexity of the scene is important (see Chapter 6). I expect that the spatial scene complexity of real baggage lies somewhere between the spatial complexity of the connected objects scene and that of the knot tracing scene (see Chapter 2). Looking forward to Chapter 5 and 6, it is shown that, in the case of detecting wires connecting objects, observer performance does not increase with the number of views if more than two views are available. With the knot tracing task observer performance still increases with the number of views when 33 views are available. These results do not conclusively determine the required number of views, and therefore I attempted to determine the definitive number of views by an experiment with real baggage (Chapter 8). The results of the experiment with real baggage did not reveal an improvement of the inspector performance with more than two views, but several explanations for this result were proposed. Given the results of the experiments of Chapters 5 and 6, and given the maximum of about 25 x-ray photos to prevent damage of the baggage contents, providing 8 or 16 views seems reasonable.

One would expect that providing the inspector with information about the materials in the suitcase would help him to form his judgement. Furthermore a high resolution image would appear to be important for recognition of small objects such as wires, detonators and batteries. Given these arguments, an x-ray scanner with a sensor line looks a better choice than a scanner exposing the entire suitcase in one shot.

The Delft Virtual Window System can couple both left-right, forward-backward and up-down movements of the observer. However, making available views in all these degrees of freedom would require a huge number of x-ray photos. Not all these views are equally useful: for many tasks a small number of views in the horizontal arc is sufficient. The connected-objects task (Chapter 5) could be done reasonably well with a front and a side view. For the knot tracing task (Chapter 6) adding viewpoints in the vertical arc to the views in the horizontal arc did not result in a performance increase of the inspectors. Many experiments concerning movement parallax provide only horizontal freedom of movement of the observer (Rogers and Graham, 1979; Todd, Akerstrom, Reichel and Hayes, 1988; Cornilleau-Pérès and Droulez, 199). Bingham and Stassen (1994) discuss some evidence that forward-backward movements are also important for apparent depth, but they suggest that these movements make sense especiallyin the case of targeted actions. There are also theoretical grounds for choosing one degree of freedom: Braunstein, Hoffman, Shapiro, Andersen and Bennett (1987) indicate that if the scene rotates around a fixed axis, a 3D reconstruction of the scene can be made with fewer views than when the axis of rotation is variable. For ergonomic reasons, movement in the horizontal arc also is preferable over vertical movement: x-ray inspection is usually done while sitting behind the monitor, and horizontal head movements are less fatiguing than vertical movements (McVey, 1970). A single axis of rotation with a fixed angle between the views is also intuitive: it tells the observer about the available views and makes clear to the observer what to do to get a particular view. Furthermore, it makes replacement of viewpoint selection via eye position by viewpoint selection via a knob possible in an intuitive way. Finally, it is expected that a single axis of rotation will allow for easier technical constructions to shoot the images than when multiple axes of rotations have to be supported.

In conclusion, a reasonable choice is to use a scanner with a sensor line to shoot 8 or 16 views. The views should have only a horizontal degree of freedom, e.g. a vertical axis of rotation, and the angle between two views has to be approximately constant.

Perspective properties- static camera

The perspective properties of x-ray images are quite complex. I start with two different perspective possibilities for a static x-ray image. Next, I will discuss two possible couplings between camera movement and image transformation. Both the static and the dynamic perspective transformation can be chosen independently for the horizontal and vertical direction of the image, giving a total of 16 possible perspective combinations. Finally, the consequences of these perspective combinations are discussed.

There are two aspects of the perspective properties of static x-ray images. First, a perspective can be either a convergent perspective (C perspective) or a parallel perspective (P perspective) (Figure 3.11). In these figures, the back of the suitcases will be shaded for clarity.

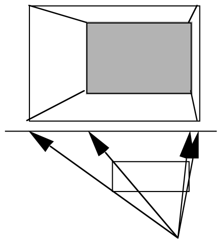


Figure 3.11a. With convergent (C) perspective the back of the suitcase is projected smaller than the front.

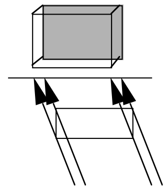


Figure 3.11b. With parallel (P) perspective, the front and the back of the suitcase are of equal size.

Second, the perspective can be different for the horizontal (H) and the vertical (V) direction in the image (Figure 3.12). Table 3.2 outlines the four possible perspective combinations for a static image. Combinations will be abbreviated, for example horizontal convergent perspective combined with vertical parallel perspective will be abbreviated as HC VP perspective. As a scanner with a sensor line always gives a convergent perspective in the direction of the sensor line, the HP VP perspective is of no interest in the present context.

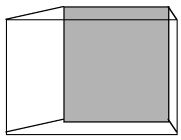


Figure 3.12a. With HC VP perspective the height of the front and back plane are equal, while the width of the front plane is greater than that of the back.

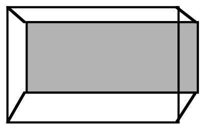
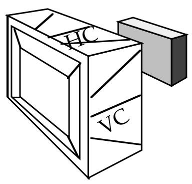
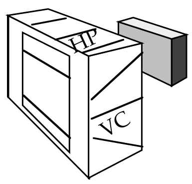
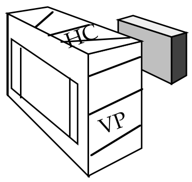
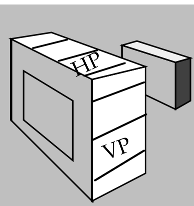


Figure 3.12b. With HP VC perspective the widths are equal, while the heights are different.

Table 3.2. Possible perspective properties for a static image. The shaded HP VP perspective combination is not relevant for scanners with a sensor line.

	Horizontal Convergent (HC)	Horizontal Parallel (HP)
Vertical Convergent (VC)		
Vertical Parallel (VP)		

X-ray images acquired with a scanner with a sensor line are images with a parallel perspective in one direction and a convergent perspective in the other direction. For example Figure 3.13, an image from a Heimann 6040-A, has HP VC perspective. In this figure, vertical convergent perspective suggests a viewpoint in the plane through the bottom of the baggage. The camera must be very close to the baggage, as the top of the suitcase is displayed almost perpendicular. However, both the left and right side of the image suggest a viewpoint in the plane through that side, indicating a camera position at infinity. Furthermore, such a combination of parallel and convergent perspective is almost never encountered in real life. Such conflicting cues may have perceptual consequences.

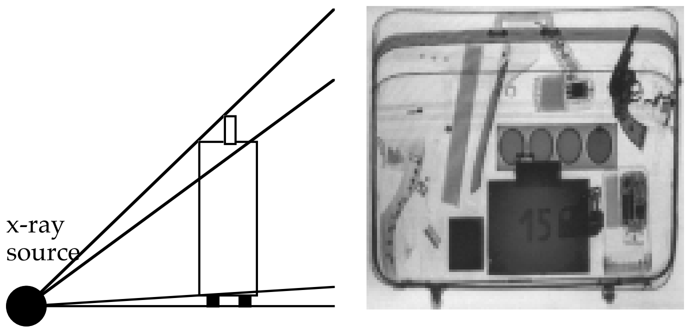


Figure 3.13. An image from a Heimann 6040-A (right) illustrates conflicting perspective. Vertically the image has convergent perspective and horizontally it has parallel perspective. The left figure shows the relation between the x-ray image and the edges of a real suitcase.

Perspective properties- moving camera

I will now discuss what happens if the camera can move. As chosen in the previous section, camera movement will be restricted to the horizontal arc. In the vertical direction a fixed angular viewing height can be chosen (Figure 3.14).

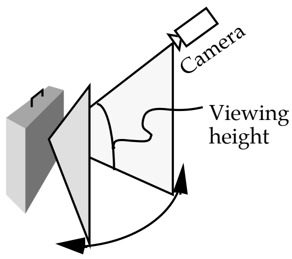


Figure 3.14. Camera movement is restricted to the horizontal arc. The angular viewing height is fixed.

A camera movement can result in a shear or in a rotation in the image (Figure 3.15). As with the parallel and convergent perspective, shear perspective (S) and rotational perspective (R) can be different for the horizontal and vertical direction in the image.

Again I will abbreviate, for example the combination of horizontal rotational perspective with vertical shear perspective is abbreviated as HR VS perspective.

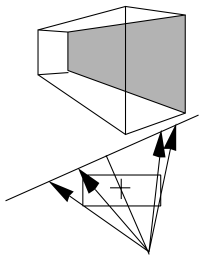


Figure 3.15a. With rotational perspective the projection plane always is perpendicular to the line from the camera through the centre of the suitcase

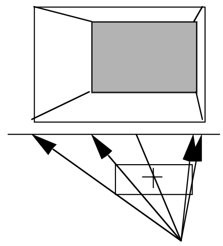


Figure 3.15b. With shear perspective the projection plane is always parallel to the suitcase.

Combined with the convergent and parallel perspective, Table 3.3 shows the possible *perspective combinations* for a moving camera.

Table 3.3. Possible perspective combinations for a moving camera. Terms are abbreviated, eg., the cell marked with '*' is indicated with HRP VSC perspective. As in Table 3.2, the shaded fields are of no interest in the present context.

			Horizontal perspective (H)			
			Rotational (R)		Shear (S)	
			Conver-gent (C)	Parallel (P)	Conver-gent (C)	Parallel (C)
Vertical Perspec-tive (V)	Rotational (R)	Convergent				
		Parallel				
	Shear (S)	Convergent		*		
		Parallel				

The perspective combinations differ in appearance: some suggest a nonrigid deformation of the baggage (Figure 3.16). The goal is to minimise the disturbing perceptual effect by choosing an appropriate perspective combination. The next section discusses the perceptual effect of the possible perspective combinations and how the disturbing effect depends on particular viewpoints.

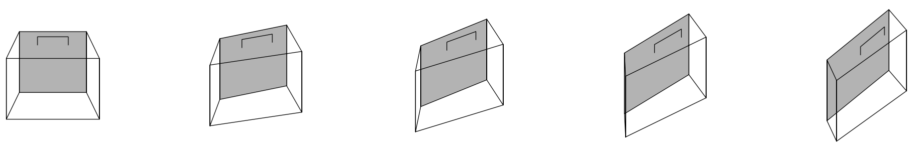


Figure 3.16. An unusual perspective combination may lead to images that suggest nonrigid deformations.

Perceptual consequences of the perspective combinations

Little is known about the perceptual consequences of the possible perspective combinations of Table 3.3, although x-ray inspectors have been working with such images for years. In order to get an idea of the perceptual consequences of the perspective combinations, a simulation was made displaying a wire frame suitcase for several perspective combinations. In the next pages I show a series of views for each combination, and I describe my impression from the simulation. The simulated combinations were inspected both with the view being selected via head tracking and with the view being selected with a knob.

I chose the viewing distance to be twice the width of the suitcase. Looking from a viewing height slightly above the suitcase (see Figure 3.14) gives a less distorting depth impression than looking at a height of 0˚. I will first discuss the perspective combinations when the viewing height is 45˚ (see Figure 3.14). A 25˚ viewing height reduces visibility of the top of the suitcase, and therefore seems to reduce disturbing visual effects. This 25˚ viewing height will be discussed next.

Figure 3.17 shows views from the perspective combinations with HR VR perspective. The first row shows images with HC VC perspective. These images look acceptable. With HP VC perspective (second row), the left and right side of the suitcase become a single line in front view. The protruding side of the suitcase distorts in these views. Perceptually, this is a highly disturbing effect. With HC VP perspective (last row), the suitcase seems to twist.

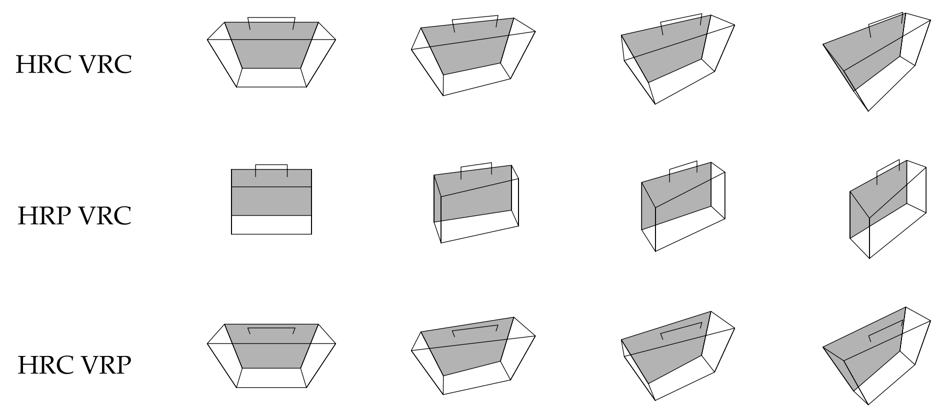


Figure 3.17. combinations with horizontal and vertical rotational (HR VR) perspective .

Figure 3.18 shows views from the perspective combinations with HR VS perspective. Images with HC VC perspective and with HP VC perspective (first and second row): here the views look acceptable. With HC VP perspective (last row), the front view is disturbing, as the heights of front and back of the suitcase are of equal height. The side views are unacceptably distorted: the sides rotate and stretch relative to each other.

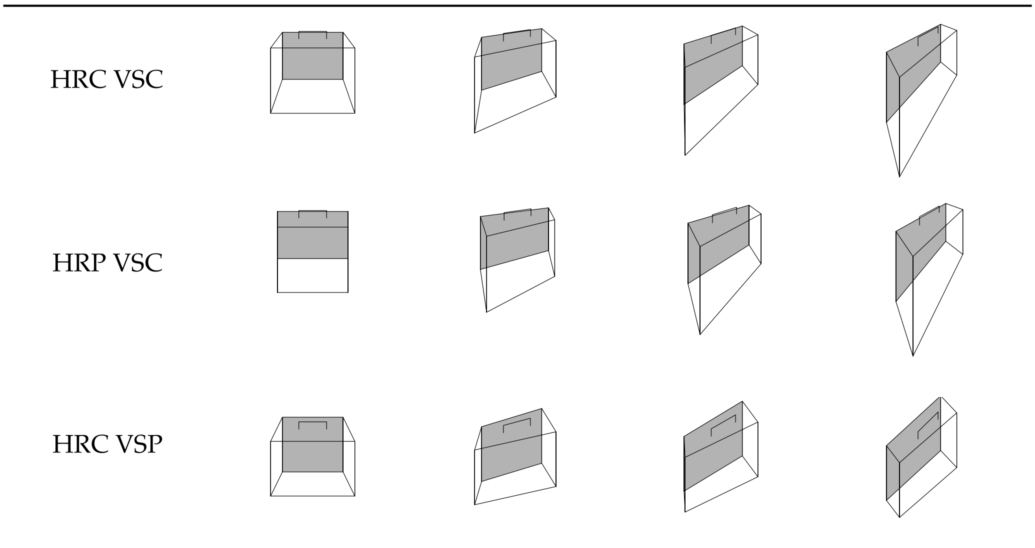


Figure 3.18. combinations with horizontal rotational and vertical shear (HR VS) perspective.

Figure 3.19 shows views with HS VR perspective. In the HC VC perspective combination (first row), the suitcase seems narrowed at the bottom. The second row shows images with HP VC perspective. The front and back of the suitcase are always of equal size. This is caused by the 45° height of the view, and is disturbing here as perceptually it suggests that the front is smaller than the back. When the viewpoint is selected manually rather than via eye position, depth reversal occurs. The last row shows the images with HC VP perspective. These images are highly distorted: the suitcase looks rotated backwards, larger at the top and smaller at the bottom. In these three cases, the baggage seems to shear if the viewpoint is selected by knob instead of by eye position.

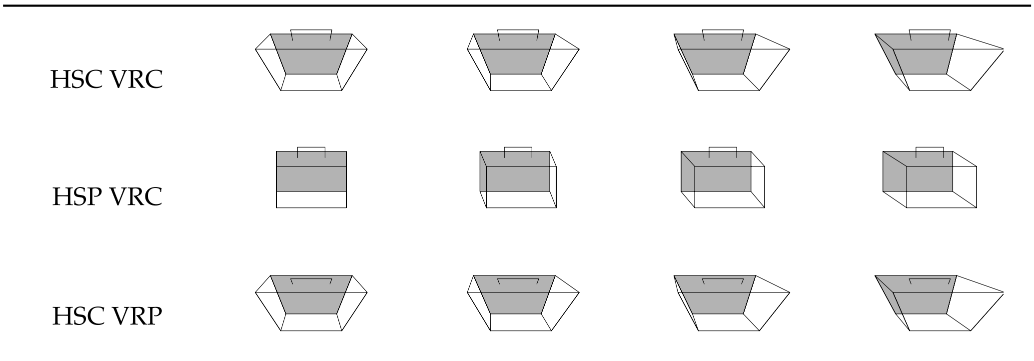


Figure 3.19. combinations with horizontal shear and vertical rotational (HS VR) perspective.

Figure 3.20 shows views with both horizontal and vertical shear perspective. With both HC and VC perspective, there is no apparent distortion when eye position is coupled accurately. Note that this combination is a perfect off-axis coupling (see Chapter 7). With HP VC perspective, the back of the baggage seems wider than the front, while with HC VP perspective the back looks higher than the front. Again, in these three cases the suitcase seems to shear if the viewpoint is selected manually instead of via eye position.

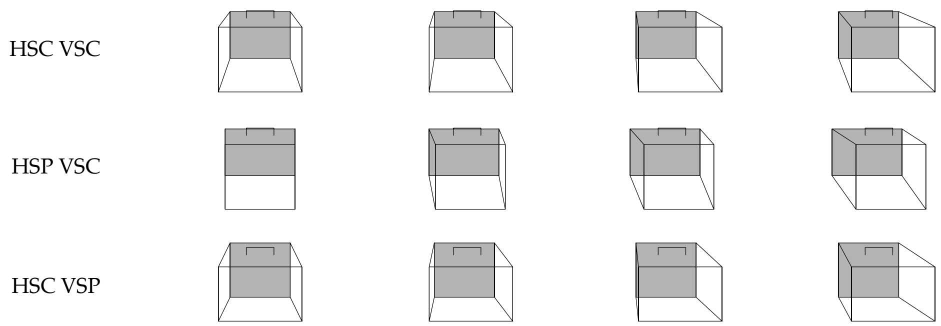


Figure 3.20. Combinations with horizontal and vertical shear (HS VS) perspective.

A number of situations can be improved by reducing the viewing height (see Figure 3.14), so that the top of the suitcase reduces almost to a single line. A viewing height of 25˚ in stead of 45˚ will achieve this. Three of the perspective combinations discussed above benefit from a lower view: the HRP VRC perspective, the HRC VSP perspective and the HSP VRC perspective combinations. Figure 3.21 shows views in these perspective combinations for a viewing height of 25˚.

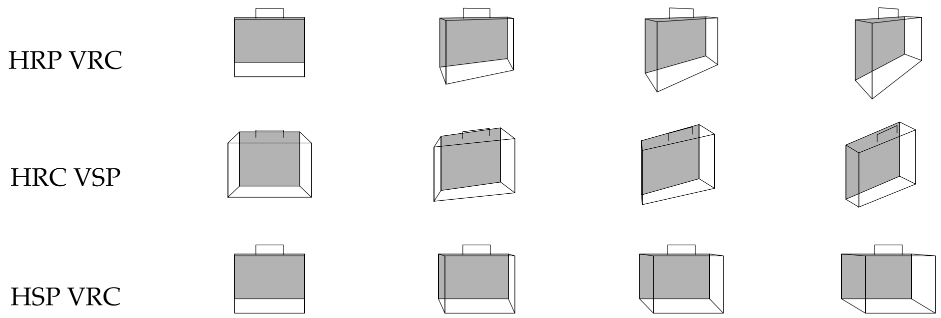


Figure 3.21. Three perspective combinations with 25˚ top view instead of 45˚. These conditions show lower distortion with this lower view.

With 25˚ viewing height, the HRP VRC perspective combination (first row) looks very much like the HRP VSC perspective with a viewing height of 45˚ (Figure 3.18), and looks acceptable. In the HRC VSP combination (second row) the nonrigid distortions are smaller than with 45˚ viewing height, but they are still disturbing. For the HSP VRC perspective combination (third row) the images look similar to the HSP VSC combination at a viewing height of 45˚ (Figure 3.20), and just as acceptable.

Discussion and conclusions

The perceived distortions in the combinations with horizontal shear perspective may be explained as follows. The observer is forced to use perspective cues because many other cues are absent. Similarly, in x-ray baggage inspection perspective cues may be important because with x-ray images several cues, such as shading, occlusion and shadows, are absent. As is discussed in Chapter 7, the perspective determines the geometrically appropriate (geometrically equivalent) viewpoint, and for horizontal shear perspective this viewpoint moves with the amount of shear. This explains the apparent distortions if the viewpoint is selected manually instead of via eye position. The distortions apparent with horizontal shear and viewpoint selection via eye position can also be explained with geometry. It can be shown that the geometrically appropriate perspective is HSC VSC perspective (called on-axis coupling in Chapter 7). This perspective is equivalent to a scaling of the objects in the scene by their distance from the observer, followed by HSP VSP perspective, with the shear depending on the viewing position of the observer. When the amount of shear is coupled correctly, *any* perspective combination is geometrically consistent both over different viewpoints, but the apparent scene is deformed. The observer can see only the consistency, and this may suggest to the observer that the perspective cues are reliable, and he therefore does an inverse object scaling, with the distance of objects from the observer extracted from the amount of shear. This explains why objects appear too large in the horizontal or vertical direction when this direction is not scaled according to the distance from the object to the observer, and why the scene looks deformed when vertical rotation instead of vertical shear is used. The distortions noticed with horizontal rotational perspective can not be explained so 'easily'.

Technical requirements were proposed indicating what images are useful for baggage inspection. A reasonable choice is to use a scanner with a sensor line to shoot 8 or 16 views. The views should have only a horizontal degree of freedom, e.g. a vertical axis of rotation, and the angle between two views has to be approximately constant.

The choice for the perspective combination proved to be more difficult. The most important result is that horizontal shear perspective will result in nonrigid deformations if the view is not selected via eye position. With horizontal rotational perspective, perspective combinations with both horizontal and vertical convergent perspective look acceptable. Furthermore the combination with vertical shear, horizontal parallel and vertical convergent perspective looks acceptable. Viewpoint selection via a knob and via eye position do not differ here. Some combinations giving a distorted impression look acceptable when viewed from such a height that the camera is approximately in the plane through the top of the suitcase. However, the choice for the perspective combination also depends on the motion mechanism, which is discussed in the next section. Afterwards, in 'Optimised concepts', the diverse requirements are combined into an optimal choice.The next section uses the technical requirements and perceptual consequences discussed in this section to make an appropriate technical concept for acquiring multiple x-ray views.

Mechanisms

This section integrates the partial analyses of the previous section into proposed mechanisms for shooting multiple x-ray images. There are a large number of possible mechanisms to shoot such images, and I start with an outline of the possibilities. Next, the requirements of the previous section are used to choose two technical concepts. These concepts are worked out in more detail. Finally a construction is described that was used for shooting the images for the real-baggage experiment.

Outline of possibilities

There are a large number of ways of making multiple x-ray images. First, there is a choice between moving the baggage, moving the x-ray source or moving the sensor line to obtain a different view (Figure 3.22). In these three concept solutions the baggage is translated through the fan-shaped x-ray beam, as in conventional baggage inspection. This means that the conveyor belt has to be reversed for each new x-ray view of the suitcase.

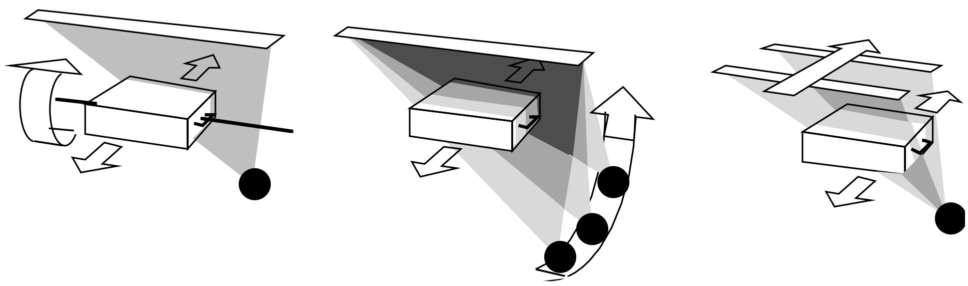


Figure 3.22. Three methods of obtaining different views. For each view, the x-ray image is taken by translating the suitcase through a static fan-shaped x-ray beam. To obtain the next viewpoint, the suitcase can be rotated (left), the x-ray source can be moved (middle) or the sensor line can be moved (right).

Second, there is a choice whether this movement to obtain the appropriate viewpoint involves a rotation around an axis in the plane of the fan-shaped x-ray beam or around an axis perpendicular to this plane (Figure 3.23). Third, a linear or rotational movement of the x-ray source or sensor line can be chosen (Figure 3.24). Even more complex movements may be used.

Figure 3.23. To acquire an appropriate view, the baggage can be rotated either around an axis parallel to the fan-shaped x-ray beam (left) or around an axis perpendicular to that beam (right).

Figure 3.24. The movement of the x-ray source or sensor line can be a translation (left) or a rotation (right).

Fourth, movements of Figure 3.22, 3.23 and 3.24 may be combined, giving a huge number of concepts for acquiring multiple views. Figure 3.25 shows an example. Finally, a movement of the x-ray source or sensor may be replaced by multiple stationary sources or sensors (Figure 3.26).

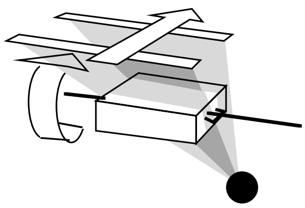


Figure 3.25. Movements can be combined. Here, a translating line scanner scans one view. For the next view, the baggage is rotated and scanned again.

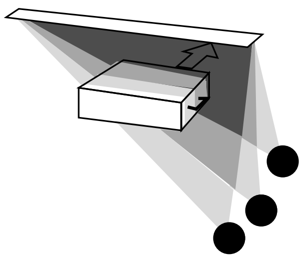


Figure 3.26. Instead of moving the x-ray source, multiple sources can be used. Here, the sources are alternated rapidly while the baggage is translated through their beams.

Acquiring multiple views with a conventional scanner

For finding the precise number of views required, an experiment was done with real x-ray images and real baggage inspectors (Chapter 8). The required x-ray images were taken from a standard scanning machine (a Heimann 9075, similar to Figure 3.6). As the source and sensor line are fixed in such a machine, it was necessary to rotate the baggage. I chose to use the parallel perspective in the horizontal direction, and the convergent perspective in the vertical direction, because the distortions this combination gives can be made acceptable by choosing a 25˚ top view of the suitcase (see 'Perceptual consequences of the perspective combinations' in the previous section). The alternative, parallel perspective in the vertical direction and convergent perspective in the horizontal direction, will show less improvement with a 25˚ top view of the suitcase. A foam construction was made (Figure 3.27) that allowed rotation of the baggage over 90˚. Polystyrene foam is nearly invisible in an x-ray image, and thus could be used to hold the suitcase. The foam was covered with paper to prevent damage.

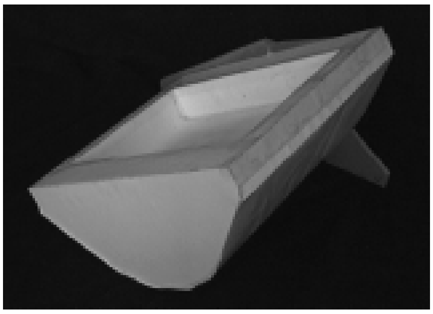


Figure 3.27. Foam construction that allows suitcases to rotate between -45° and +45°. This construction was used to make x-ray images of real baggage (see Chapter 8).

Criteria for selecting a concept

The foam construction of Figure 3.27 worked well with an existing scanner, but scanning 16 views is time consuming (more than 10 minutes), and requires manual rotation of the baggage. Therefore this concept is not effective to use in a commercial machine. The criteria that the users and manufacturers will consider are discussed below.

The first criterion is the perspective combination. The perspective cues in the image should not disturb the inspector. The perspective properties of the scanned images are determined by the orientation of the sensor line relative to the baggage and the way baggage, sensor line and source are moved.The perceptual consequences of the perspective properties were discussed in the previous section, and can be used to select a concept solution.

Another criterion is scanning speed. Some of the technical concepts are able to take multiple x-ray images in one pass of the suitcase (e.g. Figure 3.26), while other concepts require multiple passes (e.g. Figure 3.23) or halting the conveyor belt (e.g. Figure 3.25). As each image takes the conventional scan time, taking 16 images (the number I expected to be useful; see previous section) with the reversing belt strategy (Figure 3.23) would take 16 times the conventional scan time, which seems unacceptable. Stopping the belt may be acceptable, if not for too long and if stopping the belt gives no conflict with the other conveyor belts in the baggage inspection system.

A third criterion concerns reliability. For example, given the uncertainties about baggage weight and size, it seems not a good idea to rotate the baggage. Furthermore, due to the gravity the baggage contents may move if the baggage is rotated, and this would give useless images. Baggage contents may also move if the belt has to be reversed a number of times. X-ray sources may be hard to move, because they are heavy, but their movement may be simulated by moving an x-ray mirror.

Finally, the costs of implementing the concept have to be considered. Because sensor lines are expensive, it does not seem a good idea to make an array of 1000 sensor lines to shoot the image in a single pass. Concepts with more than 16 sensor lines would be extremely expensive to implement.

Chosen concepts

The criteria described above were used to select a number of concept solutions from the huge number of possible concepts for acquiring multiple views. Figure 3.28 shows the first concept solution. The baggage stops at the required position for the view. Next, the x-ray view is made by translating the sensor line over the baggage (the slit in the source has to move accordingly, in order to keep the beam aimed at the sensor line). Then the baggage is transferred to the next position by the conveyor belt. After 16 iterations the required 16 views are attained. The second concept (Figure 3.29) transfers a mirror in stead of the sensor line. Again, the baggage stops at the required position for the next view. The x-ray view is made by translating the x-ray mirror under the suitcase. This is repeated until the required number of views have been made. The varying distance between the x-ray source and sensor line, caused by the movement of the mirror away from the x-ray source, may give perspective distortions besides those resulting from the perspective combination.

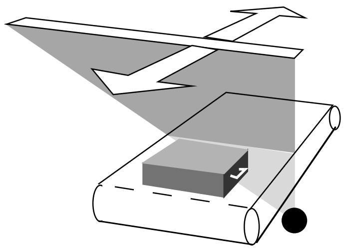


Figure 3.28. First concept solution. A single view is scanned by translating the sensor line over the suitcase. For the next view, the suitcase is translated to the next viewpoint by the conveyor belt.

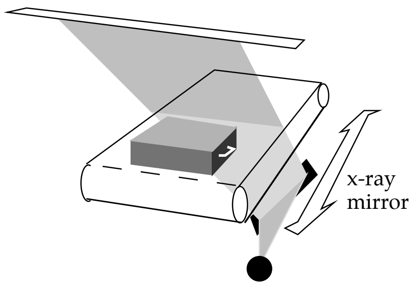


Figure 3.29. Second concept solution. A single view is scanned by translating the x-ray mirror under the suitcase. For the next view, the suitcase is translated to the next viewpoint by the conveyor belt.

Both the above concepts require the conveyor belt to be stopped for each x-ray view. This causes delays and the baggage contents may be disturbed by changing the speed of the suitcases. The next three concepts solve these problems by using multiple sources or sensors. Figure 3.30 shows such a concept. The conveyor belt just moves the baggage through the x-ray fans.

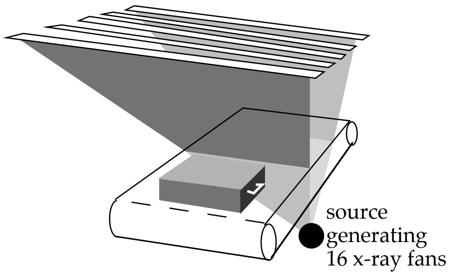


Figure 3.30. Concept that allows continuous throughput of the baggage. The source generates 16 fan-shaped beams, each giving a different view of the baggage.

Optimised concepts

Figure 3.31 shows a concept similar to Figure 3.30, but price-optimised. A combination of 4 sources and 4 sensor lines will cost about NLG 100,000 (USD 50,000) which is cheaper than a configuration with 1 source and 16 sensor lines, which will cost about NLG 150,000 (USD 75,000). The idea of this concept is that the four sources are turned on and off rapidly after each other, each exposing the four sensor lines, giving a total of 16 views. Again, the conveyor belt just moves the baggage through the pulsating x-ray fans. Rapidly turning x-ray sources on and off seems no problem: just switching the voltage between the anode and cathode should suffice. As with the previous concept, this line scanner provides images with horizontal shear perspective.

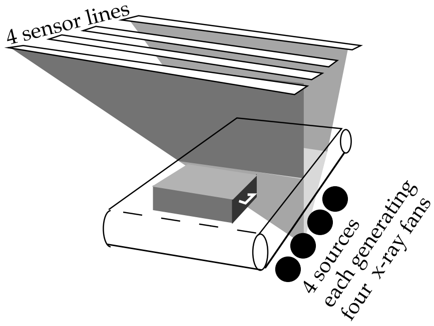


Figure 3.31. Price-optimised concept. Each of the four sources in turn generates four fan-shaped beams, giving in total 16 views. The absence of moving parts other than the conveyor belt makes the mechanism reliable.

Figure 3.32 shows a concept giving an image perspective with horizontal rotation in stead of shear. One x-ray source exposes the x-ray mirrors one after the other. The x-ray sources reflect the x-rays to the single sensor line. This way, the mirrors multiply the single real source into a number of virtual sources. To acquire 16 views, 16 mirrors are required. The conveyor belt just moves the suitcase through reflected x-ray fans. It seems possible to make the required x-ray source by rapidly rotating a lead shield with a slit around the x-ray source.

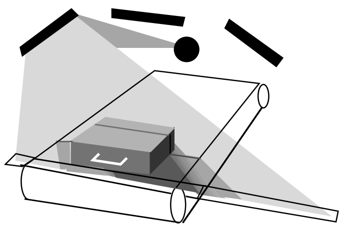


Figure 3.32. Concept giving views with horizontal rotation in stead of horizontal shear. The x-ray source exposes the x-ray mirrors one after another. In total 16 such mirrors are required to get 16 views. Only 3 are shown.

I will concentrate on the last two solutions, because they contain few or no moving parts, allowing for high reliability. Consider the concept with both multiple sources and multiple sensors (Figure 3.31). The sources and sensors can be placed at arbitrary places along the conveyor belt. The precise source and sensor positions have to be selected so that there is an equal angular distance between the acquired views. Finding a setup with exactly the same angle between the views is a hard mathematical problem, but I found a number of close solutions. One of these is shown in Figure 3.33. The precise height of the suitcase between the sources and sensor lines is important only for the required viewing angle, and can be chosen freely. Figure 3.34 shows the shear of each view (in radians), and it can be seen that the views are spread quite evenly over the viewing range. This concept will provide images with horizontal shear perspective. Therefore the acquired images should be coupled to the eye position of the inspector, but the images are not suitable for selection by a knob (see discussion under 'Perceptual consequences of the perspective combinations' in the previous section). The distance from source to suitcase is not constant, but this is geometrically appropriate (see also Chapter 7).

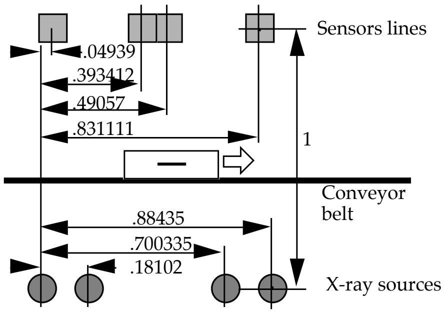


Figure 3.33. Side view of a configuration with 4 sources and 4 sensors, optimised to give an equal angle between the acquired views. Numbers indicate lengths.

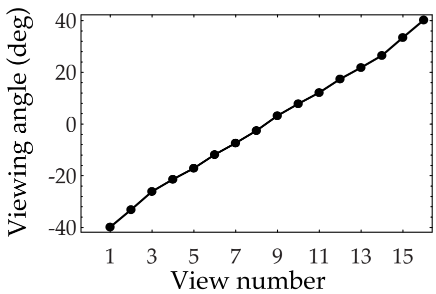


Figure 3.34. Viewing angle of the views acquired with the configuration of Figure 3.33.

The concept with x-ray mirrors (Figure 3.32) gives images with horizontal rotational perspective in stead of shear perspective . This concept is worked out in more detail in Figure 3.35. Care has to be taken that the x-ray mirrors do not overlap and that the total distance from the source via the mirror to the suitcase remains constant. The image perspective will have horizontal rotational convergent perspective and vertical shear parallel perspective. This means that these images can be presented with viewpoint selection by a knob instead of via the eye position of the observer. As x-ray mirrors are not a standard product, I am unable to estimate how much x-ray mirrors would cost. Using multiple x-ray sources may be cheaper than using x-ray mirrors.

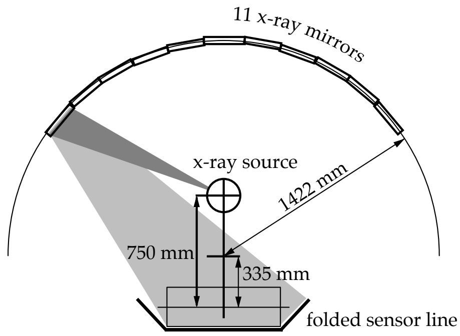


Figure 3.35. Cross-section of a configuration with x-ray mirrors. The mirrors and the sensor line are in a single plane; the x-ray source hangs 750mm in front of this plane. The total distance from the source to the centre of the suitcase is the same for all 11 views.

Conclusions

A number of concepts were proposed, considering technical possibilities, perspective properties and their perceptual consequences, price and scanning speed. The perceptual consequences of the possible perspective combinations had to be estimated, because of a lack of theoretical and experimental knowledge.

Two concepts were worked out in more detail. The proposed mechanism with multiple sources and sensors is feasible, both technically, perceptually and in terms of price. However, its images have horizontal shear perspective, where a horizontal rotational perspective is preferable. To obtain horizontal rotational perspective another concept using multiple x-ray mirrors was proposed. The price of x-ray mirrors is uncertain, but multiple sources can replace the mirrors. Although more expensive than the mechanism with multiple sources and multiple sensors, this concept still seems feasible, perceptually, technically and in terms of price.

As I have no evidence that the DVWS can improve the performance of baggage inspectors (Chapter 8), no manufacturer of x-ray machines was approached to build a prototype of a multiple-view x-ray scanner. Therefore, no technical drawings were made and no precise components were selected for the proposed concepts.

4

Detecting sharp objects

As discussed in Chapter 2, detecting sharp objects, such as knives , is a task that is relevant to x-ray baggage inspection. Especially in the hand baggage, such items are not allowed, although they may be transported in hold baggage. Usually, the inspector will recognise knives because he is familiar with most of them, but this experiment is designed merely to check the visibility of these dangerous sharp points and edges in hand-baggage. In this experiment, a sharp edge was defined as an edge sharper than 30˚. Figure 4.1a shows an object with a sharp edge, Figure 4.1b an object without such an edge. We have no x-ray scanner to make x-ray images. To match the x-ray baggage inspection task, non-familiar objects with and without sharp edges were made of transparent polyester potting resin.

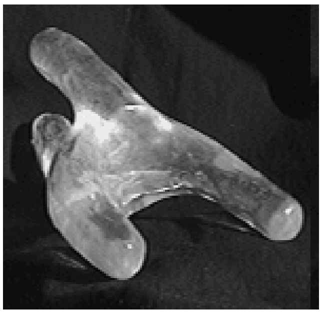


Figure 4.1a. Example of a sharp object. See Figure 2.10 for a stereoscopic depiction of this object.

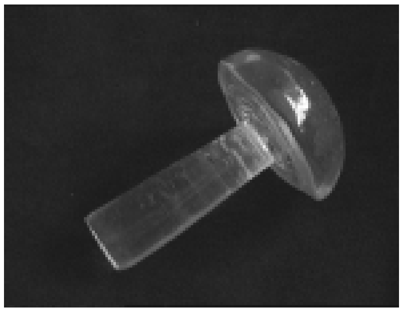


Figure 4.1b. Example of a blunt object.

If the observer can manipulate the object itself to do all the checks he wants, he is expected to be able to find such sharp edges when present. However, if the objects are inspected via a monitor he cannot manipulate the objects, feel the sharpness of the edges or choose any view he likes. Furthermore, the limited resolution and the limited number of grey levels will lower the sharpness cues, and thus his ability to see sharp edges. This expected decrease of observer performance with decreasing image quality corresponds to results from the literature (Ranadivé, 1979; Swartz, Wallace and Tkacz, 1992; Snyder, 1973).

This chapter describes three experiments. The first experiment checks the visibility of sharp edges when the objects are inspected via the DVWS, while the resolution and the number of grey levels in the views are varied between low and high settings. The second experiment tests the visibility of these edges when the objects are inspected naturally. To test the effect of the number of views on performance without the disturbing effects of varying image resolution and averaging over participants, the third experiment again tests their visibility when inspected via the DVWS while only high-resolution views are provided to the observer.

As discussed in Chapter 3, the number of available views has to be limited because there is a maximum x-ray dose to which the baggage can be exposed. Therefore, the camera will be given just one degree of freedom: the left-right movement (Figure 4.2). So movements up-down and forward-backward are allowed, but do not give him another view of the scene. The camera keeps aimed at a point in the scene: the fixation point. It moves on an arc around this point according to the eye positions of the observer. To achieve this, the eye position of the observer is continuously tracked, and the viewing angle φ_{obs} determines φ_{cam} as will be described below.

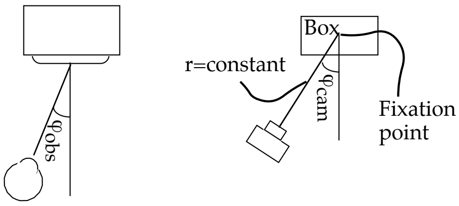


Figure 4.2. Top view of the DVWS. The angle of the camera relative to the scene φ_{cam} is adapted continuously to fit the actual φ_{obs}.

For this left-right movement only N images will be taken, with a constant angle $\Delta\varphi$ between two images. Thus, the observer will see the same view when he moves within a certain sector. This gives the situation of Figure 4.3. The lines indicate the direction from which the images were taken

The angle between the lef [N - 1] · $\Delta\varphi$)
also seems important. In ord ber of
images N, a large angle betw gle $\Delta\varphi$ gets
too large, the jerkiness of the impression
of selecting one from a numb iewing a
rigid 3D scene.

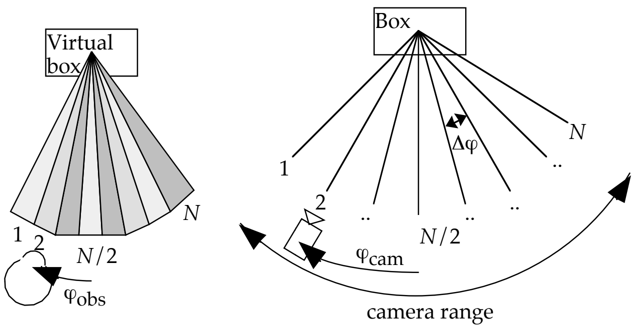


Figure 4.3. Top view of the scene. There are N available views and an angle $\Delta\varphi$ between the views.

I expect a larger camera range to be effective up to 180˚. This cannot be viewed if φ_{obs} and φ_{cam} are kept equal: to see the extreme views the observer then has to look at the display from the side. Therefore, I scaled the camera movement to a maximum comfortable head movement of ±22.5˚ (see McVey, 1970), giving a scale factor $\varphi_{cam} / \varphi_{obs} = N.\Delta\varphi / 45$. Furthermore, I expect that the observer performance will increase with the number of available views, as in earlier studies (Edelman and Bülthoff, 1992; Field, Michell, Wallis and Wilson 1995; Braunstein, Hoffman, Shapiro Andersen and Bennett, 1987).

Experiment 1- inspection via a monitor

The objective of this experiment is to test the effect of a reduced image quality and number of available viewpoints on the performance of observers in detecting sharp edges.

Method

Stimuli

The stimuli were images of mock-up baggage consisting of a transparent box with two different transparent objects as shown in Figure 4.1 in it, each possibly having sharp edges as described above. The sharpness of edges was tested during manufacturing with a wedge-shaped aperture. The box of 25 x 10.6 x 20.6 cm was made of tinted perspex. For each stimulus, two new objects were placed in new positions in the tranparent box. To get recordings of 34 mock-ups, 68 different objects were prepared. Figure 4.4 shows a sample recording of a box containing two objects.

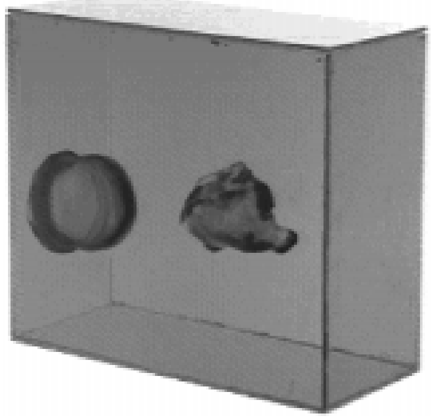


Figure 4.4. A view of a box containing two objects.

80 images of each box of objects were recorded from different viewpoints with a video camera (Sony CCDTR805E). The camera images were digitized (Archimedes real-time video digitizer from Watford Electronics) to digital images of 512x256 pixels with 16 grey levels, and stored uncompressed on a hard disc. Before recording the stimulus, the camera was tuned to use the total range of grey levels. The size of the image of the box on the screen was 12.5 x 10.9 cm in front view. The distance between camera and fixation point was 90 cm. This value is appropriate as the size of the image of the box is half of the real

box size, and the average viewing distance between observer and screen was expected to be 45 cm.

For a reduction of the number of grey levels, the original 16 grey levels were divided into 4 or 8 groups, and for each group the brightest value was taken. An informal evaluation by the experimenter indicated that the image contrast was not altered very much by this reduction. For a resolution reduction, the image pixels were grouped in 2 x 2 or 4 x 4 pixels whose intensity was averaged.

Apparatus

An Archimedes A5000 computer was used to display the images according to the eye position of the participants, and to store their responses. Preceding each trial, the appropriate viewing angles of the box were read from hard disc, reduced in number of grey levels and resolution if necessary, and stored in working memory. This caused a pause of about 10 s between trials. During the trial, the appropriate images were shown from working memory on the display (Puretek PT143D PLUS: non-glare monitor, 0.29 mm dot pitch). The screen refresh rate was 88 Hz (max. delay 11 ms), and the average light output was 150 Lux.

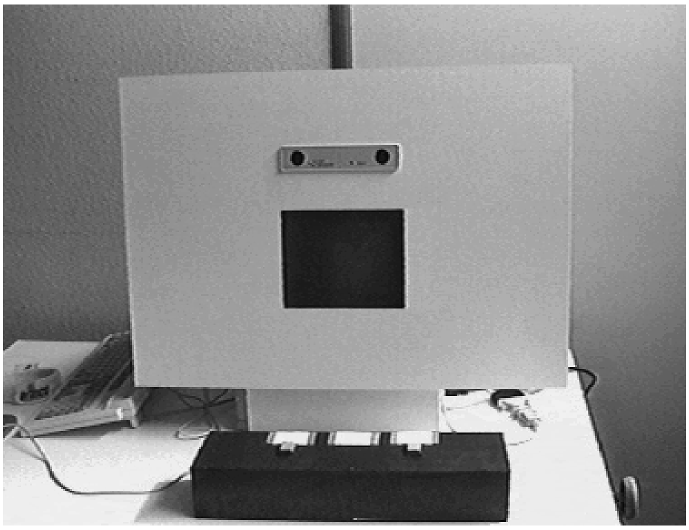


Figure 4.5. Experimental setup. The display is placed behind a reduction screen. Above the screen is the infrared eye position tracker. Below the screen is the button box with two buttons, by which the participants could make their judgements.

Figure 4.5 shows the experimental setup. To enhance the depth in the displayed scene (Gibson, 1971), a white reduction screen was placed in front of the display, making 19.3 x 16.1 cm of the display visible. The eye position was measured by a Dynasight infrared sensor (Origin Instruments,1993). This sensor was placed on top of the screen, and reported at 37 Hz (delay 27 ms) the position of a small reflector between the two eyes on a headband to the computer. This way, the eye position could be estimated with an accuracy of about 3 cm. Directly after receiving a new eye position, the corresponding image was shown on the screen (delay 16 ms). The total lag was about 11 + 27 + 16 = 54 ms. Below the screen was a button box with two buttons, by which the participants made their decision. The left button was labelled with a picture of an angle of 20˚; the right button was labelled with an angle of 40˚; between the two buttons was a label showing 30˚ angle, indicating the boundary between sharp and non-sharp.

Procedure

Participants were told that they would see a number of boxes, each with two objects in it, on the screen. Their task would be to check whether there was an object with a sharp edge in the box. It was explained that a sharp edge was a knife-like edge, sharper than 20˚: the angle indicated above the left button on the button box. They were told that they could look at most 10 seconds, and that a beep would warn for the time limit when 8 seconds had passed. If they had not made their choice after 10 seconds, the screen would turn dark, but they always had to make a choice. They made their choice by pressing a button on a button box: the left button if they detected a sharp edge, the right button if they did not. After making their choice, they had to wait for the next series of images to be loaded into main memory. After this was done, participants were warned with a beep that the next trial would start in 1 second. They were asked to respond as accurately as possible, and they were told that the participant with the most correct answers was going to be rewarded with a cake. It was explained that they could look around the box by moving their heads, and that their approximate eye positions were being tracked with an infrared tracker. The participants were asked to try the range and speed of the position tracker, in order to get used to the tracker. The sensor provided feedback by a control light which was green if the reflector was in track, and red if it was not. In this part of the training no images were shown on the display.

Prior to the experiment the participant was trained with seven trials. For the first training trial, they were allowed to look 30 seconds before the screen would go dark. Directly after the participant had made his choice, the screen showed whether he made the right choice, his response time and the range covered by his eye positions. During the actual experiment, the participant was shown 27 boxes. Overall, each experiment took about 25 minutes.

Variables, Design, Participants

The independent variables are the image resolution R, the number of grey levels in the image G, the number of available views N and the angle between two adjacent views $\Delta\varphi$. Table 4.1 gives the levels of the independent variables. To simplify the expressions for the angle between two views $\Delta\varphi$, I define the smallest angle between two images $\theta = 22.5^\circ/32$. The dependent variables were the correctness of the response C (right if they judged correctly about the sharpness of an object, or wrong) and the response time T.

Table 4.1. Independent variables and their values for experiment 1. $\theta = 22.5^\circ/32$.

Name of variable	Description	Possible values
R	Image resolution	256 x 128, 512 x 256
G	number of grey levels	4 , 8 , 16
N	number of available views	1 , 2 , 4 , 8 , 16 , 32
$\Delta\varphi$	angle between two adjacent views	θ, 2θ, 4θ, 8θ

The 3 (G) x 6 (N) x 4 ($\Delta\varphi$) x 3 (R) = 216 conditions were randomized over 8 participants, giving 27 responses per participant. This randomization was done 3 (repetition) times, for a total of 24 participants. All participants saw the 27 boxes in the same order. The participants were 24 students, mainly from the faculty of Industrial Design Engineering (10 women, 14 men) with normal or corrected-to-normal vision. They did not know the purpose of the experiment. Each participant received NLG 10 (USD 5) for taking part, and the best-performing participant received a cake.

Hypotheses

It is expected that an increase of the resolution R, number of grey levels G or the number of available views N, will increase the percentage of correct answers and decrease the response time. However, for each variable there will be a saturation point, where a higher value for that variable will not improve performance any more. For the angle between the views $\Delta\varphi$ the effect is less clear: a larger value for $\Delta\varphi$ will increase the jerkiness in the movement, but on the other hand it allows a larger range of available views with the same number of views. Our hypothesis here is that for this task, an appropriate viewpoint is more important than a smooth movement between adjacent viewpoints, and therefore that observer performance is expected to increase with the angle between the views $\Delta\varphi$. An alpha level of 0.05 was used for all statistical tests.

Results

An analysis of variance was done to find effects of the independent variables on the correctness of the response C. It shows that the angle between the views $\Delta\varphi$ is close to significance: $F(3,432)=2.61$, $p=0.051$. Figure 4.6 shows the effect: the percentage of correct answers is lower when the angle between adjacent images is 8θ than when it is smaller. The other main effects and the interactions were not significant.

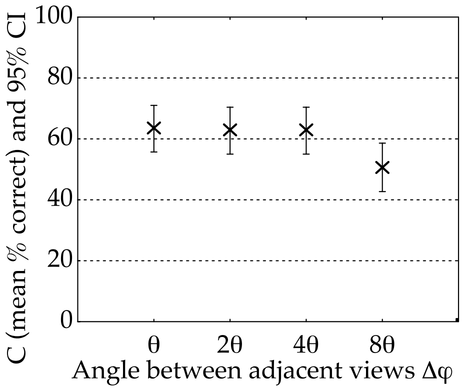


Figure 4.6. The effect of the angle between the views $\Delta\varphi$ on the mean percentage correct answers.

An analysis of variance was done to find the effects of the independent variables on the response time T. Here, the number of available views N proved to be significant: $F(5,432)=4.51$, $p=0.001$.

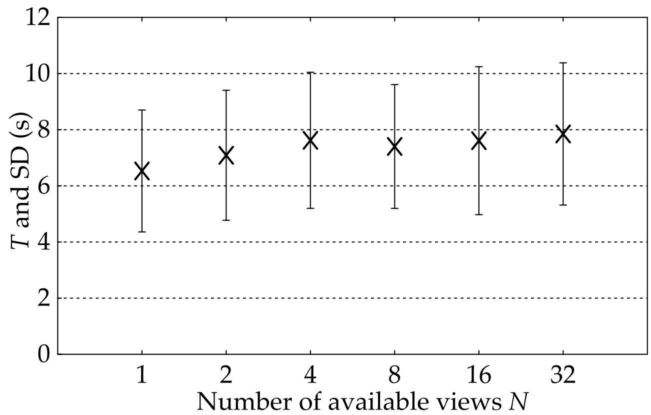


Figure 4.7. The effect of the number of available views on the mean response time.

Figure 4.7 shows the effect: response time increases with the number of views. The effect is quite small: from 1 to 32 views the response time increases from 6.5 to 8 seconds.

The two-way interaction between the number of available views and the angle between two adjacent views also proved significant: $F(15{,}432)=1.87$, $p=0.024$. Figure 4.8 shows the average response time for these conditions. This effect seems to be caused by the high response time in the condition with $\Delta\varphi=4\theta$ and $N=4$, and this value seems accidental.

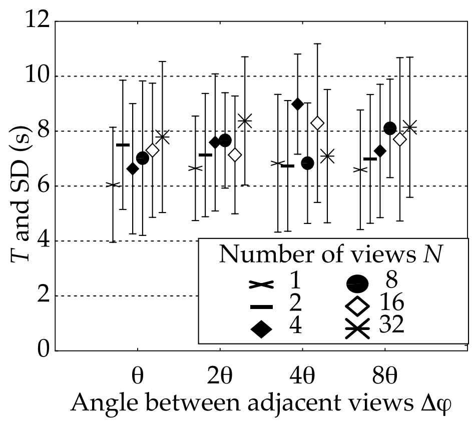


Figure 4.8. The effect of the number of available views and the angle between the views on mean response time.

A detailed analysis of the stimuli and responses indicated that there seemed to be some misinterpretation of the sharpness of edges by the participants. The objects with the highest scores for sharpness, according to the participants, are shown in Figure 4.9 and Figure 4.10. However, these objects have rounded edges, and were meant to be blunt. Instead, participants seem to judge objects as sharp if there is a thin plane at a side of the object. The last column of Table 4.3 shows for each box the percentage of participants judging one of the objects as sharp.

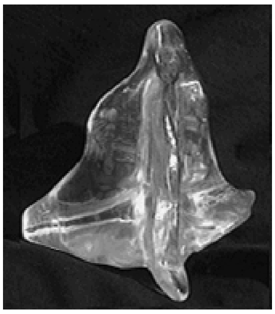


Figure 4.9. The second object from box 2. This object is judged sharp by 83% of the participants of the first experiment.

Figure 4.10. The first object from box 24. This object is judged sharp by 71% of the participants of the first experiment.

Concluding, the results are not quite as expected. First, most variables had no significant effect on the response time or percentage of correct answers. Furthermore, in contrast to our expectations, observer performance decreased with the angle between the views, and the response time increased with the number of available views. Probably, either the number of views available or the resolution was too low to perform this task. Alternatively, the task of detecting sharp edges may simply be too difficult for the participants, causing only higher response times but no performance increase when more different views are provided. Finally, the instruction may have been insufficient. I will have to pay more attention to how the participants interpret sharpness.

Experiment 2- natural inspection

Experiment 1 failed to demonstrate any advantage of the Delft Virtual Window System for finding objects with sharp edges. The resolution and number of views may have been too restricted, but alternatively the task might be too difficult, even when the participants are allowed to handle each object to explore it fully. The last hypothesis is checked in experiment 2, by giving participants the real objects and the same task.

Method

Stimuli, Procedure, Apparatus

The stimuli are real transparent objects, those that were placed in pairs in a box for the recording of the stimuli of Experiment 1. Participants was asked to detect sharp edges on the objects they would be given. It was explained that sharp edges are knife-like edges, sharper than 20˚, and that such an edge need not necessarily be straight or on the outside of the object. They were allowed to inspect each object for about 10 seconds. During inspection, they were allowed to take the objects in their hands, but they had to keep them above a cushion, as the objects break easily when dropped. Participants sat on a chair, in front of a cushion lying on a table. The room was illuminated with fluorescent light. After the inspection, they had to write down their judgement (sharp or blunt) on a form. The same three angles as the labels of the buttons of experiment 1 were printed on the form, and it was explained to participants that edges sharper than the middle angle (30˚) were to be classified as sharp, and that angles larger than 30˚ were blunt.

Prior to the experiment, the participants inspected 14 objects (the objects that also were used for training in Experiment 1). After judging an object, they were told whether their judgement was correct and why this was so.

During the experiment, the participants judged 54 objects. Now, however, they were not told about the correctness of their choice.

Variables, Participants, Design

The independent variable was the object (54 levels). The objects were in the same order as in experiment 1, but one after another instead two at once. The dependent variable was the judgement of the participant (sharp or blunt). The participants were 3 students from the faculty of Industrial Design Engineering (2 women, 1 man) with normal or corrected-to-normal vision. They did not know the purpose of the experiment. Each participant received NLG 10 (USD 5) for taking part.

Results, Discussion

Table 4.2 shows the results. Participant RH said that he felt the sharpness of the edges with his fingers, but IG and ES said they preferred looking to feeling.

There is agreement between the participants about most objects. However, the first object in box 13, 24 and 25 and the second object in box 2 seem to give problems.

There seem to be three problems that may explain the deviating answers. The first problem is misunderstanding of the instructions. For example, the first object in box 21 contained a sharp edge of a non-transparent material, but some informal talking with IG after the experiment showed that he considered this part not to belong to the object. One object in box 11 has a sharp cut in stead of a sharp edge, and ES judged the object as sharp. The second problem is caused by the rounding of the edges. For example the first object in box 24 was a bird-like object where the thin wings were rounded, but ES and RH judged them as sharp. The last problem is difficulty with estimating angles of objects. For example box 25 contains a folded starfish, and the sharpness of its edges is very difficult to estimate.

Concluding, most responses are correct, so detecting sharp edges is not too difficult if the objects themselves can be handled. A few objects cause confusion, which may be solved with more precise instructions. The poor responses of Experiment 1 must be related to inspection via the DVWS.

Table 4.2. Judgements sharp ('*') or blunt ('-') of the participants, and the physical measurements. The stimuli were presented in the same order as in Experiment 1.

	First object in box				Second object in box			
	Participants				Participants			
Box	ES	RH	IG	measured	ES	RH	IG	measured
1	-	-	*	-	*	*	*	*
2	-	-	-	-	*	*	-	-
3	*	*	*	*	-	-	-	-
4	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-
6	*	-	-	-	-	-	-	-
7	-	-	-	-	*	*	*	*
8	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-
12	-	-	-	-	*	-	-	-
13	*	*	-	-	-	-	-	-
14	-	-	-	-	*	*	*	*
15	-	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-	-
17	-	-	-	-	-	-	-	-
18	*	*	*	*	-	-	-	-
19	-	-	-	-	-	-	-	-
20	-	-	-	-	-	-	*	-
21	*	*	-	*	-	-	-	-
22	-	-	-	-	-	-	-	-
23	-	-	-	-	-	-	-	-
24	*	*	-	-	-	-	-	-
25	*	-	*	-	-	-	-	-
26	-	-	-	-	-	-	-	-
27	-	-	-	-	*	*	*	*

Experiment 3- only high resolution

Experiment 2 revealed minor problems with the task itself and with misinterpretation of the instructions. However, these results could not explain the poor responses found in experiment 1. To find out whether the changes of image resolution and number of viewpoints for each subsequent trial caused the problem with inspection via the Delft Virtual Window System, and to avoid rounding effects due to low resolution, two participants judged the boxes of experiment 1 with 32 available views with a high image quality. To allow later comparison with experiment 1 and 2, the instructions were kept the same.

Method

Variables, Design, Participants

The stimuli, apparatus and procedure were the same as experiment 1. The independent variable was the box content (27 boxes). Again, the dependent variable was the judgement of the participant ('sharp' or 'blunt'). Each box was inspected with the Delft Virtual Window System with an image resolution of 512 x 256, 16 grey levels, 32 available views and an angle between two adjacent views of 8θ. Participants were 2 students from the faculty of Industrial Design Engineering (1 woman, 1 man) with normal or corrected-to-normal vision. They did not know the purpose of the experiment. Each participant received NLG 10 (USD 5) for taking part.

Results

Table 4.3 shows the results. The participants perform worse than in experiment 2. The wrong judgements for box 2, 13, 24 and 25 correspond to similar judgements in experiment 2.

For box 7 and 27 it seems that the sharp edge can be detected only from a viewpoint that is not present in the 32 available views. If I compare the percentage of sharp judgements from the first experiment with those of this experiment, and use a difference of 40% as level of significance, boxes 1, 7,14, 18, 20 and 26 seem to be judged more similar to the measured values in the present experiment, while boxes 3, 5, and 8 are judged less similar to the measurements. It is difficult to draw conclusions from these results, as only two measurements are available for each box, but it seems that indeed the switching of the number of available views and the image quality between the trials has lowered the performance of the participants in the first experiment.

Table 4.3. Judgements of the participants. The boxes were presented in the same order as in Experiment 1. Judgements are sharp ('*') or blunt ('-'). Shaded areas indicate boxes that contain difficult objects according to the second experiment.

Box	participant VS	participant XZ	measured	%judgements sharp in experiment 1
1	*	*	*	58
2	-	*	-	83
3	-	-	*	42
4	-	-	-	8
5	-	*	-	8
6	-	*	-	17
7	-	*	*	12
8	*	*	-	25
9	-	*	-	62
10	-	-	-	12
11	-	-	-	25
12	-	-	-	21
13	-	*	-	8
14	*	*	*	29
15	*	-	-	37
16	-	-	-	29
17	-	-	-	62
18	*	*	*	37
19	-	-	-	29
20	-	-	-	37
21	*	-	*	75
22	-	-	-	29
23	-	-	-	4
24	-	*	-	71
25	-	*	-	42
26	-	-	-	46
27	-	-	*	25

Discussion and conclusions

Many results are unexpected, and explanations are difficult to find due to the low number of tested participants in the second and third experiments. The first experiment showed that for finding sharp edges, an angle of 8θ between adjacent views gives *lower* performance of the participants as compared to smaller angles between adjacent views. Apparently, the jerkiness of the movement disturbs the participants. Furthermore, the average response time *increases* with the number of viewpoints. One explanation may be that observers need extra time to interpret the extra views. Finally, Experiment 2 showed that the task can be done when participants can handle the real objects, but that there may be misunderstanding of the instructions, misinterpretation of rounded edges and difficulty with estimating angles of objects. Results from Experiment 3 suggest that limiting the available viewpoints gives additional difficulties when a particular view is required for estimating the sharpness of an edge, and that the manipulation of the parameters may have lowered observer performance in Experiment 1.

The experiments in the following chapters will again try to show the possible trade-off between the number of available viewpoints and static image quality. In order to avoid misinterpretation, another task will be used that can be explained more clearly and allows less misinterpretation by the participants.

5

Detecting connected objects[1]

Informally, experts often indicate that bombs usually consist of a battery, a timing mechanism, a detonator, explosives and wires connecting those parts. Therefore, detecting a wire connecting objects is expected to be a task relevant for a baggage inspector. This chapter describes three experiments in which participants have to detect wires and connections. The first experiment tests the trade-off between the spatial resolution and the number of grey levels of the images, and the number of available views. The second experiment tests whether manual viewpoint selection can replace viewpoint selection via the eye position that was used in the first experiment. The last experiment tests whether, for this task, the total number of views over a given camera range affects observer performance.

To limit the x-ray dose to which the baggage is exposed, the number of available views will be limited in the same way as described in Chapter 4. Again, the observer has just one degree of freedom: the left-right movement (Figure 4.2). Only N images are available, with a constant angle $\Delta\varphi$ between two images. As with the experiment described in Chapter 4, I expect observer performance to increase with increasing camera range, up to 180˚. The camera movement was scaled to a maximum comfortable head movement of ±22.5˚, giving a scale factor $\varphi_{cam} / \varphi_{obs} = N.\Delta\varphi / 45$. But if the angle $\Delta\varphi$ gets too large, observer performance may decline.

In line with results from other experiments (Ranadivé, 1979; Swartz, Wallace and Tkacz, 1992; Snyder, 1973; Uttal, Baruch and Allen, 1995; Braunstein, Hoffman, Shapiro, Andersen and Bennett, 1987), the hypotheses are that the resolution R, the number of grey levels G and the number of views N will improve the performance of the observer from a threshold up to a saturation level. When the value for a variable falls below its threshold, the task cannot be done, regardless of the levels of the other variables. For example in the task used in this experiment, the resolution threshold is about 256x128. At lower resolutions it is impossible to see the wire to be detected, even when multiple views are available. There is also a saturation level for the resolution, at which increasing the resolution brings no improvement in task performance. For most tasks, this saturation level will be well below the visual acuity. Furthermore, the threshold and saturation levels will depend on the levels of the other variables (Smets and Overbeeke, 1995). For example, for a high resolution R the threshold of the number of grey levels G will be lower, and increasing the available views will lower the resolution threshold.

Finally, it is interesting to compare the performance of an observer using eye position to select the desired viewing angle with an observer using a knob to do so. The pictures that are presented are the same in these cases, but the way they 'feel' is different: in the first case, the observer gets the impression that he is moving around a box, in the second case he gets the impression that he is turning the box indirectly. The first case seems more 'natural', and I expect participants to perform better than in the second case. If they

[1] This chapter is based on Pasman, Smets and Stappers (1997).

perform the same, it may be unnecessary to use expensive eye position trackers for the X-ray inspection apparatus.

Experiment 1- view selection with eye position

This experiment tests the effects of image resolution, number of grey levels, number of available views, and the angle between the views on the ability of the participants to detect a wire, their ability to judge whether the wire connects two objects, and their response time.

Method

Stimuli

The stimuli are very similar to those of the first experiment in Chapter 4, but now some boxes also contain a wire. Figure 5.1 shows two examples of the actual stimuli. In each box, two objects were present. Some boxes also contained a wire. In some of the boxes, the wire connected both objects. This configuration was derived from the usual construction of a bomb: a wire between a battery and a detonator. The wires used had a diameter of 0.3 mm. For recording the stimuli, the same setup as described in Chapter 4 was used.

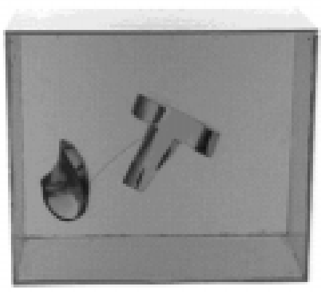


Figure 5.1a. Impression of stimuli: Two objects and a wire in a box.

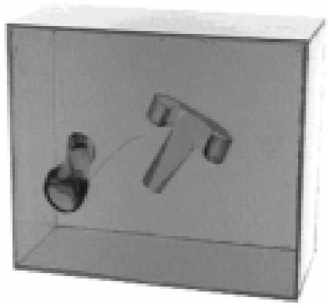


Figure 5.1b. Viewed from the right, the objects appear unconnected.

To model the possible situations in x-ray baggage inspection, I classified the contents of the boxes into three types *T*: no wire, connected and trick. Boxes of type 'no wire' did not contain any wire, only two objects. Boxes of the type 'connected' contained a wire that clearly connected both objects: in the front view the wire crossed both objects. Boxes of type 'trick' tried to fool the observer. An example 'trick' is when the second object was placed behind the first, disturbing the front view of a wire connecting the objects. Another 'trick' was to place the wire in such a way that the objects seemed to be connected in front view (Figure 5.1a), but not in side view (Figure 5.1b).

Preceding each trial, the required views of the box were read from hard disc, reduced in number of grey levels and resolution if necessary, and stored in working memory. This caused a pause of about 10 s between trials. During the trial, the appropriate images were shown from working memory on the screen. For a reduction of the number of grey levels, the original 16 grey levels were divided into 4 or 8 groups, and for each group the brightest value was taken. Informal evaluation by the experimenter indicated that this reduction had little effect on image contrast. To reduce resolution, the image pixels were grouped in 2 x 2 pixels whose intensity was averaged.

Apparatus

In Figure 5.2, an overview of the experimental setup is shown. The room was illuminated at 150 Lux by fluorescent lighting. The turn knob at the right on the table was present only in Experiment 2. The computer, display, reduction screen and viewpoint tracker were the same as in the first experiment of Chapter 4.

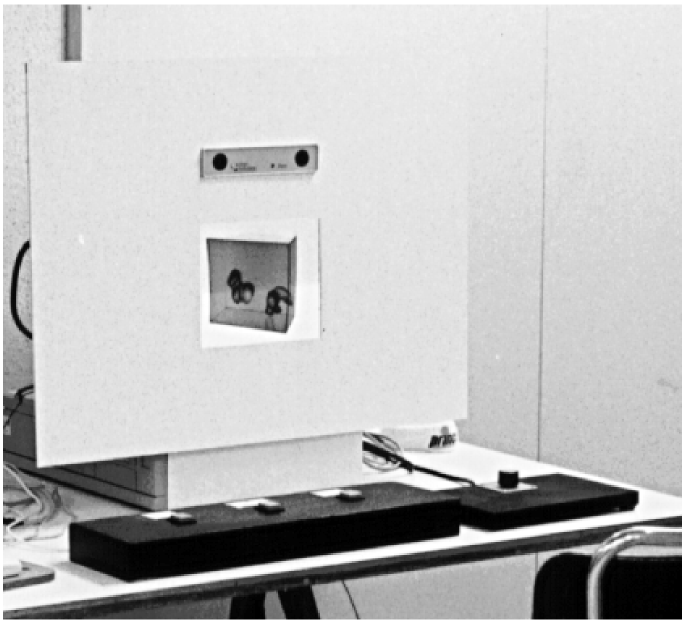


Figure 5.2. Overview of the experimental setup. Display with reduction screen in front, and the head tracker sensor on top of it. On the table on the left the button box and on the right the knob. The knob was present only in Experiment 2.

Independent variables, hypothesis

The hypotheses bear on the effects of the image resolution R, the number of grey levels G, the number of available views N, the angle between the views $\Delta\varphi$ and the type of the box contents T. Table 5.1 shows the independent variables and the tested levels. I estimated the threshold and saturation levels in some pilot sessions, and used these as lowest and highest level for the variables.

Table 5.1. The independent variables. θ = 22.5˚ / 32.

Name of variable	Description	Used values
R	Image resolution	256 x 128 pixels, 512 x 256 pixels
G	Number of grey levels	4 , 8 , 16
N	Number of available views	1 , 2 , 4 , 8 , 16 , 32
$\Delta\varphi$	Angle between the views	θ , 2 θ , 4 θ , 8 θ
T	Type of box contents	no wire, connected, trick

A higher ability of the participants to detect a connection and a lower response time were expected with increasing camera range, up to a camera range of 180˚. For this range, N = 32 views were expected to be near the saturation level. This gives an angle between the views $\Delta\varphi$ = 180˚ / 32 (In fact 33 frames are needed to reach the 180˚, so the range is slightly smaller). To simplify the notation, angles are expressed as multiples of θ = 22.5˚ / 32. Figure 5.3 gives an impression of the views for different resolutions R and numbers of grey levels G.

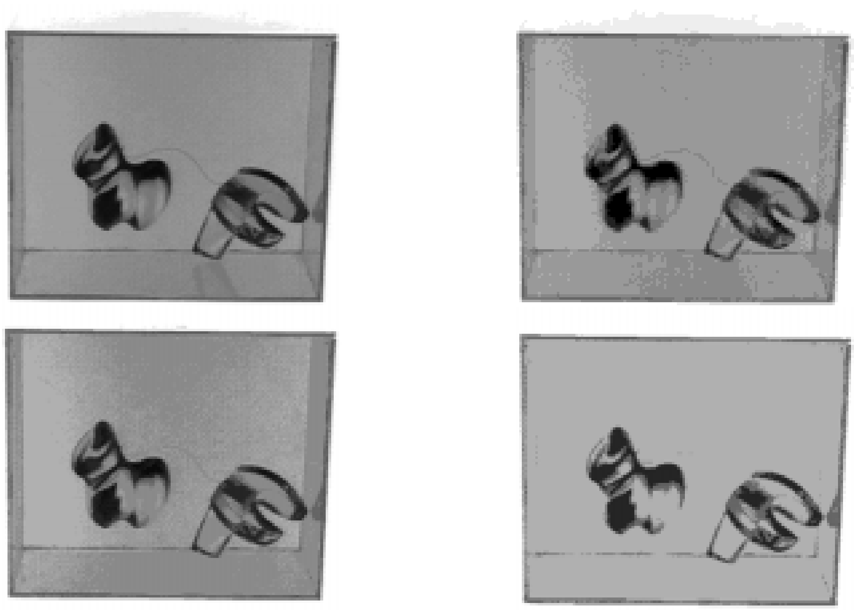


Figure 5.3. Screen impressions. Top left: R=512x256 and G=16. Top right: R=256x128 and number of grey levels G=16. Bottom left: R=512x256 and G=8. Bottom right: R=512x256 and G=4.

Dependent variables

Table 5.2 gives the measured dependent variables. The participant had to choose whether or not the two objects in the box were connected. A third choice ‘wire, but not connected’ was available. This was necessary to make a difference between seeing no wire at all and seeing a wire that does not connect the objects. Furthermore, the response time was measured and analysed to find uncertainty with difficult stimuli.

Table 5.2. Dependent variables.

Description	Possible values
chosen button	no wire, wire but not connected, connected
response time	> 0 s

Participants, Design

Each participant had to judge 30 boxes (10 of each type T). The order of presenting the boxes was the same for all participants, but randomized over all conditions. The 2(R) x 3(G) x 6(N) x 4($\Delta\varphi$)=144 conditions of each box type T were distributed randomly over 15 participants (150 judgements for each box type T). The remaining 6 judgements for each type were discarded. This way of defining the conditions for 15 participants was repeated 5 times to get 5 measurements for each condition, so 75 participants were tested.

The participants were 75 students from the Faculty of Industrial Design Engineering (25 women, 50 men). They were naive and paid volunteers. The first 15 participants received NLG 10 (USD 5) for taking part, the other 60 NLG 7.50 (a loaf of bread costs about NLG 2).

Procedure

For instruction, the participant was told that boxes would be shown on the screen, with two objects in each box. His task would be to decide whether there was a thin wire in the box and, if so, whether it connected both objects. He had to choose between 'no wire', 'wire but not connected' and 'connected' by pressing one of three buttons labelled with these words.

A participant could view the box from different sides by moving his head to the left or to the right. He was instructed to inspect the box from all sides before making a judgement, and to base his choice on the things he could see (and not the things he could imagine). During the training, the participant got a warning from the experimenter if he did not do so. He was warned that he had just 10 seconds to look at the box, but he was instructed to try to make the right choice, and that a quick response was less important.

To get used to the range and speed of the Dynasight tracker, the participant was trained without views being displayed on the monitor. The tracker provided feedback by a control light which was green if the reflector was in track, and red if it was not. At this stage one participant was found to be colour blind, but this did not pose problems.

For the training, the participant was shown 10 different boxes. One box was shown twice under different movement conditions. After the participant had made his choice, the screen showed the right choice, whether he had made the right choice, his response time and the range of his eye positions.

During the experiment, the participant was shown 30 boxes. All these boxes contained different objects and wire configurations. After the experiment, the participant was told how many stimuli were recognised correctly.

Overall, each experiment took about 25 minutes.

Results

The participants had to make two choices: 1 'is there a wire?' and 2 'if so, does it connect the objects?'. Therefore, the judgements can be split into two categories: 1 when is it possible to see a wire? and 2 when is it possible to see whether or not the wire connects both objects? Furthermore, the response time will be analysed. The results are evaluated in this order, with an analysis of variance to find the significant effects and by a graphical representation to explain the effects. An alpha level of 0.05 was used to test the significance of all effects.

From analysis of the eye movements of the first 15 participants, their average viewing distance showed 45 cm, with a standard deviation of 9 cm. Thus, the average viewing distance matched the camera distance used for recording.

Visibility of the wire

The judgement 'wire but not connected' and 'connected' both indicate that the participant saw a wire. For this analysis, only boxes of type T = 'connected' are used because for this box type the wire can be seen in the front view, and other views may be unavailable in some conditions. Table 5.3 shows the significant main effects and interactions.

Table 5.3. Significant interactions for the visibility of a wire

Interaction	F	p
G	$F(2,576)=177.38$	<0.001
N	$F(5,576)=6.59$	<0.001
R	$F(1,576)=288.72$	<0.001
$G \times R$	$F(2,576)=72.39$	<0.001
$N \times \Delta\varphi$	$F(15,576)=2.13$	<0.01
$N \times R$	$F(5,576)=4.22$	<0.001

For all independent variables except for the angle between the views, the performance increases with increasing value of that variable, as was expected. The non-significance of the angle between the views $\Delta\varphi$ is unexpected. An explanation may be that I analysed only the measurements from the T='connected' case, and that the effect would have been significant if more measurements had been done for this condition.

In Figure 5.4 - 5.6, the vertical axis shows ratios from 0 (no participant saw a wire) to 1 (all participants saw a wire). The middle point of each marker shows the average value, and the upper and lower points show the limits of the 95% confidence interval (Loosen, 1994). The shapes of the markers indicate different conditions.

Figure 5.4 shows the interaction between G and R. For 16 grey levels, the score is nearly perfect (>94% for low resolution). For fewer than 16 grey levels, the resolution has much more effect on the visibility of a wire.

Figure 5.5 shows the interaction between N and $\Delta\varphi$. Only a combination of an angle between the views larger than 4θ and a large number of views larger than 16 has a clearly positive effect on performance. This indicates that a wider range of inspection angles leads to increased visibility of the wire.

Figure 5.6 shows the interaction between N and R. At high resolution, performance is so good that a larger number of views produces no further increase.

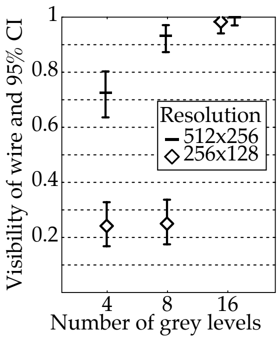


Figure 5.4. Effect of the number of grey levels and resolution on the visibility of a wire.

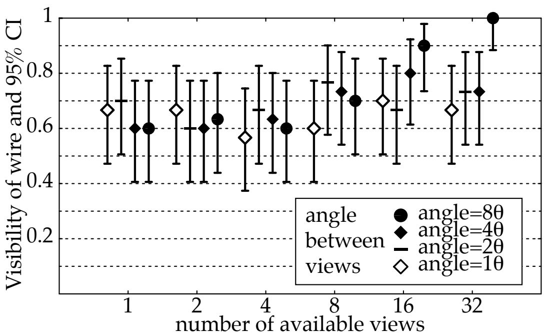


Figure 5.5. Effect of the number of views and the angle between the views on the visibility of a wire.

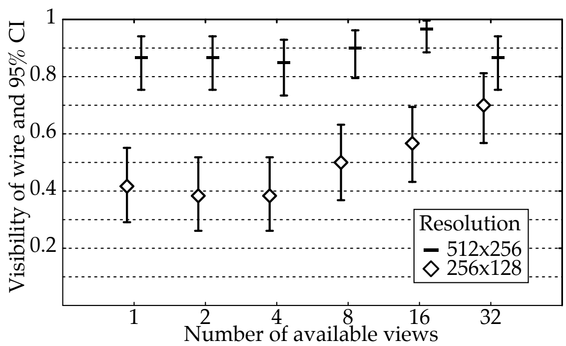


Figure 5.6. Effect of the number of views and the resolution on the visibility of a wire

Correct-ratio

Correct judgements are those judgements of the observers that match the actual situation in the box. The correct-ratio is the ratio of the correct answers to the total number of answers. In Table 5.4 the values for the main and interaction effects are shown. All main effects are significant.

For all variables except box type T, the correct-ratio increases with increasing value of that variable, as was expected. The angle between the views $\Delta\varphi$ has a very small effect (Figure 5.7). This may explain why the angle between the views had no significant effect on the visibility of the wire.

Table 5.4. Significant interactions for the correct-ratio.

Interaction	F	p
G	$F(2,1728)=53.91$	<0.001
N	$F(5,1728)=16.88$	<0.001
$\Delta\varphi$	$F(3,1728)=4.02$	<0.01
R	$F(1,1728)=121.13$	<0.001
T	$F(2,1728)=587.34$	<0.001
$G \times R$	$F(2,1728)=22.66$	<0.001
$G \times T$	$F(4,1728)=73.92$	<0.001
$N \times \Delta\varphi$	$F(15,1728)=2.88$	<0.001
$N \times T$	$F(10,1728)=3.63$	<0.001
$R \times T$	$F(2,1728)=53.91$	<0.001
$G \times N \times \Delta\varphi$	$F(30,1728)=1.70$	<0.01
$G \times R \times T$	$F(4,1728)=12.54$	<0.001

The three-way interaction between number of grey levels G, resolution R and box type T (Figure 5.8) fully explains the interactions between G and R, between G and T and between R and T. Boxes without a wire are nearly always judged correctly, maybe because a low image quality hides wires, causing a bias towards a 'no wire' judgement. For boxes of the type 'connected', the correct-ratio depends largely on the visibility of the wire: compare Figure 5.4. For boxes of type 'trick', resolution R and number of grey levels G are less effective.

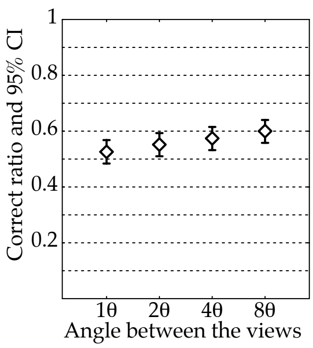


Figure 5.7. Effect of the angle between views on the correct-ratio

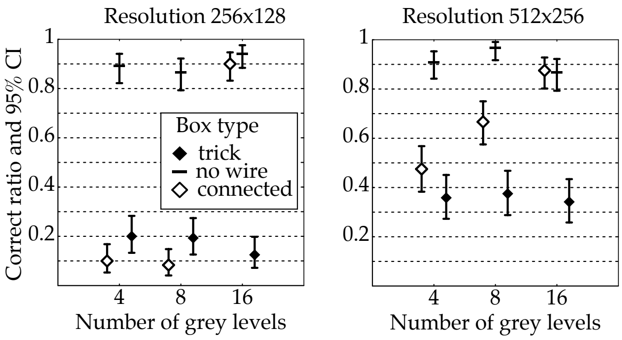


Figure 5.8. Effect of number of grey levels, resolution and box type on the correct-ratio

The interaction between the number of available views N and the box type T (Figure 5.9) shows that N has a positive effect mainly in the case of the boxes of type 'trick'. N has less effect on the correct-ratio for boxes of type 'connected', and has no effect in the case of boxes with 'no wire'.

The interaction between the number of available views N and the angle between the views $\Delta\varphi$ (Figure 5.10) shows that only a combination of a large number of views and a large angle between the views improves the correct-ratio. Only the combination of 16 or more views and an angle of at least 4θ is really effective.

I did not recognize clear patterns in the three-way interaction between G, N and $\Delta\varphi$: there are a large number of cells, and they have large confidence intervals.

Both Figure 5.9 and Figure 5.10 show an increase of the correct-ratio from up to 8 available views. This seems to indicate a threshold level for the number of views. However, in the case of trick boxes only, the task becomes impossible when the number of available views falls below its threshold level.

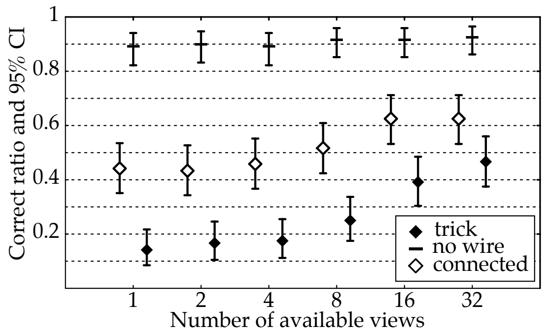


Figure 5.9. Effect of number of views and box type on the correct-ratio

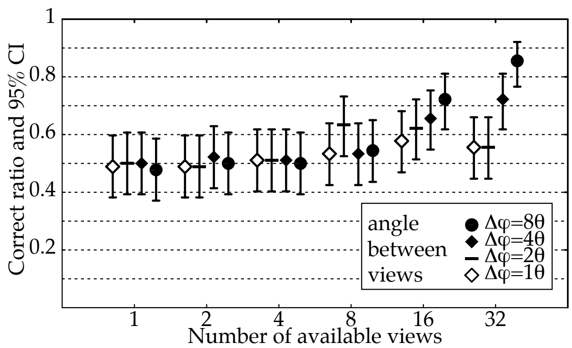


Figure 5.10. Effect of number of views and angle between views on the correct-ratio

Post-hoc analysis of camera-range

Possibly the ability for the observer to choose the view from a large range is more important than the number of views in that range and the impression of rigidity. A first indication for this hypothesis is that the angle between the views $\Delta\varphi$ is a measure for the jerkiness in the spatial impression when moving. It has a very small influence on the correct-ratio (Figure 5.7). This suggestion is strengthened by Figure 5.10, showing that a small angle between views $\Delta\varphi$ and a small number of views N has no effect, and that the situation with $N = 16$ and $\Delta\varphi = 8\theta$ has both the same correct-ratio and the same camera range as the situation with $N = 32$ and $\Delta\varphi = 4\theta$.

To test this hypothesis, a new variable expressing the camera range $(N - 1) \cdot \Delta\varphi$ is introduced. Figure 5.11 shows the camera range as a function of the angle between views and the number of views. These camera ranges were grouped into eight classes of similar value, with the range roughly doubled in each subsequent class.

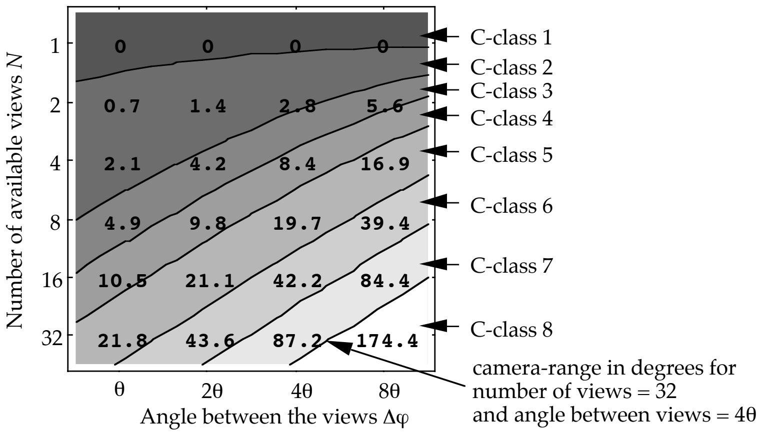


Figure 5.11. Camera range for all combinations of the number of views and the angle between views. These combinations are grouped into eight C-classes of similar range.

The independent variables are now resolution R, number of grey levels G, and camera range class (C-class) C. Because the C-classes comprise between 1 and 4 $\Delta\varphi$-N pairs, the number of measurements from the experiment is not the same for each C-class. C-class 8 is the smallest, and it contains only 5 measurements per condition. Therefore, from the other C-classes only the first 5 (random, since the conditions were in random order) measurements are taken for further analysis.

Table 5.5 shows the significant main and interaction effects according to analysis of variance. The main effect of T, G and R, and the interactions between T and G, between T and R, between G and R and between T, G and R were discussed under 'Correct-ratio'.

Table 5.5. Significant interactions for the correct-ratio using the C-class.

Interaction	F	p
T	$F(2,576)=168.11$	<0.001
C	$F(7,576)=13.43$	<0.001
G	$F(2,576)=10.72$	<0.001
R	$F(1,576)=35.26$	<0.001
$T \times C$	$F(14,576)=3.07$	<0.001
$T \times G$	$F(4,576)=15.21$	<0.001
$T \times R$	$F(2,576)=6.48$	<0.01
$G \times R$	$F(2,576)=6.77$	<0.01
$T \times G \times R$	$F(4,576)=4.75$	<0.001
$C \times G \times R$	$F(14,576)=2.15$	<0.01

The effect of the interaction between C and T is shown in Figure 5.12. If we compare the effect of N (Figure 5.9) with the effect of C (Figure 5.12) on the correct-ratio, the C-class has a much stronger effect, even with box type 'connected' for which I did not expect positive effect from a bigger camera range when compared with the number of views. There seems to be a jump upwards in the correct-ratio from C-class 6 to 7 (camera range of 45° and 90°). Thus, the threshold noticed at 8 available views (Figure 5.10), actually seems to be a threshold at a camera range of 45°.

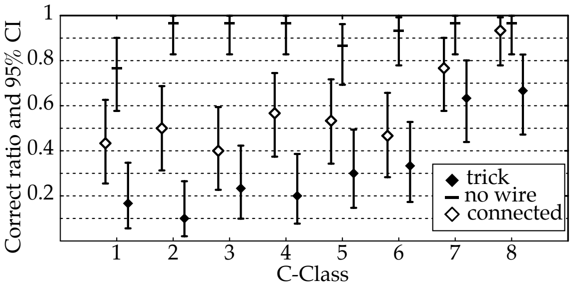


Figure 5.12. Effect of C-class and box type on the correct-ratio

The interaction between C-class, number of grey levels and resolution (Figure 5.13) shows that C-class has little effect when 16 grey levels and low resolution are used. For the other conditions, a bigger camera range improves the correct-ratio. Furthermore, at low resolution the threshold seems to lie at about C-class 6, while this is not the case for the high resolution conditions.

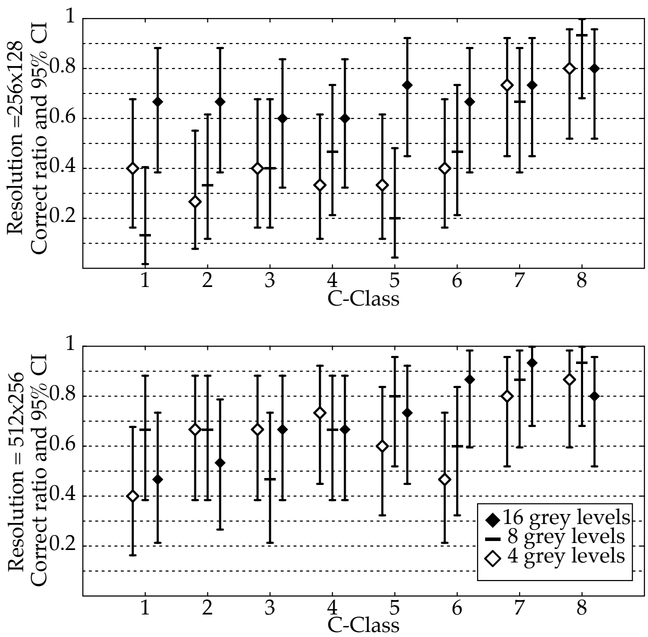


Figure 5.13. Effect of C-class, resolution and number of grey levels on the correct-ratio.

Response time

The response times were analysed with an analysis of variance (Table 5.6). Three main effects were found to be significant: the number of available views, the resolution and the box type. Interactions were not found to be significant.

Table 5.6. Significant interactions for the response time.

Interaction	F	p
T	$F(2,1728)=15.073$	<0.001
R	$F(1,1728)=11.552$	<0.01
N	$F(5,1728)=10.963$	<0.001

Figure 5.14 shows the effect of box type on mean response time. The response time for boxes of the trick type is slightly longer than for the other types. Figure 5.15 shows that response time is slightly lower in the high resolution condition than in the low resolution condition. Figure 5.16 shows that up to 4 available views, response time increases slightly with the number of available views. More than 4 views do not affect the response time.

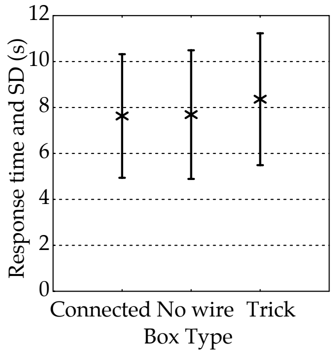


Figure 5.14. The effect of box type on mean response time.

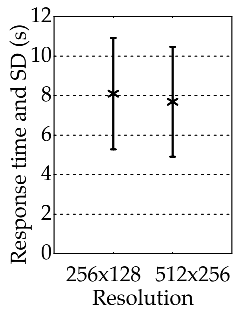


Figure 5.15. The effect of resolution on mean response time.

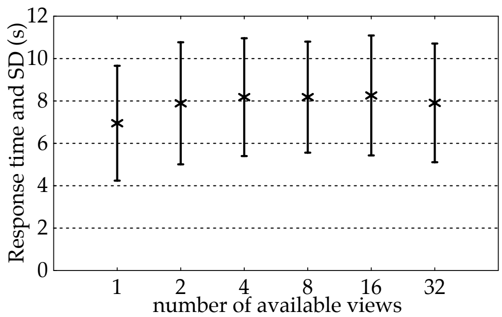


Figure 5.16. Effect of number of available views on mean response time.

As with the correct-ratio, the results for the response time might become clearer if the number of views and the angle between the views are replaced with the C-class grouping. For the C-class grouping, the effects of the variables on mean response times were tested with an analysis of variance (Table 5.7). Three main effects were found to be significant: resolution, C-Class and box type.

Table 5.7. Significant interactions for response time using C-class.

Interaction	F	p
T	$F(2,576)=5.188$	<0.01
R	$F(1,576)=8.809$	<0.01
C	$F(7,576)=4.521$	<0.001

The effects for resolution, number of available views and box type are similar to those of Figure 5.14 and Figure 5.15. The effects of C-class are shown in Figure 5.17. Up to C-class 3, that is a camera range of about 5°, response time seems to increase slightly with increasing C-class. Higher C-classes do not affect the response time.

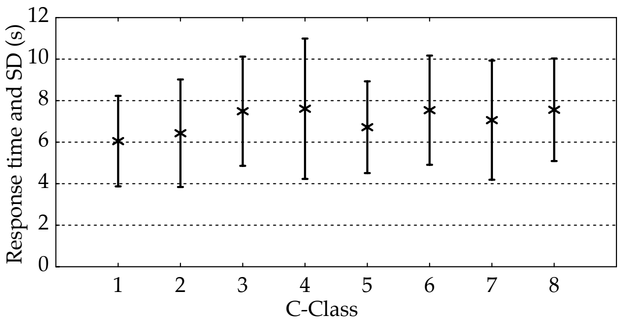


Figure 5.17. Effect of C-Class on mean response time.

The results indicated that camera range, and not the angle between views, is important for this task. This will be tested explicitly. However, the plan was to compare viewpoint selection by eye position with manual viewpoint selection, and the following experiment will test this first.

Experiment 2- knob movement

In this experiment, the viewpoint selection by eye position of experiment 1 is replaced by viewpoint selection by a knob. It was expected that selecting the view by eye position would work better than selecting it with a knob, since it seems more natural to look around a box than to turn a knob that indirectly causes a box to turn.

Method

Apparatus, Stimuli, Procedure

The same apparatus as in Experiment 1 was used. The eye position tracker was replaced by a turning knob. The turning knob was a wire-wound potentiometer with a mechanical turning angle of 270°. φ_{obs} now indicated the knob position. The camera movement was scaled to use the range of the potentiometer, giving $\varphi_{cam} / \varphi_{obs} = N.\Delta\varphi$ / 240˚. The turning knob was read out by an A/D converter. A millisecond timer was used to update the displayed view on the screen at 37 Hz, to match the update rate of Experiment 1.

The same stimuli were presented in the same order as in Experiment 1. The training and experiment ran as in Experiment 1, except that the training and explanation of the viewpoint tracker was replaced by an explanation about the turning knob.

Variables, Design, Participants

The dependent and independent variables and the design were the same as in Experiment 1. Each participant had to judge 30 boxes (10 of each type *T*). The order of presenting the boxes was the same for all participants, but randomized over all conditions. The 2(*R*) x 3(*G*) x 6(*N*) x 4($\Delta\varphi$)=144 conditions of each box type *T* were distributed randomly over 15 participants (150 judgements for each box type *T*). The remaining 6 judgements for each type were discarded.

The participants were 15 students from the Faculty of Industrial Design Engineering (4 women, 11 men). They were naive and paid volunteers, and were paid NLG 7.50 (USD 3.75) for their participation.

Results

To test for a possible difference between experiment 1 and 2, the correct-ratio of the first 15 participants of Experiment 1 is taken and compared with the results of the participants of Experiment 2 with an analysis of variance. The viewpoint selection method was not found to be significant: $F(1,567)=0.467$, $p=0.495$. Interactions with the method of choosing this angle do not prove significant either. Therefore, the hypothesis that selecting the view by eye position gives a higher observer performance than selecting the view by a knob is not confirmed.

Similarly, response times were compared with the response times of the first 15 participants of Experiment 1 with an analysis of variance. Here, the viewpoint selection method was found to be significant: $F(1,876)=53.52$, $p<0.001$. Furthermore, the resolution was found to be significant: $F(1,876)=5.74$, $p<0.05$. Figure 5.18 shows both effects: participants selecting the viewpoint by eye position work significantly faster than those selecting the viewpoint manually.

The effect on response time is surprising: it is contrary to our hypothesis, and no effect was found on the correct-ratio. It seems that the average response time of the first 15 participants in the viewpoint selection by eye position condition is faster than the average as shown in Figure 5.15. In fact, average response times as found for the manual viewpoint selection are very close to those found in Figure 5.15 for viewpoint selection by eye position. Therefore this result for the response time seems dubious.

Concluding, the findings suggest that, for the discussed task, both methods of selecting the viewpoint work equally well.

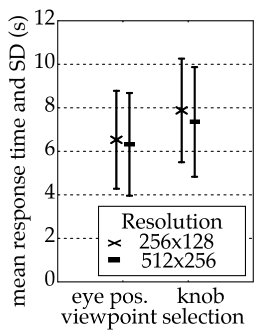


Figure 5.18. Effect of viewpoint selection method and resolution on mean response time.

Experiment 3-constant camera range

The results for the C-classes suggested that the total camera range, and not the angle between the views, is relevant for detecting connections. In the present experiment this will be tested explicitly, by manipulating the number of views over a fixed camera range of 180˚. Given the results of the previous experiments, it is expected that the angle between views will have no effect on the correct-ratio and response time of the participants.

Method

Apparatus, Stimuli, Procedure

The apparatus, stimuli and procedure were the same as in the first experiment.

Variables

The independent variables and their levels are shown in Table 5.8. The angular resolution $\Delta\varphi$ was manipulated, with possible values 8θ, 16θ, 32θ, 64θ and 128θ ($\theta = 22.5˚ / 32$, as above). The number of available views was such that the total camera range was 180˚ (Figure 5.19) (i.e., 33,17,9,5 and 3 views). The three box types *T* were the same as in the previous experiment: Boxes of type 'no wire' did not contain any wire, only two objects. Boxes of the type 'connected' contained a wire that clearly connected both objects: in front view the wire crossed both objects. Boxes of type 'trick' tried to fool the observer. The first experiment showed the biggest effect of the number of viewing angles at an image quality of 512x256 pixels with 16 grey levels (see Figure 5.13), so I used only this image quality in the current experiment. The dependent variables were the same as in the previous experiments, response choice and response time (Table 5.2).

Table 5.8. The independent variables.

Name of variable	Description	Values used
$\Delta\varphi$	angle between the views	8θ, 16θ, 32θ, 64θ, 128θ
T	Type of box contents	no wire, connected, trick

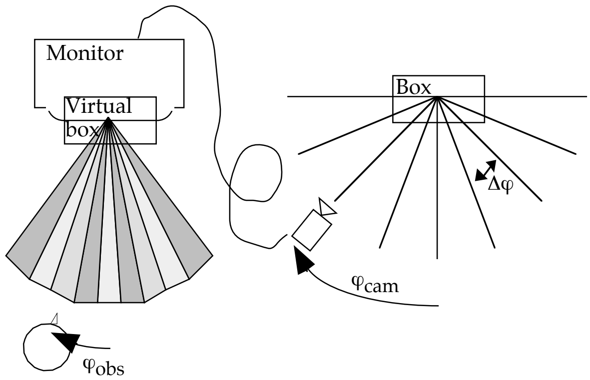


Figure 5.19. The total camera range is 180°. The angle Δφ between the available views is manipulated. The movements of the observer are magnified 4 times in the experiment.

Design, Participants

Each participant saw all 30 boxes once. The conditions were randomized over the participants in such a way that after five participants each box had been examined once in every condition, thus giving 10 measurements per condition. This procedure was repeated to obtain conditions for ten participants, of whom only the first eight were tested. To get an equal number of measurements per condition, the first 13 results for each condition were used for analysis.

The participants were 8 students from the Faculty of Industrial Design Engineering (4 women, 4 men). They were naive and paid volunteers. They were paid NLG 7.50 (USD 3.75) for their participation.

Results

An analysis of variance (Table 5.9) showed that there is no significant interaction between the correct-ratio and the angle between the views: $F(4,195)=0.690$, $p>0.5$.

Figure 5.20 shows the correct-ratio and 95% confidence interval (CI, Loosen, 1994) for each angle. The results confirm our expectations that the angle between the images has no effect on the correct-ratio.

The box type shows no significant interaction with the correct-ratio. This can be explained because the previous experiments suggested that high resolution combined with large camera range allowed high performance for all box types.

Table 5.9. Results of an analysis of variance of correct-ratio.

Interaction	F	p
$\Delta\varphi$	$F(4,180)=0.706$	0.589
T	$F(2,180)=1.412$	0.246
$\Delta\varphi \times T$	$F(8,180)=1.412$	0.190

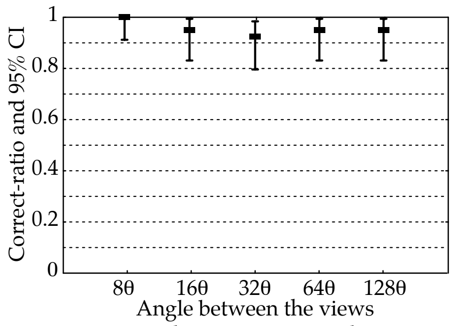


Figure 5.20. Effect of the angle between the available views on the correct-ratio.

Analysis of variance of response time (Table 5.10) showed that there is a significant interaction between response time and box type: $F(2,180)=17.176$, $p<0.001$. Figure 5.21 shows the mean response time and the standard deviations for each type. It might be expected that boxes of type 'connected' can be judged faster than boxes of type 'trick', because finding a trick-wire in a 'trick' box may imply a new search for another wire. But, surprisingly, the 'connected' type takes the longest response time.

Figure 5.22 shows the effect of the angle between the views on the mean response time. Although not significant, the tendency suggests that a smoother coupling improves observer performance. This hints that, in contrast with our expectations, observers are disturbed slightly by larger angles between the views.

Table 5.10. Results of an analysis of variance of response time.

Interaction	*F*	*p*
$\Delta\varphi$	$F(4,180)=1.227$	0.301
T	$F(2,180)=17.178$	<0.001
$\Delta\varphi \times T$	$F(8,180)=1.421$	0.190

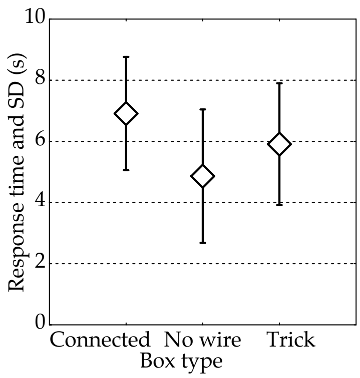


Figure 5.21. Effect of box type on response time.

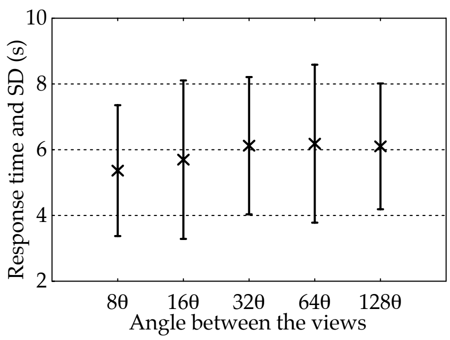


Figure 5.22. Effect of angle between views on response time.

Conclusions

As was already suggested by the results of the first experiment, a front view and two side views are sufficient to judge whether two objects are connected. A large angle between the views seems to interfere with the rigid 3D impression of the scene, and has a minor though insignificant effect on response time. However, large angles between views pose no problems in detecting connections between two objects.

General discussion and conclusions

Summarizing, the most important conclusions of the present experiments are:

1 Up to the tested range of 180˚, a larger camera range can compensate for a small number of grey levels and low image resolution.
2 Down to the tested number of 3 views over a range of 180˚, the number of views over a range is not important for detecting whether two objects are connected.
3 In order to reach a certain observer performance, a trade-off can be made between number of grey levels, resolution and camera range. This trade-off is different for a wire detection task and for judging whether objects are connected.
4 Selecting a view by eye position works just as well as selection with a turning knob.
5 For the detection of a wire, 16 grey levels is sufficient.

This experiment used mock-up baggage with simplified contents and a simplified inspection task. Nevertheless, it is expected that similar effects will occur when the DVWS is applied to baggage inspection. But it is premature to conclude that just three views (maybe even two as x-ray views have no diffraction as our transparent stimuli have) will be sufficient to do the baggage inspection task. The experiment in the next chapter investigates the effect of a small number of available views over a fixed camera range on the inspection of more complex scenes.

It is expected that it is important for a baggage inspector to be able to detect wires and to which objects these wires connect. Therefore, the present result suggest that providing 3 views over a range of 180˚ will be useful for baggage inspection. For wire detection, the usual number of grey levels in an x-ray scan (more than 256) would seem more than sufficient.

Another important result for x-ray baggage inspection via the DVWS is that viewpoint selection by eye position can be replaced with manual viewpoint selection. Viewpoint selection by eye position might tire the inspectors and thus make them less alert (McVey, 1970). Furthermore, viewpoint selection by eye position might be undesirable for aesthetic reasons: for example a number of inspectors indicated that they would not appreciate markers for a head tracker on their cap.

Finally, a lower resolution for each view may be used if multiple views are available. The recording of such images may reduce the x-ray dose required to take that image, and therefore making available N views does not necessarily imply an N-times as high exposure of the baggage to x-rays.

Need for geometric correctness of the display

Some virtual window systems, e.g. Ware, Arthur and Booth (1993), require the precise eye location of the observer to make the correct projection of the scene on the monitor screen. Deering (1992) even makes perspective corrections for the thickness of the glass of the monitor screen. Are such geometric corrections needed to do a task correctly?

The adapted Delft Virtual Window System used in the present experiment corrected the view only for the angular position of the observer, and not for his viewing distance. Furthermore, only horizontal observer movements were coupled. Because 100% correct scores are reached in the condition with 32 views and an angle between the views of 5.625°, it can be concluded that, for this task, neither correcting for viewing distance nor the ability to make vertical movements was essential for correct observer performance. Furthermore the highest correct-ratio was found at a scaling of movements camera:head=4:1, where 1:1 would match the principle of the DVWS. So this experiment does not support the necessity of geometric corrections.

It seems that providing appropriate views is more important than providing a geometrically correct presentation. Chiruvolu, Hwang and Sheridan (1991) discuss this issue and find that, for putting a peg in a hole on a moving object, a clear focus on the goal is needed. In a report from Martin Marietta Aerospace (1986) it was shown that two orthogonal views are sufficient for a module replacement task in space. To increase observer performance, making all information available that would be available in a natural situation will not always improve observer performance, but instead the display system should be optimized to the task demands.

The issue of the need for geometric correctness of the display system will be investigated in more detail in the experiment described in Chapter 7.

6

Following a wire through a knot

For enhancing task performance in spatial tasks, a good spatial impression would appear to be necessary. More specifically for the Delft Virtual Window System used here, it is expected that providing only a small number of views will interfere with the depth impression, and thus affect performance. However, the experiment described in Chapter 5 showed that for detecting wires connecting objects two oblique views are sufficient. For that task, the number of additional views between these extreme views was found to have no effect on observer performance. Apparently no spatial impression is needed for this task, and probably for most tasks. For some tasks, such as detecting the presence of an object, a single view will suffice. Other tasks, such as estimating the volume of an object, require more spatial information or a specifically advantageous viewpoint, for example when inspecting the inner surface of a cylinder. For transparent scenes, the spatial complexity of the scene seems to be important in determining the effect of the number of available views on observer performance. This hypothesis is tested by presenting a more complex scene, and comparing observer performance with the results of the experiment described in Chapter 5. Figure 6.1 shows a view of the more complex scene used in the present experiment: a transparent knot.

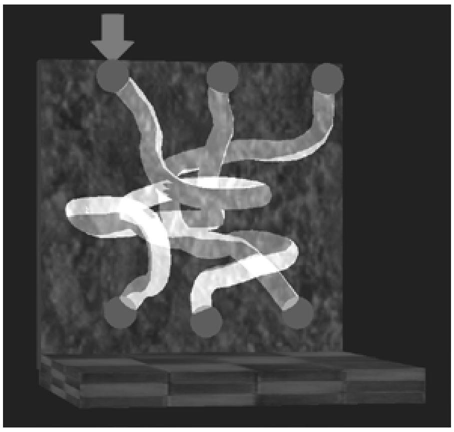


Figure 6.1 (**see colour figure on right cover flap**). Example of a knot. The arrow indicates the top end of one of the wires. Observers had to find the bottom end of the same wire.

For x-ray baggage inspection, the number of available views has to be minimized, because taking an x-ray photograph for each available view will expose the baggage to a certain level of x-rays. The number of views to be taken can be halved by presenting only a single image instead of a stereoscopic image at each viewpoint. However, looking at a single image with two eyes will provide stereoscopic cues about the flatness of the display to the observer, and these cues may impair the spatial impression gained by the observer (Koenderink, van Doorn and Kappers, 1994). The effects of looking with both eyes at a single image but in the presence of movement parallax are not well understood, and furthermore many of the existing findings may not hold for transparent scenes, as Kersten and Bülthoff (1991) showed that "vision of transparency may involve a two-way interaction with the computation of structure from motion". Therefore, the present experiment also tests whether performance is affected by monocular versus binocular viewing of transparent knots.

Previous work

The theory relating to multiple viewpoints and observer performance was discussed in Chapter 2. Both the multiple viewpoint theory (Edelman and Bülthoff, 1992; Cooper, 1989) and theories about 3D reconstruction from multiple views (Ullman, 1979; Braunstein, Hoffman, Shapiro, Andersen and Bennett, 1987) suggest that performance will increase with the number of available views, although the latter theory suggests some saturation number, where more views produce no further increase in performance. Furthermore, some viewpoints will be more informative than others (e.g., Perrett, Harries and Looker, 1992). Therefore, it is expected that for a complex scene, similar effects will exist for the number of available views.

Knot tracing tasks and similar spatial tracing tasks are often used to test the spatial impression of a scene. Especially when the wires are transparent, stereoscopic and parallax cues seem important for doing this task (see Chapter 2, 'Task 4: tracing a wire through a knot'). The wires of the knot shown in Figure 6.2 are hard to trace if you have only the front view or two separate views.

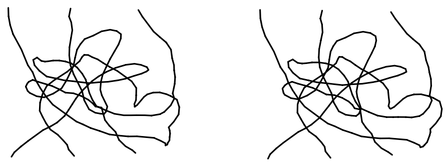


Figure 6.2. Sample knot to illustrate the importance of more than one viewpoint. With only one view, wires in a knot are hard to follow. Two views give a better spatial impression, especially when viewed stereoscopically.

Usually, views presented with the DVWS are not stereoscopic, although stereoscopic images can be combined with the DVWS. When non-stereoscopic images are presented to an observer who is using both eyes, the observer may get conflicting cues. Conflicting cues may hinder the observer's performance of the task and affect the 3D reconstruction that the observer makes according to the theory of indirect perception (Chapter 2). For example, stereoscopic cues inform him about the flatness of the display, while parallax cues match the depth in the simulated scene. For viewing static pictures, there are numerous experiments indicating a decline in performance when viewing a single image stereoscopically as compared to separate images for the left and right eye (Barfield and Rosenberg, 1995; Busquets, Parrish and Williams, 1991, Spain, 1990, Koenderink, van Doorn, and Kappers, 1994). But a coupling with the viewing position, as with the DVWS, provides movement parallax cues that often take precedence over stereoscopic cues (Ware, 1995; Norman and Todd, 1995; but see Beall, Loomis, Philbeck and Fikes, 1995) and can replace stereoscopic cues (Rogers and Graham, 1985; Cole, Pepper and Pinz, 1981; Sollenberger and Milgram, 1993; Ware, Arthur and Booth, 1993).

Experimental evidence suggests that stereoscopic cues are important for manipulation tasks, such as putting a wire through a hole (Spain and Holzhausen, 1991), putting a hook in a maze of wires (Cole, Merritt, Fore and Lester, 1990), moving a ring around a bent wire (Singh, Serra, Fairchild and Poston, 1994), a pipe-alignment task (Cole and Parker, 1989), and touching one of the wires in a knot (Voorhorst, Overbeeke and Smets, 1997). Stereoscopic cues were found to be essential for the creation of the knots for this experiment, which was done with an immersive VR system. These effects may be related to the result indicating that binocular disparity is perceived more quickly than any other visual cue (Drascic, 1991). Another explanation (Voorhorst, Overbeeke and Smets, 1997) is that observers tend to sit still while manipulating objects for their task (probably because moving interferes with the manipulation task), thus disabling movement parallax information.

However, for inspection tasks, stereoscopic cues seem of minor importance. Voorhorst, Overbeeke and Smets (1997) showed that when tracing a wire through a knot, movement parallax alone suffices. Arthur, Booth and Ware (1993) found that people can trace branches in a tree with fewer errors with a head coupled viewpoint selection than with a stereoscopic view. Sollenberger and Milgram (1991, 1993) showed that tracing a wire in a tree of wires can be done successfully with 11 viewpoints.

For transparent scenes, as with baggage inspection, most of these results may not hold. Little literature exists about depth perception in transparent scenes, but there is much evidence that the rigidity of objects and the human bias to see rigid motion is reduced as compared with non-transparent scenes (Todd, Akerstrom, Reichel and Hayes, 1988; de Poot, 1995; Kersten, Bülthoff, Schwartz and Kurtz, 1992).

Concluding, for the knot tracing task of this experiment observer performance is expected to increase with the number of available views. Parallax cues alone are expected to be sufficient for performing inspection tasks such as baggage inspection.

Experiment 1- reducing the degrees of freedom

In Experiment 1 the effects are tested of restricting the viewpoints to the horizontal arc, and making available only a limited number of views in the horizontal arc.

Method

Stimuli

Each stimulus consisted of views of a knot of three intertwined transparent wires (Figure 6.1). Each wire started at the top of the knot, and ran through the knot to one of the endpoints at the bottom of the knot. The endpoints are left, middle and right when viewed in front view. A red arrow indicated the top end of one of the three wires. The participants had to indicate the corresponding bottom end by pressing the left, middle or right button placed directly below the screen. There were 10 knots for the training and 40 knots for the experiment.

No real knot and camera were used: the views of the knot were rendered by a computer (Silicon Graphics RE Crimson). It was able to generate 37 images per second. The screen had a resolution of 1280x1024 pixels and a size of 33.5 x 28.0 cm. To enhance depth perception, a reduction screen (visible area 22.6 x 17.6 cm) was placed 12 mm in front of the monitor (Silicon Graphics Color Display CM2086A3SG). To match the viewing angle of the screen given an average distance between the observer and the screen of 52 cm, the virtual camera was chosen to have a vertical viewing angle of 30° (≈ 2*arctan[0.5 * 28/52]).

Apparatus

In the present experiment an active parallax system, the Delft Virtual Window System, is used to give the observer control over the desired viewpoint. This system is described in detail by Smets, Overbeeke and Stratmann (1987) and Overbeeke, Smets and Stratmann (1987). It consists of a monitor, an eye position sensor and a camera that looks at a scene (Figure 6.3).

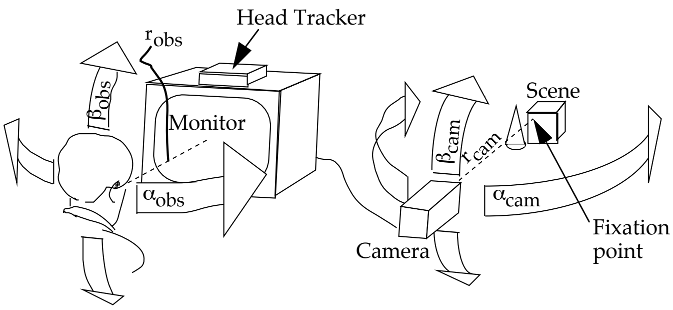


Figure 6.3. The Delft Virtual Window System consists of a monitor that displays the camera image. The camera position follows the eye position of the observer.

The monitor displays the image from the camera. The camera can rotate around the scene, but it keeps aimed at the fixation point in the scene at a constant distance r_{cam} from that point. The camera position is slaved to the rotation of the observer around the middle of the screen. That is, if the polar position of the eye of the observer relative to the screen centre is $(\alpha,\beta,r)_{obs}$, as defined in Figure 6.3, then the camera position relative to the fixation point $(\alpha,\beta,r)_{cam} = (\alpha_{obs},\beta_{obs},r_{cam})$. In the present experiment the camera range was restricted: $-90^{\circ} \le \alpha_{cam} \le 90^{\circ}$ and $-45^{\circ} \le \beta_{cam} \le 45^{\circ}$. In the monocular condition the position of the eye that was used was tracked, in the binocular condition the average of the two eyes.

Figure 6.4 shows an overview of the experimental setup. To record the response of the observer a button box with 3 buttons was placed below the screen. For tracking eye position, a 3D position sensor (a Dynasight from Origin Instruments) was used. It tracked a small reflector on the spectacle frame the participants wore. The tracker sent information about the angular eye position of the participants to the computer with 37Hz and an accuracy of less than 1° (given the usual motions of the participants). Because an observer is unable to see the screen when looking at the ±90° position of the monitor, the angular eye position was scaled 4 times, thus $(\alpha,\beta,r)_{cam} = (4\alpha_{obs},4\beta_{obs},r_{cam})$. Such a scaling of the movements of the observer is expected not to have a disturbing effect, as the experiment described in Chapter 5 showed that with constant observer movement range, increased camera range increased the performance of the observer.

Figure 6.4. Overview of the experimental setup: a monitor with the eye position tracker on top of it and a reduction screen in front. The button box is below the monitor.

Procedure

For participants in the monocular condition the dominant eye was determined prior to the experiment. To test this, the experimenter closed one eye, held his thumb between the opened eye and, alternating, the left or right eye of the participant. The participant was asked to indicate when the thumb appeared to hide the eye of the experimenter. The eye covered by the thumb at that moment, as seen by the experimenter, was taken as the dominant eye. Some subjects were found to have no dominant eye, and in that case an eye was chosen at random. They wore a patch covering the other eye. Participants in the binocular condition looked with both eyes at the same screen image. The participants were instructed to follow the indicated wire and to indicate the bottom end of that wire by pressing the left, middle, or right button below that end. It was explained that they could look around the knot by moving their heads. They were asked to move their heads at the start of each trial to find out whether the control was of use to them. This was done because a pilot study indicated that a large angle between views tends to demotivate participants from moving in subsequent stimuli. During the trials they held their hands near the button box.

The participants had 10 seconds to inspect each knot, after which the screen turned dark. A beep was sounded after 8 seconds to warn for the time limit. Participants were instructed to strive for a correct and fast answer, and that correctness was more important than speed. They could think as long as they wanted, even after the screen went dark, but usually they made a choice well before the beep.

The participants were trained with 10 knots in advance of the experiment. Immediately after they pressed a button, they were informed whether they made the right choice, and how long it took to make the choice. During the experiment they had to judge 40 knots, each knot under different viewing conditions, and received no feedback. Between their response and the start of the next trial there was a one second delay. In the instructions, participants were told that their score would be shown after the experiment.

Variables, participants, design

The independent variables were the viewing condition *V* (monocular or binocular) and the motion condition *M* (Figure 6.5). The motion condition describes the available views. In the hvc motion condition (Figure 6.5a), both horizontal and vertical continuous viewpoints are available. For the hc motion condition (Figure 6.5b), the viewpoints are restricted to the horizontal arc, setting the vertical angular position of the camera $\beta_{cam} = 0$. In the h33 condition (Figure 6.5c), 33 viewpoints in the horizontal arc are available. The 33 viewpoints are spread evenly over the full camera range of 180˚. This corresponds to the conditions of Experiment 3 of Chapter 5. The dependent variables are correctness of choice *C* and response time *T*.

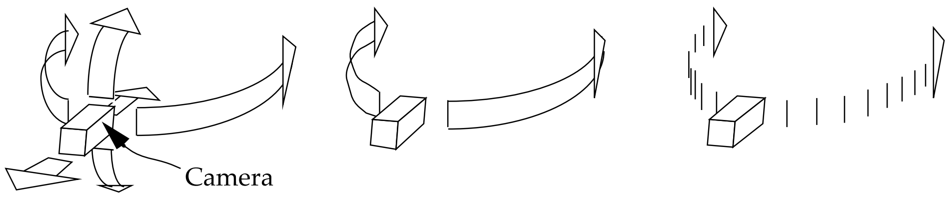


Figure 6.5a. In the unrestricted DVWS the number of available views is infinite in 3 dimensions.

Figure 6.5b. To limit the number of views, the camera motion is restricted to the horizontal arc.

Figure 6.5c. To make the number of views finite, the camera motion is made discrete.

The participants were 12 students from the Faculty of Industrial Design Engineering (4 women, 8 men). All participants were naive volunteers with normal or corrected-to-normal vision. All had stereo vision, in the sense that they were able to recognise a figure in a random-dot stereogram (Appendix A). Each person participated about 25 minutes, and received NLG 7.50 (USD 3.75) for doing so (one loaf of bread costs about NLG 2.00).

All participants judged the same 40 knots, but conditions and knots were presented in random order. For each knot the 3 motion conditions were randomized over three participants, and this was done 2 (*V*) x 2 (replications) times to obtain the stimuli for 12 participants. The starting point for each knot (Left, Right or Middle) was randomly chosen. At random, 6 participants were assigned to the monocular and binocular viewing condition respectively.

Hypothesis

Restricting the viewpoints to the horizontal arc reduces the information that can be retrieved and makes the image movement less natural, as humans never move in a perfect arc around a point of interest. Making the viewpoints discrete will introduce image jumps as the observer moves to the next viewing zone. It is expected that both manipulations will affect the spatial impression and reduce performance.

The second hypothesis addresses the difference between binocular and monocular viewing conditions. In the binocular condition, both eyes look at the same picture on the display, as no stereoscopic views were used. This may reduce the depth impression and therefore is expected to reduce the number of correct responses and increase the response time as compared to monocular viewing.

Results

Many participants made enthusiastic comments about the 3D impression and operational comfort offered by the Delft Virtual Window System.

A repeated-measures analysis of variance was done to test for the effect of the conditions on the percentage of correct answers (Kirk, 1968). An alpha level of 0.05 was used for all statistical tests. For the percentage of correct answers, chance level is 33%.

None of the effects was found to be significant: for the viewing condition, $F(1,10)=1.17$, $p=0.306$; for the motion condition $F(2,20)=1.46$, $p=0.255$; and for the interaction $F(2,20)=0.02$, $p=0.997$. Figure 6.6 shows the percentage of correct answers and the 95% binomial confidence interval (CI) (Loosen, 1994) as a function of the viewing and motion condition. Although the effect of the motion condition is not significant, Figure 6.6 suggest a lower percentage of correct answers in the h33 condition than in the hc and hvc condition, and a saturation point in the hc condition.

A repeated-measures analysis of the response times showed that the viewing condition is not significant: $F(1,10)=2.21$, $p=0.168$. The motion condition was significant: $F(2,20)=6.88$, $p<0.01$. A post-hoc Tukey-HSD (Kirk, 1968) test showed that the response time in the h33 condition is significantly higher than in the hc and the hvc condition ($p<0.05$). The response times between the hc and hvc condition did not differ significantly. Figure 6.7 shows the effect of the motion and viewing condition on response time.

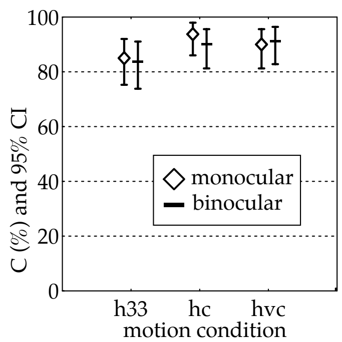


Figure 6.6. Percentage of correct answers and 95% confidence interval.

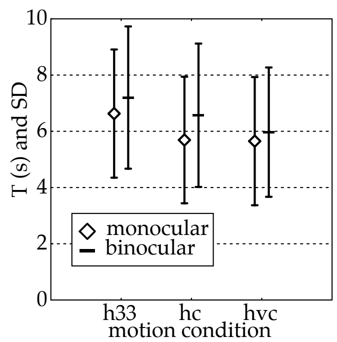


Figure 6.7. Response time and standard deviation.

The response time decreased significantly from the 33 viewpoints to the hc condition, but it did not differ between the hc and hvc condition. This indicates the saturation point for the response time at the hc condition. The percentage of correct answers did not change over these motion conditions. Figure 6.7 suggests that monocular observers perform faster than binocular observers, as was hypothesised, but this effect is not significant.

Experiment 2- restricting the number of views

It is expected that the percentage of correct answers will decrease when still fewer viewpoints are provided. To test this, the number of discrete viewpoints is reduced from 33 to 3, to find the saturation point for the percentage of correct answers. The same effects as for Experiment 1 are expected, so the percentage correct answers will increase and the response time will decrease as the number of viewpoints increases. Again, monocular observers are expected to perform faster and with a higher percentage of correct answers than binocular observers.

Method

Stimuli, apparatus, procedure, variables

The same stimuli, apparatus and procedure were used as in Experiment 1. The independent variables are the same as in Experiment 1, but the levels for the motion conditions *M* are now halved four times, from 33, 17, 9 and 5 down to 3 views in the horizontal arc, because I need an odd number of viewpoints to get a symmetric camera range and one front view. As the total camera range was always 180˚, the angle between two available viewpoints was 180˚/(*M*-1).

Design, participants

All participants judged the same 40 knots, but conditions and knots were presented in random order. For each knot the 5 motion conditions were randomized over three persons, and this was done 2 (viewing condition) x 2 (replications) times to obtain the stimuli for 20 participants. The starting point for each knot (Left, Right or Middle) was randomly chosen. At random, 10 participants were assigned to the monocular and binocular condition each.

The participants were 20 students from the Faculty of Industrial Design Engineering (6 women, 14 men). All were naive volunteers with normal or corrected-to-normal vision. Each person participated for about 25 minutes, and received NLG 7.50 (USD 3.75) for doing so.

Results

A repeated-measures analysis of variance was done to test for the effect of the motion and viewing condition on the percentage of correct answers (Table 6.1).

Table 6.1. Results of a repeated-measures analysis of variance of the number of correct answers.

Interaction	F	p
V	$F(1,18)=0.82$	0.376
M	$F(4,72)=6.57$	<0.001
$V \times M$	$F(4,72)=0.35$	0.845

Only the motion condition was found to be significant: $F(4,72)=6.57$, $p<0.001$. Figure 6.8 shows the effect of the variables on the percentage of correct answers and the 95% binomial confidence interval (CI) (Loosen, 1994). The percentage of correct answers increases with increasing motion condition, i.e. with increasing numbers of available views over 180˚.

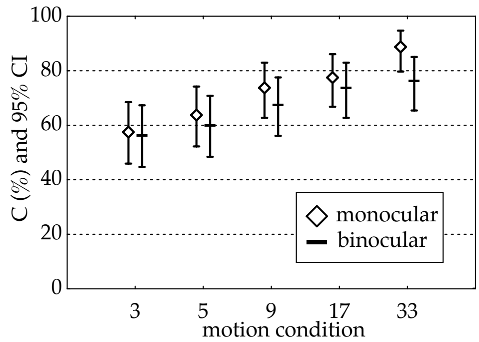


Figure 6.8. Percentage of correct answers and 95% confidence interval.

A post-hoc Tukey-HSD test showed that the number of correct answers in the 33 viewpoint condition is significantly higher than in the 3- and 5 viewpoints condition, and that the number of correct answers in the 17 viewpoints condition is significantly higher than in the 3 viewpoints condition (p<0.05). The figure suggests that monocular observers indeed perform slightly better than binocular observers, as was hypothesized, but this effect is not significant.

The effect of the variables on response time was also tested with a repeated-measures analysis of variance (Table 6.2). As with the percentage of correct answers, only the motion condition has a significant effect on response time: $F(4.72)=16.85$, $p<0.001$. Figure 6.9 shows the effect of the variables on response time. The response time clearly decreases with the motion condition.

Table 6.2. Results of a repeated-measures analysis of variance of response time.

Interaction	F	p
V	$F(1,18)=0.08$	0.921
M	$F(4,72)=16.85$	<0.001
V x M	$F(4,72)=0.80$	0.532

A post-hoc Tukey-HSD test showed that response times in the 3 and 5 viewpoint condition are significantly larger than in the 33, 17 and 9 viewpoint condition, and that response time in the 9 viewpoint condition is significantly larger than in the 33 viewpoint condition (p<0.05). An ad-hoc analysis of variance of the number of time-outs (a reaction time of more than 10 s) showed a significant effect of the motion condition, but not of the viewing condition.

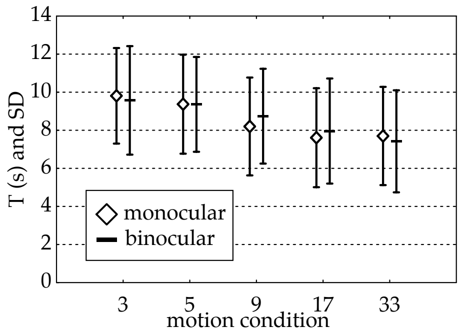


Figure 6.9. Mean response time and standard deviation. People use significantly more time with more restricted viewing conditions.

Summary of the results of Experiment 1 and 2

As an overview, Figure 6.10 combines the percentage of correct answers of Figure 6.6 and 6.8. As expected, the percentage of correct answers increases with the number of viewpoints. Saturation for the percentage of correct answers is reached at the hc condition. Although the data suggests a slight advantage of monocular over binocular observers (as expected), the effect was not significant.

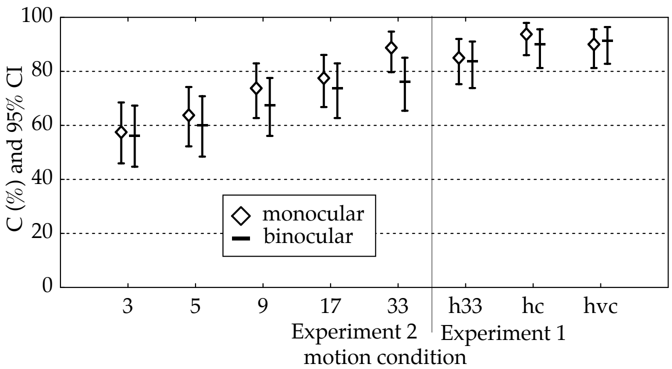


Figure 6.10. Percentage of correct answers and 95% confidence interval of the two experiments, as a function of the viewing and motion condition. Up to the hc condition, the percentage of correct answers increases.

Figure 6.11 combines the response times of Figure 6.7 and 6.9. As expected, response times decrease with increasing numbers of viewpoints. Compared with the hc condition, the hvc condition improves neither the response time nor the percentage of correct answers, so for this task the saturation point is somewhere between the h33 and the hc condition. Again, the figures suggest that monocular observers perform faster than binocular observers, but this effect was not shown to be significant.

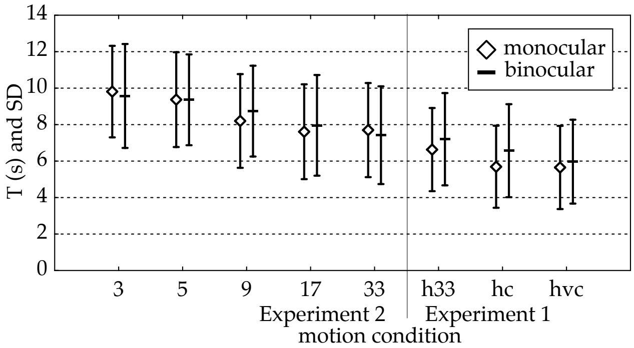


Figure 6.11. Mean response time and standard deviation of the two experiments. Observers make faster decisions when more viewpoints are available.

Detailed discussion and conclusions

Both Sollenberger and Milgram (1993) and Arthur et al. (1993) tested observer performance for a tree tracing task, which is similar to knot tracing. For continuous horizontal and vertical viewpoint selection they found an average of 96% correct answers. This is slightly higher than the 91% I found. A larger difference is in the effect of halving the number of viewpoints. In the present experiment, halving the number of viewpoints causes a reduction of the correct percentage of about 7%. Arthur et al. (1993) experimentally tested the effect of the number of viewpoints on the observer performance. For a reduction of the number of viewpoints from 11 to 6 (over 30° camera range) they find a drop of about 74% to about 55% correct. This is larger than the largest performance drop in our task, and may indicate that their scene is more complex than our scene. There are a number of explanations for this difference. First, they used two different motions in the two viewing conditions: cycling forward and backward in the 11 viewpoints condition and presenting each frame for 5 seconds in the 6 viewpoints case. Second, they used a tree-tracing task where I used a knot tracing task. Third, they used a different projection method for their spatial display. This issue is worked out further in Chapter 7. Finally, their camera range was 6 times smaller than ours.

For the present task, saturation occurs somewhere between 33 viewpoints and continuous horizontal viewpoints. For the less complex task of detecting a wire between objects (Chapter 5), I found a saturation for 3 available viewpoints. Concluding, the hypothesis that the required number of viewpoints depends on the complexity of the scene is affirmed. In general (for an unknown task), more viewpoints will improve observer performance, but for specific tasks a saturation point will exist.

Looking with two eyes at a single picture gives the observer a cue to the flatness of the picture, which is in conflict with the movement parallax cues. This conflict of cues was expected to degrade the performance of the observer. But the results do not affirm such disturbing effects, although the figures suggest that it does exist. Similar to our results, Arthur et al. (1993) also failed to find significant differences for the response time and the error rate between monocular and binocular observers. Apparently, conflicting cues do not harm the ability to follow wires through a knot. This experiment merely tested the differences between observers using both eyes and observers using only one eye. The experiment described in Chapter 7 will explore the distorting effect of the viewing and projection conditions in more detail.

Summarizing, in a knot task observer performance increases with the number of viewpoints, but saturation occurs with continuous viewpoints in the horizontal arc. For less complex scenes saturation can occur with fewer available viewpoints. As the complexity of x-ray images of baggage seems to lie between the stimuli of the experiments described here and those in Chapter 5, the saturation point for the number of views can be expected to lie above three for x-ray baggage inspection. However, I have no exact measure for this scene complexity, and the effects of scene complexity on human performance will depend on the task. Furthermore, there is an important informal finding: large angles between the views tend to demotivate people from looking around the scene, which also suggests that more than three views have to be provided to elicit x-ray inspectors to use the available views. More research is needed to find the precise saturation point for x-ray baggage inspection and to test the effect of different projection methods for spatial displays. Experiments for answering these questions are described in Chapters 7 and 8. A more thorough investigation of scene complexity would seem to lie outside the scope of this thesis.

7

Bump height matching[1]

The spatial impression the observer gets from an image (the *apparent layout*) does not necessarily match the geometry of the original scene as could be measured with a measuring rod (*Euclidean layout*). I will refer to such a mismatch between apparent and Euclidean layout as *distortion*. It is known that distortion is affected by the viewing position of the observer relative to the image (his *viewpoint*). For example, Smith and Gruber (1958) showed that apparent distances in a photograph of a corridor increase with viewing distance. In virtual window displays, where the displayed picture is adapted continuously to match the actual viewpoint of the observer, such distortions may be corrected. Whether distortions occur in virtual window displays will depend on the relation between the viewpoint of the observer and the camera settings (its position, rotation and viewing angle). I will refer to this relation between the observer's viewpoint and the camera settings as the *coupling method*. In virtual window displays, distortion may also be caused by inaccurate implementation of the coupling, for example due to delays in the update cycle or errors in the measurement of the viewpoint. The experiment described in this chapter investigates the effect of the coupling method and viewpoint measurement inaccuracies on distortion, using a height matching task. Distortions in virtual window displays may reduce task performance, and may therefore be important for eliminating such distortions in a baggage inspection system based on the DVWS.

If the visual angle subtended by objects affects the apparent depth of the scene, it can be expected that the optical angles in the photograph should match the optical angles in the real scene in order to give the same apparent depth (Pirenne, 1975). Only one viewpoint fulfils this constraint (is *geometrically equivalent*). This geometrically equivalent viewpoint matches the position of the lens of the camera relative to the recording plate, and scales with enlargement of the picture. Figure 7.1 illustrates this geometrically equivalent viewpoint for viewing a picture made with a normal camera. As the lens was 5 cm in front of the centre of the picture when the photograph was made, and the photograph was enlarged two times, the geometrically equivalent viewpoint lies 2*5=10 cm in front of the centre of the picture.

[1] Part of this work was presented in Pasman, Stappers and Smets (1997).

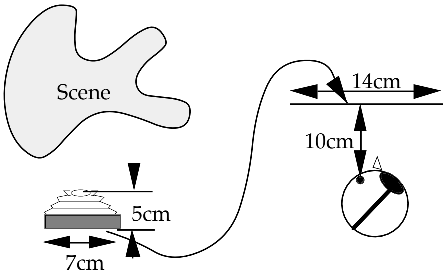


Figure 7.1. A geometrically equivalent viewpoint is a viewpoint of the observer where all visual angles subtended by the displayed objects match those of the real objects. See text.

Figure 7.2 illustrates that a geometrically inequivalent viewpoint can lead to distortion. The geometrically equivalent viewpoint lies at about 8 cm in front and 2 cm above the middle of the picture. You should slant the book 45° to get the horizon of the picture (actually above the photo) at eye level. If the picture is viewed more from the bottom (but at an equal perpendicular distance from the picture) the 2 barrels at the right, in particular, appear shorter. This effect is called apparent depth compression. If the picture is viewed more from the left, the tops of the barrels appear to move leftward relative to the floor, especially when the picture is also viewed from below. This effect is called apparent shear.

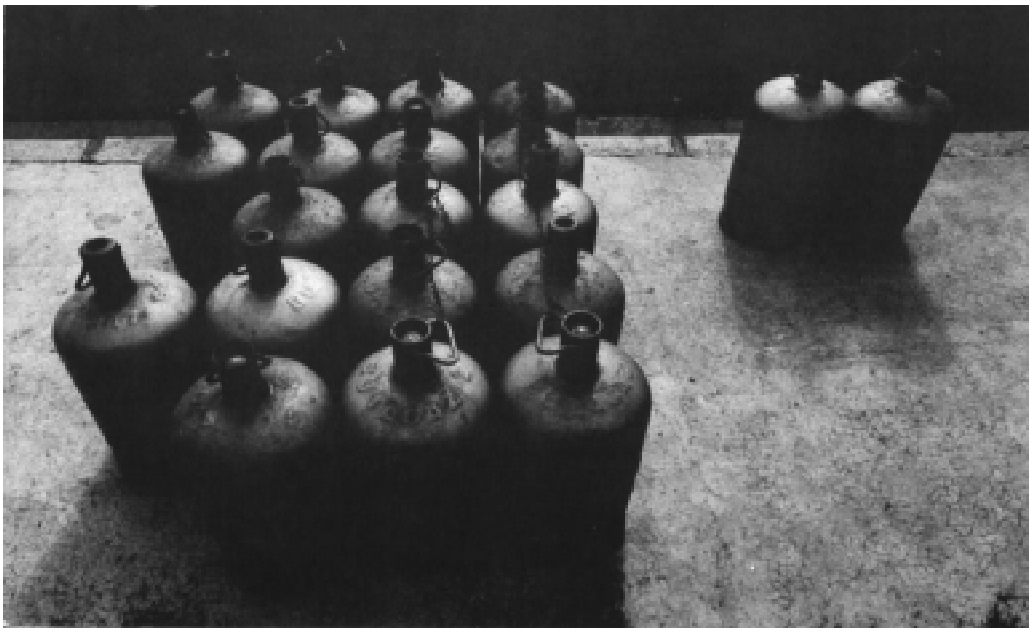


Figure 7.2. The barrels appear distorted at geometrically inequivalent viewpoints. See text. (From Petzold, 1973).

Similar effects are known for portraits, where the eyes of the depicted person seem to follow the observer, and in pictures of someone pointing his finger out of the picture, where his finger appears to keep pointing at the observer as he moves (Zorin, 1995).

For most images, such distortions are inconspicuous (Cutting, 1987). This may be related to the task: for example recognition of your family members or admiration of a beautiful landscape need not be hindered by such distortions. There is evidence that human observers actively compensate for the distortions caused by a geometrically inequivalent viewpoint, but only if the cues informing the observer that he is looking at a picture are strong enough. Such cues may consist of cues to the flatness of the image, such as a clearly visible frame around the picture, and a raster in front of the image (e.g., when the picture is displayed with pixels on a monitor).

In virtual window displays the displayed image is adapted to the viewpoint of the observer. This may reduce or even eliminate the distortions. To do this, the viewpoint has to be measured accurately, and a failure to do so (a *viewpoint measurement error*) may cause distortion. Furthermore, if the observer actively compensates for his non-perpendicular (*oblique*) viewpoint, adaptation of the display to his viewpoint might be counterproductive.

In the sections below, I will describe the various coupling methods and previous work that has been done on distortion in picture perception. Next, an experiment investigating distortions in virtual window displays will be described. Finally, the results and their implications for virtual window design will be discussed.

Coupling methods

Virtual window displays can be implemented by different coupling methods. Here, on-axis and off-axis coupling methods are considered (Figure 7.3). As the coupling method is similar for both eyes, it is sufficient here to discuss the case of only one eye. This coupling causes objects to shift relative to each other as the observer moves (*movement parallax*), and movement parallax is an important depth cue (Chapter 2).

With *on-axis coupling* (Figure 7.3a), as in the Delft Virtual Window System (Chapter 1), the lens of the camera always stays before the middle of the recording plate (thus making *on-axis images*). For the DVWS case considered here, the camera's motions are constrained so that it rotates at a fixed distance around the point of interest in the 3D scene, the *fixation point*. The angular position of the camera is controlled by the angular position of the observer. As the camera keeps aimed at the fixation point, the fixation point is always in the centre of the image.

With *off-axis coupling* (Figure 7.3b), such as the fishtank VR system (Ware, Arthur and Booth, 1993; Castle, 1995) both the camera and the lens position are coupled to the position of the observer. The camera translates with the observer, but does not rotate (thus making *off-axis images*). The lens translates relative to the recording plate as the observer translates relative to the display. Consequently, all visual angles subtended by objects in the image correspond to their angles in the real scene when viewed from the same viewpoint. With real cameras, off-axis images might also be acquired with a wide-angle image of which only a small part is used at a time, or with a camera making on-axis images and real-time image processing. With computer generated perspective images, both on-axis and off-axis coupling are easy to achieve.

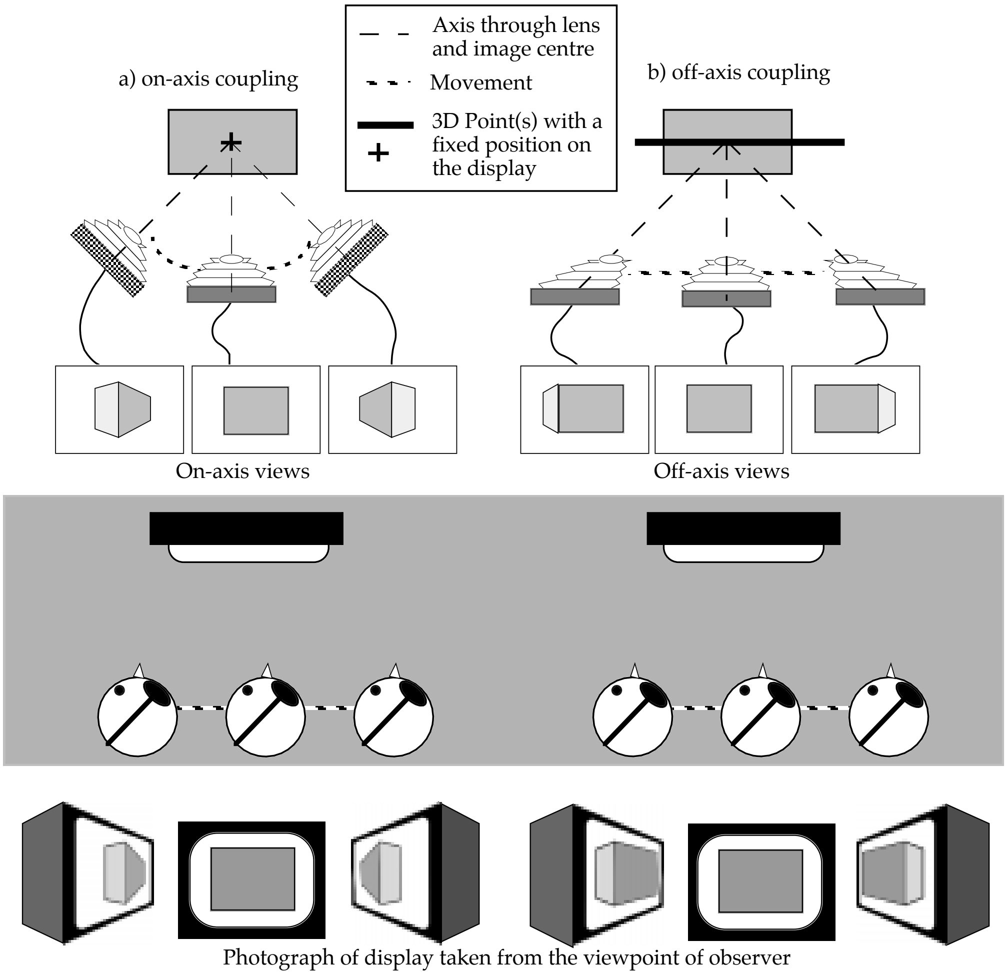


Figure 7.3. Two different coupling methods used for virtual window displays. With on-axis coupling the camera rotates around the fixation point as the observer rotates around the middle of the display. With off-axis coupling the camera translates with the movements of the observer. To select the appropriate part of the scene, the lens has to shift relative to the image sensor inside the camera.

Previous work

There are many visual cues that can cause distortion if presented inappropriately. Some general guidelines for avoiding distortions give a general idea about what cues are important to consider. Gibson (1971) suggested that for optimally viewing perspective pictures, "the observer should look with one eye at the correct viewing distance, the picture should be upright and perpendicular to the line of sight, and there should be a reduction screen in front of the picture hiding the rest of the world" (p. 30). Pirenne (1970), Gourneri (1859, see Cutting, 1986) and Jones and Hagen (1978) note that many artists have strategies to minimise distortions: "(1) stress the depiction of the texture of surfaces, (2) avoid extreme perspective convergence and (3) choose a station point [viewpoint, in our terminology] that presents a characteristic view of the scene." (Jones and Hagen, 1978, pp. 191-192). As discussed in the introduction, the coupling method, the geometric equivalence of the viewpoint, the accuracy of the measurement of the observer's viewpoint and cues to the flatness of the image may all have an impact on the apparent layout of the scene. If the observer mentally compensates for distortions expected at an oblique viewpoint, this may also affect the distortion. Furthermore, several other factors, such as the layout of the scene and delays in the system, may cause distortion. Although a lot of experiments have been done on distortion in picture perception due to geometric inequivalence (Smith and Gruber, 1958; Pirenne, 1970, 1975; Rosinski and Farber, 1980; Cutting, 1986; Halloran, 1989), little is known of what happens when the depicted image is coupled to the movements of the observer, as in virtual window displays. The following sections discuss previous work that has been done on these issues.

Geometric inequivalence

Much research has been done on the effect of geometric inequivalence on the apparent layout of static, on-axis images. Nearly all of these studies assume that humans make a 3D reconstruction of the viewed scene (see Chapter 2) using only the rules of geometry. This is tempting, as perspective cues from a single view and a few assumptions allow a precise depth reconstruction, but this is not how actual human perception operates, and none of the theories discussed in Chapter 2 suggests that the apparent layout depends on perspective cues only. Nevertheless, geometry is a useful explanation for distortion because it allows prediction of distortions in a mathematical way, and because it connects to most literature concerning distortions.

Figure 7.4 illustrates how geometric inequivalence may lead to distortion. If the viewing distance is smaller than geometrically equivalent (Figure 7.4b), a 3D reconstruction using geometry will be compressed in depth as compared to the original scene.

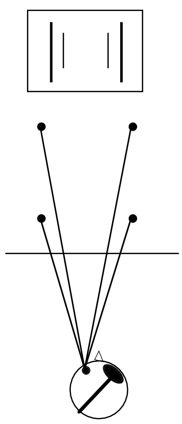


Figure 7.4a. Top: a view. Bottom: 3D reconstruction from perspective, at the geometrically equivalent viewpoint.

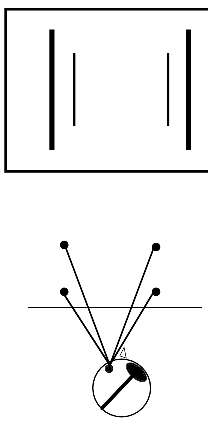


Figure 7.4b. If the viewpoint is too close, the view appears twice as large.
A geometric reconstruction becomes compressed in depth.

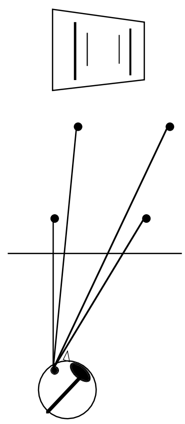


Figure 7.4c. Geometric reconstruction gives shear distortion if the viewpoint is displaced parallel to the display (from Cutting, 1986)

If the observer looks from a distance d' rather than the camera distance d, geometry suggests that the apparent depth will be scaled by a factor d'/d. If the viewpoint has been displaced parallel to the display, the 3D reconstruction from perspective cues will be sheared (Figure 7.4c) (Rosinski and Farber, 1980; Adams, 1972; Lumsden, 1980).

Numerous experiments show that apparent depth increases with increasing viewing distance, i.e. that human perception is sensitive to geometric cues. For example, Smith and Gruber (1958) showed that apparent distances in a photograph of a corridor increase with increasing viewing distance, though the effect is smaller than predicted by geometry. Adams (1972) displayed a floor and background of tiles. Participants matched the width-height ratio of tiles on the background to those on the wall. The effects matched geometric predictions, although the size of the effect is smaller than predicted. Purdy (1960), Smith and Gruber (see Rosinski and Farber, 1980) and Braunstein and Payne (1969) found similar results for the perception of a projected grid of fine opaque lines, that was magnified and slanted. An inconvenient point about this evidence for the effect of geometric inequivalence on apparent depth is that most of the stimuli used did not contain other depth cues (such as texture, shading, stereoscopic, shadow and accommodation cues), thus disabling the participants to use cues other than perspective. Therefore, these results may not be generalized to practical applications such as x-ray baggage inspection via a virtual window display.

A geometric reconstruction of an on-axis image viewed obliquely will be sheared. There is some evidence that shear distortion of this kind also occurs in human perception. Hochberg (1986) indicated in the cinema situation that distortion is usually not noticeable from oblique viewpoints, but that when relative sizes and velocities become important, as in dance, the distortions become obtrusive. Cutting (1987) displayed rotating wireframe cubes on a slanted display. With a slant of 45° the apparent rigidity of cubes is lower than with 67° or 90° (head on) display slant. For television viewing, McVey (1970) indicated that oblique viewing positions up to 45° oblique produce images with an 'acceptable' amount of distortion, but that informational displays (e.g. graphs) become problematic. For movies, telelenses may reduce distortion for viewers looking obliquely at a film (Cutting, 1989).

In our literature search, I did not encounter any experiment dealing with the effect of geometric inequivalence in virtual window displays. In all cases, only static viewpoints were considered.

Human insensitivity to geometric inequivalence

In contrast with the evidence above, many researchers report that viewing from a geometrically inequivalent viewpoint usually does not lead to distortions. For example paintings usually do not appear distorted when one looks at them from an oblique viewpoint, and viewing distances also have little effect (Pirenne, 1970, 1975; Rosinski and Farber, 1980). Again, objects on a television do not appear to deform when the camera moves (Cutting, 1989). Perkins (1973) showed that observers looking obliquely at 26° and at 41° from the image normal can usually distinguish depicted rectangular corners from nonrectangular corners.

At least three mechanisms have been proposed that might allow such undistorted information to be extracted from a geometrically inequivalent viewpoint. The first mechanism (Rosinski and Farber, 1980; Pirenne, 1970; Pirenne, 1975) assumes that the image has been made with an on-axis camera, and thus an estimation can be made of the geometric equivalent viewpoint. The geometrically equivalent viewpoint is in front of the middle of the image, and Rosinski and Farber (1980) suggest that the viewing distance may be estimated from the width and height of the image, or from image clarity. The distance between the actual and the estimated geometrically equivalent viewpoint can be used to correct for the oblique viewpoint[2]. The second mechanism suggests that the geometrically equivalent viewpoint can be recovered from perspective cues in the image, perhaps using assumptions about the Euclidean layout. As with the first mechanism, this allows a correction for the oblique viewpoint. The third mechanism (Cutting, 1987) proposes that only depth cues that are relatively unaffected by the viewpoint be used.

There is evidence that humans may use the first mechanism, which assumes that the geometrically equivalent viewpoint lies perpendicular to the display. Rosinski and Farber (1980) suggest that Perkins (1973) enabled this mechanism by not using a reduction screen, but that Smith and Gruber (1958) and Purdy (1960) (see 'Geometric inequivalence') did not enable it by using a reduction screen. Rosinski and Farber (1980) did two experiments. In the first, observers making slant judgements in photographs of a slanted, striped surface did not compensate for geometric inequivalence when looking with one eye and through an aperture, but they did when they looked with both eyes, without aperture and in a well-lit environment. In the second, they made computer rendered slanted lattices with a correct viewpoint at 112 cm. The observers made their slant judgements from various distances. The apparent depth matched that of a geometric reconstruction of the scene from a viewpoint 56 cm in front of the display. Rosinski and Farber suggest that human observers estimate the geometrically equivalent viewing distance to be twice the height of the display[3] (27 cm in their setup).

According to mechanism 2, the geometrically equivalent viewpoint is reconstructed from perspective cues in the image. Geometrically, such a reconstruction can be made (Halloran, 1989; Ramsey and Sleeper, 1988 (p. 798); Sedgwick 1980; 3D builder, 1996). However, there is not much evidence for the use of this mechanism in human perception. If human perception used such a mechanism, one would not expect apparent depth to depend on the presence of a frame, as described for the first mechanism. Then again, most of the stimuli used there contained only perspective cues, and therefore results may not hold for richer pictures.

The third mechanism does not need to compensate for a geometrically inequivalent viewpoint, as it uses only depth cues that are more or less viewpoint independent. For example Hagen (1974) suggests that "presumably the high-order information specifying an individual's face consists of such relations as the length or sharpness of the forehead curve relative to the length or sharpness of the nose curve, the width of the eyes relative to the length of the nose, and so on" (p. 475). Since about 1980, there has been growing experimental evidence supporting such a mechanism, and against the other two mechanisms. Cutting (1987) displayed rotating cubes projected onto a simulated plane which was slanted relative to the visible monitor display. Rigidity judgements on slanted planes without cues to slant did not differ from judgements with unslanted planes. He concludes that the orientation of the observer relative to the projection plane does not influence apparent rigidity, and that local distortions are so small as to be unregistered by

---

[2]For pictures in pictures, this mechanism would require a double correction.

[3]Rosinski and Farber (1980) argue that "this viewing distance is (. . .) optimal in terms of providing a maximal signal-to-noise ratio for conventional TV displays. It may be that with video displays, an observer learns an assumed correct viewing point on the basis of optical picture clarity" (p. 171). However, McVey (1970) suggests that the optimal viewing distance for TV displays is 6.25 times the width of the display, which is much further away than the twice the height of the display as suggested by Rosinski and Farber.

the visual system with a 67° slanted projection plane. Halloran (1989) tested judgements of the orientation of parts of a rowing-boat in pictures seen from geometrically inequivalent viewpoints, and concludes that pictorial relations between the image contents and the border influence the judgements. Furthermore, he replicated the experiment with Rosinski and Farber (1980), but with even more extreme oblique views, and showed that a geometrically inequivalent viewpoint does have an effect at extreme angles.

Finally, the ability to compensate for a geometrically inequivalent viewpoint may depend on age (Hagen and Elliott, 1976; Jones and Hagen, 1978). Haber (1980) points out that an adult looking at a picture sees both its flatness and the depicted spatial scene, but that 3 year-olds treat pictures as if they were looking at real scenes.

Concluding, for static images it seems that distortions usually do not occur for oblique views up to about 70° from head on. Humans may compensate for geometrically inequivalent viewpoints. The distortions predicted by the three mechanisms differ for different coupling methods. If humans use the first mechanism (thus assuming an on-axis image), huge distortions can be expected with off-axis coupling, and no distortion at all with on-axis coupling. If humans used the second mechanism (thus completely compensating for oblique viewing), no difference would be expected between on- and off-axis coupling. With the third mechanism, distortions will be as predicted with geometry.

Cues to the flatness of the image

Cues to the flatness of the image may trigger a compensation mechanism for oblique viewing (as discussed above). Also, they may flatten the 3D reconstruction made by the observer (see Chapter 2). Such flatness cues can consist of binocular parallax cues, the presence of a visible frame emphasizing the image borders, the grain of the surface (e.g. the pixel size of the display), reflections of the image surface, colour flatness, and convergence and accommodation cues. Following Gibson's advice (see above, under ‘Previous work’) will reduce these flatness cues.

Hochberg (1986) suggested that the presence of flatness cues reduces the apparent depth in pictures. For example, accommodation is such a cue, especially at small distances (Gooding, Miller, Moore and Kim, 1991; Schlosberg, 1941). For static images, Koenderink, van Doorn and Kappers (1994) found that monocular observers perceive a larger pictorial depth than binocular observers. When viewing through a synopter, an optical device that places both eyes optically at the same viewpoint, they paradoxically found that even larger pictorial depth is perceived. Reduction of the visible area with the synopter as compared to the other viewing conditions may have caused this effect. The results of the experiment described in Chapter 6 suggested a tendency towards monocular observers working faster and making fewer mistakes than binocular observers in tracing wires through a knot on a virtual window display. In the pilot research for the present experiment (see below, ‘Pilot experiments’) I noticed that, with off-axis coupling and binocular instead of monocular viewing while only a monocular image is displayed, a sphere appears as an egg, resembling the distortions described by Pirenne (1970). The presence of a frame around the image may be another flatness cue, and may trigger some compensation mechanism. The experiment of Rosinski and Farber (1980), described under ‘Human insensitivity to geometric inequivalence’, showed a difference between judgements with and without a frame indicating the slant of the display. Eby and Braunstein (1995) found compression of the apparent layout of a spatial scene caused by a frame in front of the image.

On the other hand, for virtual window displays Arthur, Booth and Ware (1993) compared performance in a tree tracing task on a virtual window display. They found no significant differences between monocular and binocular viewing, and 71% of the subjects preferred non-stereoscopic viewing (but using both eyes) to monocular viewing. For a depth-height matching task in a static image, Adams (1972) too found no differences between monocular and binocular viewing. This is surprising, as binocular viewing gives the observer a cue about the flatness of the image, and this cue conflicts with the other cues. Possible explanations for this are (a) that monocular viewing improved the apparent depth, but another factor reduced the apparent depth (monocular viewing may be less comfortable because subjects were asked to close or cover the left eye, or alternatively some subjects may have used their non-dominant eye), and the summed effect was zero, or (b) that it depends on the task whether binocular cues are being used by the observer.

Cues from parallax shifts

There are a number of theories which seek to recover the spatial layout from motion parallax cues geometrically. These theories differ in the specific parallax information that is used. Most theories recover the spatial layout from a few views taken from the infinite number of available views. Theoretically, two perspective views from different viewpoints provide sufficient cues for a geometrical reconstruction up to a scaling factor (Ullman, 1979; Longuet-Higgins, 1981), but more views may be useful in less restricted situations (Braunstein, Hoffman, Shapiro, Andersen and Bennett, 1987).

It is not clear to what extent human perception uses parallax cues. Human depth perception from parallax seems to use other information besides a few ‘snapshots’. Braunstein et al. (1987) showed that, for same-or-different judgements of a few spatial points, human performance still increases when the observer is presented more viewpoints than geometrically required. A knot tracing task (Chapter 6) showed similar evidence. Todd, Akerstrom, Reichel and Hayes (1988) found, for shifting random dot clouds, complex effects of the number of available views and timing aspects of the image on apparent rigidity. Thus, the theories seem not to fit actual human performance. Furthermore, these theories are unclear about how human perception resolves conflicts. This issue will be discussed below, under ‘Coupling method’.

Most experiments with parallax shifts have been done with shifting random dot clouds. These suggest that parallax shifts are an important cue for human perception. Tittle, Todd, Perotti and Norman (1995) investigated such clouds viewed from a small viewing distance and with the shifts not coupled to the viewpoint of the observer. They found that for matching the width and the depth of a cylinder and for setting two planes perpendicular, apparent depth appears expanded as compared with geometric depth. Rogers and Graham (1983, 1985) suggest that apparent depth from motion parallax is about the same as apparent depth from stereoscopic cues. Norman and Todd (1995) showed that apparent depth from motion parallax easily overrules depth from stereoscopic cues, and stereoscopic cues are usually considered to be strong depth cues (that is, they have a major impact on a 3D reconstruction). Therefore, the viewpoint dependency of apparent depth as found for static images may be completely different for virtual window systems.

However, other results suggest that motion parallax is a weak cue, probably about as strong as accommodation cues (Gogel and Tietz, 1979), and perspective cues may take precedence over motion parallax (Wickens, 1990). Braunstein, Hoffman, Shapiro, Andersen and Bennett (1987) showed that, for same-or-different judgements of a few spatial points, human performance still increases when the observer is presented more viewpoints than geometrically required. For motion parallax, there may be a tendency for apparent depth to match apparent width. For example, Durgin, Proffitt, Olson and Reinke

(1995) found for oscillating real cones a tendency of observers to judge the height equal to the width. This is in line with similar findings for depth judgements of shifting contours of concentric circles (Caudek and Proffitt, 1993). The apparent depth may be influenced by the shape of objects, for example Ono and Steinbach (1990) found for random dot patterns that a sine shape is perceived to be of higher amplitude than a sawtooth-shape. These results may well be affected by the relatively poor stimuli used in these experiments, and therefore it may be inappropriate to generalize them to complex images as encountered in x-ray baggage inspection. The strength of parallax cues may depend on the richness of the scene.

Concluding, it is not clear to what extent motion parallax cues will influence apparent depth, and how they interact with perspective cues. Furthermore, for human perception motion parallax cues may require support from other depth cues. Therefore it is difficult to predict apparent depth from motion for real scenes presented with the DVWS.

Other factors possibly causing distortion

Several other factors may cause distortion. These factors have to be considered in order to make appropriate stimuli for our experiment, as the experiment described in this chapter will compare performance with Euclidean measures.

The scene layout may affect distortion. Kjelldahl and Prime (1995) showed for relative depth estimation of computer rendered shaded dishes that people make larger errors in judging distance if the separation between the dishes is vertical rather than horizontal. The texture used in the scene also influences depth perception, especially in the case of local details (Cumming, Johnston and Parker, 1993). Texture density has an effect on apparent depth (Börjesson and Lind, 1996). The colour also influences depth, as Claessen (1996) showed that a real blue object appears to protrude further than a yellow one. Kappers, van Doorn and Koenderink (1994) showed that the shading of real objects can also cause distortion.

If the camera image is presented on a television monitor, both the thickness of the glass and the curvature of the monitor will deform the image. Glass in front of a picture, for example in front of the television monitor, will magnify the picture. In television monitors, the picture is not flat but curved. This will effectively reduce the picture. Deering (1992) indicates that both reduction and magnification can displace the geometrically equivalent viewpoint up to 2 cm, but that the combined effects will not cancel out. He derives a formula giving a correction factor, given some point of interest on the display.

The delay and the refresh rate of the system can cause distortion. If the displacement of objects gets larger than 0.15° per update (Padmos and Milders, 1992), the image movement will appear shaking. Furthermore, for a moving observer a delayed image does not fit his actual viewpoint. This causes striking deformations of objects when the observer changes the speed of his movement. In the pilot tests I noticed with off-axis coupling that shear became more striking with increasing observer speed and with decreasing refresh rate. This shear distortion was also noticed in a prototype setup of a virtual window system with three displays instead of one (Djajadiningrat, Smets and Overbeeke, 1997), and seems typical for virtual window systems. The delay effect is discussed in more detail in Chapter 9.

Finally, the task of the observer, and the relation between the observer and his environment, determine whether distortion is a problem or not. As mentioned earlier, paintings and objects on a television screen do not appear distorted when the viewer is dislocated. But in these cases there are few consequences for the spectator if he misperceives the depth in the image. But if the observer has to manipulate objects in such a distorted environment, he needs some time to adapt to the distortions. For example, one needs to get used to a new pair of spectacles. More extremely, Kohler (1962) reported that observers wearing distorting goggles will correct many distortions such as colour fringes, line curvature and reversed left and right. In contrast with the immediate correction that takes place when looking at pictures, the time required to get used to these distortions takes from days to some weeks.

Many other factors may exist that influence the depth in virtual window displays. For example, Drascic (1991) found slow learning effects for parallax displays. The factors discussed here will be taken into account in the setup of the following experiment, but one can only hope not to miss an essential cue.

Coupling method

As far as I know, no research has been done on the effect of on- and off-axis coupling on apparent layout in virtual window displays. Both for perspective cues and for parallax cues, a single geometric spatial reconstruction given multiple views from a virtual window system seems possible only for off-axis coupling. For on-axis coupling, several inconsistencies may occur, such as vertical movement of parts of objects while the observer moves only horizontally, geometrically suggesting viewpoint dependent (nonrigid) distortions. Therefore, distortions are expected with on-axis coupling.

With off-axis coupling, the viewpoint is always geometrically equivalent to the displayed view, provided that the viewpoint is measured accurately. In the presence of a viewpoint measurement error, distortions can occur. If human observers use perspective cues in a geometric way, the distortions will cause shear and depth compression as discussed under ‘Geometric inequivalence’. With on-axis coupling, the geometrically equivalent viewing position is always in front of the middle of the display, at the fixed camera distance. However, with on-axis coupling the observer needs to move away from this position in order to acquire another view. Therefore, geometry suggests that distortion will occur for most viewpoints, and this distortion is expected to increase with increasing distance from the geometrically equivalent viewpoint. However, if the observer compensates for his geometrically inequivalent viewpoint, the effects for on- and off-axis coupling may be different, as discussed under ‘Human insensitivity to geometric inequivalence’.

For on-axis coupling, motion parallax cues again do not fulfil geometric assumptions given a rigid spatial scene. For example, a viewpoint movement along a line through the centre of the display does not cause a different image to be displayed, as the camera stays at a fixed distance from the fixation point. This is a cue to the flatness of the image. Furthermore, with on-axis coupling it is possible that the displayed objects move downwards while the observer moves only sidewards, and this is in conflict with geometric models of motion parallax. It is not clear how human perception deals with these inconsistencies, and to what extent theoretical models match human performance. Therefore it is difficult to predict apparent depth in on-axis virtual window displays.

Off-axis coupling may be more effective in giving a depth impression than on-axis coupling. Hayashibe (1991) showed that the apparent depth from shifting random-dot patterns is mainly caused by the relative speed between two shifting areas, and that the coupling to head movement only reduces the number of depth reversals. Similarly, Caudek and Proffitt (1993) argue that the depth perceived in a cone, protruding out of the display and oscillating around the vertical axis, is caused only by the relative shifts of object contours and not by foreshortening. This suggests that the rotational part of the on-axis coupling might not contribute to the apparent depth, but only to apparent rotational motion. As Ono and Steinbach (1990) found for random dot patterns that perceived object motion increases as perceived depth decreases, off-axis coupling may be more effective in suggesting depth.

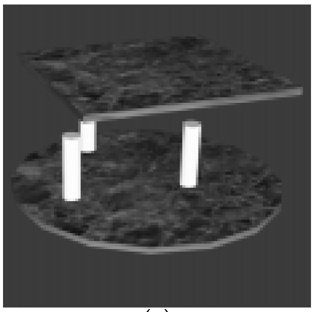
Concluding, in virtual window displays with off-axis coupling, perspective cues and motion parallax cues agree geometrically, and the parallax shifts may be more effective than with on-axis coupling. With on-axis coupling, perspective cues and motion parallax cues are geometrically inconsistent with each other, and in themselves in most cases. Geometry suggests that even distortions that vary with the movement of the viewpoint (nonrigid distortions) may be expected for on-axis coupling. It is not clear how human perception deals with these inconsistencies, and therefore it is very difficult to prediction distortions in on-axis virtual window displays.

Pilot experiments

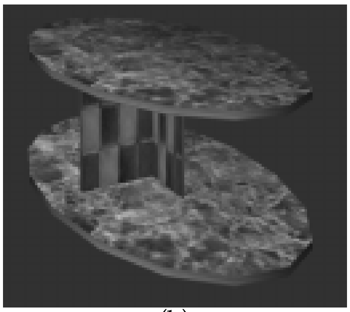
I did thirteen pilot tests, to gain an impression of the effects of the task and scene layout on the apparent layout and task performance. For example, in one pilot experiment observers adjusted one of three pillars to form an equilateral triangle (Figure 7.5a), and in another pilot experiment observers rotated a block-textured plane perpendicularly to another similarly textured plane (Figure 7.5b). For the last task, the effect of scene complexity was also investigated (Figure 7.5c). In another pilot test similar to that of Figure 7.5b, observers had to adjust the width of the textured plane to the width of the other plane.

Most effects could be explained from compression of the apparent depth with on-axis coupling. For example, the angles were adjusted smaller than 90˚ with on-axis coupling, and planes protruding out of the display were adjusted wider than planes parallel to the display. Paradoxically, shear distortion was noticeable only with *off*-axis coupling in spheres, triangles, and also of the smoother landscapes of the present experiment, especially when the objects protruded out of the display and were viewed with both eyes. With on-axis coupling no distortions were noted, neither changing with the viewpoint nor static.

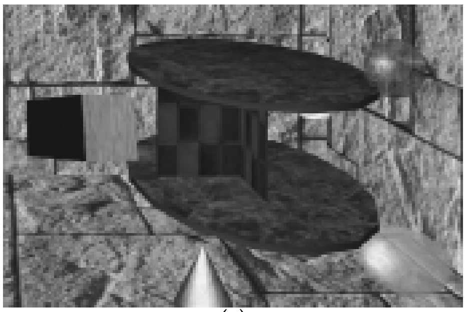
Summarizing the long introduction, off-axis coupling seems necessary to avoid distortion, if human perception uses depth cues in a geometric way. It seems that humans notice only certain forms of distortion, and that these distortions do not occur with on-axis coupling but may occur with off-axis coupling. On the other hand, the behaviour as found from task performance can be explained with distortions as predicted by geometry. I am mainly interested in task performance, and less in noticeable distortions.



(a)



(b)



(c)

Figure 7.5. Views from three pilot experiments. See text. Distortion could be noticed with off-axis couplings (b, c) but not with on-axis couplings (a). On the other hand, the results could be explained from a compressed apparent depth with on-axis coupling.

Experiment

As discussed, this experiment is a first exploration of the effects of geometric inequivalence and cues to the flatness of the display on the apparent layout in virtual window displays. Although previous work (see above) is not conclusive about the mechanisms involved in depth perception, my working hypothesis was the use of a mechanism compensating for the effects of a geometric inequivalent viewpoint (mechanisms one and two in the section ‘Human insensitivity to geometric inequivalence’). By diminishing other cues to the flatness of the display, I hoped to be able to manipulate the occurrence of compensation by having participants look with only one or with both eyes. With on-axis coupling, the geometric equivalent viewpoint always lies perpendicular to the display, but not with off-axis coupling. If the compensation mechanism supposes the viewpoint to lie perpendicular to the display, with on-axis coupling apparent depth would be unaffected by a geometric inequivalence. Therefore, we can distinguish from the experimental results what kind of compensation mechanism is being used.

Method

Variables, subjects, design

The independent variables (Table 7.1) were the coupling method *C* (on- or off-axis), the viewing condition *V* (monocular or binocular), a viewpoint measurement error parallel to the display plane *DXY* (0 or +10 cm in a random direction parallel to the display plane) and a viewpoint measurement error perpendicular to the screen plane *DZ* (0 or 10 cm out of the display plane). Due to a mistake, *DXY* was always to the right, relative to the observer (see Figure 7.13) instead of in all directions.

The dependent variables were the height ratio *H* (subject's height setting / measured height of foam model) and the total response time for each landscape *T*. All head positions were recorded during the experiment to check whether the average viewing distance matched the fixed camera distance from the fixation point for on-axis coupling (see 'Coupling method').

At random, six subjects were assigned to the monocular viewing condition and six to the binocular viewing condition. Each of the subjects judged 16 landscapes each of which contained 5 bumps (Figure 7.6a, b). The 2(*P*) x 2(*DXY*)x 2(*DZ*) x 2(repetitions)=16 conditions were randomized over these 16 landscapes, and the landscapes were presented in random order. For each subject, the 10 (2 landscapes x 5 bumps) height setting/measured height ratios he made in each condition were averaged, to get his height ratio *H*=virtual height/measured height. This gives a mixed analysis of variance (SPF-p.qru) design (Kirk, 1968).

The subjects were 12 naive volunteer students (6 male, 6 female) with normal or corrected-to normal vision. All subjects had stereoscopic vision, as they were able to recognize a figure hidden in a random-dot stereogram (Appendix A). They were paid NLG 7.50 (a loaf of bread costs about NLG 2).

Table 7.1. Independent variables and their levels.

Variable	Description	Levels
C	Coupling method	on-axis, off-axis
V	Viewing condition	monocular, binocular
DXY	viewpoint measurement error parallel to the display plane	0 cm, 10 cm in a random direction parallel to the display plane.
DZ	viewpoint measurement error perpendicular to the display plane	0 cm, 10 cm out of the display plane

Stimuli

The stimuli consisted of pairs of real (foam) and virtual (computer simulated) landscapes. Figure 7.6 shows a picture of both. The virtual stimuli were 228x228 mm and consisted of a gently sloping landscape (+5 .. -5 mm) with 5 bumps of height 10 .. 25 mm. They were rendered with 1280x1024 24-bit colour pixels by a Silicon Graphics Reality Engine. For on-axis coupling, the camera distance to the fixation point was 450 mm. This distance was chosen to match the expected average viewing distance of the participants. No spotlights were added in the virtual scene, so there was no shading information in the virtual stimuli. This was necessary to prevent the participants from adjusting the shadows instead of the bumps. The texture on the virtual bumps was made by scanning a real texture on a flat piece of foam. The real stimuli were milled in polyurethane foam from the computer data with an accuracy of 0.25 mm, and covered with a granite-effect structure spray (Plasti-kote Fleckstone 35: serpentine marble). They were lighted with a 20W halogen spot set to 26W to get a daylight-like colour. The spot was 1.8m higher and slightly from the back of the landscape, to get a light shading on the bumps at the side of the observer. The texture of the virtual stimuli was carefully matched to the real stimuli.

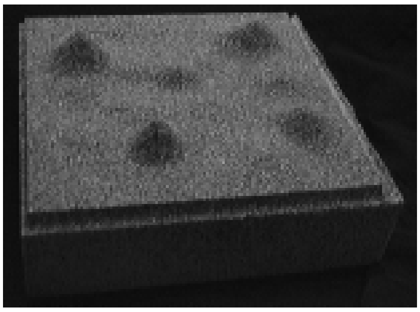


Figure 7.6a (**see colour figure on right cover flap**). Real landscapes were milled in foam and had textured paint sprayed on the surface.

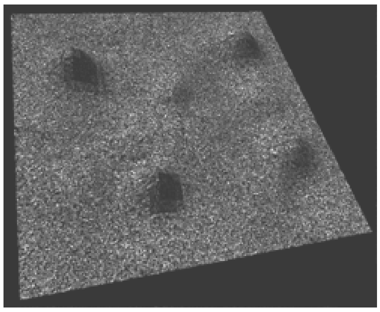


Figure 7.6b (**see colour figure on right cover flap**). Rendered virtual landscape. Shadows were added for this picture, to make the bumps visible.

Apparatus

Figure 7.7 shows an overview of the experimental setup. In the middle is the monitor display for the virtual landscape. On its left is the real landscape, on its right the button box and the mouse for adjusting the bumps. Behind the monitor is the eye position tracker. A trade-off was made between the image complexity and the update rate.

Figure 7.7. An overview of the experimental setup. The monitor with the virtual landscape is in front of the subject. The real landscape is on the left, the head tracker behind and the mouse and button box to the right of the monitor. A cardboard cover attached to the display served as a reduction screen.

I chose to use approximately 1800 triangles and an update rate of 29 Hz. The triangle density on the bumps was 2 times as high as the landscape between the bumps. While the subject was adjusting a bump, the update rate dropped to 14 Hz. The 21 inch screen (Taxan Ergovision 2100LR; 0.3 mm dot pitch; anti-glare and anti-reflection coating; refresh rate 100 Hz) was placed parallel to the floor. This setup was designed to elicit extremely oblique viewpoints from the subjects (see Chapter 2). The short side of the monitor was next to the subject, to prevent the bumps going off the display when the observer was looking at extreme viewpoints. In front of the screen, a reduction screen of 25 x 32 cm hid the borders of the screen to reduce conflicting visual cues between the views on the screen and the environment. The visual cues from the environment were further reduced by a dark background and dimmed lighting. The screen was warmed up at least 30 minutes in advance of each trial to prevent colour changes during the experiment. The 843x843 pixels of the virtual landscape matched the size of the real landscape. A box with two buttons, a green one labelled 'volgende bult' (next bump) and an orange one labelled 'landschap OK' (landscape finished), allowed the subject to select the next bump or to finish the landscape. An infrared tracker (Dynasight from Origin Instruments) tracked a small reflector on a spectacle frame that the subjects wore. Of monocular observers, the position of the eye they used was tracked, of binocular observers the average of the two eyes was tracked. As the reflector was not exactly in the eye position, the position of the eye contained an error of at most 3 cm, but for normal viewpoints the error was estimated to be about 0.5 cm.

Procedure

It was explained to the participants that they would be presented a real and a virtual landscape, and that they had to adjust the height of the 5 peaks on the virtual landscape to match the heights of the corresponding peaks on the real landscape. They were told to increase the height of the selected bump by moving the mouse away, and to decrease the height by pulling the mouse towards them. With the button labelled 'volgende bult' (next bump) they could select the next bump. If they pressed this button again at bump 5, bump 1 was selected. This way, they could jump through the bumps to select one for adjusting the height. If they were confident about their settings, they had to press the button labelled 'landschap OK' (landscape OK). Then the virtual landscape disappeared, the real landscape was replaced by a new landscape and the new landscape appeared on the screen.

About the head tracker they were told that it followed a small reflector on a spectacle frame. Subjects in the binocular viewing condition received a spectacle frame with a small reflector. Subjects in the monocular condition were told that they have to look with one eye, and their dominant eye was determined, as described in Chapter 6. These subjects received a spectacle frame with both a reflector and an eye patch covering their non-dominant eye. It was explained to the subjects that an LED on the tracker would turn from green to red if they moved too fast or out of range, and that the image on the screen would shake in that case because it could not be adjusted in the correct way.

Participants were told not to be obsessed with a perfect match, but to match the height in a few seconds. They were asked to try to set the correct height of each bump the first time they selected it and to take about 10 seconds for each bump, as it would take too much time to re-adjust all the bumps later. There was no strict time limit, but they were warned during the training (and in two cases during the experiment) if they took more than 20 seconds for each bump. Subjects were asked not to touch the screen or the real bumps.

The light was then turned off, except for the spotlight for the real landscape. Scattered light allowed the subjects to see the button box and the mouse. For training, they were asked to adjust a landscape twice. Some subjects noticed that they could look around the virtual landscape during the first presentation of the landscape, but not during the second (the virtual landscape was projected on-axis during the first presentation, and off-axis during the second presentation). If participants asked about this, they were told that one of the aims of the experiment was to test this difference in ability to look around the scene. They were not given feedback about the correctness of their settings, as this might have frustrated the subjects with on-axis coupling. After the training subjects were allowed to ask questions, and they were asked to work faster if necessary.

Before the experiment, participants were informed that they had to adjust 16 landscapes (none of the subjects noticed that there were only 8 landscapes, each presented twice but rotated 90°). In total, each trial lasted about 40 minutes.

Results

Initially it was planned to analyse the difference between the matched virtual height and the measured height of the bump. However, this would give negative values if the subject adjusted lower than the measured height. Furthermore, the variance of this value was found to depend linearly on the height of the real bump. Therefore, the height ratio H (virtual bump height as matched by the subject / measured bump height) was analysed. The ratio H solves both problems, and moreover it is a dimensionless ratio with an optical interpretation, allowing direct comparison of the depth scaling due to geometric inequivalence d'/d (see 'Geometric inequivalence'). For example, a setting giving a ratio H=2 indicates that the virtual bump was adjusted twice as high as the real bump to appear to be of equal height, or to put it in another way the apparent depth in the virtual scene was half as great as the apparent depth in the real scene.

The setup was successful in eliciting extreme viewpoints. Especially for inspecting the real landscape, a lot of subjects looked at a at nearly-zero height over the landscape to see the bumps in side view. They usually tried the same with the virtual landscape, although they had to take a slightly higher viewpoint to avoid clipping of the bumps by the screen border.

A repeated-measures analysis of variance (Kirk, 1968) on the height ratios (Table 7.2) shows significant main effects of all variables. The mean height ratio H for on-axis coupling is 1.79, for off-axis coupling 1.31. A post-hoc t-test (Hays, 1981) shows that the average H is larger than 1 with on-axis coupling ($p<0.01$). Another t-test showed that for off-axis coupling the average ratio H does not differ significantly from 1.

Table 7.2. Significant effects for the analysis of variance performed on the height ratio (height setting/measured height) H.

Interaction	F	p
V	$F(1,10) = 5.35$	<0.05
C	$F(1,10) = 119.21$	<0.001
DXY	$F(1,10) = 19.88$	<0.01
DZ	$F(1,10) = 58.81$	<0.001
C x DZ	$F(1,10) = 5.47$	<0.05
V x DXY x DZ	$F(1,10)=9.88$	<0.01

With monocular viewing, the mean height ratio H is 1.71, with binocular viewing 1.39. The hypothesis that binocular viewing reduces performance only for off-axis coupling is not confirmed: the V x C interaction is not significant, and binocular viewing improves the performance with both coupling systems.

A viewpoint measurement error perpendicular to the screen DZ gives an increase from 1.405 to 1.694 of the mean height ratio as compared to the condition without measurement error. An error parallel to the display plane DXY *decreases* the mean ratio from 1.601 to 1.497. Figure 7.8 shows the effects of the independent variables on the mean height ratio H. A post-hoc analysis of variance was done to check the absence of interactions with off-axis coupling as suggested by Figure 7.8. The analysis showed that this is indeed the case, as for off-axis coupling only the viewpoint measurement errors have a significant effect on the height ratio. For a viewpoint measurement error parallel to the display DXY $F(1,10)=9.08$, $p<0.05$ and for a viewpoint measurement error perpendicular to the display DZ $F(1,10)=63.1$, $p<0.001$.

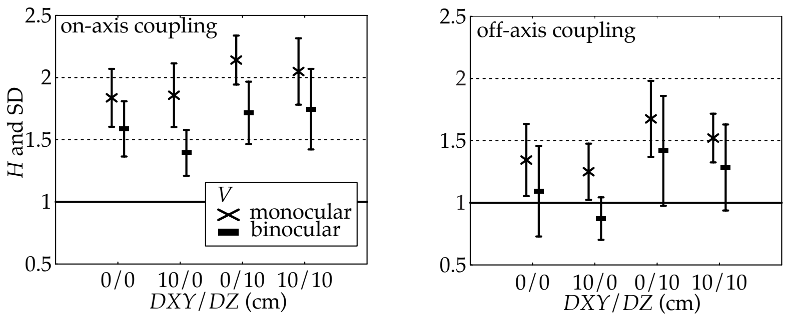


Figure 7.8. The effects of the coupling method P, the viewing condition V, a viewpoint measurement error parallel to the display plane DXY and perpendicular to the display plane DZ on the mean height ratio H.

The analysis of variance of the mean error on the height ratios H (Table 7.1) also showed an interaction between DZ and the coupling method C. Figure 7.9 shows this interaction: with on-axis coupling, with DZ = 0 cm H=1.67, and with DZ = 10 cm H=1.91. The increase of H with off-axis coupling is larger: from 1.14 to 1.48. Thus, off-axis coupling is more sensitive to viewpoint measurement error perpendicular to the display. For a viewpoint measurement error parallel to the display DXY, this sensitivity difference does not exist, as there is no significant interaction between the coupling method P and DXY. Furthermore, because H gets closer to 1 in the presence of a viewpoint measurement error parallel to the display plane DXY as compared to absence of such an error, it seems that the apparent layout gets closer to the Euclidean layout with DXY than without DXY.

The three-way interaction V x DXY x DZ is shown in Figure 7.10. It shows that the difference between monocular and binocular viewing is much larger at a geometrically inequivalent viewpoint than at the geometrically equivalent viewpoint.

A Pearson product-moment correlation (Norusis, 1993) did not show a correlation between the distance of the bump to the centre of the screen and the height ratio H.

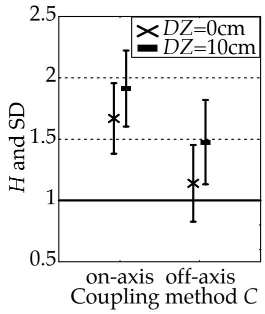


Figure 7.9. The effects on the mean height ratio *H* of the coupling method *C* and a viewpoint measurement error perpendicular to display plane *DZ* .

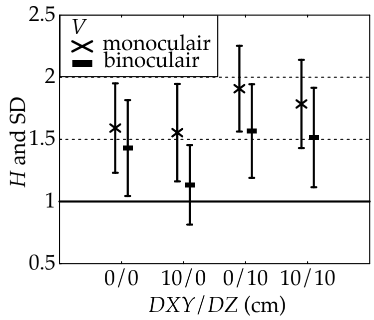


Figure 7.10. The effects on the mean height ratio *H* of the viewing condition *V*, a viewpoint measurement error in the display plane *DXY* and perpendicular to the display plane *DZ*.

A repeated-measures analysis of variance of the response time *T* shows a main effect for *DXY* (p<0.05) and for the three-way interaction *V* x *DXY* x *DZ* (p<0.05). Figure 7.11 shows the effect of the variables on the mean response time *T*. The average response time is 75.5 s (15.5 s for each bump). In contrast with the error on the height ratio *H*, the response time for the *DXY* = 10 cm condition is larger than in the *DXY* = 0 cm condition. This indicates that some difficulties with shear may be present. As with the effect of the same interaction on the height ratio *H*, the difference between monocular and binocular observers is much larger when a viewpoint measurement error *DXY* or *DZ* is present than when it is not.

The difference between the observers was checked graphically. Figure 7.12 shows that the average height ratios *H* are different for different observers, and that one observer has nearly the same response with on- and off-axis coupling.

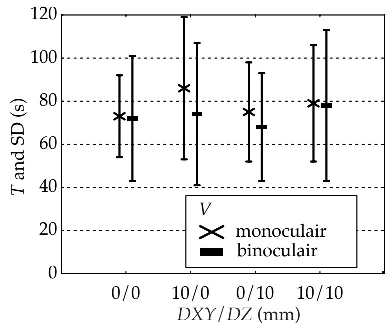


Figure 7.11. The effects of the viewing condition *V*, a viewpoint measurement error in the display plane *DXY* and perpendicular to the display plane *DZ* on the mean response time *T*.

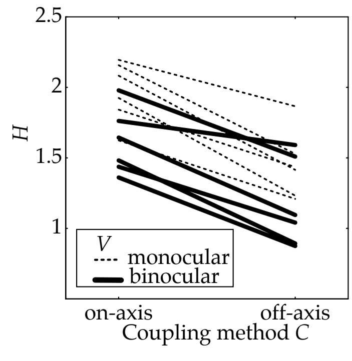


Figure 7.12. The effects for each observer of the viewing condition *V* and the coupling method *C* on his mean height ratio *H*.

An analysis was done to test for an effect on the height ratio *H* of the direction of the viewpoint measurement error parallel to the display plane *DXY*. Figure 7.13 shows the errors on the height ratio *H* for each tested direction of the error. For an observer sitting normally in front of the display (Figure 7.7), the positive x direction is backwards and the positive y direction is to the right. Erroneously, the direction of each error was chosen randomly between 0 and π instead of between 0 and 2π.

The effects of the x- and y- components of the direction on the height ratios were tested separately. A Pearson product-moment correlation (Norusis, 1993) was done for both coupling methods. All correlations were found to be significant. For on-axis coupling the ratio decreases with increasing x (r=-0.2, p<0.01) and with increasing y (r=-0.33, p<0.001). For off-axis coupling the ratio decreases with increasing x (r=-0.34, p<0.001) and increases with increasing y (r=0.2, p<0.01).

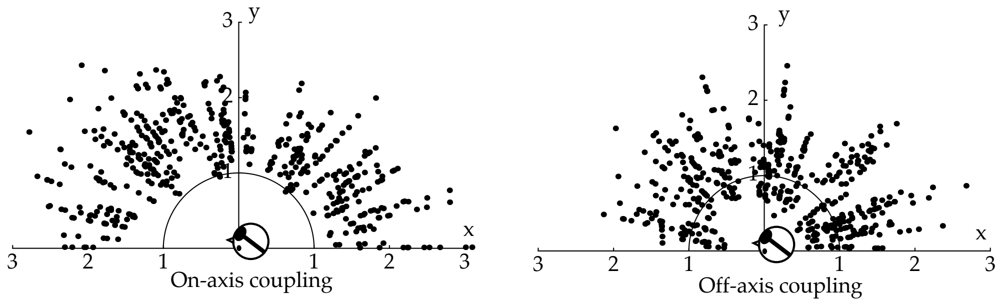


Figure 7.13. The effects on the mean height ratio H of the direction of a viewpoint measurement error parallel to the plane of the display DXY and the coupling method C. The angular position of a dot indicates the direction of the viewpoint measurement error, and H is represented by its distance from the origin. For on-axis coupling, H is outside 1 and elliptical. For off-axis coupling, H is around 1 and circular.

To check the fixed distance of the camera to the fixation point for on-axis coupling, the head positions of the subjects were averaged to find the average viewpoint. This was found to be 394 mm to the bottom, 44 mm to the right and 274 mm above the display, so the average viewing distance to the fixation point was 482 mm. Therefore, the fixed camera viewing distance for on-axis coupling (450 mm) had been well chosen.

Summarizing, with off-axis coupling the apparent height of the virtual bumps is equal to the apparent height of the real bumps. With on-axis coupling the virtual bumps have to be 1.79 times as high as the real bumps in order to give an equal apparent height, indicating that the apparent height is lower than the virtual height with on-axis coupling. As compared with on-axis coupling, with off-axis coupling the matching is more sensitive to viewpoint measurement errors perpendicular to the plane of the display. Similarly, monocular viewing reduces apparent depth as compared with binocular viewing in the presence of a viewpoint measurement error. For a viewpoint measurement error parallel to the display, the direction of the error also influences the apparent depth.

Discussion

The finding that geometric inequivalence affects apparent depth with on-axis coupling is in conflict both with a human compensation mechanism that works by assuming the correct viewpoint perpendicular to the display and with a compensation mechanism that extracts the correct viewpoint from pictorial cues. Instead, the results suggest that no compensation mechanism is used at all.

Many results can be explained by geometry. The height ratios *H* can be interpreted as how much the apparent depth from the scene on the display is compressed as compared to the apparent depth of the real scene. For the coupling method, an important result is that on-axis coupling causes an apparent layout corresponding to a depth-compressed Euclidean layout, while with off-axis coupling the apparent layout of a displayed scene corresponds to the Euclidean layout, provided that the viewpoint is measured accurately. The average depth compression for on-axis coupling was 1.79. This compression can be explained geometrically, as follows. The average (perpendicular) distance of the observers to the plane of the display was 274 mm. As the camera for on-axis coupling was at a constant distance of 450 mm, the scaling of the depth in a geometric reconstruction is 274/450 = 0.61 (see 'Geometric inequivalence'). To compensate for this scaled depth, the bump would have to be 1/.61=1.64 times the normal height. This is slightly lower than the 1.79 from the experiment, so perhaps the observers move slightly closer to the display while comparing the precise heights of the bumps, and move away again during the adjustment. If a compensation mechanism assumed the correct viewing distance to be twice the height of the display (see 'Human insensitivity to geometric inequivalence'), 50 cm would be taken as the correct viewing distance. 50 cm is close to the actual camera distance of 45 cm, and therefore no distortion would be expected if human perception used such a mechanism. But this is not in agreement with our results. For off-axis projection, a viewpoint measurement error of 10 cm increases the average viewing distance from 27.4 cm to 37.4 cm. Geometry predicts a height ratio of 37.4/27.4 = 1.36, which is close to the actual results. For off-axis projection, geometry predicts independent effects of a viewpoint measurement parallel to the display and such an error perpendicular to the display, which also was shown to be the case for the experimental results. Concluding, the results, both for on- and for off-axis coupling, are in agreement with the compression predicted by uncompensated geometry.

A last argument against a compensation mechanism assuming a correct viewing position perpendicular to the display comes from the results for monocular and binocular viewing. It was hypothesized that binocular viewing provides a flatness cue that triggers a compensation which, in its turn, worsens performance with off-axis coupling but not with on-axis coupling. Furthermore, binocular viewing was expected to introduce a conflict between parallax shifts and stereoscopic cues. Experimentally, though, binocular viewing was found to improve performance as compared to monocular viewing, especially if the observer is dislocated. Why do binocular observers perform better? Consider the explanations mentioned in 'Cues to the flatness of the image'. In contrast with the experiment of Arthur, Booth and Ware (1993), observers in the monocular condition in the present experiment always looked with their dominant eye. Reduced comfort or conflicting information from the two eyes in the monocular condition (see 'Cues to the flatness of the image') is unlikely, given the results of the experiment described in Chapter 6. But there are other differences between the present experiment and the experiments that indicate an advantage for monocular observers. In both the experiment of Koenderink, van Doorn and Kappers (1994) and that of Chapter 6 the environment was clearly visible, although a reduction screen in front of the monitor was used in the experiment of Chapter 6. Koenderink et al. (1994) tested the depth perceived in photographs by stationary observers at the geometrically equivalent viewpoint. For the knot tracing task of Chapter 6, the observers could move, and the observer's angular position was scaled by a factor 4. Furthermore, extreme viewpoints were provoked by the present experiment, but not in the experiments of Koenderink et al. (1994) and that of Chapter 6. All these differences can affect the apparent depth, as was discussed in the 'Previous work' section.

Finally, one result suggests a compensation mechanism that assumes a correct viewing position perpendicular to the display: on-axis coupling was found to be less sensitive to a viewpoint measurement error perpendicular to the display than off-axis coupling. However, the sensitivity effect may also be caused by the indirect effect of a viewpoint measurement error perpendicular to the display for on-axis coupling: the camera is always at 450 mm from the fixation point and only its viewing direction is affected by a viewpoint measurement error.

There are two unexpected results. The first is that a viewpoint measurement error parallel to the display decreases the error on the height settings. Geometry indicates that this causes a shear distortion, which should not affect the height of the bumps. One explanation is that subjects may match the axis of the bump instead of the perpendicular height of the bumps (Figure 7.14), and that therefore a bump appears to be higher if a shear distortion away from the observer or sideways is present. At an average viewing distance of 27.4 cm, a viewpoint measurement error of 10 cm would enlarge the diagonal of the bump 1.065 times. This fits the actual enlargement of 1.601/1.497=1.069 amazingly well.

The last unexpected result is that for on-axis coupling the position of the bump relative to the fixation point does not influence the error on the height settings. Apparently, the movements of bumps far from the fixation point do not disturb the subjects.

Figure 7.14. Instead of the perpendicular height of the bumps (left), subjects may have used the axis of the bumps (right) in making their judgements.

Most subjects seemed not even to perceive movement of bumps, they just noticed that they were able to look further around the scene. Also, no effect of the coupling mechanism on response time was found. These results suggest that response time depends on distortions of which the observer is conscious, while observer performance is affected by other distortions of which the observer is less conscious.

Conclusions

Most results could be explained from distortions as predicted by geometry. The literature suggested that cues to the flatness of the display might trigger a perceptual mechanism compensating for the resulting distortions. But the major part of the present results could be explained by the absence of such a compensation. This result is in agreement with more recent literature. Absence of compensation geometrically implies that off-axis coupling gives an apparent depth matching Euclidean depth, and that on-axis coupling gives a compressed apparent depth if the observer is closer to the display than geometrically equivalent given the fixed camera distance to the fixation point. It can also explain why on-axis coupling is less sensitive than off-axis coupling to a viewpoint measurement error perpendicular to the display. The fact that with on-axis coupling observer performance is independent of the distance of the bumps from the fixation point (the centre of the display) is unexpected.

There are several factors to consider when choosing the appropriate coupling method. When the apparent depth of a displayed scene must match some Euclidean measures, for example for enhanced reality setups where a displayed world is projected over the real world, off-axis coupling is required[4]. On the other hand, on-axis coupling also has advantages over off-axis coupling. The technical implementation of an on-axis coupling is cheaper than that of an off-axis coupling, especially when a real camera is used. As the experiment described in Chapters 5 and 6 indicated, head movements can be scaled with on-axis coupling, allowing the observer to look further around the object. In contrast with off-axis coupling, with on-axis coupling other observers looking along with the moving observer do not see large shear distortion but an 'ordinary closed-circuit TV picture'. Furthermore, it may be possible to correct in a simple way for the reduced perceived depth, because the depth compression seems quite constant over a large range of viewpoints. Finally, on-axis coupling was found to be less sensitive to errors in the distance of the eye to the display. For many tasks, these advantages may be more important than geometric equivalence. For a given application and task, a trade-off has to be made on these points. It is possible to combine on- and off-axis coupling, by rotating the scene depending on the angular position of the observer and using an off-axis coupling, but the perceptual aspects of such a combination remain to be examined.

Concluding, we found that subjects' behaviour was very well predicted by the geometric model: people did not correct for a geometrically inequivalent view. Human perception notices some distortions, while other, unnoticed, distortions disturb task performance. The results support the conclusions of Halloran (1989), and are in perfect accord with our ideas about the task dependency of the required views: "when partial cues conflict, the choice among them will depend on the requirements of the perceptual task at hand. To account for performance, it seems necessary and sufficient to know two things: the partial geometries being projected, and the observer's perceptual decision strategy" (p. 478). If I had time for more research, I would concentrate on this 'observer's perceptual decision strategy' and its relation to his task.

[4] Another argument for choosing off-axis coupling might be that on-axis coupling gives an undesired vertical disparity when it is used to present separate pictures stereoscopically to the observer's two eyes (Castle, 1995). However, vertical disparity is no real problem as it also occurs with natural viewing. Vertical disparity may be even used to recover depth (Bishop, 1996), but it may be not used as such by humans (Cumming, Johnston and Parker, 1991) or only with large displays (Rogers and Bradshaw, 1993).

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.

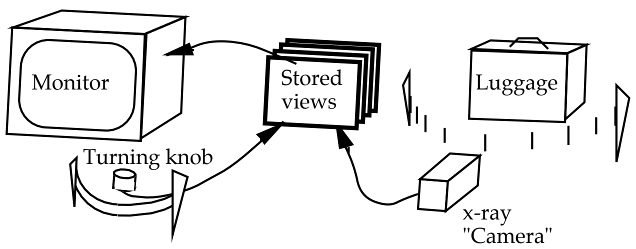


Figure 8.1. Scheme of the setup. A turning knob is used to select the closest available view, which will be displayed. The available views were taken earlier.

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).

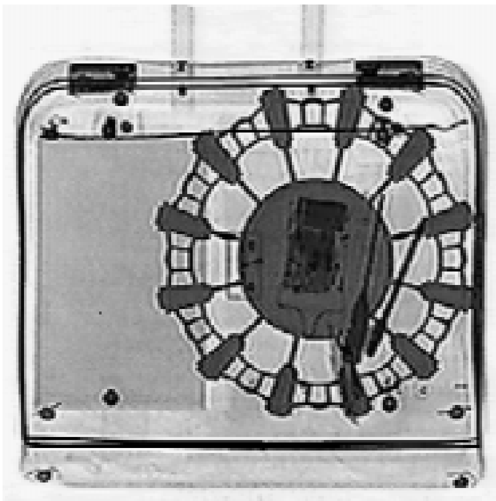


Figure 8.2 (**see colour figure 1.1 on left cover flap**).
Example of x-ray scan of suitcase containing a clock, a book and a tin opener. Colours (see text) indicate the materials.

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.

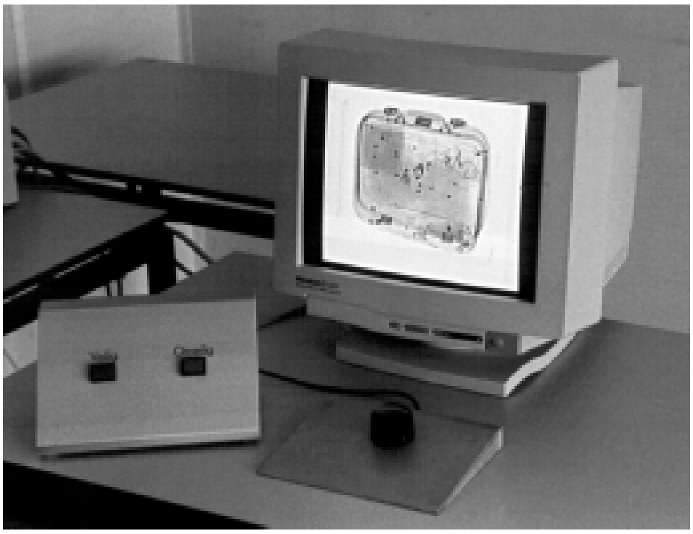


Figure 8.3. Overview of the experimental setup. One of the available views of the suitcase could be selected with the turning knob. The decision for 'safe' or 'unsafe' was made with the button box.

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*.

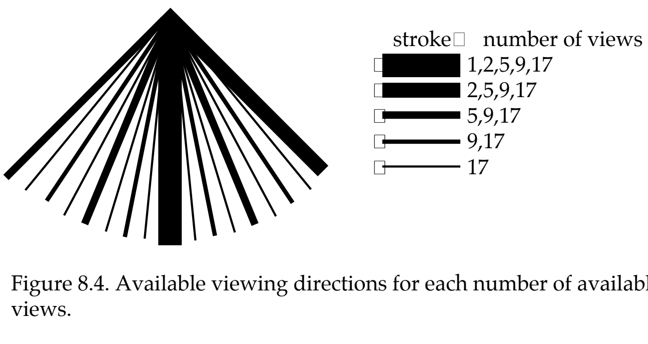


Figure 8.4. Available viewing directions for each number of available views.

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.

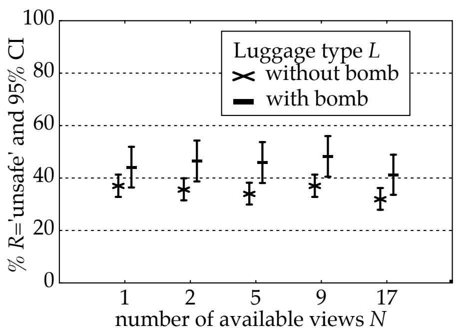


Figure 8.5. Percentage responses R='unsafe' of the participants for different baggage types and number of available views.

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: $F(4,3290)=0.46$, $p=0.769$.

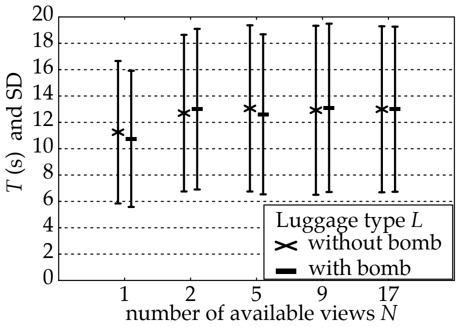


Figure 8.6. Average response times of the participants in the 5 conditions for different baggage types and number of available views.

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.

9

Implications of the findings

This chapter discusses the relevance of the findings of this thesis for x-ray baggage inspection, virtual window displays, Industrial Design Engineering and perceptual theories.

X-ray baggage inspection

The results of the experiments with tasks that I thought to be relevant for x-ray baggage inspection indicated that the DVWS can be used to improve baggage inspection. The experiments showed that multiple views can compensate for low image resolution and a low number of grey levels. Furthermore, the DVWS allows the observer to select a useful viewpoint, which may enhance performance on several tasks. It was expected that the image complexity of real baggage would fall somewhere between the images used for the connected-objects task (Chapter 5) and the images used for the knot-tracing task (Chapter 6). It was therefore expected that the performance of the observer would also increase with the number of viewpoints, at a rate somewhere between the results for the connected-objects task and the knot-tracing task.

The results of the connected-objects task indicated that this task can be done without a decrease in observer performance when the viewpoint selection by eye position is replaced by viewpoint selection via a knob. This result is important as tracking the eye position of an observer is difficult, requires expensive apparatus and, for reliable operation, still requires the operator to wear a distinctive marker near his eye. This last point, in particular, was expected to meet resistance from baggage inspectors. Thus, replacing the viewpoint selection by eye position with viewpoint selection via a knob seems an appropriate choice for an x-ray baggage inspection system based on the DVWS.

The analysis of Chapter 3 and the bump-matching experiment of Chapter 7 indicated that if such viewpoint selection is done by knob instead of by eye position, on-axis coupling, such as that provided by the DVWS, gives less distortion than off-axis coupling. With off-axis coupling shear distortion will occur, and this may lower performance on spatial tasks. On the other hand, the bump-matching experiment indicated that with on-axis coupling perceived depth may be compressed as compared with the real depth by a factor 2. But such a compressed depth seems acceptable, as estimating real depth sizes is expected to be of minor importance for x-ray baggage inspection.

However, given the results of the real-baggage experiment (Figure 9.1; see Chapter 8) I was unable to prove the usefulness of the DVWS for x-ray baggage inspection. No effect of the number of available views on the judgements was found. This is surprising given the promising results of earlier experiments. On the other hand, the usefulness of the DVWS has not been disproved, as there are a number of alternative explanations for the results of the last experiment.

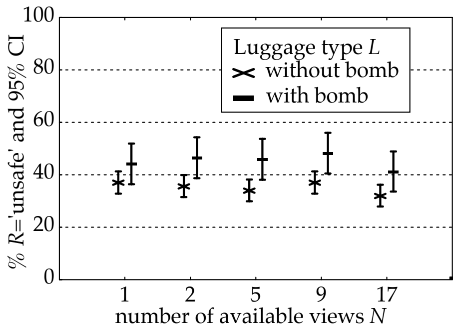


Figure 9.1. Judgements of baggage with and without a bomb. With more than one viewpoint the judgements of bomb- and non-bomb baggage differ with 5% significance (repeated from Figure 8.5).

As discussed in Chapter 8, the most serious problem is that the baggage inspection task is not operationalized clearly. Baggage inspectors are unable to explain precisely what they do to determine whether a suitcase is safe. Usually they say that they are looking for suspicious items, such as batteries and explosives. However, there are a lot of suitcases that contain batteries and may contain explosives, but are not indicated as dangerous. Furthermore, potentially dangerous objects may have an unusual appearance, not previously encountered: for example a recent development is an all-plastic battery (Simon, 1997). The minimal cues that have to be presented to the baggage inspector cannot be determined without such a clear operationalization of his task. In the case of medical tasks there is a more extensive literature about inspection of x-ray photographs. In the case of medical x-ray inspection tasks there is an uncertainty about what good inspectors are looking for (Bass and Chiles, 1990). This is similar to the findings for x-ray baggage inspection in this thesis. Future research about baggage inspection should start by operationalizing the baggage inspection task.

Another problem that might explain the results is that, with their experience and long trainingtraining, the inspectors were so used to working with a single front view that they failed to pick up the extra spatial cues provided by the DVWS. For example, they may have based their judgement on a single view. If the inspectors really fail to use spatial cues for baggage inspection, a remedy would be to train the inspectors to use this information. Because of the long experience of the inspectors, such training cannot be given just by giving a training session in advance of an experiment. Instead, this possible defect in the experiment can be evaded by giving a few inspectors intensive training. This is expensive and is probably attainable only in close co-operation with a manufacturer and users of baggage inspection apparatus.

Concluding, experimental results for tasks that were proposed as relevant for x-ray baggage inspection indicated that observer performance will increase with the number of available views. However, the experiment with real baggage showed no performance increase with increasing numbers of available views. Future research concerning x-ray baggage inspection should describe a clear operationalization of the baggage inspection task first first.

Virtual window displays

This section discusses implications of the properties of a virtual window display on the performance of an observer. The image quality (resolution and number of grey levels), cues to the flatness of the display and the coupling method (on- or off-axis coupling) were shown to affect the task performance of an observer working with a virtual window display. Furthermore, the intuitiveness of the viewpoint selection mechanism and the delay between the movement of the eye of the observer and the corresponding update of the display are important. These aspects will be discussed below, except for intuitiveness, which will be discussed under 'Industrial Design Engineering'.

Image quality

Much is known about the requirements for a static image, given some visibility requirements of objects in the depicted scene (e.g. Snyder, 1973; Olzak and Thomas, 1975; Gille, Samadani, Martin and Larimer, 1994). The availability of multiple views may compensate for low spatial image resolution and for a low number of grey levels, as was shown in the connected-objects experiment of Chapter 5. A similar effect seems to play a role with television: a film on television gives a much higher impression of the image quality than when single frames from the film are inspected. With virtual reality via helmet-mounted displays, the static image quality is far lower: typical VR image resolutions are 320 x 200 colour pixels for each eye (Holloway and Lastra, 1993). Similarly, MPEG video compression adjusts the resolution of a particular frame to the amount of difference between that frame and the next (see also Gonzalez, 1995). In spite of the common use of this effect and the extensive theoretical and technical literature about it, little perceptual investigations have been done on this effect.

Cues to the flatness of the display

Cues to the flatness of the display may have perceptual consequences. Such flatness cues can be (1) stereoscopic cues, if the display does not provide stereoscopic cues about the scene, while the observer looks at the display with both eyes, (2) a frame around the display, (3) a grid laid over the display, caused by the pixels of raster displays and (4) absence if or inappropriate shadows or shading.

According to the indirect theory (see Chapter 2), such flatness cues will flatten the depth in the 3D reconstruction made by the observer, thus affecting his performance if he needs depth cues for his task. Furthermore, flatness cues may trigger some mental mechanism compensating for viewing pictures obliquely (see Chapter 7).

In this thesis, only the effect of using both eyes while the display does not provide stereoscopic cues about the scene were investigated. X-ray images do not contain shadow cues, and the grid caused by a raster display is unavoidable given the choice to use a sensor line (see Chapter 3). A reduction screen hiding the environment and the frame of the display was used in all our experiments, because I felt that it improved the apparent depth in the scene, but I did not test this effect experimentally.

For the tasks discussed in this thesis, looking with both eyes at a non-stereoscopic display did not hinder the observers. In the bump height matching task (Chapter 7) observers using both eyes performed even better than those using one eye. This finding has an important ergonomic and aesthetic consequence: observers working via a non-stereoscopic virtual window display do not need to work with an eye-patch in order to achieve high performance.

The effect may depend on the task, as other experiments, such as described in Chapter 6 and that of Arthur, Booth and Ware (1993), showed no advantage to observers using both eyes over observers using one eye. One explanation for this difference using the direct theory is that observers used the stereoscopic cues differently for the two tasks. For the knot tracing task, the observers may have tried to use stereoscopic cues for separating the wires. As there were no stereoscopic cues in the display, performance in the two-eyes viewing condition did not improve as compared to the one-eye condition. As discussed in the task analysis (Chapter 2), the stereoscopic cues might be less relevant for the bump matching task in a natural situation (where real instead of simulated bumps were to be adjusted). Therefore the observers may have used the images from the two eyes for reducing noise instead of using its stereoscopic cues. For example, Bradshaw and Rogers (1996) indicate that two images can be used to reduce noise in the images by a factor 1.4. The ability to reduce noise may explain the higher performance of observers using both eyes as compared to observers using one eye. Thus, although the observers know that they are looking at a monitor display, the way the observers use binocular cues may be identical to what they would do if they were doing the task via natural inspection.

Another explanation, using the indirect theory, is that monocular observers get an *increased* height impression of the real bumps as compared to binocular observers. For example, absence of stereoscopic cues might place too much weight on shadow cues, causing an exaggerated impression of the height of the bumps.

Summarizing, no disturbing effects of looking with both eyes at a non-stereoscopic display were found: binocular viewing may even improve the performance of the observer as compared to monocular viewing.

Coupling method

The choice of the coupling method was shown to have important consequences, especially for distortions. Given the results of the experiment described in Chapter 7, it seems reasonable to distinguish between distortions that are seen as such (*noticed* distortions) and distortions that are usually not noticed, but nevertheless influence performance (*unnoticed* distortions). A similar discrepancy between judged display quality and actual performance given some display quality also occurs in the case of static image quality (Overveld, 1994): subjective quality ratings of a static image are largely determined by noise and blur, while these are of minor importance for performing a visual task with these images. It is only if the target contrast is extremely low that contrast is a prime determinant of the visibility of targets (Vyborny, 1997).

On-axis coupling has several advantages over off-axis coupling. I start with the perceptual advantages. Distortions that occur with on-axis coupling are usually unnoticed, while noticed distortions occur with off-axis coupling. The unnoticed distortions seem strongly related to the fact that on-axis images always are 'regular television' images. Probably the distortions in these images are not noticed because we are used to such distortions, as they occur under normal viewing conditions in ordinary television images, photographs and paintings. First, the observer will not notice shear distortion with on-axis coupling if his viewpoint is measured inaccurately, as is the case with off-axis coupling. Such shear distortion may be disrupting, for example in recognition tasks and tasks involving the use of visual angles. Second, with on-axis coupling people looking along with the observer controlling viewpoint selection do not perceive highly distorting views, but they will see large distortions with off-axis coupling. Third, on-axis coupling does not require all degrees of freedom of the observer movements to be imitated by camera movement. As many tasks do not require all directions to be coupled, this may save the expense apparatus capable of tracking observer movements in all these directions. Fourth, with on-axis coupling the camera movements can be scaled relative to the movements of the observer, apparently without causing noticeable distortion. Such scaling can be done to increase the visible range of views, for example to improve the ability to look around an object. Such a scaling is not natural, as it will cause a conflict between the parallax cues and the proprioceptive cues (the data about body movement provided by muscle tension sensors and the equilibratory senses). Nevertheless, a scaled camera motion may be advantageous for some tasks. Fifth, in the presence of a delay no distortions are noticed, as occurs with off-axis coupling which is discussed below under 'Delay effect'. This point is discussed in the next paragraph. Finally, for some tasks viewpoint selection by eye position can be replaced with viewpoint selection via a knob without decreasing the performance of the observer.

However, unnoticed distortions do occur with on-axis coupling, even if the viewpoint is measured accurately. The results of the experiment described in Chapter 7 suggest that the fixed distance between camera and the fixation point should agree with the average (perpendicular) distance between the observer and the display. If this is not the case, the displayed scene is scaled in depth as compared to the real scene. This effect may be offset if depth scaling with on-axis coupling can be shown to be systematic, which is suggested by geometry. Such compensation can be achieved by expanding the depth in the scene before projecting it, or by manipulating the camera viewing distance and viewing angle. An advantage of on-axis coupling with regard to unnoticed distortions is that on-axis coupling is less sensitive than off-axis coupling to a mismatch between the distance of the camera and the actual distance of the observer's eye to the plane of the display.

On-axis coupling also has technological advantages over off-axis coupling. First, on-axis coupling is less sensitive to inaccurate viewpoint measurements. This may be explained by the fixed distance between camera and fixation point. Therefore, a less precise and probably cheaper eye position tracker and display system can be used. Furthermore, standard cameras provide an on-axis image, and therefore on-axis views can be acquired more easily and cheaper with real cameras than off-axis views. Off-axis coupling with a real camera will require either selection of a part of the image from a high-resolution camera with a large viewing angle, real-time image processing, or a special 'perspective correction' camera lens that shifts relative to the image plane, depending on the camera position (Figure 9.2). Such cameras with a lens that can shift relative to the image plane are used often in architecture (Figure 9.3).

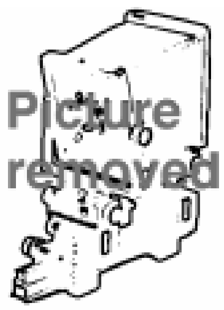


Figure 9.2. 'Perspective correction' camera capable of slanting the back containing the photographic plate. Such a camera can be used to record views for off-axis coupling (from Abraben, 1994).

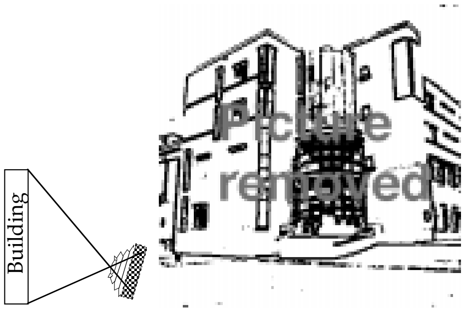


Figure 9.3a. The camera is slanted backwards to show all of the building. Vertical lines in the building run towards a vanishing point.

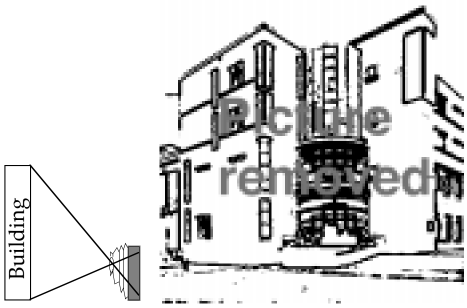


Figure 9.3b. If the back of the camera is rotated to be parallel to the vertical lines the building, these vertical lines will run parallel in the photo (Abraben, 1994).

On the other hand, off-axis coupling also has advantages over on-axis coupling. The most obvious advantage is that with off-axis coupling the displayed world forms a rigid whole with the real world around the monitor if the viewing position of the observer is coupled accurately to the displayed view. Therefore the displayed world can be fitted into the real world, or even mixed through it (mixed or augmented reality, see Drascic and Milgram, 1996). For example consider the height of the horizon in the displayed scene. With off-axis coupling the height of the displayed horizon stays at eye height when the observer moves. With on-axis coupling the height of the displayed horizon depends on the visual angle of the observer's eye relative to the display, and this does not depend directly on his absolute eye height. The direct theory suggests that this has large consequences for tasks where the horizon is used. For example, consider the stair climbing task (see Figure 2.4). Figure 9.4a shows an observer sitting perpendicular to the centre of the display, and therefore the camera is not rotated relative to the scene. Figure 9.4b shows an observer who moved away from such a perpendicular position. This causes the displayed horizon to deviate from the horizon of the environment of the display. The direct theory suggested that the observer uses the visual angles α and β (see Figure 2.4) directly to determine his ability to climb the stair. But with on-axis coupling there are two horizons, causing a conflict. However, as long as the observer's viewpoint is approximately in front of the middle of the display and at a constant distance from it, augmented reality displays can also be made with on-axis coupling (Overbeeke and Stratmann, 1988).

A less obvious advantage of off-axis coupling over on-axis coupling is that off-axis coupling seems free of unnoticed distortions, if the coupling is calibrated correctly. With on-axis coupling, such distortions affect performance if precise depth estimations or slants in depth have to be estimated (see Chapter 7).

Concluding, both on- and off-axis coupling have advantages and disadvantages. With real cameras, on-axis coupling is cheaper to implement than off-axis coupling. An important effect is that off-axis coupling and accurate calibration are necessary if no distortions can be accepted, for example if values such as apparent depth or slant have to match the apparent depth of a real scene. With on-axis coupling the apparent depth of a scene is compressed in depth as compared to the real scene, when the observer takes extremely oblique viewpoints. The distortions that occur with on-axis coupling are usually not noticed by observers, even with large inaccuracies in the measurement of the viewpoint. This may introduce distortions, but these distortions are usually not noticed. For example, the viewing range may be enlarged to enable the observer to reach more extreme views. Such manipulations may improve observer performance. On the other hand, off-axis coupling has to be calibrated accurately to avoid distortions that will be noticed by observers. The appropriate choice for the coupling method will depend on the task of the observer.

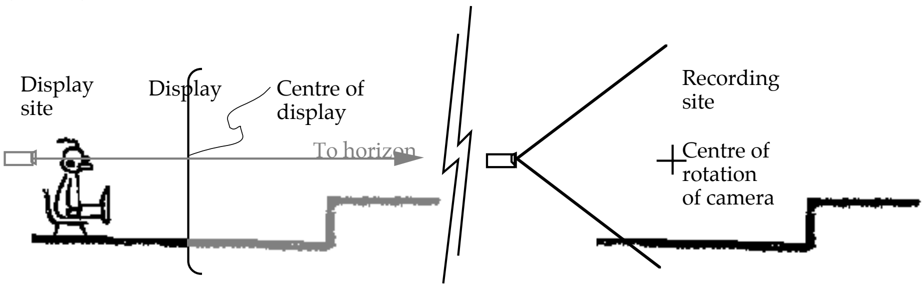


Figure 9.4a. With on-axis coupling, the camera is not rotated relative to the scene as long as the observer's viewpoint is in front of the centre of the display .

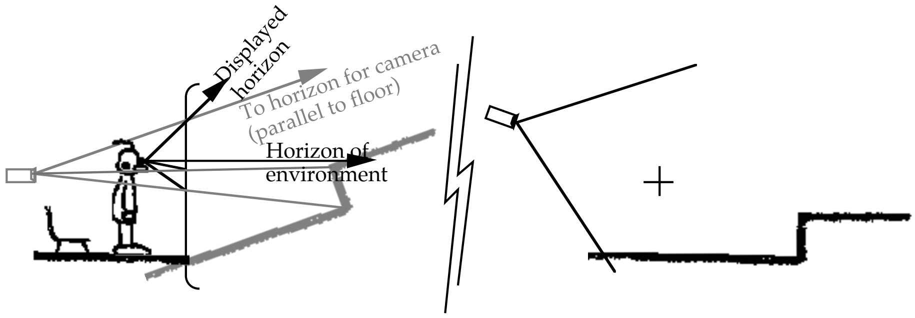


Figure 9.4b. However, as the observer moves away from this viewpoint the camera rotates according to his new position. With on-axis coupling this causes the displayed horizon to deviate from the horizon of the environment. According to the direct theory this may have large consequences, e.g. for stair climbing (compare Figure 2.4).

Delay effect

Both in the system as described in Chapter 7, in several pilot setups and in a setup with three displays instead of one (Djajadiningrat, Smets and Overbeeke, 1997) distortions caused by delays were noticed. As discussed under ‘Coupling method’, distortions are noticed only with off-axis coupling, and can be explained geometrically, given a delayed view at some viewpoint. I did not test experimentally the implications of such delays, because delays during viewing can be minimized, and therefore their relevance for x-ray baggage inspection is small. Given the available time for the inspector to make his judgement and the concepts of Chapter 3, the x-ray views of the baggage will have to be stored and displayed when required given the viewpoint of the inspector. In such a configuration, the delay is very small. Nevertheless, distortions due to delays were noticed in several setups, and are relevant for virtual window displays. Wloka (1995) discusses the sources and possible solutions for delays in detail.

The effect of a delay D is clearly noticeable in virtual window displays, and is different for on- and off-axis coupling. With off-axis coupling, the tops of all bumps seem to run ahead of the ground, following the observer with their distortion. As soon as the observer stops moving, they swing back to their perpendicular orientation. For example, suppose an observer is looking at the tops of two bumps that are projected with off-axis coupling. As long as he does not move (Figure 9.5a) the bumps appear perpendicular to the floor. When he moves to the right, the picture that shows the bumps straight up appears when the observer's eye has already passed the position perpendicular to the bump (Figure 9.5b). The observer interprets this image as if the bump is pointing towards him, and thus the bump must be sheared relative to the ground. When he moves in the other direction (Figure 9.5c) the perceived shear of the bumps reverses. I have never encountered this delay effect in the literature, but nor I have searched the extensive literature on delay effects systematically.

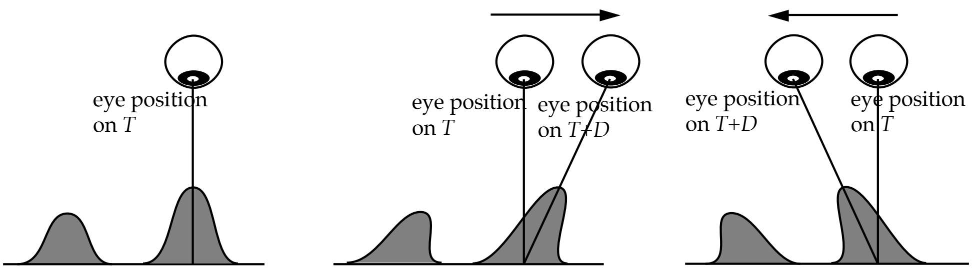


Figure 9.5a. Suppose the observer is looking at the top of a bump at moment *T*. If the observer does not move, the delayed refresh of the display *D* has no consequences.

Figure 9.5b. If the observer moves the bumps appear sheared, as off-axis coupling amounts to shearing the scene and the shear is delayed relative to his eye movement.

Figure 9.5c. The shear of the bumps reverses as the observer changes the direction of his motion.

The distortion caused by a delay gets worse if the observer moves faster and if the objects in the displayed scene get further away from the display plane, because the velocity of the parallax shifts increases with those factors. The delay effect had only small effects on the bump matching task, because the bumps were low and because there was no need for the observers to move fast.

With on-axis coupling, the overall movement of landscape is delayed, but the landscape stays a rigid whole. However, as described under ‘Coupling method’, unnoticed distortions may exist.

A number of factors that affect the performance of an observer working via a virtual window display were discussed. The coupling method and delays between the movement of the observer and the corresponding update of the display have important consequences. With off-axis coupling, distortions can be noticed, but as shown by task performance the distortions are small provided that the coupling is accurate. With on-axis coupling the distortions are not noticed, but have considerable effect on task performance at oblique viewpoints.

Industrial Design Engineering

Industrial design engineers have learned to solve problems concerning product development in a systematic way. They consider a number of solutions and select the most appropriate solution, considering technical, ergonomic, aesthetic, environmental and economical aspects (Smets, 1992).

For industrial design engineers the results of this thesis are important for two reasons. First, in industrial products computer-controlled interfaces providing a spatial impression are growing more important (Bouwmeester, 1996; Louwerse, 1996; van Bueren, 1997). Spatial displays are currently used in medical, military, training and data analysis applications, and are of growing importance in other applications such as entertainment, safety, public notice boards and advertising. Second, CAD systems with spatial displays are becoming a usual tool in product development. The design process itself is usually aided by displayed impressions of the planned product. Thus, both designers themselves and their customers use spatial displays. The results are important because they concern the intuitiveness of the user interface and the distortions in spatial displays.

For choosing the appropriate solution for designing a product and for presenting a spatial impression to an observer, the design engineer should consider the perceptual requirements of the task in hand, the distortions in the perceived scene that are caused by a geometrically inequivalent viewpoint and by the coupling method, and the intuitiveness of the system.

Distortions due to coupling method

One consideration when choosing a virtual window display is the coupling method to be used. As discussed under 'Coupling method', off-axis coupling is required when the displayed world has to be linked to or mixed with objects in the real world. For example if one wants to use the display itself as a piece of paper and draw on it with a pen, on-axis coupling is not useful as the displayed 'piece of paper' rotates into and out of the screen as the observer moves (Figure 9.6a). Off-axis coupling solves this problem (Figure 9.6b). However, using off-axis coupling will result in distortions for other people not coupled to the display. If other observers also have to look at the screen, for example for training or for attending a presentation, on-axis coupling seems to be the appropriate choice.

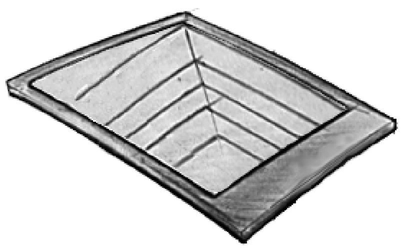


Figure 9.6a. Suppose that the observer wants to write on the displayed top plane. If on-axis coupling is used, that plane will be slanted relative to the plane of the real display when the observer is not directly in front of the middle of the display.

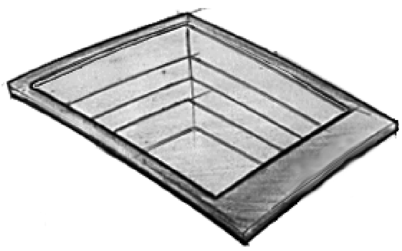


Figure 9.6b. With off-axis coupling the displayed planes stay parallel to the plane of the real display. In such cases where the real and displayed world are closely linked, off-axis coupling is more suitable than on-axis coupling.

Intuitiveness of the user interface

Virtual window displays can be used to provide a more intuitive design environment. Especially with a large number of degrees of freedom, viewpoint selection via a knob may be unintuitive.

For selecting a view given only one degree of freedom, as was the case with the wire detection task (Chapter 5), a slider or turning knob is sufficient and intuitive. Apparently for doing the wire detection task successfully, and probably for most tasks via a spatial display, the observer requires control over the displayed view, but he does not need to relate the parallax in the display to his proprioceptive cues. Selecting a viewpoint by eye position may even be inappropriate from an ergonomic point of view. Rotating a knob can be done faster and more easily than moving the head.

Current CAD applications offer many degrees of freedom (3 rotational axes, 3 translational axes, viewing angle of the camera, wire frame versus solid rendering, etc.) for manipulating the view, but this is usually done by providing a slider for each axis. This is not intuitive, especially if the current view is not the front view (Figure 9.7a and 9.7b).

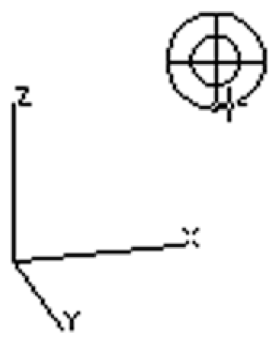


Figure 9.7a. Selecting a viewpoint with Autocad 13 and earlier versions. Moving the cross within the circle rotates the axes. These axes are related to the scene orientation. The z-axis always stays vertical, and the viewing distance has to be adjusted with another command.

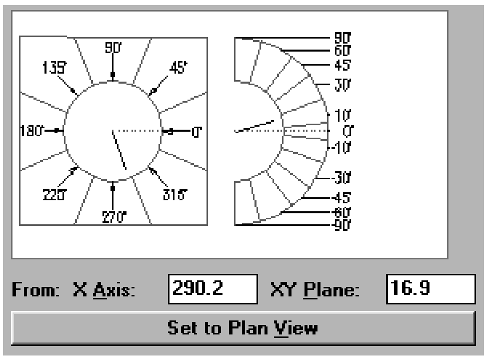


Figure 9.7b. Part of the new viewpoint selection window of Autocad 13. The element sizes in the azimuth setting (right) seem irregular. The effect of the settings on the view are hard to imagine, especially if the current view is not the front view.

There are more intuitive ways of changing the view. The interface of the 'Scene Viewer' utility from Silicon Graphics is a good example. If the cursor (the hand icon) is near the middle of the window (Figure 9.8a), dragging the hand horizontally will rotate the scene around the vertical axis. If the hand is near the bottom of the window (Figure 9.8b), dragging the hand horizontally will rotate the scene around the axis out of the display. However, the use by the Scene Viewer program of different mouse buttons to select zooming and moving forward and backward is less intuitive.

Thus, when a large number of degrees of freedom are available for selecting the viewpoint, it is hard to keep the interface intuitive (see also Djajadiningrat, Overbeeke and Smets, 1997). Coupling the viewpoint to the eye position of the observer is more intuitive.

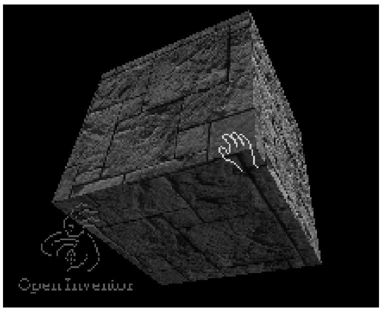


Figure 9.8a. A more intuitive way of selecting a view. Dragging the hand horizontally while it is in the middle of the window rotates the scene about the vertical axis.

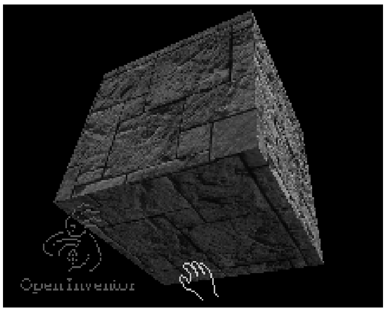


Figure 9.8b. If the hand is near the bottom of the window, dragging it horizontally results in a rotation of the scene around the axis out of the display.

Distortions due to geometric inequivalence

The distortions in the scene caused by the display system are an important consideration when choosing or using a virtual window display. Most computer aided design (*CAD*) systems do not take the viewpoint of the observer into account. Therefore the displayed scene will be distorted depending on such things as the coupling method, viewing position and camera position. This holds for both single perspective renderings and virtual window displays. For example, Figure 9.9a shows a close-up photograph of a lunchbox. It looks nearly square and higher than it would from a larger viewing distance (Figure 9.9b). Such views are generated easily with CAD programs, and in CAD programs there is usually no indication of viewing distance, which is closely related to such distortions. Such distortions are often used on purpose in advertising, and give the observer a misleading impression of spaciousness. For virtual window displays, close-up views may be appropriate when the observer is close to the display. For displays without coupling of the display to eye position, such as photographs and normal television images, a telelens may prevent such perceived distortions (Cutting, 1987). The degree to which distortions are obtrusive in static on-axis large-angle images may be minimized by image processing (Zorin, 1995; Buchroeder, 1995).

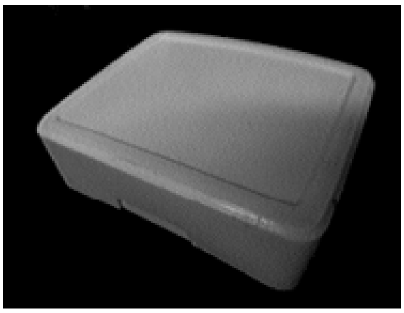


Figure 9.9a. Close-up photograph of a lunchbox. The top looks almost square, and the box looks almost half as thick as its width.

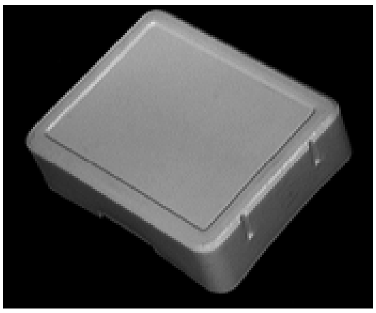


Figure 9.9b. Lunchbox photographed with a telelens. For most people, this gives a less distorted impression of the width-height-thickness ratios.

For several tasks such distortions may be unimportant, but distortion effects can be expected to be important when aesthetic judgements come into play, for example in advertising, entertainment, public notice boards and designs made on a virtual window.

As described in the previous section under 'Coupling method', both on- and off-axis coupling can cause distortions, although they are usually noted by the observer only with off-axis coupling. This makes for a difficult choice for the designer: should he convince his customers by providing a subjectively convincing image without noticeable distortions, or should he choose a configuration that is without unnoticed distortions? One aspect providing the answer is the task in hand.

Task in hand

To choose an appropriate solution for presenting a spatial impression to an observer, the task in hand should be considered. For example, virtual window displays that rely exclusively on parallax shifts cannot be used for all applications, as parallax shifts are available only when the observer moves or when objects in the virtual world move relative to the observer. It is known that with precise manipulation tasks the observer tends and probably needs to minimize his movements relative to the object to be manipulated (Voorhorst, Overbeeke and Smets, 1997). In this case stereoscopic cues may be added to the virtual window display to give the observer depth cues if the task requires this.

If a display has to provide images giving a spatial impression to the user, the design engineer can choose from a range of displays, most notably virtual window displays and head mounted displays. With head mounted displays, the observer has 2 small displays in front of his eyes. These systems are growing more important as their price is decreasing rapidly. Such systems can simulate a complete world instead of only a window with a simulated scene. Therefore such immersive VR systems are useful for designing large objects, for example in architecture. Another advantage is that in such completely simulated worlds it is possible to alter the laws of physics. For example, the body of a patient may appear transparent.

However, most applications concern a task in the real world. For such tasks, a spatial display has to provide the observer with additional data, alongside the directly visible cues from his task. In such a situation the observer has to see the real world as well, and immersive VR systems cannot be used. Mixed VR systems, where an image is merged through the real world with a half-transparent mirror, try to combine the advantages of virtual window systems with those of immersive VR. However, immersive VR still requires the observer to wear a helmet with a half-transparent mirror, and this may hinder his activities.

Many other task aspects can lead to the choice for a specific display. A number of ways of analysing the task in hand are given by Kirwan and Ainsworth (1992). To start with natural inspection, as was done in Chapter 2 for the tasks in this thesis, is possible to find the useful perceptual cues for a task, but this only holds as long as the required cues for the task can also be acquired with natural inspection. Concluding, the task in hand is of critical importance for the choice of the appropriate display system, but currently I am unable to be more precise beyond indicating some factors influencing the choice.

Concluding, the results of this thesis are relevant for an industrial design engineer, both for the designing process itself and for the users of the products designed. Important considerations for choosing an appropriate way of providing a spatial impression of a product or scene are the matching of the display to the task in hand, the disturbing effects of distortions, and the intuitiveness of the method of selecting a viewpoint.

Perceptual theories

This thesis has described some experiments in sparsely investigated areas. The compensatory effect of viewpoint multiplicity on image quality was tested. Furthermore the experiments dealt with perception and performance on technical tasks with a transparent scene. The results may give new grounds for judging and correcting theories concerning human information extraction from transparent scenes.

For an analysis of the spatial cues required for a task (see Chapter 2) the direct theory was found to be more useful than the indirect theory. The direct theory indicates that the question about the required information is urgent as it will drive the explorative behaviour of the observer, while the indirect theory places more interest on the extraction of 3D structure from the light from the environment.

The need for a complete reconstruction, as suggested by the indirect theory, is questionable. A complete reconstruction seems impossible given only some pixels in a view representing a wire. With normal baggage inspection, where only a single x-ray view is available, a reliable reconstruction can again not be made. However, these and the other tasks discussed in this thesis could be achieved with limited depth cues. As Tittle, Todd, Perotti and Norman (1995) suggested: “Most perceptual judgements required in natural vision do not require an explicit knowledge of Euclidean metric structure and can be performed accurately on the basis of ordinal or topological relations”. Why should one check all cues if one of them is sufficient? Building a complete reconstruction of the scene is a waste of energy, and I suspect that humans only do things if they have a good reason (though not necessarily a logical reason) to do so.

On the other hand, the direct theory seems to oversimplify the extraction of information from the scene. Biological evidence indicates that neural cells in the eye do indeed extract zero-crossings from the light falling into the eye, as suggested by the indirect approach (see Figure 2.6). The direct theory does not explain why finding zero-crossings is essential for finding task-specific information. The suggestions of the direct approach, for the bump matching experiment described in Chapter 2, are an interesting example of such an oversimplification. The direct theory suggested that the width/height ratios of the bumps can be compared, but one needs the contours of the bumps to find the width and the height. In order to see these contours one needs motion parallax or texture cues.

Neither the direct nor the indirect theory explain *how* the task in hand steers exploratory behaviour. Knowledge about this relation is essential for building efficient interfaces and for understanding how spatial cues can be substituted for other spatial cues, in order to provide the observer with the information required for his task.

A

Stereo test

In some of the experiments, the subjects were tested for their ability to use stereoscopic information. To test this, they were asked to describe the figures hidden in a random-dot stereogram. Figures A.1 and A.2 give an impression of the two random-dot stereograms that were used. To facilitate viewing, the stereogram was presented via a commercially available stereo viewer bought in a toy store. The stereograms were printed in 300 dpi on a transparent sheet that was the viewed in the stereo viewer.

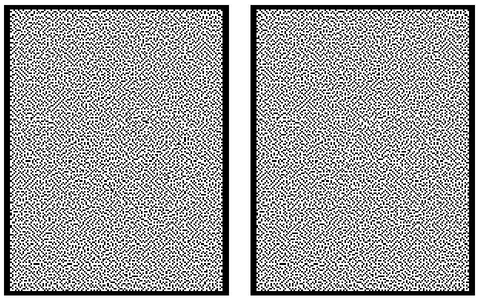


Figure A.1. Random-dot stereogram: pacman. This stereoscopic image can be viewed by looking at the left image with the left eye, and at the right with the right eye.

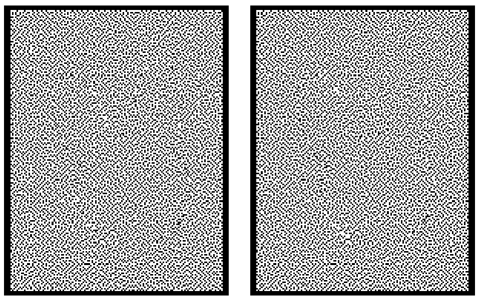


Figure A.2. Random-dot stereogram: rabbit. See Figure A.1.

Bibliography

3D builder [Computer software] (1996). 3D construction company, Elizabethton, Tennessee, USA.
Abraben, E. M. (1994). *Point of view: the art of architectural photography.* New York: Van Nostrand Reinhold.
ACI (1995). Heathrow's £90m cost of comprehensive hold baggage screening. *ACI Europe airport business, June/July*, 25.
Adams, K. R. (1972). Perspective and the viewpoint. *Leonardo 5*, 209-217.
Arthur, K. W., Booth, K. S., & Ware, C. (1993). Evaluating 3D task performance for fish tank virtual worlds. *ACM Transactions on Information Systems, 11*, 239-265.
Attree, J. (1996). Explosive detection - update on reality. *Southeast Asia Fire & Security, March*, 17-20.
Barfield, W., & Rosenberg, C. (1995). Judgments of azimuth and elevation as a function of monoscopic and binocular depth cues using a perspective display. *Human Factors, 37* (1), 173.
Bass, J. C., & Chiles, C. (1990). Correlation with detection of solitary pulmonary nodules. *Investigative Radiology, 25* (September), 994-998.
Beall, A. C., Loomis, J. M., Philbeck, J. W., & Fikes, T. G. (1995). Absolute motion parallax weakly determines visual scale in real and virtual environments. *Proceedings of the SPIE, 2411*, 288-297. Bellingham, WA: SPIE.
Beurs, M. de (1994). *Ruimtelijke afbeeldingen en kleurgebruik op beeldschermen.* (Report No. S-697, 19 december). Delft, The Netherlands: Faculty of Mechanical Engineering and Marine Technology, Delft University Of Technology.
Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.
Biederman, I., & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff (1995). *Journal of Experimental Psychology: Human Perception and Performance, 21* (6), 1506-1514.
Bingham, G. P., & Stassen, M. G. (1994). Monocular egocentric distance information generated by head movement. *Ecological Psychology, 6*, 219-238.
Bishop, P. O. (1996). Stereoscopic depth perception and vertical disparity: neural mechanisms. *Vision Research, 36* (13), 1969-1972.
Börjesson, E., & Lind, M. (1996). The effect of polar projection on the perception of euclidean structure from motion. *Perception & Psychophysics, 58* (6), 871-882.
Bouwmeester, R. (1996). *Ergonomic design of an overview display in a nuclear power plant control room: Integrated process status overview.* (Report No. 001507). Unpublished master's thesis, Faculty of Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands.
Bradshaw, M. F., & Rogers, B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Research, 36* (21), 3457-3468.
Braunstein, M. L., & Payne, J. W. (1969). Perspective and form ration as determinants of relative slant judgments. *Journal of Experimental Psychology: Human Perception and Performance, 81* (3), 584-590.
Braunstein, M. L., Hoffman, D. D., Shapiro, L. R., Andersen, G. J., & Bennett, B. M. (1987). Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception and Performance, 13* (3), 335-343.

Buchroeder, R. A. (1995). Distortion correction for oblique projection onto a curved screen. *Proceedings of the SPIE, 2537*, 96-104. Bellingham, WA: SPIE.
Bueren, L. L. van (1997). *Alternate scan schemes for intraluminal ultrasound imaging.* (Report No. 001670). Unpublished master's thesis. Faculty of Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands.
Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences USA, 89* (January), 60-64. Available Internet: www.mpik-tueb.mpg.de/people/personal/people.html.
Busquets, A. M., Parrish, R. V. & Williams, S. P. (1991). Effects of alternate pictorial pathway displays and stereo 3-D presentation on simulated transport landing approach performance. *Proceedings of the SPIE, 1457*, 91-102. Bellingham, WA: SPIE.
Castle, O. M. (1995). *Synthetic image generation for a multiple-view autostereo display* (Tech. Rep. No. 382, 59-77). Unpublished doctoral thesis, Computer lab, University of Cambridge, Cambridge, England. Available Internet: http://www.cl.cam.ac.uk/Research/Rainbow/publications/index.html.
Caudek, C., & Proffitt, D. R. (1993). Depth perception in motion parallax and stereokinesis. *Journal of Experimental Psychology: Human Perception and Performance, 19* (1), 32-47.
Chiruvolu, R., Hwang, V., & Sheridan, T. (1991). Visual display aids for teleoperation. *Proceedings of the SPIE, 1612*, 299-310. Bellingham, WA: SPIE.
Claessen, J. P. (1995). *Shaped by colour: A study on the effect of colour on the perceived shape of objects*. Unpublished doctoral dissertation, Delft University of Technology, Faculty of Industrial Design Engineering, The Netherlands.
Cole, R. E., & Parker, D. L. (1989). Stereo TV improves manipulator performance. *Proceedings of the SPIE, 1083*, 18-27. Bellingham, WA: SPIE.
Cole, R. E., Merritt, J. O., Fore, S., & Lester, P. (1990). Remote manipulator task impossible without stereo TV. *Proceedings of the SPIE, 1256*, 255-265. Bellingham, WA: SPIE.
Cole, R. E., Pepper, R. L., & Pinz, B. L. (1981). *The influence of head movement parallax on perceptual performance under direct- and TV displayed conditions (*Technical Report No 678, May). San Diego, CA: Naval Ocean Systems Center.
Cooper, L. A. (1989). Mental models of the structure of visual objects. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure & process.* (pp. 91-119). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
Cornilleau-Pérès, V., & Droulez, J. (1993). Stereo-motion cooperation and the use of motion disparity in the visual perception of 3-D structure. *Perception & Psychophysics, 54* (2), 223-239.
Cumming, B. G., Johnston, E. B., & Parker, A. J. (1991). Vertical disparities and the perception of three-dimensional shape. *Nature, 349*, 411-413.
Cumming, B. G., Johnston, E. B., & Parker, A. J. (1993). Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Research, 33* (5/6), 827-838.
Cutting, J. E. (1986). *Perception with an eye for motion*. Cambridge, MA: MIT Press.
Cutting, J. E. (1987). Rigidity in cinema seen from the front row, side aisle. *Journal of Experimental Psychology: Human Perception and Performance, 13* (3), 323-334.
Cutting, J. E. (1989). On the efficacy of cinema, or what the visual system did not evolve to do. In *Spatial displays and spatial instruments* (Report number N90-22918/8, PC A99/MF A04). NASA, Ames Research Center, July, 17-1 - 17-8.

Deering, M. (1992). High resolution virtual reality. *Computer Graphics, 26* (2), July, 195-202.
Diner, D. B., & Venema, S. C. (1989). Graphic overlays in high-precision teleoperation: Current and future work at jpl. *Proceedings of the NASA Conference on Space Telerobotics, JPL 89-7 (3)*, 511-520.
Djajadiningrat, J. P., Overbeeke, C. J., & Smets, G. J. F. (in press). The importance of simultaneously accessible degrees of freedom for rotation of objects. To appear in *Behaviour and Information Technology.*
Djajadiningrat, J. P., Smets, G. J. F., & Overbeeke, C. J. (1997). Cubby: a multiscreen movement parallax display for direct manual manipulation. *Displays, 17*, 191-197.
Doorn, A. J. van, & Koenderink. J. J. (1996). How to probe different aspects of surface relief. *Proceedings of the fourth european workshop on ecological psychology* (Zeist, the Netherlands, July 2-5), 115-130. The Netherlands: Delft University Press.
Dorey, F. C. (1983). *Aviation security.* London: Granada Publishing Limited.
Dotzler, D. (1996). Invision enters strategic alliance with EG&G astrophysics, to develop new explosive detection system. *Insight (newsletter of InVision Technologies), 1* (3), December, 1.
Drascic, D. (1991). Skill acquisition and task performance in teleoperation using monoscopic and stereoscopic video remote viewing. *Proceedings of the Human Factors Society 35th Annual Meeting,* 1367-1371. Santa Monica, CA: Human Factors Society.
Drascic, D., & Milgram, P. (1996). Perceptual issues in augmented reality. *Proceedings of the SPIE, 2653*, 123-134. Bellingham, WA: SPIE.
Durgin, F. H., Proffitt, D. R., Olson, T. J., and Reinke, K. S. (1995). Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance, 21* (3), 679-699.
Eby, D. W., & Braunstein, M. L. (1995). The perceptual flattening of three-dimensional scenes enclosed by a frame. *Perception, 24* (9), 981-993.
Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32* (12), 2385-2400.
Edgar, G. K., & Bex, P. J. (1995). Vision and displays. In K. Carr, & R. England, *Simulated and virtual realities: Elements of perception.* London: Taylor and Francis.
EG&G Astrophysics (1996). *Z-scan: Automatic explosives and contraband detector.* Long Beach, CA: EG&G Astrophysics.
Europscan (1993). *Radioscopic control of luggage and small freight controlix vision.* Rungis Cedex, France: Europ Scan S.A.
Evans, P., Godber, S. X., & Robinson, M. (1994). Three-dimensional X-ray imaging techniques. *Proceedings of the SPIE, 2177*, 161-165. Bellingham, WA: SPIE.
Field, S., Michell, M. J., Wallis, M. G. W., & Wilson, A. R. M. (1995). What should be done about interval breast cancers?: Two view mammography and possibly a shorter screening interval. *British Medical Journal, 310* (28 January), 203-204.
Gaver, W., Sellen, A., Heath, C., & Luff, P. (1993). One is not enough: multiple views in a media space. *Conference proceedings of the Interchi '93*, 335-341. Amsterdam, The Netherlands: IOS.
Gibson, J. J. (1971). The information available in pictures. *Leonardo*, 4, 27-35.
Gibson, J. J. (1986). *The ecological approach to visual perception.* Hillsdale, New Jersey: Lawrence Erlbaum Associates. (Original work published 1979).

Gille, J., Samadani, R., Martin, R., & Larimer, J. (1994). Grayscale/ resolution tradeoff. *Proceedings of the SPIE, 2179*, 47-59. Bellingham, WA: SPIE.
Gogel, W. C., & Tietz, J. D. (1979). A comparison of oculomotor and motion parallax cues of egocentric distance. *Vision Research*, 19, 1161-1170.
Gonzalez, R. (1995). *Video coding for hypermedia based on a computer graphics model.* Telecommunications Software Research Centre, University of Wollongong, Australia. Available Internet: http://easgonzales.eas.gu.edu.au/phd/phd.html.
Gooding, L., Miller, M. E., Moore, J., & Kim, S. (1991). The effect of viewing and disparity on perceived depth. *Proceedings of the SPIE*, *1457*, 259-266. Bellingham, WA: SPIE.
Gourneri, J. de la (1859). Traité de perspective linéaire contenant les tracés pour les tableaux plans et courbes, les bas-reliefs et les décorations théatrales, avec une théorie des effets de perspective. Paris: Dalmont et Dunod, Mallet-Bachelier.
Haber, R. N., & Herhenson, M. (1974). *The psychology of visual perception.* UK: Redwood Burn limited.
Hagen, M. A. (1974). Picture perception: toward a theoretical model. *Psychological Bulletin*, *81*, 471-497.
Hagen, M. A., & Elliott, H. B. (1976). An investigation of the relationship between viewing condition and preference for true and modified linear perspective in adults. *Journal of Experimental psychology: Human Perception and Performance, 2* (4), 429-490.
Halloran, T. O. (1989). Picture perception is array-specific: Viewing angle versus apparent orientation. *Perception & Psychophysics*, *45* (4), 467-482.
Harmon, L. D. (1973). The recognition of faces. *Scientific American, November*, 71-82.
Hayashibe, K. (1991). Reversals of visual depth caused by motion parallax. *Perception, 20*, 17-28.
Hays, W. L. (1981). *Statistics* (3rd ed.). New York: CBS College Publishing.
Heimann (1987). *Heimann: Hi-mat mehrenergie-verfahren* [Brochure]. Wiesbaden, Germany: Heimann systems GmbH.
Heimann (1994). *Röntgenprüfgerät Hi-scan 7555-A: Sicherheit durch Qualität* [Brochure]. Wiesbaden, Germany: Heimann Systems GmbH.
Heimann (1996). *Automatic explosives detection system hi-scan 10050 EDS* [Brochure]. Wiesbaden, Germany: Heimann Systems GmbH.
Heimann (1997). *Hi-scan 10050 EDS Automatic explosives detection system* [On-line]. Available Internet: www.heimannsystems.com/eds.htm.
Henderson, B. W. (1990). Airport x-ray ct scanner finds small amounts of explosives in baggage. *Aviation Week & Space Technology, 8* (October), 34.
Hochberg, J. (1986). Presentation of motion and space in video and cinematic displays. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Chapter 22). New York: John Wiley & Sons.
Holloway, R., & Lastra, A. (1993). *Virtual environments: A survey of the technology.* (Report TR93-033). Department of computer science, University of North Carolina, Chapel Hill, NC.
Hughes, D. (1989). X-ray backscatter equipment provides automatic screening for explosives. *Aviation Week & Space Technology, 17* (April), 65-67.
Imatron (1991). Explosives detection scanner system. *Aerospace World, 1*, 45.
InVision (1997). *Explosive detection solutions* [Brochure]. Foster City, CA: InVision Technologies.

Invision & EG&G (1997). *CTZ-Scan: EG&G Astrophysics/Invision alliance: product plan, overview and discussions* [Brochure]. Foster City, CA: InVision Technologies.
Isorad (1987). *SDS 400 series fluoroscopic x-ray inspection systems* [Brochure]. Yavne, Israel: Isorad.
Jones, R. K., & Hagen, M. A. (1978). The perceptual constraints on choosing a pictorial station point. *Leonardo, 11*, 191-196.
Kersten D., & Bülthoff, H. H. (1991). *Apparent opacity affects perception of structure from motion* (A.I. Memo 1285; C.B.I.P. memo 34). Cambridge, MA: Massachusetts Institute of Technology, Artificial Intelligence Laboratory. Available Internet: www.mpik-tueb.mpg.de/ people/personal/people.html.
Kersten, D., Bülthoff, H. H., Schwartz, B. L., Kurtz, K. J. (1992). Interaction between transparency and structure from motion. *Neural Computation, 4*, 573-589.
Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont, California: Wadsworth.
Kirwan, B., & Ainsworth, L. (Eds.). (1992). *A guide to task analysis*. London: Taylor and Francis.
Kjelldahl, L., & Prime, M. (1995). Study on how depth perception is affected by different presentation methods of 3D objects on a 2D display. *Computers & Graphics, 19* (2), 199-202.
Koenderink, J. J., Doorn, A. J. van, & Kappers, A. M. L. (1994). On so-called paradoxical monocular stereoscopy. *Perception, 23*, 583-594.
Kohler, I. (1962). Experiments with goggles. *Scientific American, 206* (5), 63-72.
Kotowski, A. (1986). A review of modern x-ray screening devices. *ICAO Bulletin, October*, 22-23.
Kundel, H. L., & Follette, P. S. la (1972). Visual search patterns and experience with radiological images. *Radiology, 103* (3), 523-528.
Linkenbach, H., & Stein, K. U. (1981). Luggage control with x-ray eyes. *Siemens Review XLVIII* (6), 12-15.
Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature, 293*, 133-135.
Loosen, F. (1994). *Statistische tabellen* (6th ed.). Leuven, Belgium: Acco.
Louwerse, W. (1996). *Virtual Attraction.* (Report No. 001573). Unpublished master's thesis (in Dutch), Faculty of Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands.
Lumsden, E. A. (1980). Problems of magnification and minification: An explanation of the distortions of distance, slant, shape, and velocity. In M. Hagen (Ed.), *The perception of pictures: Vol. 1. Alberti's window: The projective model of pictorial information* (pp. 91-135). New York: Academic Press.
Macrae, D., & Taverna, M. (1990). Special report: Explosives detection raises complex issues. *Aerospace World, 11*, 26-32.
Magal (1994). *The Aisys 370B: An automatic bomb detection system* [Brochure]. Yahud, Israel: Magal Security Systems Ltd.
Maharay, H. P. (1989). Stray radiation from baggage x-ray equipment: results and implications. *Health Physics, 57* (1), July, 141-148.
Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W. H. Freeman and Company.
Martin Marietta Aerospace (1988). *Teleoperator human factors study: Final report* (Report No. NASA-CR-178930). Denver, Colorado: Martin Marietta Aerospace.

McVey, G. F. (1970). Television: Some viewer-display considerations. *AV Communications Review, 18* (3), fall, 277-290.
Meister, R. (1966). The iso-deformation of images and the criterion for delimitation of the usable areas in cine-auditoriums. *Journal of the Society of Motion Picture and Television Engineers, 75* (March), 179-182.
Metzger, W. (1975). *Gesetze des Sehens.* (3rd ed). Frankfurt am Main, Germany: Verlag Waldemar Kramer.
Nodine, C. F., & Kundel, H. L. (1987). Using eye movements to study visual search and to improve tumor detection. *RadioGraphics, 7* (6), november, 1241-1250.
Nodine, C. F., Kundel, H. L., Lauver, S. C., & Toto, L. C. (1996). Nature of expertise in searching mammograms for breast masses. *Academic Radiology, 3*, 1000-1006.
Norman, J. F., & Todd, J. T. (1995). The perception of 3-D structure from contradictory optical patterns. *Perception and Psychophysics, 57* (6), 826-834.
Norusis, M. J. (1993). *SPSS for windows base system user's guide release 6.0.* Chicago, IL: SPSS Inc.
Olzak, L. A., & Thomas, J. P. (1975). Seeing spatial patterns. In E. C. Carterette, & P. Friedman (Eds.), *Handbook of perception* (Chapter 7). New York: Academic Press.
Ono, H., & Steinbach, M. J. (1990). Monocular stereopsis with and without head movement. *Perception & Psychophysics, 48* (2), 179-187.
Origin Instruments (1993). *The DynaSight sensor: Developer manual.* Grand Prairie, Texas: Origin Instruments.
Ouden, R. den (1995). *Betrouwbaarheid en snelheid van bagagecontrole apparatuur op Schiphol [Reliability and speed of luggage inspection apparatus at Schiphol].* Unpublished master's thesis, Delft University of Technology, Faculty of Industrial Design Engineering, Delft, The Netherlands.
Overbeeke, C. J., & Stratmann, M. H. (1988). *Space through movement.* Unpublished doctoral dissertation, Faculty of Industrial Design Engineering, Delft University of Technology, The Netherlands.
Overbeeke, C. J., Smets, G. J. F., & Stratmann, M. H. (1987). Depth on a flat screen II. *Perceptual and Motor Skills*, 65, 120.
Overveld, W. M. C. J. (1994). Design considerations for X-ray: trade-offs in noise, contrast and blur affect image quality and performance. *IPO Annual Progress Report, 29*, 72-80. Eindhoven, The Netherlands: Eindhoven University of Technology.
Padmos, P., & Milders, M. V. (1992). Quality criteria for simulator images: A literature review. *Human Factors, 34* (6), 727-748.
Pasman, W., Smets, G. J. F., & Stappers, P. J. (1997). Effects of Image Quality, Number of Selectable Viewpoints, and Way to Select the Viewpoint in X-Ray Luggage Inspection. *Presence*, 6 (3), 268-281.
Pasman, W., Stappers, P. J., & Smets, G. J. F. (1997). Distortion in on- and off-axis movement-coupled perspective displays. *Proceedings of the 9th international conference on perception and action* (july 20-27, University of Toronto, Scarborough, Canada), 77. New Jersey: Lawrence Erlbaum Associates.
Perkins, D. N. (1973). Compensating for distortion in viewing pictures obliquely. *Perception & Psychophysics, 14* (1), 13-18.
Perrett, D. I., Harries, M. H., & Loker, S. (1992). Use of preferential inspection to define the viewing sphere and characteristic views of an arbitrary machined tool part. *Perception, 21*, 497-515.
Petzold, P. (1973). Effects and experiments in photography. New York: Focal Press.

Pirenne, M. H. (1970). *Optics, painting, and photography*. London: Cambridge University Press.
Pirenne, M. H. (1975). Vision and art. In E. C. Carterette, & P. Friedman (Eds.), *Handbook of perception* (Chapter 12). New York: Academic Press.
Poot, H. J. G. de (1995). *Monocular perception of motion in depth.* Unpublished doctoral dissertation, Faculty of Biology, University of Utrecht, Utrecht, The Netherlands.
Purdy, W. C. (1960). *The hypothesis of psychophysical correspondence in space perception* (NTIS R60ELC56). New York: General Electric Technical Information Series.
Ramsey, C. G., & Sleeper, H. R. (1988). *Architectural graphic standards* (8th ed.). New York: Wiley.
Ranadivé, V. (1979). *Video resolution, framerate, and gray scale tradeoffs under limited bandwidth for undersea teleoperation.* Unpublished master's thesis, Faculty of Electr. Engineering and Computer Science, Boston, Massachusetts Institute of Technology, MA.
Rogers, B. J. , & Graham, M. E. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8*, 125-134.
Rogers, B. J. , & Graham, M. E. (1985). Motion parallax and the perception of three-dimensional surfaces. In D. J. Ingle, M. Jeannerod, & D. N. Lee (Eds.), *Brain mechanisms and spatial vision* (NATO ASI Series), (pp. 95 - 111). Dordrecht, The Netherlands: Martinus Nijhoff.
Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature, 361*, 253-255.
Rogers, B., & Graham, M. (1983). Anisotropies in the perception of three-dimensional surfaces. *Science, 221*, 1409-1411
Rosinski, R. R., & Farber, J. (1980). Compensation for viewing point in the perception of pictured space. In M. Hagen (Ed.), *The perception of pictures: Vol. 1. Alberti's window: The projective model of pictorial information* (pp. 137-176). New York: Academic Press.
Schlosberg, H. (1941). Stereoscopic depth from single pictures. *American Journal of Psychology, 54*, 601-605.
Schweers, J., & Vianen, P. van (1982). *Natuurkunde op corpusculaire grondslag.* Den Bosch, The Netherlands: Malmberg.
Sedgwick, H. A. (1980). The geometry of spatial layout in pictorial representation. In M. Hagen (Ed.), *The perception of pictures: Vol. 1. Alberti's window: The projective model of pictorial information* (pp. 33-90). New York: Academic Press.
Sedgwick, H. A. (1986). Space perception. In Boff, K. R., Kaufman, L., & Thomas, J. P. (Eds.), *Handbook of perception and human performance* (Chapter 21). New York: John Wiley & Sons.
Simon, M. H. (1997). Plastic power: Lightweight batteries show their muscle in demonstrations. *Scientific American, April*, 31.
Singh, G., Serra, L., Fairchild, K., & Poston, T. (1994). Visual creation of virtual design environments and virtual worlds research at ISS. *Presence, 3* (1), 94-107.
Smets, G. (1992). Integration of technology assessment and humanities at the department of industrial design engineering Delft university of technology. *Universiteit & Hogeschool, 38* (6), 263-270.
Smets, G. J. F. (1995). Designing for Telepresence: The Delft Virtual Window System. In P. Hancock, J. Flach, J. Caird & K. Vicente (Eds.), *Local applications of the ecological approach to human-machine systems* (pp. 182-207). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Smets, G. J. F., & Overbeeke, K. J. (1995). Trade-off between resolution and interactivity in spatial task performance. *IEEE Computer Graphics and Applications*, 15 (5), 46-51.
Smets, G. J. F., Overbeeke, C. J., & Stratmann, M. H. (1987). Depth on a flat screen. *Perceptual and Motor Skills*, *64*, 1023-1034.
Smets, G. J. F., Stratmann, M. H., & Overbeeke, C. J (1990). Method of causing an observer to get a three-dimensional impression from a two-dimensional representation. *European Patent 0189233*.
Smets, G. J. F., Stratmann, M. H., & Overbeeke, C. J. (1988). Method of causing an observer to get a three-dimensional impression from a two-dimensional representation. *U.S. Patent 4,757,380*.
Smith, O. W., & Gruber, H. (1958). Perception of depth in photographs. *Perceptual and Motor Skills, 8*, 307-313.
Snyder, H. L. (1973). Image quality and observer performance. In L. M. Biberman (Ed.), *Perception of displayed information* (pp. 87-118). New York: Plenum Press.
Sollenberger, R. L., & Milgram, P. (1991). A comparative study of rotational and stereoscopic computer graphic depth cues. *Proceedings of the Human Factors Society 35th Annual Meeting*, 1452-1456. Santa Monica, CA: Human Factors Society.
Sollenberger, R. L., & Milgram, P. (1993). Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. *Human Factors, 35* (3), 483-499.
Spain, E. H. (1990). Stereo advantage for a peg-in-hole task using a force-feedback manipulator. *Proceedings of the SPIE, 1256*, 244-254. Bellingham, WA: SPIE.
Spain, E. H., & Holzhausen, K. P. (1991). Stereoscopic versus orthogonal view displays for performance of a remote manipulation task. *Proceedings of the SPIE, 1457*, 103-110. Bellingham, WA: SPIE.
Stappers, P. J. (1992). *Scaling the visual consequences of active head movements: A study of active perceivers and spatial technology.* Unpublished doctoral dissertation, Delft University of Technology, Delft, the Netherlands.
Swartz, M., Wallace, D., & Tkacz, S (1992). The influence of frame rate and resolution reduction on human performance. *Proceedings of the Human Factors Society 36th Annual Meeting*, 1440-1444. Santa Monica, CA: Human Factors Society.
Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21* (3), 663-678.
Todd, J. T., Akerstrom, R. A., Reichel, F. D., & Hayes, W. (1988). Apparent rotation in three-dimensional space: Effects of temporal, spatial, and structural factors. *Perception & Psychophysics, 43*, 179-188.
Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 48* (5), 419-430.
Tsacoumis, T. P. (1983). The effectiveness of passenger security screening - an overview. *Journal of Testing and Evaluation, 66* (6), November, 366-367. Philadelphia: ASTM.
Ullman, S. (1979). *The interpretation of visual motion* (3rd ed.). Cambridge, MA: MIT Press.
Uttal, W. R., Baruch, T., & Allen, L. (1995). Combining image degradations in a recognition task. *Perception & Psychophysics, 57* (5), 682-691
Vivid (1990). *Effective detection today* [Brochure]. Woburn, MA: Vivid Technologies.
Voorhorst, F. A., Overbeeke, C. J., & Smets, G. J. F. (1997). Using movement parallax for 3D laparoscopy. *Medical Progress through Technology, 21*, 211-218.

Vyborny, C. J. (1997). The AAPM/RSNA physics tutorial for residents: Image quality and the clinical radiographic examination. Imaging & *Therapeutic Technology, 17* (2), 479-498.
Wald, N. J., Murphy, P., Major, P., Parkes, C., Townsend, J., & Frost, C. (1995). UKCCCR multicentre randomised controlled trial of one and two view mammography in breast cancer screening. *British Medical Journal*, *311*, 1189-1193.
Ware, C. (1995). Dynamic stereo displays. *CHI'95 conference proceedings: Human Factors in Computing Systems*, 310-316.
Ware, C., Arthur, K., & Booth, K. S. (1993). Fish tank virtual reality. *Proceedings of the INTERCHI '93*, 37-42. Amsterdam, the Netherlands: IOS.
Wickens, C. D. (1990). Three-dimensional stereoscopic display implementation: Guidelines derived from human visual capabilities. *Proceedings of the SPIE, 1256*, 2-11. Bellingham, WA: SPIE.
Wloka, M. M. (1995). Lag in multiprocessor virtual reality. *Presence, 4* (1), winter, 50-63.
Woolley, D. (1986). Airports build for future traffic amid new security concern. *Interavia, 7* (July), 763-770.
Zorin, D. (1995). *Correction of geometric perceptual distortions in pictures.* Unpublished master's thesis, Department of Computer Science (Caltech), California Institute of Technology, Pasadena, CA. Available Internet: www.gg.caltech.edu:80/~dzorin.

Summary

This thesis investigates the possibility of improving x-ray baggage inspection by presenting the inspector a spatial impression of the baggage. The Delft Virtual Window System (DVWS) is used to give the inspector such a spatial impression of the baggage on a normal monitor display. This spatial impression of the baggage is obtained by coupling the position of the x-ray camera to the viewing position of the observer. For example if the inspector moves to the right, an image is displayed that shows the suitcase more from the right.

Each image taken exposes the baggage to an x-ray dose, and the maximum allowable x-ray dose is reached after about 25 images. Therefore we have to be careful about which viewpoints the inspector can investigate. Furthermore, given the state of the art techniques of making sharp x-ray pictures with a low x-ray dose, the resulting pictures have unusual perspective properties: they contain convergent perspective in, e.g., the horizontal direction of the picture, but parallel perspective in the vertical direction. Considering the costs, we prefer to use the existing techniques. Therefore, the perspective properties of the views have to be chosen carefully in order to get views which are acceptable when presented interactively with the DVWS. Chapter 3 outlines the possibilities of shooting such multiple x-ray views efficiently, using the current technologies of baggage inspection.

Thus, the question about the usefulness of the DVWS in the context of baggage inspection expands to a number of questions:

1 What exactly does 'useful' mean for baggage inspection, and how do we test it?
2 What are the useful images (what image quality; what viewpoint; how many viewpoints)?
3 How should the inspector control the image he views?
4 How should the images be presented, for optimum inspector performance?

The first of these questions is difficult to answer. Applying the DVWS to an x-ray inspection system is useful if it enhances the ability of the inspectors to find suspicious items as compared to x-ray inspection systems without the DVWS. As a baggage scanner based on the DVWS was expected to be used essentially for hold baggage, where the threat of bombs is most serious, the 'suspicious items' were reduced to 'bombs'. Usually, a bomb consists of a battery, a detonator, a timing mechanism, a wire connecting these parts, and explosives, and it seems reasonable to assume that detecting such parts and connections is important for baggage inspection. Still, the replacement of the vague term 'suspicious' by 'bomb' does not help us much in analysing the baggage inspection task scientifically, as most bomb parts do not have fixed shapes and as some parts are not always present in a bomb. For example, explosives and batteries can be shaped in any form, and wires may be omitted. But the importance of these problems was not clear when starting this project. I started to investigate a number of tasks that seemed relevant for x-ray baggage inspection, using perception theories as a criterion.

To find out what sort of images would be useful and how the observer should control the image he views, I started with an experimental investigation of image quality, required number of viewpoints, and the way the observer selects the viewpoint.

Experiments 1 to 3 (Chapters 4-6) deal with these questions. Next I investigated a more difficult question, i.e. the best way to present the images (Chapter 3 and Chapter 7). The results of these experiments interested an airport and a manufacturer of x-ray scanners (Heimann GmbH). We cooperated on testing the effect of providing multiple viewpoints on real baggage inspection (Chapter 8).

In the first experiment (Chapter 4), I tried to show that the DVWS can improve the ability to see sharp edges. This task is more relevant for hand baggage than for hold baggage, but at that time I had not decided yet to concentrate on hold baggage. I found many unexpected results here: for example, response time increased and performance decreased with the available number of views.

In the second experiment I showed that, for detecting wires connecting two objects, performance increases with increasing *camera range* (the angular distance between the extreme available views). Furthermore, a reduced image quality (resolution and number of grey levels) can be offset by increasing the number of available views. It was shown that three extreme views are sufficient for this task, and that increasing the number of views within this range does not improve observer performance. I concluded that for x-ray baggage inspection it is necessary to provide extreme views to the inspector. Another welcome result was that, for this task, observers performed just as well when selecting the view manually instead of via their head movements. Selecting a view with the knob is less tiring than moving the head, and eliminates the need for expensive head position tracking. Although three views were found sufficient to detect wires connecting objects, I suspected that the availability of only three views would give a limited spatial impression and therefore might be insufficient for a task in a more complex scene.

In the third experiment it was shown that a large camera range is not sufficient for following a wire through a semi-transparent knot. Performance increased with the number of available views within a fixed horizontal range, up to continuous views. Providing both horizontal and vertical continuous views did not improve performance as compared with horizontal continuous views only. Here, response times decreased with the number of available views. Thus, the required image quality and number of available views seem to depend on the spatial complexity of the scene. For x-ray baggage inspection, the number of required views was expected to lie somewhere between that required for this task and the three views required for detecting a wire between two objects.

The best way to present the images – the optimum configuration of various perspective and display possibilities – was investigated in Chapter 3 and Chapter 7. In Chapter 3 a large range of perspective and display possibilities is explored, but this range was too large to investigate completely in an experiment.

The fourth experiment (Chapter 7) tested the effects on observer performance of a viewpoint measurement error and of the way the camera settings are coupled to the viewpoint of the observer. There are at least two ways to make an image given some viewpoint: one can keep the camera aimed at some point in the scene (*on-axis coupling*) or alternatively one can shift the camera to the new viewpoint without rotating the projection plane (*off-axis coupling*). The DVWS is an on-axis coupling. Geometrically, off-axis coupling seems the correct choice if the camera position is coupled to the eye position, because it is only with off-axis coupling that the objects represented subtend the same optical angles as objects in a real scene would subtend. Furthermore, measurement inaccuracies of the actual viewing position of the observer may cause the scene to appear different from a real scene (*distortion*). Both the coupling method and viewpoint measurement errors may decrease observer performance. It was shown that such distortions do occur as predicted by geometry, although usually observers do not notice them. However, human observers also use other than geometric cues for their task, and the distortions found seem less relevant for x-ray baggage inspection. For baggage inspection, on-axis coupling (i.e. the DVWS) seems the right choice, especially when the view is being selected via a knob.

The last experiment (Chapter 8) tested the effect of the number of available viewpoints on 'real' x-ray baggage inspection. An expert from an airport packed 68 suitcases, hiding complete bombs in 15 of them. The suitcases were scanned on an x-ray baggage scanner. The acquired views were presented with the DVWS to experienced baggage inspectors at the airport, and they were asked to detect bombs. The results showed no effect of the number of available views on the judgement of the inspectors, although the response time increased when two viewpoints were provided instead of one. These results suggest that the inspectors need a thorough training to interpret the spatial impression of the baggage. Probably, as happens frequently with new technology, it may even be necessary to use new inspectors with no experience of traditional x-ray inspection.

In conclusion, for baggage inspection and related tasks I found the following answers to the four questions posed at the start of this abstract:

1 In general, it is essential to clearly operationalize the task in perceptual terms.
2 Depending on the task, two views are sufficient or continuous views are required.
3 For some tasks, a knob is sufficient to select the required view, while for other tasks a coupling with the actual viewpoint of the observer can improve performance.
4 The perspective properties of the views can disturb the observer, depending on the choices made for (3).

Samenvatting

Dit proefschrift onderzoekt de mogelijkheid om röntgen-bagageinspectie te verbeteren door de inspecteur een ruimtelijke indruk van de bagage te geven. Het Delft Virtual Window System werd gebruikt om de inspecteur een zo'n ruimtelijke indruk van de bagage te geven op een gewoon monitorbeeldscherm. Het systeem geeft de inspecteur een ruimtelijke indruk van de bagage door de positie van de röntgencamera te koppelen aan de kijkpositie van de waarnemer. Bijvoorbeeld, als de inspecteur naar rechts beweegt wordt een beeld getoond dat de koffer meer van rechts laat zien.

Elke opname stelt de bagage bloot aan een bepaalde röntgendosis, en de maximaal toegestane dosis is bereikt na zo'n 25 opnames. Daarom moeten we zorgvuldig overwegen welke aanzichten de inspecteur kan checken. Verder, gegeven de state of the art technieken om scherpe röntgenbeelden te maken met een lage röntgendosis, hebben de aanzichten ongebruikelijke perspectief eigenschappen: ze bevatten convergent perspectief in, zeg, de horizontale richting in het beeld, maar parallel perspectief in de vertikale richting. Gezien de kosten prefereren wij het gebruik van bestaande technieken. Daarom zullen de perspectief-eigenschappen van de beelden zorgvuldig gekozen moeten worden om acceptabele aanzichten te krijgen als de beelden interactief met het DVWS gepresenteerd worden. Hoofdstuk 3 schetst de mogelijkheden om efficiënt meerdere van zulke röntgenbeelden op te nemen met de huidige bagage-inspectie technologieën.

Zo komen we vanuit de bruikbaarheid van het DVWS voor bagage-inspectie tot de volgende vragen:

1 Wat betekent ‘nuttig’ precies voor bagage-inspectie, en hoe testen we dat?
2 Wat zijn de nuttige beelden (welke beeldkwaliteit; welk aanzicht, hoe veel aanzichten)?
3 Hoe moet de inspecteur het gewenste aanzicht controleren?
4 Hoe kunnen we de beelden het best aanbieden, voor optimale prestaties van de inspecteur?

Vooral de eerste vraag is moeilijk te beantwoorden. Toepassing van het DVWS is nuttig als het de vaardigheden van de inspecteurs in het vinden van verdachte voorwerpen verhoogt in vergelijking met röntgen-inspectiesystemen zonder het DVWS. Omdat verwacht werd dat een bagagescanner op basis van het DVWS vooral gebruikt zal gaan worden voor ruimbagage, waar bommen de belangrijkste bedreiging zijn, werden de ‘verdachte voorwerpen’ beperkt tot ‘bommen’. Gewoonlijk bestaat een bom uit een batterij, een ontsteker, een tijdmechanisme, een draad die deze delen verbindt en explosieven, en we kunnen aannemen dat het detecteren van zulke onderdelen belangrijk is voor bagage-inspectie. Maar de precisering van de vage term ‘verdacht’ tot ‘bom’ helpt ons niet veel bij een wetenschappelijke analyse van de bagage-inspectietaak, omdat deze bom-onderdelen geen vaste vorm hebben en omdat deze onderdelen niet altijd in een bom zitten. Bijvoorbeeld, explosieven en batterijen kunnen elke vorm hebben, en draden kunnen achterwege blijven.

Om het aantal nuttige aanzichten en de manier waarop de waarnemer het gewenste aanzicht selecteert te bepalen, begon ik met een experimenteel onderzoek naar de beeldkwaliteit, het benodigde aantal aanzichten, en de manier waarop de waarnemer het aanzicht moet selecteren. Experimenten 1 tot 3 (Hoofdstuk 4-6) gaan over deze vragen. Vervolgens bekeek ik een moeilijker vraag, namelijk de beste manier om de aanzichten af te beelden (Hoofdstuk 3 en 7). De resultaten van deze experimenten trokken de belangstelling van een luchthaven en een producent van röntgenscanners (Heimann GmbH). Wij werkten samen om het nut van aanbieden van meerdere aanzichten voor echte bagage-inspectie te testen (Hoofdstuk 8).

In het eerste experiment (Hoofdstuk 4) probeerde ik aan te tonen dat het DVWS de zichtbaarheid van scherpe kanten verhoogt. Deze taak is relevanter voor handbagage dan voor ruimbagage, maar op dat moment had ik nog niet besloten om mij te concentreren op ruimbagage. Ik vond veel onverwachte resultaten, bijvoorbeeld dat de reactietijd groeide en dat de prestaties daalden met het aantal beschikbare aanzichten.

In het tweede experiment toonde ik aan dat, voor het detecteren van verbindende draden tussen twee objecten, de prestatie stijgt met het *camera-bereik* (de hoekafstand tussen de extreme beschikbare aanzichten). Verder kan een lagere beeldkwaliteit (resolutie en aantal grijswaarden) gecompenseerd worden door het aantal beschikbare aanzichten te verhogen. Voor deze taak werd aangetoond dat drie extreme aanzichten voldoende zijn, en dat het verhogen van het aantal aanzichten binnen een bereik de prestaties niet verhoogt. Ik concludeerde dat het voor bagage-inspectie nodig is om de inspecteur extreme aanzichten aan te bieden. Een ander bruikbaar resultaat was dat, voor deze taak, waarnemers even goed presteerden als ze het aanzicht met de hand kozen in plaats van via hun hoofdbewegingen. Een aanzicht kiezen met een knop is minder vermoeiend dan met hoofdbewegingen, en maakt dure hoofdpositie-bepalers overbodig. Ondanks dat drie aanzichten voldoende bleken om verbindende draden tussen objecten te detecteren verwachtte ik dat de beschikbaarheid van slechts drie aanzichten een beperkte ruimtelijke indruk geeft, en daarom onvoldoende kan zijn voor een taak in een meer complexe scene.

Het derde experiment toonde aan dat een groot camerabereik niet voldoende is om een draad door een halftransparante knoop te volgen. De prestaties verbeterden met het aantal beschikbare aanzichten binnen een vast horizontaal bereik, tot horizontaal continue aanzichten. Het aanbieden van zowel horizontaal als verticaal continue aanzichten verbeterde de prestaties niet in vergelijking met alleen horizontaal continue aanzichten. De reactietijden daalden met het aantal beschikbare aanzichten. De benodigde beeldkwaliteit en aantal aanzichten lijkt af te hangen van de ruimtelijke complexiteit van de scene. Voor bagage-inspectie werd verwacht dat het aantal benodigde aanzichten ergens tussen het aantal nodig voor deze taak en het aantal nodig voor de detectie van draden tussen objecten zal liggen.

De beste manier om de beelden te presenteren – de optimale configuratie van verschillende perspectief- en afbeeldingsmogelijkheden – werd onderzocht in Hoofdstuk 3 en Hoofdstuk 7. In Hoofdstuk 3 werd een groot aantal perspectief- en afbeeldingsmogelijkheden onderzocht, maar het aantal was te groot om grondig experimenteel te onderzoeken.

Het vierde experiment (Hoofdstuk 7) testte de effecten van een fout in de gemeten kijkpositie en van de manier waarop de camera-instellingen gekoppeld zijn aan het kijkpunt van de waarnemer op de prestaties van de waarnemer. Er zijn minstens twee manieren om een aanzicht te maken gegeven een kijkpunt: men kan de camera gericht houden op een punt in de scene (*on-axis koppeling*) of men kan de camera zonder te roteren naar het nieuwe kijkpunt verschuiven (*off-axis koppeling*). Het DVWS is een on-axis koppeling. Geometrisch gezien is off-axis koppeling de juiste keus, omdat alleen met off-axis koppeling objecten dezelfde optische hoek maken als echte objecten. Verder kunnen meetfouten op de kijkpositie van de waarnemer ervoor zorgen dat de scene er anders uitziet dan een echte scene (*distortie*). Zowel de koppelingsmethode als fouten in de gemeten kijkpositie kunnen de prestaties van de waarnemer verlagen. Ik toonde aan dat vervormingen optreden zoals geometrisch voorspeld, ondanks dat waarnemers ze meestal niet opmerken. Echter, menselijke waarnemers gebruiken ook andere dan geometrische diepte-informatie voor hun taak, en de gevonden distorties lijken minder relevant voor röntgen-bagageinspectie. Voor bagage-inspectie lijkt on-axis koppeling (dus, het DVWS) de juiste keus, vooral als het aanzicht met een knop geselecteerd wordt.

Het laatste experiment (Hoofdstuk 8) testte het effect van het aantal beschikbare aanzichten op 'echte' röntgen-bagageinspectie. Een expert van een luchthaven vulde 68 koffers, en verstopte complete bommen in 15 ervan. De koffers werden gescand op een röntgen bagage-scanner. Deze aanzichten werden met het DVWS gepresenteerd aan ervaren bagage-inspecteurs op de luchthaven, en hen werd gevraagd om bommen te detecteren. Uit de resultaten bleek geen effect van het aantal beschikbare aanzichten op de beslissing van de inspecteurs, hoewel de denktijd toenam als twee in plaats van een aanzicht beschikbaar was. Deze resultaten suggereren dat de inspecteurs een grondige training nodig hebben om de ruimtelijke indruk van de bagage te interpreteren. Mogelijk, en dit is niet ongebruikelijk met nieuwe technologieën, is het zelfs noodzakelijk nieuwe inspecteurs zonder ervaring met de traditionele röntgen-bagageinspectie te gebruiken.

Concluderend vond ik voor bagage-inspectie en gerelateerde taken de volgende antwoorden op de vier vragen die gesteld werden in het begin van deze samenvatting:

1 In het algemeen is het essentieel om de taak duidelijk te operationaliseren in perceptuele termen.
2 In afhankelijkheid van de taak zijn twee aanzichten voldoende of continue aanzichten benodigd.
3 Voor sommige taken is een knop voldoende om het benodigde aanzicht te selecteren, terwijl voor andere taken een koppeling met het huidige kijkpunt van de waarnemer zijn prestaties kan verbeteren.
4 De perspectief-eigenschappen van de aanzichten kunnen de waarnemer storen, afhankelijk van de keuzen die bij punt (3) gemaakt werden.

Acknowledgements

Gerda Smets, Kees Overbeeke and Pieter Jan Stappers had a major impact on the theoretical aspects and the precise experimental set-ups. Gerda started this project and steered the project, yet left much to my own initiative. Gerda, Kees and Michael Stratmann laid the foundations for the Delft Virtual Window System. Kees was particularly helpful with statistical issues, and made valuable comments on all my texts. Pieter Jan was my daily guide, gave huge amounts of comments on early texts and experimental set-ups, often from unusual (especially upside-down) points of view.

The production of many transparent objects (Chapters 4 and 5) would have taken much more time without the help of Guus Mansveld, Erik Ulijn, Toni Subroto, Rudolf Wormgoor, and the people at the faculty of Chemistry. Erik Ulijn also helped with programming the Acorn computer. Frans Loosen (Catholic University of Leuven) helped with the statistics of the experiment of Chapter 5. The construction of virtual knots (Chapter 6) was made possible with the virtual reality software of Pieter Jan and Arnold Paalder. Alexander van Elsas gave useful hints for the programming of the Silicon Graphics computer (Chapters 6 and 7). Paul Locher (Montclair State University, NJ) gave valuable general and statistical comments on the experiments described in Chapters 6 and 7. For the experiment with the real x-ray baggage, I want to thank the people at Heimann (Germany) for their help with scanning the baggage, especially Wolfgang Glaßner for his fast-working scanning program. Toni Subroto helped with the design and construction of the baggage rotation apparatus. For this last experiment, lots of people cannot be mentioned, but nevertheless their help was essential. Onno van Nierop helped with several layout issues and with taking photographs.

Tom Djajadiningrat, one of the other two research students on my project, helped me with the analysis of the concept solutions described in Chapter 3 and made mock-ups for some of the solutions. As a Macintosh-computer fan, he gave numerous useful hints. Fred Voorhorst was the other research student on my project. He did the analyses of the eye positions of the participants in my experiments. In spite of our common interests, the integration of our projects was not as large as anticipated.

The less formal contacts were also necessary. Contact with Pieter Jan, in particular, reduced my distance to the group. Furthermore, I had lots of informal discussions with Arnold and Toni concerning numerous aspects of information and aircraft technology.

Finally, this work has been made possible by the Dutch Technology Foundation (STW), as they supported this research under grant DIO22-2732. Particular thanks to the secretary, Piet Winkel, but also to the members of the 'gebruikerscommissie' (user committee), especially Prof. Casper Erkelens and Dr. Alexander Wertheim, for their critical remarks.

Illustration Acknowledgements

The author thanks the following organizations and authors for their kind permission to reproduce some of their figures in this thesis:

Figure 2.3 and 2.4
From Stappers (1992) with permission from P. J. Stappers.

Figure 3.4
From Isorad (1987), Isorad Limited did not respond to requests for reproduction.

Figure 3.6
From Heimann (1987) with permission from Heimann Systems GmbH.

Figure 3.7 and 3.8
From EG&G Astrophysics (1996) with permission from EG&G Astrophysics.

Figure 3.10
From Invision (1997) with permission from InVision Technologies, Inc.

Figure 3.13
From Heimann (1987) with permission from Heimann Systems GmbH.

Figure 7.2
From Petzold (1973). The author of this book could not be found.

Figure 7.4
From Cutting (1986) with permission from J. E. Cutting.

Figure 9.2 and 9.3
From Abraben (1994). Removed as Van Nostrand Reinhold did not give permission for reproduction in this electronic version.

Curriculum Vitae

Ik werd geboren op 7 maart 1968 te Soest. Tussen 1987 en 1991 deed ik de informatica opleiding aan de Universiteit van Amsterdam, waar ik in 1991 cum laude afstudeerde bij prof dr.. Klint met als afstudeerrichting programmatuur. Het afstudeeronderwerp betrof het maken van een incrementele parser voor het ASF+SDF systeem, een systeem om algebraïsche specificaties te maken en te evalueren. Tussen 1991 en 1993 studeerde ik muziekwetenschap aan de Universiteit van Utrecht, waar ik mijn propedeuse en het eerste jaar van het doctoraal afrondde. In 1993 begon ik als onderzoeker in opleiding (OIO) mijn onderzoek aan dit proefschrift aan de Technische Universiteit te Delft.

Index