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.

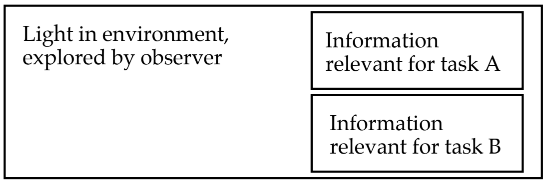

Figure 2.1. The direct theory proposes that the light in the environment contains all information necessary for the task.

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.

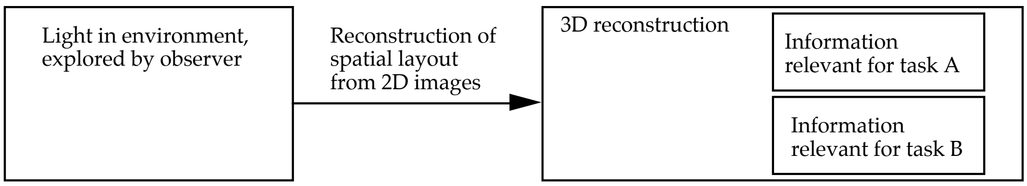

Figure 2.2. The indirect theory proposes that part of the information is lost and has to be recovered into a 3D reconstruction.

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*.

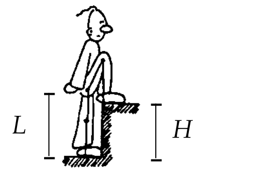

Figure 2.3. Observer stepping on a stair. *H* must be smaller than his leg length *L*.

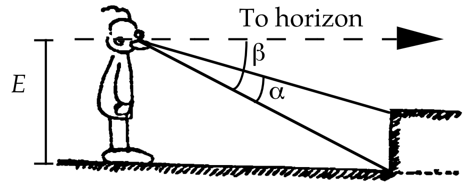

Figure 2.4. The ability to climb a stair may be detected by checking whether $\alpha/\beta < L/E$. *E* is the eye height of the observer (from Stappers, 1992).

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.

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

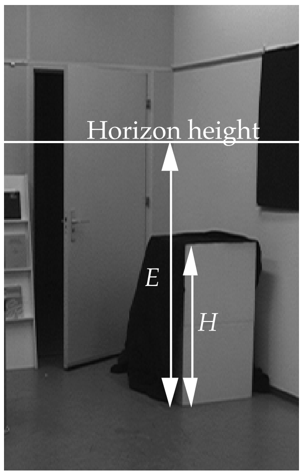

Figure 2.5. According to the direct theory, the observer may determine his ability to step on a stair from the ratio *H*/*E*. The horizon is indicated by gravitational, perspective and texture cues.

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.

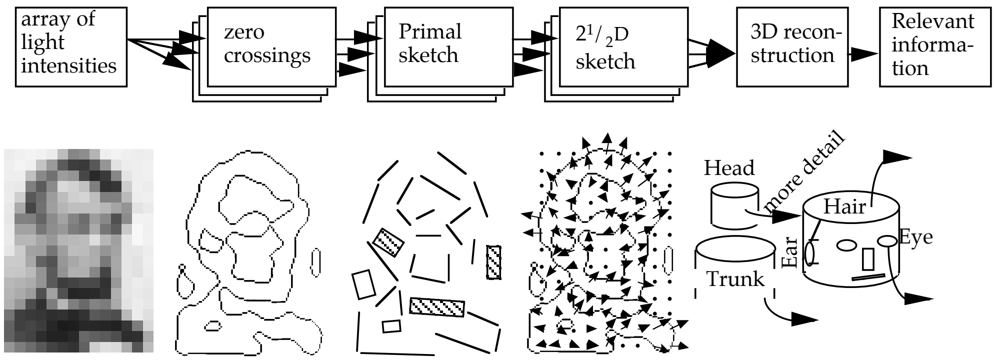

Figure 2.6. Outline of the computational model of vision as described by Marr. The image of the eyes is converted in a number of steps to a hierarchical 3D reconstruction. The first row shows these steps, the second typical structures in these steps

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.

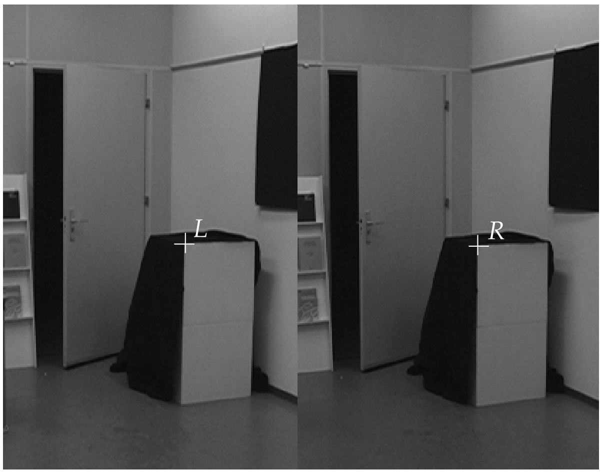

Figure 2.7. Image taken at left and right eye position. *L* and *R* are a matched position in these images, and they are expected to indicate a single position *S* in the spatial scene. This is a stereoscopic image, and can be viewed by looking at the left image with the left eye and at the right image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

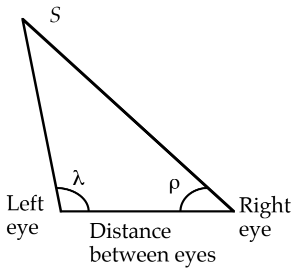

Figure 2.8. Depth reconstruction of point *S* from the direction from both eyes to that point.

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.

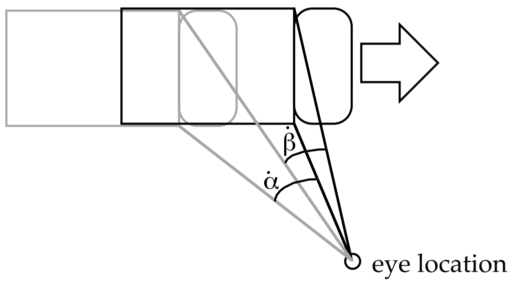

Figure 2.9. Top view of a truck passing an observer, as an example of motion parallax as a depth cue. Relative to the observer's eye location, the side of the truck near the observer has a higher angular speed than the side further away.

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.

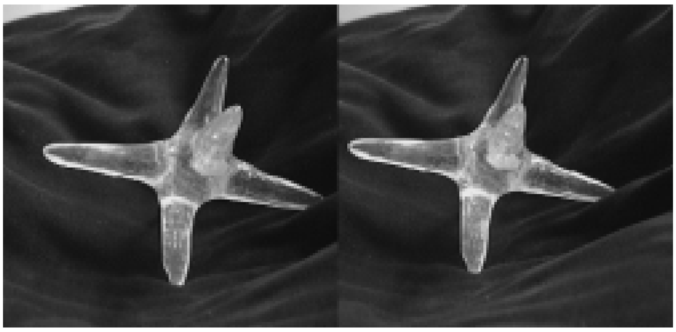

Figure 2.10. Stereoscopic image of an object with possibly sharp edges.

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.

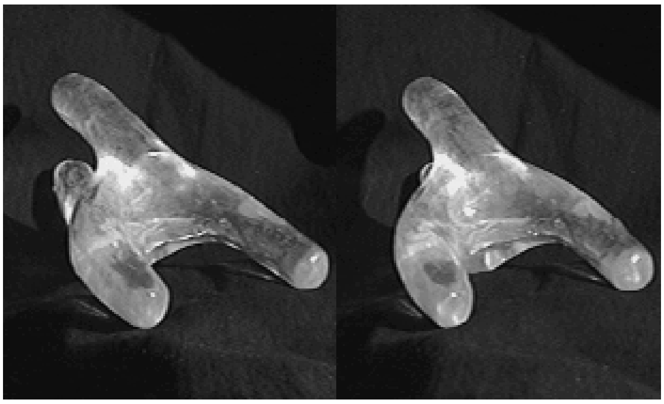

Figure 2.11. Stereoscopic image of an object with a sharp edge in its inner curve.

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.

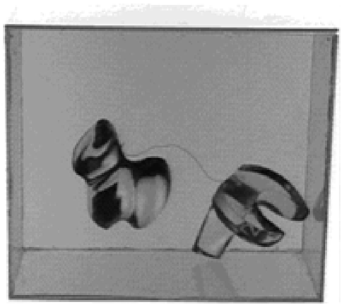

Figure 2.12. Wire connecting two objects in a box.

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.

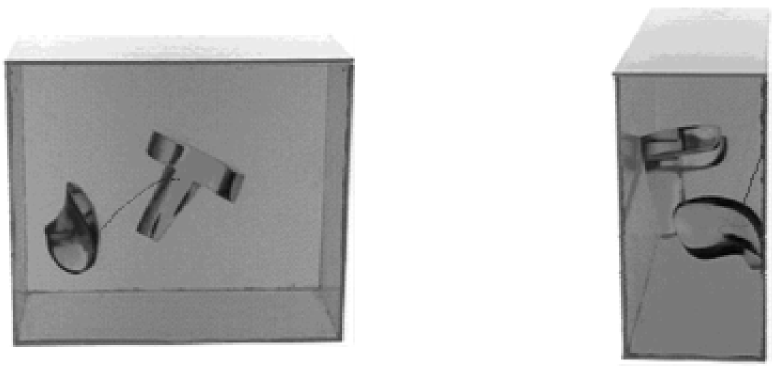

Figure 2.13. Front view of wire (left), possibly connecting to the objects. From an appropriate viewpoint (right) the wire is found not to touch the object on the right.

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.

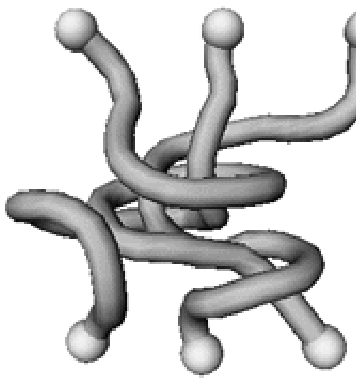

Figure 2.14. For tracing a knot through a wire, one view is sufficient.

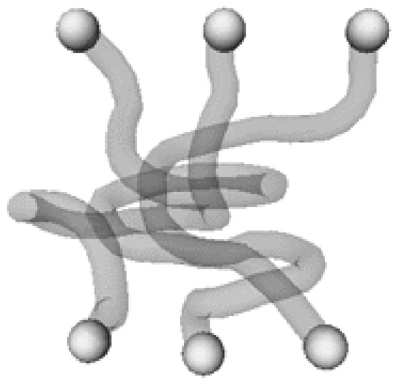

Figure 2.15. If the knot is transparent, additional depth cues are required.

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.

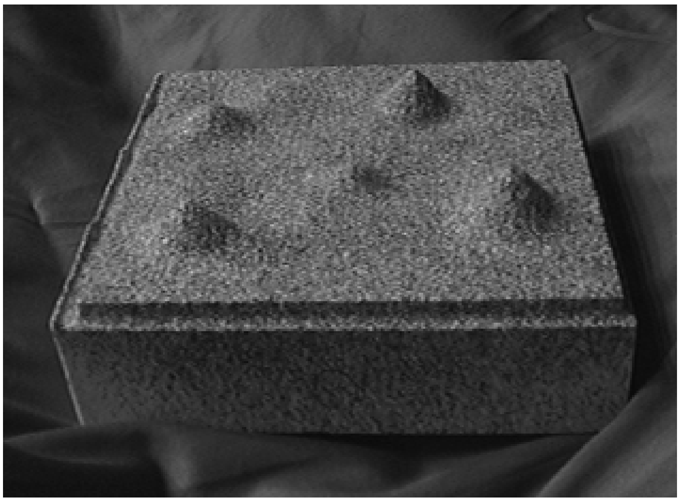

Figure 2.16. **See also figures on right cover flap.** Bumps in a landscape. Subjects had to adjust the height of the bumps to match a second landscape that had bumps in the same positions.

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.

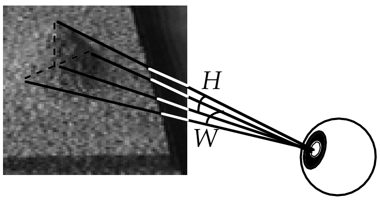

Figure 2.17. From a viewpoint perpendicular to the bumps, the bump heights can be compared by comparing the visual angle H or the ratio H/W.

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.

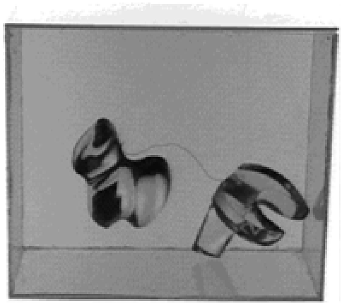

Figure 2.18a. A wire is visible when enough grey levels are displayed.

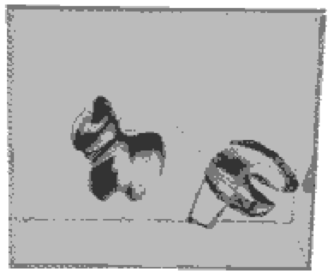

Figure 2.18b. With reduced number of grey levels, the wire may become invisible.

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.

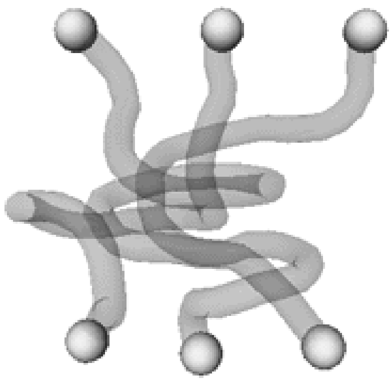

Figure 2.19a. Front view of a transparent knot.

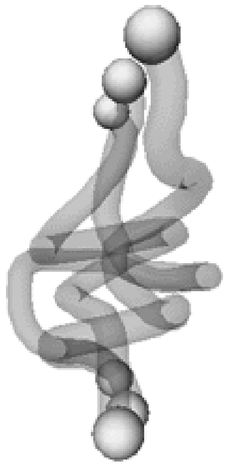

Figure 2.19b. Side view of the same knot. It is difficult to match the two views.

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.

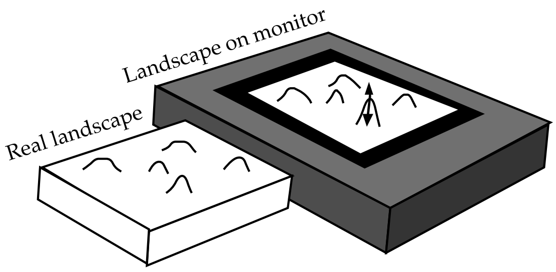

Figure 2.20: The plane between the bumps is placed parallel to the monitor display to induce the observer to take viewpoints that are oblique relative to the display.

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.