Perception, 1996, volume 25, pages 1027–1042

Visual inspection of three-dimensional objects by human observers

Tobias Niemann, Markus Lappe, Klaus-Peter Hoffmann
Allgemeine Zoologie und Neurobiologie, Ruhruniversität, D-44780 Bochum, Germany;
e-mail: kph@neurobiologie.ruhr.uni.bochum.de
Received 28 July 1995, in revised form 5 March 1996

Abstract. Eye movements are an important aid in active visual exploration of the environment and in performing behavioural tasks. Eye movements might also play a role in human perception of three-dimensional (3-D) objects. Eye-movement strategies were investigated when humans inspected and memorised 3-D objects. Subjects were instructed to memorise the 3-D structure of parts of statues of human figures placed on a turntable free to rotate through 360°. Eye movements and turning behaviour were recorded. Different turning and eye-movement strategies could be observed. Subjects showed individual turning behaviours that were reproducible between trials. Turning strategies ranged from focusing on only a limited number of perspective views to almost continuously rotating the object with only short stops. On average twelve–thirteen views were inspected during memorising. Eye movements also revealed individual strategies. Fixation locations within each inspection view ranged from either closely spaced on isolated parts of the object to distributed over the whole view with large saccades in between. Eye movements were often directed to the same details from different perspectives. The differences in turning and viewing strategy also resulted in differences in the ability to recognise parts of the object later on. In general, successful later recognition required that the subject actually fixated the part to be recognised. A strategy of thoroughly inspecting the object with a series of closely spaced fixations from only a limited number of viewpoints led to best recognition rates. This was especially true for two subjects trained in fine arts with prior experiences in modelling. The results support models of viewpoint-dependent object recognition with viewer-centred, two-dimensional representations of 3-D objects.

1 Introduction

In biological visual systems, recognising and memorising three-dimensional (3-D) objects is an essential capacity. Eye movements facilitate the visual exploration of the environment and are also likely to be involved in obtaining the visual features necessary for memorising and recognising objects. The need to recognise objects from all possible viewpoints requires a system that can match the retinal image of an object to a stored representation of the object in memory. Two classes of models of information processing for object recognition have been proposed (see Ullman 1989). In the object-centred or 3-D representation model, only one object description encoded in a viewpoint-independent fashion is represented in memory. The stored object representation is compared directly with similar invariant descriptions computed from the retinal input (Marr and Nishihara 1978; Lowe 1987). In a different concept, the two-dimensional (2-D) viewer-centred representation model, a small number of viewer-centred descriptions of an object from a set of particular views are stored. A direct comparison between the input shape and a stored description is no longer possible because of a likely misorientation between the two which makes a normalisation process by a 3-D transformation necessary (Huttenlocher and Ullman 1987; Ullman 1989). A recently proposed approach of viewpoint-dependent object recognition interpolates novel views of a 3-D object between viewer-centred, 2-D representations of the 3-D object (Poggio and Edelman 1990; Edelman and Weinshall 1991).

Psychophysical experiments investigating strategies of memorising and recognising 3-D objects in human and nonhuman primates support the idea of storing a few 2-D descriptions of an object and interpolating between these viewer-centred representations (Rock and DiVita 1987; Bülthoff and Edelman 1992; Logothetis et al 1994). In memorising 3-D objects, inspection time seems to be not evenly distributed across all views. Particular views may be used for representation according to the type of object (eg geometric forms or faces). Prototypical views seem to be also important for memorising 3-D objects (Perrett and Harries 1988; Harries et al 1991). On the other hand, Cutzu and Edelman (1994) could not find any evidence for universally valid canonical views. Instead they suggest subject-specific, context-dependent, and task-dependent views for representation.

Neurons in the inferior temporal cortex (IT) seem to be involved in object recognition. IT neurons are sensitive to global features of objects, such as their shape, have large receptive fields (eg Gross et al 1972; Desimone et al 1984; Tanaka et al 1991), and might support mechanisms for invariant object recognition (Lueschow et al 1994). Perrett et al (1991) described cells in the temporal cortex sensitive to faces and heads that reveal a tuning to a range of different views of the head. After training monkeys to memorise a certain object, Logothetis (1994) could identify neurons in IT that revealed viewpoint-specific responses to only the object used for training, thus providing additional support for a viewer-centred object-recognition system in IT.

Eye movements are likely to be involved in obtaining the visual features used for memorising and recognising 3-D objects. They bring the retinal area of highest acuity, the fovea, onto the location of attention in order to process detailed information from the visual array (Breitmeyer 1986; Rayner and Pollatsek 1992). As a possible link between memorising and recognising 3-D objects and the execution of exploratory eye movements, the frontal eye field may play a role. It is involved in the control of voluntary saccades during exploration of natural visual scenes (Burman and Segraves 1994).

We investigated whether voluntary eye movements might aid the human perception of 3-D objects. Features like identifiable points, corner angles, or segment lengths are important inputs for models of 3-D-object recognition (Poggio and Edelman 1990), especially for identification of same object parts from different perspective views. The spatial distribution of fixations is supposed to indicate important features in a stimulus. Research in human scan paths and fixation locations during picture viewing in the past years suggested that eye movements are used to evaluate parts of a stimulus which contain valuable information for scene analysis (see Viviani 1990). However, a variety of different types of information may be important for an analysis, depending on the type of stimulus and instruction. Simple features like contours or angles seem to be valuable for fixation as well as features involving higher cognitive functions (see eg Mackworth and Morandi 1967; Yarbus 1967; Antes 1974; Loftus and Mackworth 1978; Stark and Ellis 1981; Guez et al 1994). We recorded eye movements in humans during the inspection of 3-D objects while the observers were free to rotate the object around the vertical axis. In combination with the subjects' selection of perspectives of the 3-D object during the memorising procedure we tried to determine possible perceptual strategies and stimulus features in obtaining information important for memorising 3-D objects.

2 Methods

2.1 *Subjects*

The main experiments employed sixteen human subjects, the control experiments four subjects. These eleven male and nine female subjects, 22–33 years old, were tested with normal or corrected-to-normal vision. All subjects were naive towards the purpose of the task. Two of the subjects (AT and TI) represented a special group in that they had prior experience in painting, drawing, and modelling.

2.2 *Experimental setup*

3-D objects were presented on a turntable in front of the observer. Viewing distance was 96 cm. The turntable could be rotated interactively by the observer to allow different perspective views of the object. The angular position of the turntable was recorded every tenth of a second with a resolution of 1°, and was stored on a PC for later off-line analysis. The objects were illuminated by diffuse daylight from an angle of 45°.

2.3 *Eye-movement-recording system*

Eye movements of the observer were monitored with an infrared eye tracker (Ober2). The eye-movement data were stored on a 486 PC and could be analysed off-line by software that we have developed. The sampling rate of the analogue-to-digital converter was 150 Hz.

2.4 *Eye-movement superposition*

Eye position was superimposed on-line onto a videotaped view of the object. This was achieved by an optical-superposition system that we ourselves have developed: in order to record the observer's view of the objects we used an identical second turntable which contained an identical copy of the object. The second turntable moved in synchrony with the original turntable. A video camera pointed towards the second turntable from a position identical to the position of the observer's eye relative to the original turntable. The resulting video image was a picture of the rotating object as seen by the observer. For the eye-movement superposition, a semisilvered mirror was positioned at a 45° angle between the video camera and the object on the second turntable. An on-line display of the observer's eye position was presented on a computer monitor positioned on the side of the mirror so that eye movements and stimulus were superimposed on the video image. Appropriate calibrations ensured that the actual fixation on the object coincided with the position of the on-line display of the observer's eye position on the object. Any trials with inappropriate calibrations were discarded from analysis. The video signal of the superimposed signals was stored on a video recorder with a frame rate of 50 frames per second for later analysis.

2.5 *Objects*

The 3-D objects we used were chosen in cooperation with partners of the ESPRIT INSIGHT project in order to provide a unified set of stimuli for various vision experiments ranging from psychophysics to neurobiology and computational vision [eg Koenderink et al (1996)]. The experiments presented here were performed with models of human figures. We tested four different model torsos differing in sex, posture, and surface texture. The size of the torsos was approximately 90 cm × 50 cm × 25 cm (figure 1).

2.6 *Experimental procedure*

All subjects inspected all torsos. The subjects did not see the torsos beforehand and inspected them only once during the experimental session. They were instructed to inspect thoroughly and memorise the global 3-D structure of the object. A recognition test was performed to encourage and force the subjects to concentrate on 3-D memorising during inspection. Subjects were told in advance about the form of the recognition test. They were told to concentrate only on the upper/shoulder part of the torso. Allowed inspection time was 2 min. During this time, the eye movements of the subject and the angular position of the object were recorded. Afterwards, the subjects were presented successively with sixteen photographs showing parts of the torso in the upright position in detail (approximately 9 deg × 6 deg) from eight possible perspectives (0°, 45°, 90°, 135°, 180°, etc) (see figure 1 as an example). Eight photographs from the torso just inspected were mixed with eight photographs from a different torso as distractors. The photographs of the torsos were taken from the same set of possible positions. During recognition the photographs were shown at a

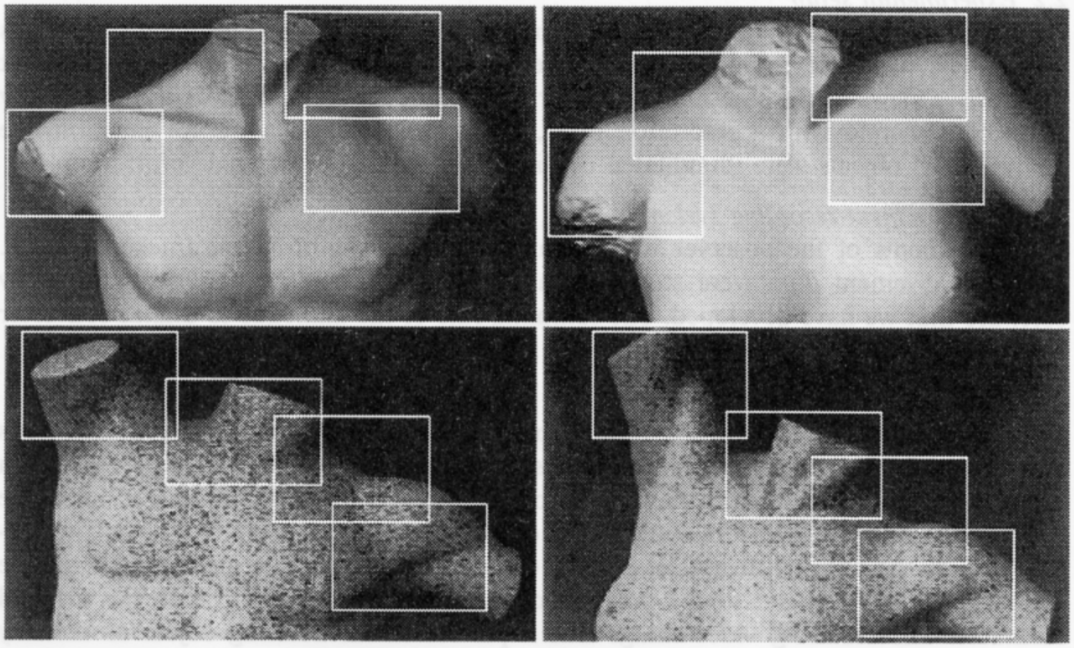


Figure 1. Frontal view (0°) of the 3-D objects used in the experiments, differing in sex, posture, and surface texture. Only the upper/shoulder part of the torso, approximately within the parts of the torsos shown in this figure, were relevant for the experiments. The white frames within the torsos show examples of sections presented on photographs used for recognition. At least three–four photographs of parts of the object from eight possible perspectives were taken during preparation of the experiments (0°, 45°, 90°, 135°, 180°, etc).

distance with the size of the details appropriate to the size of the details of the torso just inspected. The subjects were asked to judge whether the previously inspected object was shown on the photograph or not.

3 Results

The measurements were analysed with respect to the choice of perspective views, the eye movements, and the recognition rates.

3.1 *Viewing strategy and inspection time*

For all subjects the distribution of the time they spent on different perspectives of the objects was found to be nonuniform (chi-squared, $p \leqslant 0.0001$).[1] However, subjects showed different turning strategies in choosing the perspective views of the object. Two strategies are reflected by the position change of the object during measurements (figures 2a and 2b). Some subjects preponderantly rotated the object during inspection and only seldom kept a perspective for some period of time (figure 2b). Other subjects focused their inspection on particular views and quickly moved the object from one perspective view to the next (figure 2a).

[1] To rule out the possibility that a nonuniform distribution of perspectives results from the impossibility of rotating the object over 360° at a constant speed, a control experiment was performed. Subjects had to rotate the turntable as smoothly and constantly as possible while measuring the angular position of the turntable. When the distribution of different angular positions was divided in bins of 12° or more the distribution was found to be uniform (chi-squared test). Thus the distribution of perspectives of the original experiments were divided in 12° bins. The finding of a nonuniform distribution thus indicates that subjects indeed spent different amounts of time on different views.

To see more clearly how much time the subjects devoted to particular perspectives of the object, time-distribution graphs are shown in figures 2c and 2d. The percentages of time spent looking at particular views as the stimuli were rotated through 360° are displayed. Subject TI, investigating only particular views, had twelve preferential views ($p \leqslant 0.01$, binomial distribution) during inspection.

Subject WW, turning the object almost continuously during inspection, revealed ten preferential views ($p \leqslant 0.01$, binomial distribution). However, these ten views comprised less than half of the total viewing time. Most of the time was spent continuously turning the object (see also figure 3).

In order to investigate whether the turning behaviour fell into different classes, the ratio 'turntable stationary' to the total measurement time was determined (figure 3). No distinct classes of turning behaviour but rather a continuous distribution was observed. However, the individual turning strategies of the subjects remained stable over the different objects. Interestingly, the two subjects (TI and AT) trained in fine arts displayed the highest ratio and investigated the stationary objects most (~ 80% of total inspection time) while other subjects spent only 30% of the total inspection time on stationary views.

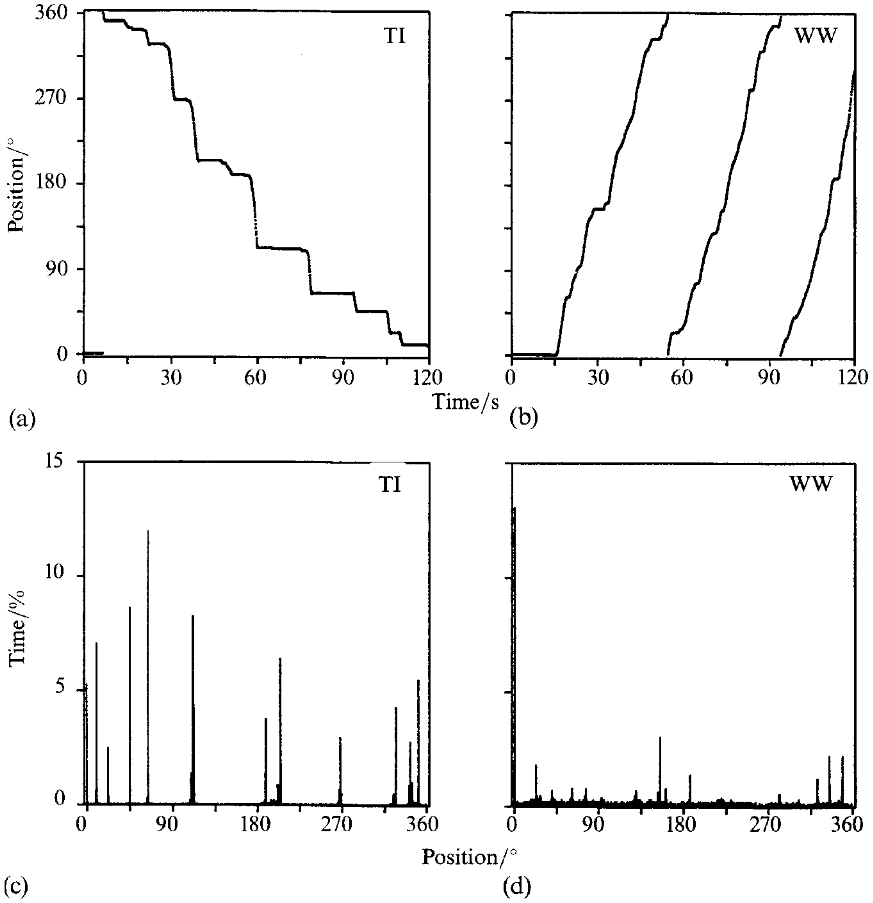


Figure 2. (a) and (b) Change of position of the torso during the measurements (120 s) for two subjects (TI and WW). Subject TI turned the object counterclockwise once during the time of measurement. Subject WW turned it clockwise nearly three times. The sampling rate of the angular position was 10 Hz. (c) and (d) Relative distribution of inspection time for different views of the torso for the same subjects. Bin width was 1°.

Table 1. Mean saccadic amplitude, fixation duration, recognition rate, and preferred views for different subjects. As the distributions of saccadic amplitude and fixation duration were skewed, the mean and standard deviation were calculated from the logarithms of the data. The saccadic

	Subject						
	TI	AT	KS	TF	BH	BW	RD
Saccadic amplitude/deg							
mean	1.29	1.35	1.43	1.60	1.62	1.62	1.70
+SD	2.94	3.29	3.09	4.63	4.41	3.93	3.72
−SD	0.56	0.55	0.66	0.55	0.59	0.66	0.77
Fixation duration/ms							
mean	184.8	269.7	247.1	233.4	221.7	292.3	358.2
+SD	276.9	438.1	404.4	357.9	354.0	498.3	580.1
−SD	123.4	166.1	150.9	152.2	138.8	171.5	221.2
Recognition rate/%	75.0	78.1	65.6	68.7	56.3	62.5	53.1
Number of preferred perspectives							
mean	11.5	10.0	7.3	14.3	15.5	11.7	13.5
SD	±4.1	±1.4	±4.6	±3.7	±2.1	±3.9	±4.9

The preferred views of each subject and for each trial during inspection of 3-D objects were analysed. They were evaluated with $p \leqslant 0.01$ (binomial distribution) and averaged (see table 1). All subjects had several preferred views during inspection of 3-D objects. Moreover, subjects seemed to investigate the objects using on average twelve–thirteen perspectives. To investigate whether subjects had similar or the same preferred views (canonical or prototypical views) or whether individual choices of preferred views existed, we cumulatively plotted the turning data of all subjects and experiments in one graph to look for peaks or clusters of positions which could be universally important for memorising 3-D objects (figure 4). There is a clear and significant peak representing the frontal view (0°) indicating that this was the preferred view for all subjects ($p \leqslant 0.01$, binomial distribution) [see figures 1 or 5 for frontal view (0°)].

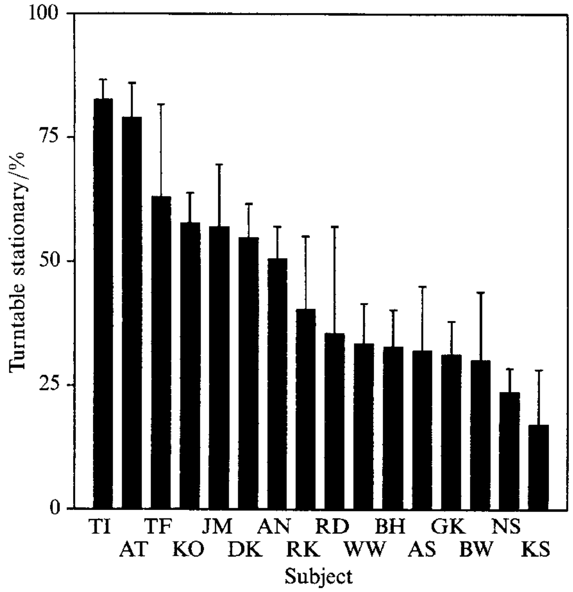


Figure 3. Mean 'turntable stationary' as a percentage of total measurement time (120 s) for different subjects. 'Turntable stationary' was defined and evaluated with $p \leqslant 0.01$ (binomial distribution). The percentage was determined for each trial and then averaged for each subject.

amplitude and fixation duration were calculated from at least 1000 values, the recognition rate and preferred views from 4 values. The preferred views were defined and evaluated with $p \leqslant 0.01$ (binomial distribution).

									Mean
JM	GK	DK	NS	KO	RK	AS	WW	AN	
1.72	1.81	1.86	1.89	1.98	2.13	2.15	2.50	2.60	
4.52	4.08	4.52	5.62	4.58	5.82	5.54	7.77	5.96	
0.65	0.81	0.76	0.69	0.85	0.78	0.83	0.80	1.13	
222.9	244.1	343.0	261.6	234.9	292.0	287.0	220.5	318.6	
357.5	369.9	633.0	427.6	371.7	475.0	483.6	345.5	531.0	
138.9	161.1	185.8	160.0	148.4	179.5	170.3	140.7	191.1	
65.6	65.6	62.5	46.9	68.8	79.0	62.5	68.7	68.7	
19.2	12.2	12.5	11.5	14.5	6.7	16.0	11.7	12.5	12.5
±4.1	±4.9	±1.9	±3.4	±2.0	±2.9	±4.9	±3.3	±1.3	±3.1

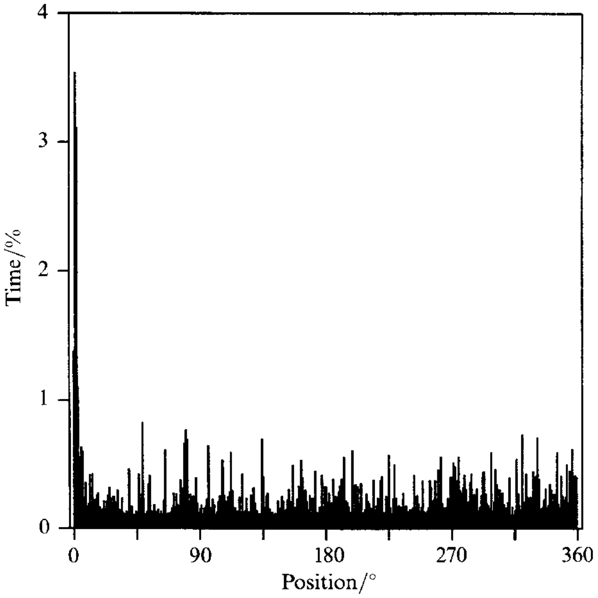
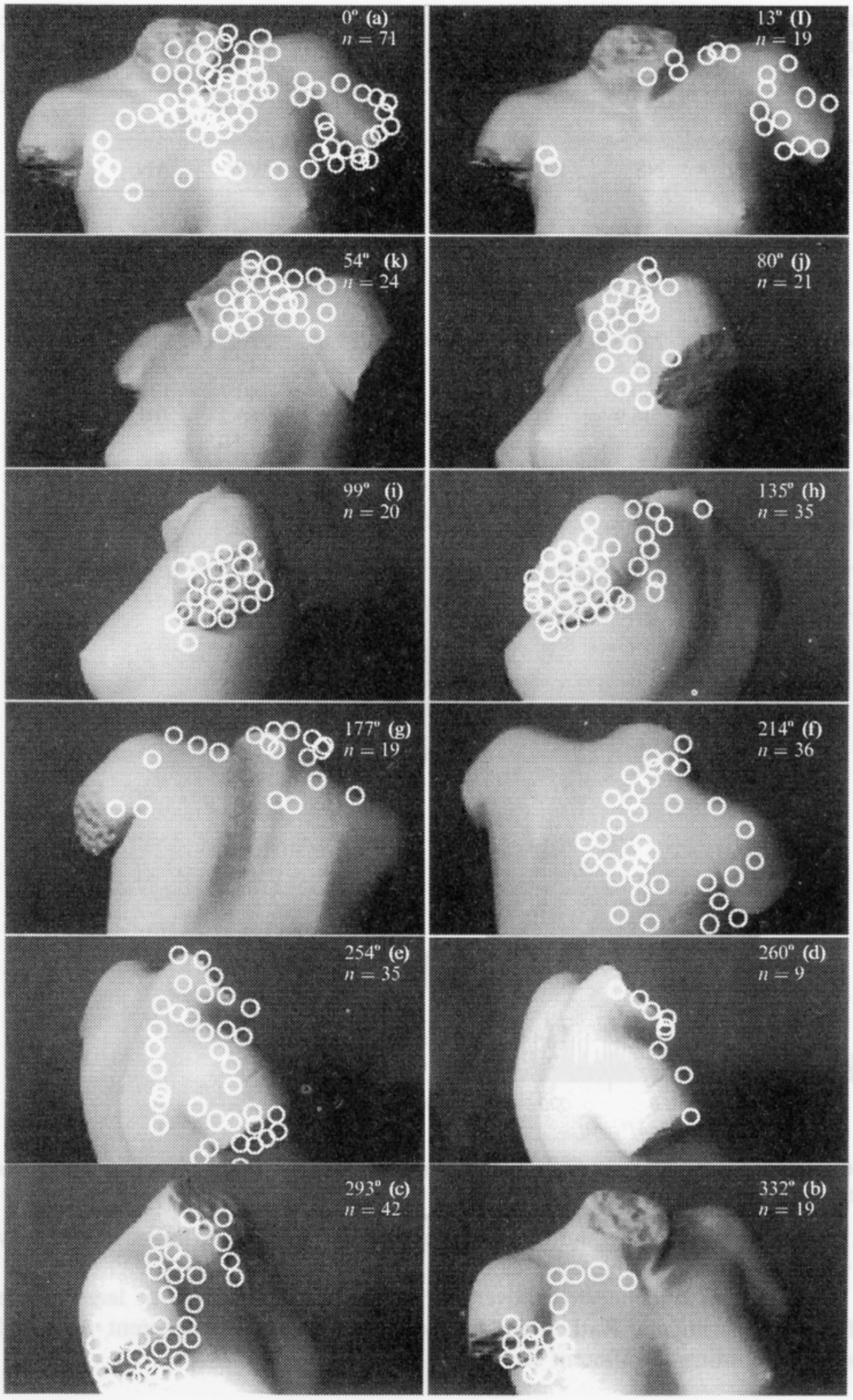


Figure 4. Cumulative relative distribution of inspection time for different views for all torsos and subjects ($n = 48\,571$). Bin width was 1°.

But apart from the frontal view, no other distinct peaks that would indicate generally preferred views were found to be significant.[2]

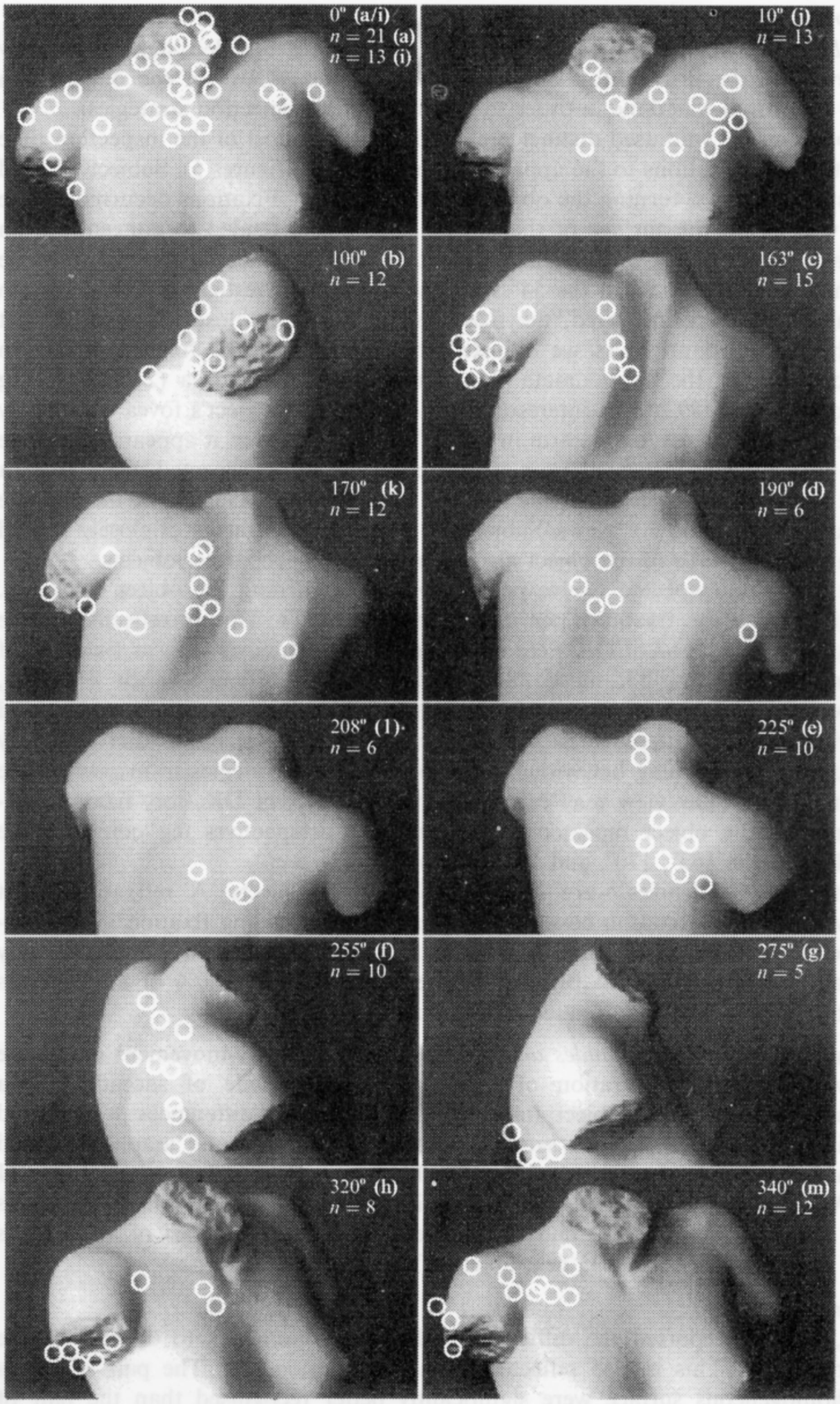
Thus, we found no evidence for universally valid canonical views, at least in the context of these experiments. Rather, the subjects had individually different preferred views, possibly according to different strategies.

[2] One reason for the preferred viewing of the frontal (0°) view could be the standardised experimental procedure, since all measurements started with showing the frontal view (0°) first. In order to test whether subjects investigated the starting position preferentially, we performed control experiments that started from different views. Subjects showed the preference for the frontal view of the torsos in these experiments also.



Subject TI

Figure 5. Distribution of fixations and preferred perspectives during inspection of a torso for two subjects (TI and DK). Each white circle represents a single fixation. The diameter of the circle is 2 deg, approximating the diameter of a human fovea. The angular position (in °), the



Subject DK

Figure 5 continued.
number of fixations (n), and the sequence of inspection (alphabetical order a, b, c . . .) is given for each perspective view. Fixations during rotation are not shown.

3.2 *Eye movements*

3.2.1 *Distribution of fixations.* We analysed the superimposed video signal frame by frame to evaluate the distribution of fixations on different views of the objects. An analysis by individual was done in order to find individual inspection strategies.

For those subjects who used distinct preferred views for most of the inspection time we could attach the fixations to the appropriate perspectives (figures 5). Subject TI used twelve perspective views, turning the object counterclockwise. Fixations occurred on the outline as well as on the surface of the object. The subject tended to concentrate on different parts of the object in different perspectives. For instance, in the 99° view the arm was investigated whereas in the 54° view the neck was investigated. When certain parts of the object were inspected, the fixations tended to be closely spaced, almost covering the inspected area. As we adjusted the diameter of the fixation marks in figure 5 according to the approximate diameter of the human fovea (2 deg) (see eg Rayner and Pollatsek 1992), it is interesting to note that the subject's fovea completely covered the area of interest (eg the arm in the 99° view). Moreover, it appears that some parts that were already inspected from a previous view were inspected again from a subsequent perspective: for instance, the left arm in the 99° and 135° views, the left side of the neck in the 54° and 80° views. Subject TI retained this strategy of closely spaced fixation on certain details of the object also when inspecting the other objects.

Subject DK also inspected twelve perspective views, turning the object clockwise. He showed many fewer fixations per view than subject TI. This is reflected in the fixation duration. Fixations of DK were on average nearly twice as long as fixations of TI (DK, mean 343.0 ms; TI, mean 184.8 ms; significant difference with $p \leqslant 0.01$, U-test). Fixations of DK were less concentrated on particular parts of the object and more distributed. The difference between these two subjects suggests that not only different strategies in turning behaviour but also individual strategies in inspecting the objects from a particular view may exist. Nevertheless, subject DK, too, fixated individual points on the object repeatedly from different perspectives (eg points of the shoulder blade in the 163°, 170°, and 190° views).

The refixation tendencies were investigated for all subjects. A refixation event occurred when the new fixation covered the area of the preceding fixation in a different perspective or covered an area very close to it (at not more than approximately 1 deg distance). Thus, a quantitative analysis revealed that over 50% of all fixations were directed to details already fixated in a different perspective.

3.2.2 *Analysis of saccadic amplitudes and fixation durations.* Eye-movement data were analysed with respect to duration of fixations and amplitude of saccades. Data were accumulated for each subject (table 1). Interindividual differences were found both for fixation duration and for saccade amplitude. Duration and amplitude could differ by a factor of nearly two (duration: TI 184.8 ms, RD 358.2 ms; amplitude: TI 1.29 deg, AN 2.60 deg; significant differences with $p \leqslant 0.01$, U-test). The two subjects trained in fine arts (AT and TI) had the shortest saccadic amplitudes of all subjects (TI, mean 1.29 deg; AT, mean 1.35 deg; no significant difference).

3.3 *Recognition*

In general, subjects reported the difficulty of the recognition task to depend on the pair of torsos used. This is also reflected in the recognition rate. The pair of torsos having a homogeneous surface were significantly better recognised than the pair of torsos with a textured surface ($p \leqslant 0.01$, U-test; figure 6).

Recognition rates showed individual differences but the subjects trained in fine arts (AT and TI) were among those with the highest rates (table 1).

Usually recognition was better from photographs that showed parts of the object which the subject had actually fixated during memorising. To analyse this relationship further,

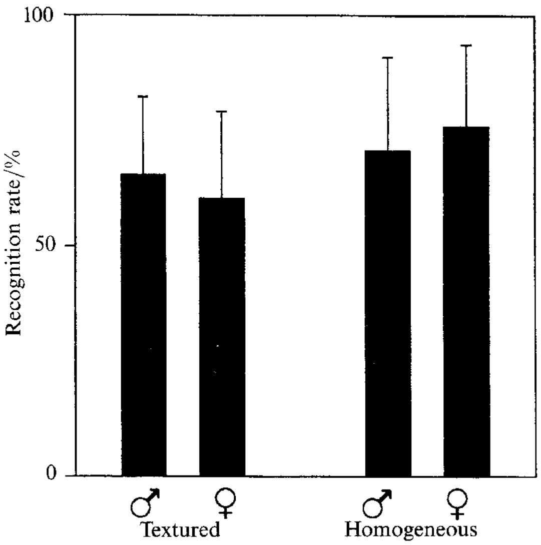


Figure 6. Mean recognition rates (+SD) for all subjects for different torsos (♂ male torso; ♀ female torso; textured, torso with 'textured' surface; homogeneous, torso with 'homogeneous' surface). Level of chance is 50%.

recognition data for all subjects and objects were pooled and distributed into three categories: (A) the subject actually fixated the part of the object that was presented on the photograph; (B) the subject did not fixate the part presented on the photograph from any perspective during inspection; (C) the subject fixated the part of the object presented on the photograph, but from a different perspective. In most of the recognition tasks the subjects actually fixated that part of the object presented on the photograph [(A) 83% of all cases]. Recognition here was successful, with a rate of 67.5%. In only 7% of all recognition tasks a section of the object was presented on a photo which actually was not fixated from any perspective during inspection (B). Recognition here was only 41.7%, dropping to the level of chance (50%). In 10.2% of all recognition tasks the subjects were inspecting a photograph showing a detail of the object that they had fixated under a different perspective (C). However, recognition in this condition was again successful, with a rate of 71.4%. Cases A and B and cases B and C were significantly different, with $p \leqslant 0.01$, whereas cases A and C were not significantly different (fourfold table test).

In order to test whether successive fixations are important for memorising and recognising 3-D objects, a control experiment was performed in which naive subjects had to suppress eye movements during inspection. Recognition rate was again investigated. Four subjects took part. A spot was placed just above the object in order to present a stable fixation target during rotation of the object. Subjects had to fixate the spot during measurement. Permanent fixation was controlled with the eye tracker, the inspection time was again 120 s. Recognition, which was performed again with the photographs, was about 10% worse compared with the rate when eye movements were used, but still significantly above chance level ($p \leqslant 0.01$, binomial distribution). The control experiment indicates that successive fixations are beneficial for memorising and recognising 3-D objects.

4 Discussion

We found the following characteristics in human visual inspection of 3-D objects.

(i) The distribution of perspective views chosen by a subject was nonuniform. Subjects showed different kinds of behaviour for the choice of perspective views, ranging from an inspection of a limited number of viewpoints to a nearly continuous turning behav-

iour with only short stops in between. The average number of inspected viewpoints was about thirteen.
(ii) In any particular view, gaze was directed to only some parts or details of the object. Moreover, eye movements were often directed to the same points of the object from different perspectives. Subjects showed different strategies when memorising 3-D objects, as reflected in the fixation positions, saccadic amplitude, and fixation duration.
(iii) In general, successful recognition required that the subject actually fixated the part to be recognised. However, this fixation did not need necessarily to be from the same perspective, but could also occur from a neighbouring perspective.

4.1 *Viewing strategy*

For all subjects the distribution of the time spent in different perspective views of the objects was found to be nonuniform. This behaviour of preferential inspection was also found in other studies involving 3-D geometrical objects, opaque smooth objects, model heads, or widgets for visual inspection (Perrett and Harries 1988; Harries et al 1991; Perrett et al 1992). According to the type of object and the degree of freedom in rotation, different characteristic views were found in these studies. Regularly faceted tetrahedra and opaque smooth objects were typically viewed from perspectives where the principal axis was parallel or perpendicular to the line of sight. Heads were viewed preferentially from a full-face view and a view close to the profile. Widgets were viewed from perspectives where faces of the object were orthogonal to the line of sight (plan views). However, in all of these studies preferential views were consistently similar across all subjects. In our experiments only the frontal view was consistently inspected by all subjects. There was no indication of a coherence between perspectives with principal axis parallel or perpendicular to the line of sight and the preferential view as described by Perrett and Harries (1988). Possibly the structure of our objects was too complex for this simple relation to apply. On the other hand, Harries et al (1991) also found specific prototypical views for the inspection of heads, such as the full-face and close-to-profile view. When human torsos were inspected, the frontal view was found to be prototypical, since it was inspected by all subjects. The preference for a full-face view in the case of model heads may be due to the presence of facial features which have social relevance for recognition. Human bodies may also contain such socially relevant features. Although these features were not important in the experimental context, it may have led to general preference for the frontal view of the torsos. There was no other perspective which could be regarded as prototypical. Instead every subject had his or her own set of preferential views. These individual preferences support a concept of memorising 3-D objects in the subject-specific context and task-dependent manner previously proposed by Cutzu and Edelman (1994). The tendency to inspect only a limited set of views of the object lends support to models in which 3-D objects are stored as multiple viewer-centred representations in memory (Koenderink and van Doorn 1979; Ullman 1989; Poggio and Edelman 1990; Edelman and Weinshall 1991). Our subjects used approximately twelve–thirteen views for memorising a 3-D object for a full 360° turn. On average, every 30° a 2-D viewer-centred image was taken for a complete representation of the object. When human and nonhuman primates are trained to a particular view of an object, recognition of this object drops considerably when the object is rotated by about 40° (Rock and DiVita 1987; Bülthoff and Edelman 1992; Logothetis et al 1994). Thus for recognising a 3-D object in every possible perspective of 360°, at least ten perspectives must be stored. In this case, specific views lying in between the stored viewer-centred representations can successfully be interpolaied (see Poggio and Edelman 1990; Edelman and Weinshall 1991). Our result on the viewing strategy seems to support such a process of memorising for recognition.

However, some subjects showed a different viewing strategy. They continuously rotated the object with only short stops at specific perspectives. One reason might be that these subjects may be inexperienced in memorising 3-D objects. We included two subjects who had received some training in fine arts and had experience in memorising 3-D objects, for instance in the context of modelling. These subjects readily showed a viewing strategy with a limited number of perspectives (about ten – twelve) and quick moves from one view to the next. Possibly, this kind of strategy develops with practice and depends on prior experience. However, turning the object in that manner yielded better recognition scores, thus indicating a more successful strategy.

4.2 *Eye movements*

Subjects showed individual inspection strategies but similarities across objects. As the experimental instructions in each trial were the same, although the type of torso changed, subjects probably always adopted their individual strategy. This type of inspection behaviour was previously observed in investigations of human scan paths and fixation locations during inspection of 2-D pictures (eg Yarbus 1967; Noton and Stark 1970; Gale and Findlay 1983; Groner et al 1984; Ellis and Stark 1986). Inspection of 3-D stimuli results in similar fixation positions as with inspection of 2-D stimuli. It is assumed that the distribution of fixations indicates what part of the stimulus is considered to be an important feature or contains valuable information. Fixations tend to cluster at, for example, angles, line ends, contours, unpredictable details, or areas with contrast information or socially important features, depending on the type of stimulus and instruction (eg Mackworth and Morandi 1967; Yarbus 1967; Antes 1974; Loftus and Mackworth 1978; Stark and Ellis 1981; Burman and Segraves 1994; Guez et al 1994). Individual analysis revealed that the outline of the object was fixated as well as the inner parts. As the objects in our experiments were rather complex (compared with, for example, simple geometrical objects) it is difficult to decide whether the location of fixations refers to simple features such as contrast borders, or to complex features such as the arm or neck. Eye movements could be controlled by bottom-up or top-down processes or by a mixture of both (see Viviani 1990).

As the subjects were presented with photographs showing parts of the object during the recognition test, they might have adopted an inspection strategy scrutinising the objects for detailed discriminatory features within the object set. Thus, they may have looked for and memorised detailed minutiae and irregularities of the material rather than memorise the global 3-D structure of the object. Although we cannot rule out the possibility, we do not believe in such an inspection strategy. Subjects were told in advance to concentrate on and memorise the global or general 3-D structure. The selection of parts of the objects on the photographs were chosen in order to show general 3-D structures but not striking details or irregularities. The subjects knew in advance about the form of the recognition test but it was impossible for the subjects to predict which part would be presented from the possible set of photographs during the recognition test. Thus to concentrate on certain details or irregularities of a certain part in a certain perspective might not be successful. Also, the time for inspection was limited and probably not long enough to concentrate on every detail in every possible perspective. For those reasons, we believe that the inspection strategies shown are more concerned with memorising general 3-D structures of the object, some more effective, some less effective.

However, in our experiments about half of all fixations were directed to refixate the same points of the object from different perspectives. For models of object recognition from perspective views (Poggio and Edelman 1990) feature inputs such as corner angles, segment lengths, local colour, or texture of the object are required. These feature inputs would be needed at first to build up an internal representation of the object and later for object recognition. For full information about the 3-D structure of the object, corresponding features from different views have to be identified (Ullman 1979). This might explain the high amount of refixations in our experiments. In memorising the 3-D structure of an object for later recognition subjects may use an eye-movement strategy that seems to be in accordance with this proposed model to build up an internal representation of the object.

The analysis of fixation duration and saccadic amplitudes revealed individual differences of up to a factor of two. These differences were previously observed also in different experimental contexts (see Suppes 1990; Viviani 1990). Loftus (1972) investigated the influence of duration of fixation on recognition memory in a study of natural scenes inspected by freely moving eyes. No effect was found on recognition accuracy (see also Loftus 1981). Our result confirms this finding. Recognition rates did not depend on the fixation duration in our experiments: subjects TI and RK, who were both among those with the highest recognition rates, differed distinctly in their fixation duration. Rather, we suppose a link between the saccadic amplitude and the recognition rate. The two subjects (AT and TI) with among the highest recognition rates had the smallest saccadic amplitude. Moreover, these subjects inspected the object from a limited number of views. Their *combination* of viewing and oculomotor strategy seems to have advantages for memorising and recognising 3-D objects.

4.3 *Recognition*

The quality of recognition was in good agreement with scores obtained in other studies (Warrington 1982; Perrett and Harries 1988; Harries et al 1991). Yet recognition rate depended on the type of object and subjectively experienced difficulty. Further analysis of recognition rates indicated that the parts of the object recognised had actually to be fixated from the same or a different perspective. Thus successive fixations were necessary during memorisation of 3-D structures. This finding is in line with that of Loftus and Bell (1975) investigating the recognition performance in picture memory. In that study subjects had a considerably better recognition performance when encoding and reporting an informative detail than when a detail was not reported. Moreover, the experiments of Parker (1978) suggest that a correct identification of details usually needs a fixation beforehand. This was also controlled in an experiment where the eye movements were suppressed. Recognition was worse than when eye movements were allowed. However, since the score was still above chance level when eye movements were restricted, peripheral vision also contributes to the process of memorising 3-D objects.

4.4 *Neurophysiology*

In physiological studies in the temporal cortex of macaque monkeys cells were identified that were responsive to faces in particular views (Desimone et al 1984; Hasselmo et al 1989; Perrett et al 1991) and sensitive to changes in perspective. Recently, cells selective to bodies were found (Wachsmuth et al 1994) which responded optimally to particular perspective body views. Perrett et al (1991) observed more head-selective cells coding face, profile, and back views than cells coding in-between views, which coincided with their observation of consistent preferential views of 3-D-model head in humans (Harries et al 1991). A majority of cells tuned to a particular view may also be the reason for the general preferential frontal view in the case of 3-D torsos. Body-selective cells like those described by Wachsmuth et al (1994) may be tuned mainly to the frontal or face view, which could be due to the importance of this view in a social context (ie body features). Logothetis (1994) identified neurons in the IT of monkeys which are only selective to a trained perspective of an abstract object. Thus a preponderance of cells tuned to particular views (eg face view) may not be a general function of the system but rather depend on the context and type of object. However, the physiological studies of Perrett and coworkers and of Logothetis (1994) support the concept of a memory-based, 2-D, viewer-centred object-recognition system. Eye movements seem to be important for memorising and recognising 3-D objects. Burman and Segraves (1994) demonstrated that the frontal eye field is involved in generation of voluntary saccades during natural scanning eye movements in monkeys. The frontal eye field is probably also involved in the generation of scanning eye movements during the memorising of 3-D objects, thus being part of a complex system guiding eye movements for later successful recognition.

Acknowledgements. We thank Ines Kubigsteltig and Julia Strehse for their assistance. This work was supported by ESPRIT INSIGHT II.

References

Antes J R, 1974 "The time course of picture viewing" *Journal of Experimental Psychology* **103** 62 – 70

Breitmeyer B G, 1986 "Eye movements and visual pattern perception", in *Pattern Recognition by Humans and Machines* volume 2, Eds E C Schwab, H C Nusbaum (Orlando, FA:Academic Press)

Bülthoff H H, Edelman S, 1992 "Psychophysical support for a 2D view interpolation theory of object recognition" *Proceedings of the National Academy of Sciences of the USA* **89** 60 – 64

Burman D D, Segraves M A, 1994 "Primate frontal eye field activity during natural scanning eye movements" *Journal of Neurophysiology* **71** 1266 – 1271

Cutzu F, Edelman S, 1994 "Canonical views in object representation and recognition" *Vision Research* **34** 3037 – 3056

Desimone R, Albright T D, Gross C G, Bruce C, 1984 "Stimulus-selective properties of inferior temporal neurons in the macaque" *Journal of Neuroscience* **4** 2051 – 2062

Edelman S, Weinshall D, 1991 "A self-organizing multiple-view representation of 3D-objects" *Biological Cybernetics* **64** 209 – 219

Ellis S R, Stark L, 1986 "Statistical dependency in visual scanning" *Human Factors* **28** 421 – 438

Gale A G, Findlay J M, 1983 "Eye movement patterns in viewing ambiguous figures", in *Eye Movements and Psychological Functions: International Views* Eds R Groner, C Menz, D F Fisher, R A Monty (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 145 – 168

Groner R, Walder F, Groner M, 1984 "Looking at faces: Local and global aspects of scanpaths", in *Theoretical and Applied Aspects of Eye Movement Research* Eds A G Gale, F Johnson (Amsterdam: North Holland/Elsevier) pp 523 – 533

Gross C G, Rocha-Miranda C E, Bender D B, 1972 "Visual properties of neurons in inferotemporal cortex of the monkey" *Science* **166** 1303 – 1306

Guez J-E, Marchal P, Le Gargasson J-F, Grall Y, O'Regan J K, 1994 "Eye fixations near corners: Evidence for a centre of gravity calculation based on contrast, rather than luminance or curvature" *Vision Research* **34** 1625 – 1635

Harries M H, Perrett D I, Lavender A, 1991 "Preferential inspection of views of 3-D model heads" *Perception* **20** 669 – 680

Hasselmo M E, Rolls E T, Baylis G C, 1989 "The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey" *Behavioural Brain Research* **32** 203 – 218

Huttenlocher D P, Ullman S, 1987 "Object recognition using alignment" *Proceedings of the 1st International Conference on Computer Vision* (Washington, DC: IEEE) pp 102 – 111

Koenderink J J, Doorn A J van, 1979 "The internal representation of solid shape with respect to vision" *Biological Cybernetics* **32** 211 – 216

Koenderink J J, Doorn A J van, Kappers A M L, 1996 "Pictorial surface attitude and local depth comparisons" *Perception & Psychophysics* **58** 163 – 173

Loftus G R, 1972 "Eye fixations and recognition memory" *Cognitive Psychology* **3** 525 – 551

Loftus G R, 1981 "Tachistoscopic simulations of eye fixations on pictures" *Journal of Experimental Psychology: Human Learning and Memory* **7** 369 – 376

Loftus G R, Bell S M, 1975 "Two types of information in picture memory" *Journal of Experimental Psychology: Human Learning and Memory* **104** 103 – 113

Loftus G R, Mackworth N H, 1978 "Cognitive determinants of fixation location during picture viewing" *Journal of Experimental Psychology: Human Perception and Performance* **4** 565 – 572

Logothetis N K, 1994 "Shape representation in the inferior temporal cortex of the monkey" *Abstracts of the 17th Annual Meeting of the European Neuroscience Association, Supplement 7* 40.02 (Oxford: Oxford University Press)

Logothetis N K, Pauls J, Bülthoff H H, Poggio T, 1994 "View-dependent object recognition by monkeys" *Current Biology* **4** 401 – 414

Lowe D G, 1987 "Three-dimensional object recognition from single two-dimensional images" *Artificial Intelligence* **31** 355 – 395

Lueschow A, Miller E K, Desimone R, 1994 "Inferior temporal mechanisms for invariant object recognition" *Cerebral Cortex* **5** 523 – 531

Mackworth N H, Morandi A J, 1967 "The gaze selects informative details within pictures" *Perception & Psychophysics* **2** 547 – 552

Marr D, Nishihara H K, 1978 "Representation and recognition of the spatial organization of three dimensional structure" *Proceedings of the Royal Society of London, Series B* **200** 269 – 294

Noton D, Stark L, 1970 "Scanpaths in saccadic eye movements while viewing and recognizing patterns" *Vision Research* **11** 929 – 942

Parker R E, 1978 "Picture processing during recognition" *Journal of Experimental Psychology: Human Perception and Performance* **4** 284 – 293

Perrett D I, Harries M H, 1988 "Characteristic views and the visual inspection of simple faceted and smooth objects: 'tetrahedra and potatoes'" *Perception* **17** 703 – 720

Perrett D I, Harries M H, Looker S, 1992 "Use of preferential inspection to define the viewing sphere and characteristic views of an arbitrary machined tool part" *Perception* **21** 497 – 515

Perrett D I, Oram M W, Harries M H, Bevan R, Hietanen J K, Benson P J, Thomas S, 1991 "Viewer-centred and object-centred encoding of heads by cells in the superior temporal sulcus of the rhesus monkey" *Experimental Brain Research* **86** 159 – 173

Poggio T, Edelman S, 1990 "A network that learns to recognize three-dimensional objects" *Nature (London)* **343** 263 – 266

Rayner K, Pollatsek A, 1992 "Eye movements and scene perception" *Canadian Journal of Psychology* **46** 342 – 376

Rock I, DiVita J, 1987 "A case of viewer-centred object perception" *Cognitive Psychology* **19** 280 – 293

Stark L, Ellis S R, 1981 "Scanpath revisited: Cognitive models direct active looking", in *Eye Movements: Cognition and Visual Perception* Eds D F Fisher, R A Monty, J W Senders (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 193 – 226

Suppes P, 1990 "Eye-movement models for arithmetic and reading performance", in *Eye Movements and their Role in Visual and Cognitive Processes* Ed. E Kowler (Amsterdam: Elsevier) pp 455 – 477

Tanaka K, Saito H, Fukada Y, Moriya M, 1991 "Coding visual images of objects in the inferotemporal cortex of the macaque monkey" *Journal of Neurophysiology* **66** 170 – 189

Ullman S, 1979 *The Interpretation of Visual Motion* (Cambridge, MA: MIT Press)

Ullman S, 1989 "Aligning pictorial descriptions: An approach to object recognition" *Cognition* **32** 193 – 254

Viviani P, 1990 "Eye movements in visual search: Cognitive, perceptual and motor control aspects", in *Eye Movements and their Role in Visual and Cognitive Processes* Ed. E Kowler (Amsterdam: Elsevier) pp 353 – 393

Wachsmuth E, Oram M W, Perrett D I, 1994 "Recognition of objects and their component parts: Responses of single units in the temporal cortex of the macaque" *Cerebral Cortex* **5** 509 – 522

Warrington E K, 1982 "Neuropsychophysical studies of object recognition" *Philosophical Transactions of the Royal Society of London, Series B* **298** 15 – 33

Yarbus A L, 1967 *Eye Movements and Vision* (New York: Plenum)