The Royal Society

Biases in three-dimensional structure-from-motion arise from noise in the early visual system

M. A. Hogervorst* **and R. A. Eagle**

Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK

The projected pattern of retinal-image motion supplies the human visual system with valuable information about properties of the three-dimensional environment. How well three-dimensional properties can be recovered depends both on the accuracy with which the early motion system estimates retinal motion, and on the way later processes interpret this retinal motion. Here we combine both early and late stages of the computational process to account for the hitherto puzzling phenomenon of systematic biases in three-dimensional shape perception. We present data showing how the perceived depth of a hinged plane ('an open book') can be systematically biased by the extent over which it rotates. We then present a Bayesian model that combines early measurement noise with geometric reconstruction of the three-dimensional scene. Although this model has no in-built bias towards particular three-dimensional shapes, it accounts for the data well. Our analysis suggests that the biases stem largely from the geometric constraints imposed on what three-dimensional scenes are compatible with the (noisy) early motion measurements. Given these findings, we suggest that the visual system may act as an optimal estimator of three-dimensional structure-from-motion.

Keywords: human vision; motion; depth; perspective; noise

1. INTRODUCTION

It has been shown (Wallach & O'Connell 1953; Rogers & Graham 1979) that the changes in the retinal image caused by relative movement between an observer and a viewed object can create a vivid impression of a three-dimensional object. To recover three-dimensional structure from two-dimensional retinal motion, assumptions have to made about the viewed scene. One common assumption limits the three-dimensional solutions to rigidly moving bodies (Ullman 1979). However, even with the use of the rigidity assumption, often more than one solution exists. Geometrically, a stimulus under two-view orthographic projection is compatible with a one-parameter family of scenes (Huang & Lee 1989; Koenderink & van Doorn 1991). This is illustrated in figure 1, where two different scenes have been arranged to produce identical image motion. It is noteworthy that despite the large range of objects consistent with such two-view displays, human reports of perceived shape are found to be consistent over time (Todd & Bressan 1990; Liter *et al*. 1994).

If more views are added, or the scene is viewed under perspective projection, then there is information available that uniquely specifies the three-dimensional structure and movement (Longuet-Higgins 1981; Ullman 1979). Figure 1*c*,*d* shows the flow-fields resulting from the objects depicted in figure 1*a*,*b* under three-view perspective projection. The addition of the third view yields differences in the acceleration component in the image motion for 'matched points' (points at the same image location on the two structures). This is true even in the centre of the displays, where orthographic and perspective projection yield similar results. For more peripheral locations, where orthographic projection is no longer a good approximation, the speeds and directions of matched points become increasingly different. Intriguingly, even though recent evidence suggests that humans can make use of both of these sources of information when discriminating two three-dimensional shapes defined by motion (Eagle & Blake 1995; Eagle & Hogervorst 1997), reports of misperceptions of the underlying scene exist in the literature even for these displays (Braunstein *et al*. 1993; Caudek & Proffitt 1993; Liter *et al*. 1994; Tittle *et al*. 1995).

This paper has two aims. First, to provide systematic measurements of these perceptual biases for a range of three-dimensional scenes, in which objects related by a linear stretching in depth are rotated over different extents. Second, to provide a computational model to investigate the source of these biases.

2. METHODS

Subjects viewed computer-generated structure-from-motion (SFM) stimuli that simulated rigidly connected, vertically hinged planes that rotated back and forth smoothly about a vertical axis (see figure 1). The stimuli were viewed under perspective projection with the eye in the centre of projection. The subjects' task was to set the dihedral angle of a probe

*Author for correspondence (maarten.hogervorst@psy.ox.ac.uk).

Received 15 April 1998 *Accepted* 11 May 1998

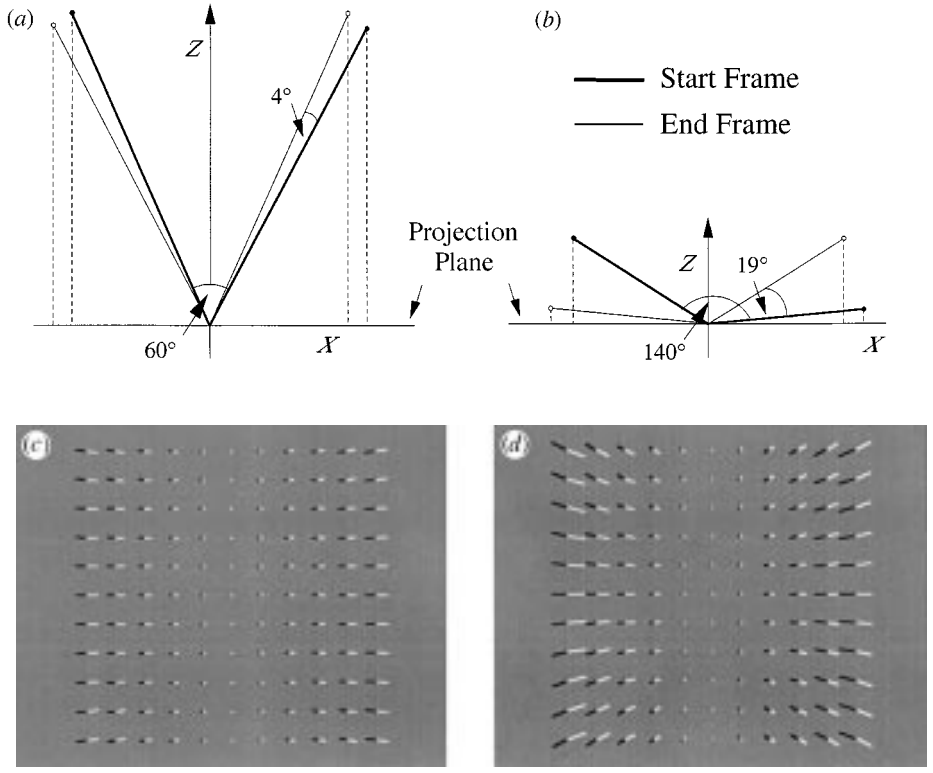

Figure 1. (*a*, *b*) Plan-views of two rotating hinges that, under two-view orthographic projection, yield identical image motion. (*a*) Dihedral angle of 60° rotating over 4°; (*b*) dihedral angle of 140° rotating over 19°. (*c*, *d*) Flow-fields produced by these stimuli under three-view perspective projection. White shows the flow from the first to the middle view and black shows the flow from the middle to the last view. The scale of the flow-fields is that of the large-field stimuli used in the experiment: 33° × 33°.

stimulus defined by binocular disparities, texture and motion to match that of the test stimulus.

(a) *Stimuli*

The stimuli were generated by a Silicon Graphics Indy Workstation on a 19 inch (48.3 cm) Silicon Graphics monitor with a screen resolution of 1280 pixels × 1024 pixels. The simulated objects consisted of a number of points, shown in the projection as green dots at high contrast against a dark background, in which standard anti-aliasing techniques (four bits for each colour) provided by the system were used to arrive at sub-pixel resolution. Image sizes of 8° × 8° and 33° × 33° were used with dot densities of 2.4 and 0.63 points per degree, respectively. To reduce static depth information, the objects were created by back-projecting points with random position in the projection plane onto the objects, and projection was clipped outside the upper and lower extents of the viewing window (i.e. ±4° and ±15.5°). The stimuli were updated at the frame rate of 76 Hz.

A set of 66 stimuli were generated simulating 11 different dihedral angles, ranging from 35° to 168°, and six rotation angles, ranging from 2° to 58°. In an experiment run before the main one, one subject adjusted the angular speed for each stimulus to find the setting that maximized the ease at which the shape judgement task could be performed. The matched frequency of oscillation was found to vary between 3.4 Hz (small displacements) and 0.34 Hz (large displacements) according a power function of the average displacement with a constant of −1/3.

The probe was defined by binocular disparities (appropriate to each subject's inter-ocular separation), texture (circles of various sizes defined by compression, density and perspective), motion (rotation over 45° at 1.5 Hz) and visible outlines; all consistent with a real object viewed from the subject's position. For such stimuli, veridical judgements of three-dimensional shape have been reported (Johnston *et al*. 1994) and we have found that observers can accurately set this probe to a right-angle. The probe consisted of two hinged planes, each of which spanned 15.5° × 33°. As the probe rotated, subjects could change the dihedral angle between these planes by moving the mouse.

(b) *Procedure*

Subjects sat in a darkened room with their head supported by a chin-rest 33 cm from the screen. Subjects wore red–green anaglyph glasses (cross-talk *ca*. 6%) to render the motion stimulus monocular and the probe binocular. The subjects judged the simulated dihedral angle of the SFM stimulus by adjusting the dihedral angle of the probe, which replaced the SFM stimulus when the mouse button was held down. Subjects made six settings for each structure and motion combination, one in each of six sessions.

3. RESULTS

Figure 2 shows the settings for one of the observers and the average over three observers as a function of simulated dihedral angle and rotation angle for both large- and small-field stimuli. The results show a systematic dependency of perceived structure on rotation angle. For small dihedral angles and rotation angles, the perceived dihedral angle was severely overestimated (depth underestimated). Note too that the error bars are small, showing that this is a systematic bias. This result is consistent with findings for stimuli viewed under orthographic projection (Liter *et al*. 1994). Increasing the rotation angle led to more veridical settings, although the perceived dihedral angle was now slightly underestimated. Furthermore, increasing the dihedral angle led to a corresponding increase in the perceived dihedral angle, and the settings for the largest dihedral angles were independent of rotation angle.

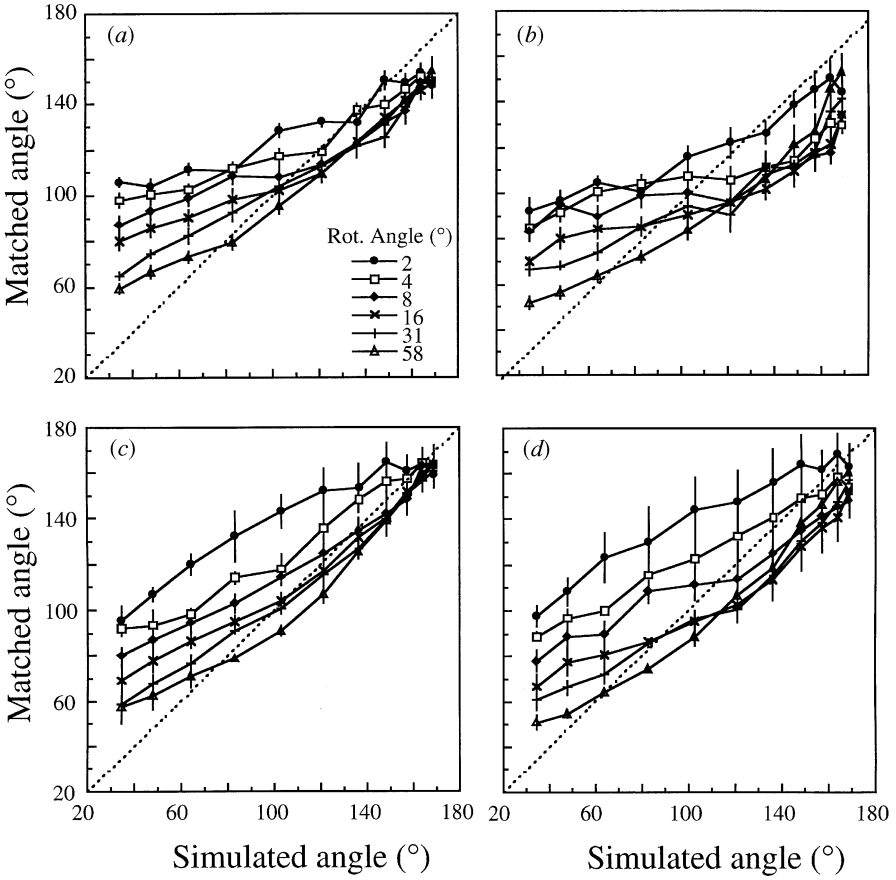

Figure 2. The settings for naive subject J.H., illustrating perceived three-dimensional structure, are shown for the large- and small-field conditions in (*a*) and (*b*), respectively, with error bars depicting the s.e.m. The mean settings for three subjects are shown for the large- and small-field conditions in (*c*) and (*d*), respectively, with error bars depicting the s.e. across the subjects' means.

An immediate concern was that the severe underestimations of depth may have been due to 'flatness cues' specified by other visual sources. Therefore, in a large-field control condition, the stimuli were generated with texture density cues appropriate to the scene structure and were viewed through pinholes of 1 mm to remove any screen-depth cues due to accommodation. Performance was not systematically different, indicating that these biases were not due to these cue conflicts.

4. MODEL

In an attempt to understand these biases, we developed a Bayesian model for the recovery of three-dimensional SFM, as this allows for a combined treatment of the task as one of perceptual inference based on the stimulus information together with prior knowledge of the structure of the world (e.g. Knill & Richards 1996). A key notion is that noise on the retinal motion measurements leads to a distribution of possible underlying scenes, rather than the unique solution obtained under noise-free conditions. Bayes' equation formulates the probability that a given set of measurements $\boldsymbol{I}$ originates from a three-dimensional scene S, $p(\boldsymbol{S}|\boldsymbol{I})$, in terms of the probability of obtaining those measurements given a scene, $p(\boldsymbol{I}|\boldsymbol{S})$, and the prior probability of encountering a scene, $p(\boldsymbol{S})$:

$$p(\boldsymbol{S}|\boldsymbol{I}) \propto p(\boldsymbol{I}|\boldsymbol{S})p(\boldsymbol{S}). \quad (1)$$

(a) *Implementation*

We assume that the system makes six independent measurements of each point, which indicate the average location $\boldsymbol{x}$, average velocity $\boldsymbol{v}$, and average acceleration $\boldsymbol{a}$. The uncertainty in the measurements of the position $\boldsymbol{x}$ are assumed to be negligible relative to that of the flow components. (Discrimination thresholds for performing a three-line bisection task have been shown to follow Weber's law, with Weber fractions of around 2% under optimal conditions (Westheimer & McKee 1979).) We assume that the noise in the remaining four measurements follow Gaussian distributions. The width of these distributions are derived from existing human sensitivity data for extracting speed and direction (De Bruyn & Orban 1988) and temporal changes in speed and direction (Snowden & Braddick 1991; Werkhoven *et al.* 1992). Note that under the assumption of Gaussian noise, the width of the noise distribution is $\sqrt{2}$ smaller than the threshold of 84%.

From the results of De Bruyn & Orban (1988), the noise in the measurement of the speed S, for speeds up to $64^\circ\,\mathrm{s}^{-1}$ (the range available in the stimuli), is well-characterized by

$$\sigma_S = 0.049 + 0.035S(^\circ\,\mathrm{s}^{-1}), \quad (2)$$

whereas the noise in the direction measurement is well-characterized by

$$\sigma_\phi = 2.83/S + 1.06(^\circ), \quad (3)$$

in which S is the mean speed of the point.

Snowden & Braddick (1991) found that, over a large range of speeds and temporal frequencies, square-wave modulations in speed could be detected when the difference exceeded 30% of the lower speed. For our stimuli, a triangle-wave modulation is a better approximation and

Werkhoven *et al*. (1992) have shown that thresholds are $\sqrt{3}$ larger in this case. Given this, and the additional finding that thresholds increase for low velocities, the noise in the overall speed change δS is estimated as

$$\sigma_{\delta S} = 0.058 + 0.36S(^\circ\,\mathrm{s}^{-1}). \quad (4)$$

Werkhoven *et al*. (1992) also measured thresholds for detecting modulation in the velocity direction, although over a more limited range of mean speeds. Under the assumption that sensitivity to direction modulation varies with speed in the same manner as does sensitivity to speed modulation (i.e. increases proportionally to the Weber fraction for speed change, $\sigma_{\delta S}/S$), the noise in the measurement of the change in direction $\delta\phi$ is given by

$$\sigma_{\delta\phi} = 2.32/S + 14.5(^\circ). \quad (5)$$

The likelihood or the probability of obtaining a set of measurements $\boldsymbol{I}_k = (S, \phi, \delta S, \delta\phi)_k$, $(k = 1, \ldots, \mathcal{N})$ from a stimulus with stimulus parameters $\tilde{\boldsymbol{I}}_k = (\tilde{S}, \tilde{\phi}, \delta\tilde{S}, \delta\tilde{\phi})_k$ containing $\mathcal{N}$ points and simulating a scene S with dihedral angle d and rotation angle r is modelled by (modulo normalization):

$$p(\boldsymbol{I}|d,r) \propto \exp\left\{ -\frac{1}{2\mathcal{N}\lambda^2}\sum_{k=1}^{\mathcal{N}}\left(\frac{(\Delta S)^2}{\sigma_S^2} + \frac{(\Delta\phi)^2}{\sigma_\phi^2} + \frac{(\Delta\delta S)^2}{\sigma_{\delta S}^2} + \frac{(\Delta\delta\phi)^2}{\sigma_{\delta\phi}^2}\right)_k \right\}, \quad (6)$$

in which Δ is the difference between the measured property and the property of the stimulus (indicated by a tilde symbol), i.e. $\Delta S = S - \tilde{S}$. Without the factor $1/\mathcal{N}\lambda^2$ this would be a simple multiplication of the probabilities of $4\mathcal{N}$ independent measurements (i.e. using probability summation). The factor $1/\mathcal{N}$ accounts for the lack of improvement in performance of the visual system with an increase in number of points as observed in uniform motion experiments (De Bruyn & Orban 1988; Snowden & Braddick 1991; Werkhoven *et al*. 1992). Effectively, the improvement with number of points due to probability summation is counterbalanced by an increase in the noise level in the individual measurements by a factor of $\sqrt{\mathcal{N}}$. In this form the weight of each point remains proportional to the amount of information it carries. Furthermore, we divide the sum by λ^2. The free parameter λ allows the noise to be λ times larger for this task than in the uniform motion experiments.

Regarding the prior probability function $p(\boldsymbol{S})$, we assume that only rigidly rotating hinges could occur, as these formed the set of probe states from which subjects could choose from. All structures were perceived as convex, so the probability of occurrence of a dihedral angle outside the range 0–180° was set to zero. Rotation angles below 2° (outside the stimulus range) were also assigned a probability of zero, to deal with situations in which many of the texture elements were perceived as stationary. A pattern of dots with very little relative motion is seen as having no depth, and this prior effectively forces the model into interpreting such stimuli as having a large dihedral angle (small depth). All scenes within the non-zero range were assigned an equal probability of occurrence, i.e. within this range the model had no preferences for any three-dimensional dihedral angle over any other. On any given trial, the mean of the posterior distribution, $p(d,r|\boldsymbol{I})$, was taken as the predicted percept. This decision strategy is preferred over taking the maximum, another common choice, as the maximum is not a robust estimator for the extended, relatively flat posterior distributions found in our case (see figure 3). Moreover, the mean is closer to the actual value than the maximum on any given trial (in terms of the square-root deviation between the actual and estimated dihedral angle).

The matching process is modelled in the following way. In each trial the stimulus parameters of each point are measured with noise of magnitude $\lambda\sqrt{\mathcal{N}}\sigma$ (in which σ stands for σ_S, σ_ϕ $\sigma_{\delta S}$ or $\sigma_{\delta\phi}$, see equations (2–5)). The Bayesian framework is then used to find the optimal solution given these measurements. The predicted mean dihedral angle for a given stimulus is calculated by averaging over many sets of measurements (in the simulations 40 samples are used). In addition, the model gives a prediction of the variation in the responses.

In principle, a similar analysis has to be carried out to estimate the dihedral angle of the probe stimulus. However, because the probe stimulus is well defined, we can assume that the uncertainty in the dihedral angle of the probe is negligible. This is in accordance with the experiments of Johnston *et al*. (1994) who showed that perception of depth is highly veridical under similar circumstances.

(b) *Results*

Figure 3 shows two examples of posterior probability functions for two stimuli used in the experiment. Only dihedral angles greater than 8° are shown, as the dihedral angle must exceed the angular extent of the stimulus to comply with the laws of projective geometry. For smaller dihedral angles, the likelihood function, and therefore the posterior probability, is zero because of the high precision for localization. When both the dihedral angle and rotation angle are small (left), the distribution spreads along an iso-displacement contour (the contour that links the family of scenes producing identical image motion under two-view orthographic projection) (see also Bennett *et al*. 1996). This spread occurs for two reasons. First, because the additional information provided by perspective projection and multiple views—which can be used to uniquely identify which member of the family is actually being simulated—is unstable around this scene, i.e. the directions and accelerations of imaged, textured elements are similar to those produced by nearby scenes. Second, the visual system has much higher sensitivity to first-order speed information than to second-order acceleration information. The mean of the distribution is shifted to a larger dihedral angle from that simulated, yielding a biased response. Thus, even with zero-mean Gaussian noise, the spread of the likelihood function is not necessarily symmetrical about the scene corresponding to the measured flow-field. The right part of the graph shows the posterior distribution for a scene with a larger dihedral angle and rotation angle. Here, the distribution is more tightly constrained and the shift in the mean is both smaller and in the opposite direction. The reason for this is that for scenes with both a large

dihedral angle and rotation angle, the directions and accelerations of imaged, textured elements change rapidly for nearby scenes. Clearly then, even for the same level of noise the spread of the posterior varies significantly for different scenes.

Figure 4 shows that the model can account for the mean settings made by subjects over the entire range of scenes tested. Even for individual cases where some departure from the line $y=x$ exists, the data still fall on a single line, accounting for the dependence of perceived dihedral angle on rotation angle. For the small-field condition, λ, the free parameter, was set to 1, whereas for the large-field condition, λ was set to 1.4. The justification for this is that even though observers were free to make eye movements, SFM processes must integrate the motions of points many degrees apart, and it is known that motion discrimination thresholds rise for peripheral stimuli (McKee & Nakayama 1984). The uniform scaling of sensitivity to all moving points by λ (instead of an eccentricity-related scaling) must be an over-simplification. However, as the information is not uniformly distributed across the stimulus, but increases towards the periphery, this simplification is a good approximation to a more realistic scaling.

In addition to the good fit for the means, the variances in the data are well predicted by the model. Average standard deviations in the data and the model were 11° and 10°, respectively, for the large-field stimuli, and 13° and 13° for the small-field stimuli. For comparison, the maximum of the posterior distribution produced standard deviations of 22° and 29°, respectively.

5. DISCUSSION

It is important to demonstrate that the good performance of the model in predicting biases does not arise simply from the fact that low rotation angles and low dihedral angles are discarded in the analysis. Simulations indeed showed that a decrease in the lowest possible rotation angle leads to smaller predicted biases for objects rotating over small angles, but no change in the predictions for large rotation angles. (A change to 0.25° lowers the predictions for 2° and 4° rotations by 16° and 6° for the large-field stimuli and by 15° and 11° for the small-field stimuli.) However, these differences were small relative to the magnitude of the biases. That these cut-offs are of minor consequence is also clear from an inspection of figure 3: the distribution becomes very narrow for small rotations and dihedral angles, such that the contribution from this region of scene-space to the posterior distribution is very limited. Our conclusion from this is that the biases arise primarily from the nonlinear transformation that occurs between measurement space and the space of three-dimensional scenes, under which Gaussian blobs become distorted. On the other hand, simulations showed that changes in λ from their optimal value had a relatively large impact on the predictions, indicating that a good estimate of the magnitude of the noise is essential. (A 30% change in λ led to a 65% increase in the average-square root deviation between predicted and measured mean settings (i.e. averaged over subjects) for the large stimuli and a 25% increase for the small stimuli.)

To account for consistent biases in perceived three-dimensional SFM, some investigators have proposed the use of perceptual heuristics (Braunstein 1994): for instance, the assumption that objects tend to be as deep as they are wide (Caudek & Proffitt 1993); that a rotating structure is frontoparallel in the frame when its image is longest (Johansson & Jansson 1968); or that depth is proportional to the amount of relative image motion (Caudek & Proffitt 1993; Liter *et al*. 1994). Moreover, the reliance on such priors has been thought to preclude the use of visual information due to perspective and/or multi-frame displays as such usage would seem inconsistent with biased percepts. In contrast, we have shown that biases in recovered three-dimensional dihedral angle can be explained from an analysis in which all geometric relevant information is used without relying on prior preferences for particular dihedral angles. We have also developed a simple model of the noise on early motion measurements to explore the consequences for three-dimensional shape recovery which, with some exceptions (Nakayama 1985; Eagle & Blake 1995; Werkhoven & van Veen 1995; Eagle & Hogervorst 1997), have largely been ignored. Thus, the biases emerge naturally when a realistic model of SFM estimation is considered.

The model also provides a good quantitative account of the variability in subjects' settings, and moreover provides a possible explanation of the consistency of subjects' percepts even for two-view orthographic displays (e.g. Liter *et al*. 1994). First, not all scenes along the iso-displacement lines have equal likelihood, because: (i) the stimulus motion will not be consistent with scenes that yield strong perspective effects (i.e. non-parallel flow); and (ii) there are no accelerations in two-view stimuli, so the stimulus motion will not be consistent with scenes that yield large accelerations. Second, the model bases its output on the mean of the posterior distribution, a relatively stable decision strategy (compared, say, to one based on the mode) that will tend to yield a response near to the centre of the possible range of scenes.

In our model, the search-space of possible objects is restricted to rigidly hinged planes, as this formed the set of probe states from which subjects could choose. Under natural viewing conditions the visual system will encounter many more kinds of scenes, so that the priors cannot be set to zero outside some limited range. However, values in the likely distribution also approach zero for scenes that do not yield similar image motions. Figure 3 shows that the scenes under consideration can be restricted to those lying close to the iso-displacement contour that contains the stimulus. This finding is consistent with empirical data showing that three-dimensional shape discrimination is good for objects that differ in affine structure (van Damme & van de Grind 1993), but poor for affine-equivalent structures (e.g. Todd & Bressan 1990; Eagle & Blake 1995). This means that it should be possible to generalize the model to more complex scenes, and to cases in which the prior distribution is not constrained. A challenge that remains is to attach prior probabilities of occurrence for the family of shapes compatible with the noisy flow-field. Ultimately it will be necessary to take measurements from natural scenes to establish these, as has been done with colour and luminance distributions (e.g. Field 1987). Although it

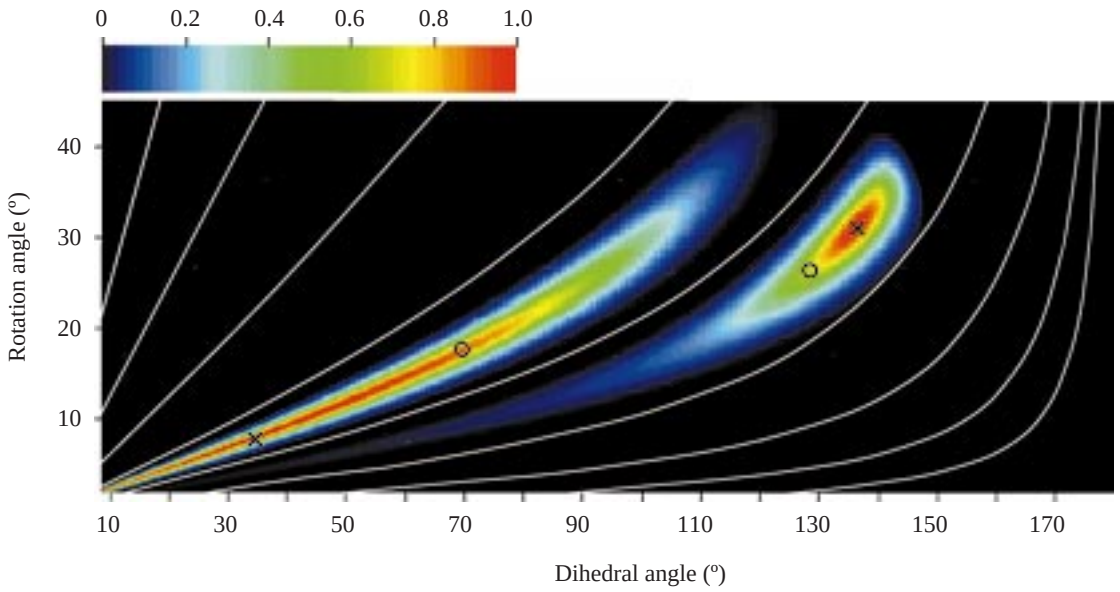

Figure 3. The posterior probability distributions for two small-field stimuli at a single trial on which the measurements are taken equal to the stimulus parameters. The probability of a scene is specified by the colour at that point on the graph. Also shown are iso-displacement contours connecting scenes that produce identical image-motion under two-frame orthographic projection, with lines towards the upper-left indicating larger displacements. On the left is depicted a scene with a dihedral angle of 35° and a rotation angle of 8°. On the right is depicted a scene with a dihedral angle of 157° rotating over 31°. Crosses indicate simulated scenes and circles indicate predicted scenes.

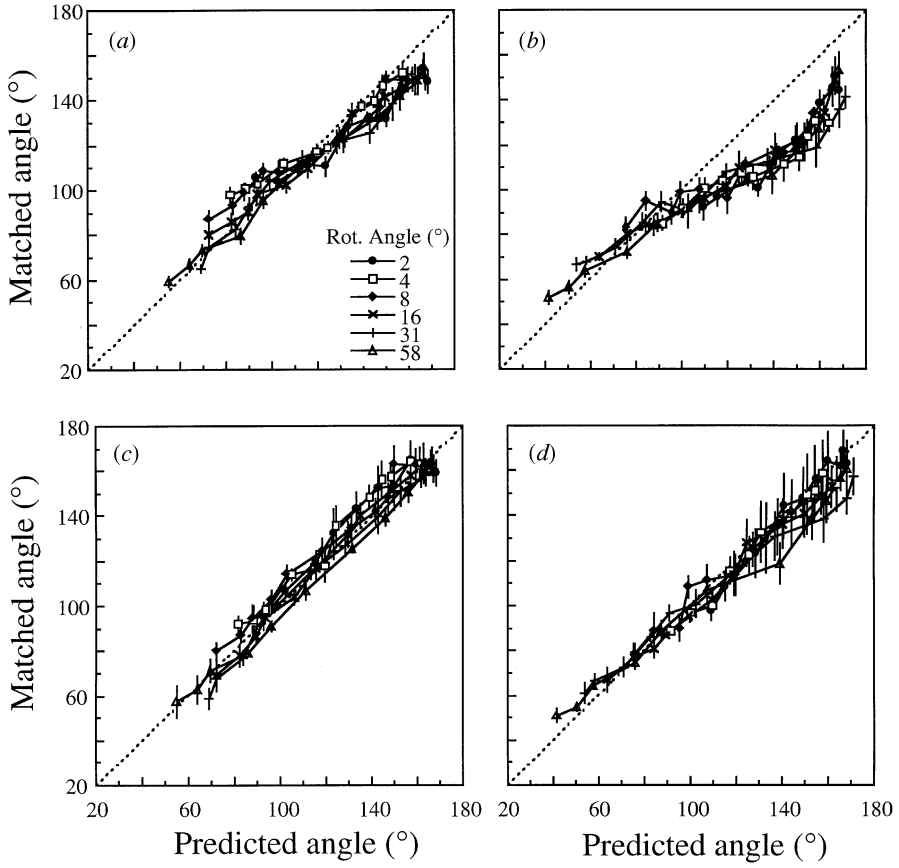

Figure 4. Replot of the data shown in figure 2, but now against the model's predicted settings. The settings for subject J.H. are again shown for the large- and small-field conditions in (*a*) and (*b*), with the means for three subjects shown in (*c*) and (*d*).

might not be trivial to determine which assumptions are used by the visual system, this topic should form an integral part of human perception research. The techniques described in this paper may apply generally to other domains, such as stereopsis, where systematic biases in three-dimensional shape judgements exist (e.g. Glennerster *et al*. 1996).

We gratefully acknowledge advice from André Noest regarding the development of the model. The work was funded by a

project grant from the BBSRC (No. 43/SO5036) and the Royal Society.

REFERENCES

Bennett, B. M., Hoffman, D. D., Prakash, C. & Richman, S. N. 1996 Observer theory, Bayes theory and psychophysics. In *Perception as Bayesian inference* (ed. D. C. Knill & W. Richards), pp. 163–212. Cambridge University Press.

Braunstein, M. L. 1994 Structure from motion. In *Visual detection of motion* (ed. A. T. Smith & R. J. Snowden), pp. 367–394. London: Academic Press.

Braunstein, M. L., Liter, J. C. & Tittle, J. S. 1993 Recovering 3-D shape from perspective translations and orthographic rotations. *J. Exp. Psychol. Hum. Percept. Perform.* **19**, 598–614.

Caudek, C. & Proffitt, D. R. 1993 Depth perception in motion parallax and stereokinesis. *J. Exp. Psychol. Hum. Percept. Perform.* **19**, 32–47.

De Bruyn, B. & Orban, G. A. 1988 Human velocity and direction discrimination measured with random-dot patterns. *Vision Res.* **28**, 1323–1335.

Eagle, R. A. & Blake, A. 1995 Two-dimensional constraints on three-dimensional structure from motion. *Vision Res.* **35**, 2927–2941.

Eagle, R. A. & Hogervorst, M. A. 1997 The role of perspective information in structure-from-motion judgements. *Invest. Ophthalmol. Vis. Sci. Suppl.* **38**, 75.

Field, D. J. 1987 Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am.* A **4**, 2379–2394.

Glennerster, A., Rogers, B. J. & Bradshaw, M. F. 1996 Stereoscopic depth constancy depends on the subject's task. *Vision Res.* **21**, 3441–3456.

Huang, T. & Lee, C. 1989 Motion and structure from orthographic projections. *IEEE Trans. Patt. Anal. Mach. Intell.* **11**, 536–540.

Johansson, G. & Jansson, G. 1968 Perceived rotary motion from changes in a straight line. *Percept. Psychophys.* **4**, 165–170.

Johnston, E. B., Cumming, B. C. & Landy, M. S. 1994 Integration of stereopsis and motion shape cues. *Vision Res.* **34**, 2259–2275.

Knill, D. C. & Richards, W. (eds) 1996 *Perception as Bayesian inference*. Cambridge University Press.

Koenderink, J. J. & van Doorn, A. J. 1991 Affine structure from motion. *J. Opt. Soc. Am.* A **8**, 377–385.

Liter, J. C., Braunstein, M. L. & Hoffman, D. D. 1994 Inferring structure from motion in two-view and multiview displays. *Perception* **22**, 1441–1465.

Longuet-Higgins, H. C. 1981 A computer algorithm for reconstructing a scene from two projections. *Nature* **293**, 133–135.

McKee, S. P. & Nakayama, K. 1984 The detection of motion in the peripheral visual field. *Vision Res.* **24**, 25–32.

Nakayama, K. 1985 Extraction of higher order derivatives of the optical velocity field: limitations imposed by biological hardware. In *Sensory experience, adaptation and perception* (ed. L. Spillman & B. R. Wooten), pp. 59–71. Hillsdale, NJ: Lawrence Erlbaum.

Rogers, B. J. & Graham, M. E. 1979 Motion parallax as an independent cue for depth perception. *Perception* **8**, 125–134.

Snowden, R. J. & Braddick, O. J. 1991 The temporal integration and resolution of velocity signals. *Vision Res.* **31**, 907–914.

Tittle, J. S., Todd, J. T., Perotti, V. J. & Norman, J. F. 1995 Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 663–678.

Todd, J. T. & Bressan, P. 1990 The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Percept. Psychophys.* **48**, 419–430.

Ullman, S. 1979 *The interpretation of visual motion*. Cambridge, MA: MIT Press.

van Damme, W. J. M & van de Grind, W. A. 1993 Active vision and the identification of 3-dimensional shape. *Vision Res.* **33**, 1581–1587.

Wallach, H. & O'Connell, D. 1953 The kinetic depth effect. *J. Exp. Psychol.* **45**, 205–217.

Werkhoven, P. & van Veen, H. A. H. C. 1995 Extraction of relief from visual motion. *Percept. Psychophys.* **57**, 645–656.

Werkhoven, P., Snippe, H. P. & Toet, A. 1992 Visual processing of optic acceleration. *Vision Res.* **32**, 2313–2329.

Westheimer, G. & McKee, S. P. 1979 What prior uniocular processing is necessary for stereopsis? *Invest. Ophthalmol. Vis. Sci.* **18**, 614–621.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.