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.

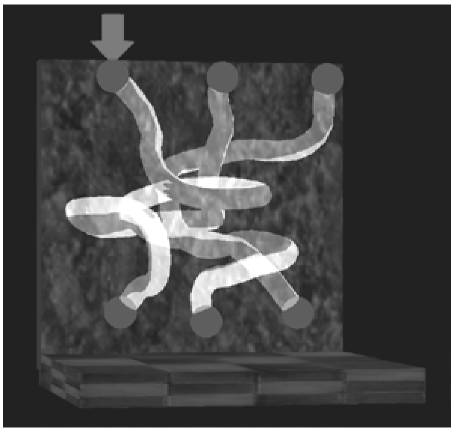

Figure 6.1 (**see colour figure on right cover flap**). Example of a knot. The arrow indicates the top end of one of the wires. Observers had to find the bottom end of the same wire.

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.

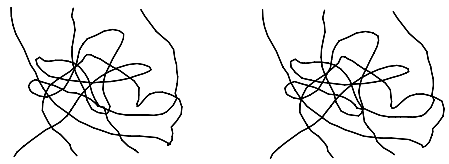

Figure 6.2. Sample knot to illustrate the importance of more than one viewpoint. With only one view, wires in a knot are hard to follow. Two views give a better spatial impression, especially when viewed stereoscopically.

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).

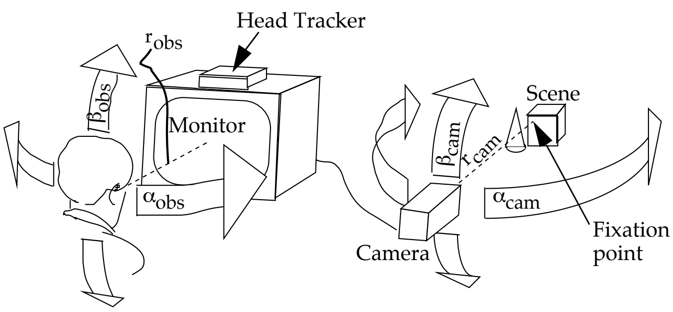

Figure 6.3. The Delft Virtual Window System consists of a monitor that displays the camera image. The camera position follows the eye position of the observer.

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*.

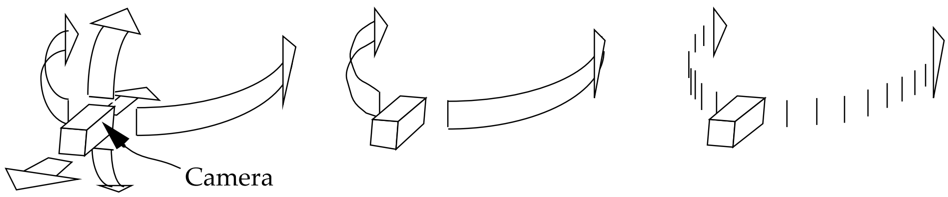

Figure 6.5a. In the unrestricted DVWS the number of available views is infinite in 3 dimensions.

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.

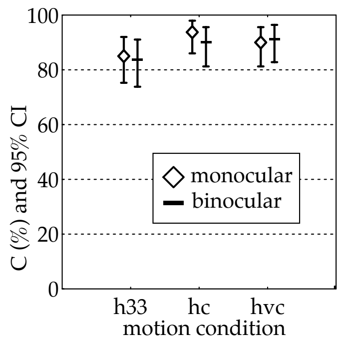

Figure 6.6. Percentage of correct answers and 95% confidence interval.

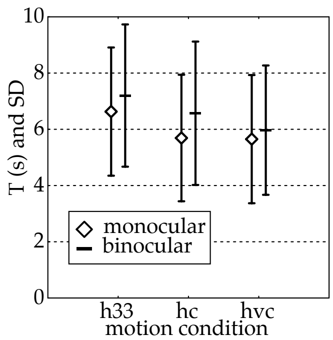

Figure 6.7. Response time and standard deviation.

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 x 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˚.

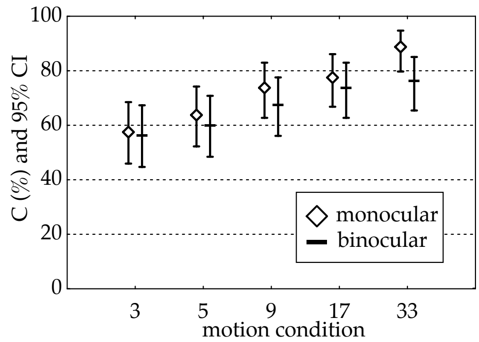

Figure 6.8. Percentage of correct answers and 95% confidence interval.

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.

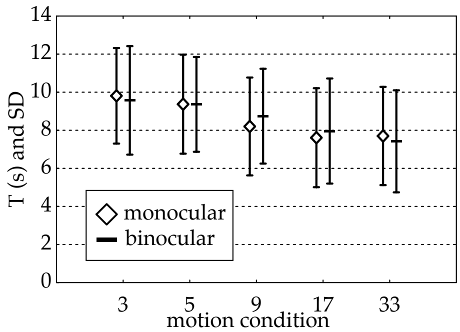

Figure 6.9. Mean response time and standard deviation. People use significantly more time with more restricted viewing conditions.

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.

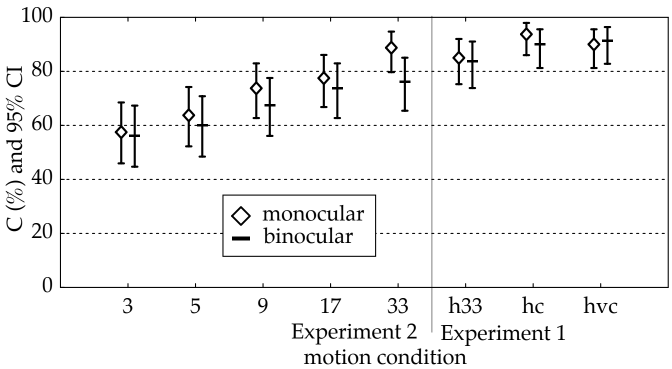

Figure 6.10. Percentage of correct answers and 95% confidence interval of the two experiments, as a function of the viewing and motion condition. Up to the hc condition, the percentage of correct answers increases.

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.

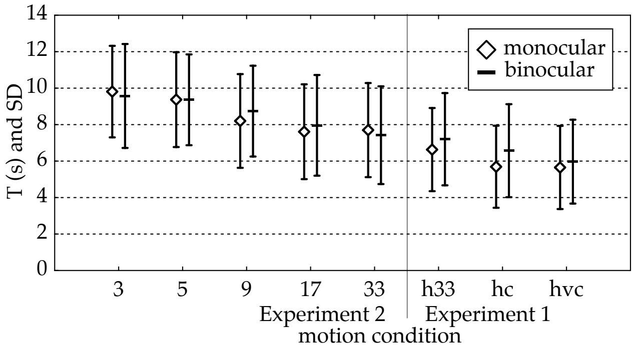

Figure 6.11. Mean response time and standard deviation of the two experiments. Observers make faster decisions when more viewpoints are available.

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.